

AGRADECIMIENTOS

Honrar honra.

JOSÉ MARTÍ

Quisiera agradecer, en primer lugar, a mi director de tesis por haberme brindado la oportunidad de transitar este camino, por ser guía y amigo.

A todos los nuevos amigos que he hecho. Eskerrik asko.

A mis compañeros de la Facultad de Informática de la Universidad de Cienfuegos por todo su apoyo.

A la UPV/EHU y la UCf por haber dado el soporte para el desarrollo de esta investigación.

Por último, y no menos importante, al apoyo incondicional, comprensión y estímulo de toda mi familia que ha sido de un valor incalculable.

Dedico especialmente esta tesis a mis padres Elda y Eduardo.

PRÓLOGO

Los métodos de clustering, o clasificación automática, constituyen un grupo importante de herramientas de análisis de datos. Estos métodos son utilizados para el análisis de imágenes médicas con el objetivo de identificar y categorizar las formas o funciones de diferentes tipos de tejidos u órganos, proceso éste conocido también como segmentación de imágenes.

Sin embargo, la mayoría de los métodos de clustering existentes se basan en una noción general de semejanza sin tener en cuenta ninguna información adicional del dominio como, por ejemplo, las relaciones espaciales entre los objetos.

En la presente tesis se aborda el problema de la segmentación de imágenes desde el punto de la clasificación automática. Se destaca la idea de que la segmentación de una imagen se puede ver como la partición de los píxeles.

Además se apoya la idea de que los modelos que describan adecuadamente las relaciones espaciales entre los objetos pueden servir como herramientas útiles para la clasificación. De manera implícita se asume que en el caso de la segmentación de imágenes los píxeles cercanos espacialmente poseen características similares y deben estar, por tanto, en la misma clase. Los métodos de clasificación automática que traten de preservar la estructura espacial constituyen un objeto de interés en esta investigación.

Este trabajo trata los temas de la segmentación de imágenes y los métodos de clustering. Por esa razón en la Parte I se incluyen una serie de conceptos y métodos que a nuestro juicio resultan de interés para facilitar la comprensión de los problemas que se analizan en la tesis. Así en el Capítulo 1 se analizan los principales conceptos relacionados con las imágenes digitales, particularmente las imágenes médicas. El Capítulo 2 brinda una visión acerca de los métodos de clustering clásicos. A los métodos de segmentación de imágenes se dedica el Capítulo 3.

En La Parte II se investigan las relaciones espaciales entre los objetos de un conjunto interpretadas a la luz del concepto de autocorrelación espacial. La interpretación descriptiva asumida en este trabajo postula que si los valores relativamente altos (bajos) de la magnitud analizada en una determinada localización están acompañados de valores relativamente altos (bajos) en las localizaciones vecinas, entonces puede hablarse de la existencia de autocorrelación positiva. Si por el contrario, los valores relativamente altos (bajos) se alternan con valores relativamente bajos (altos) en las localizaciones vecinas, la autocorrelación será negativa.

Se estudian modificaciones a los índices de autocorrelación clásicos, se proponen nuevos índices, y se estudian sus propiedades y relaciones. La aplicación del índice de Geary al estudio de las particiones de un conjunto de objetos con relaciones espaciales ha determinado la definición de un índice de autocorrelación espacial asociado a particiones, C_B , que toma en cuenta únicamente las fronteras entre las clases de una partición.

La Parte III está dedicada a los métodos de clustering con restricciones. En el Capítulo 5 se presenta el estudio teórico de un método de clustering jerárquico aglomerativo con restricciones de contigüidad. Se propone una función de desemejanzas que toma en cuenta las relaciones de contigüidad impuestas sobre los datos. Se demuestra que la jerarquía de clusters que se obtiene con el método propuesto al utilizar una estrategia aglomerativa restringida a la fusión de clusters adyacentes, es libre de inversiones.

En el Capítulo 6 se evalúa el método propuesto tomando como área de aplicación la segmentación de imágenes médicas, específicamente imágenes del cerebro. Los objetos a clasificar en este caso son los píxeles de una imagen y las relaciones de contigüidad tienen en cuenta la disposición espacial. En esta evaluación se han utilizado tanto imágenes simuladas como imágenes reales. Estos experimentos contribuyen a ilustrar las potencialidades y límites del método propuesto, así como a apoyar con evidencia empírica los resultados teóricos.

Finalmente, en la Parte IV se presentan nuestras conclusiones. Como líneas futuras de trabajo se proponen varias cuestiones interesantes que emanaron del proceso de investigación realizado.

TABLA DE CONTENIDOS

PARTE I.....	1
1 INTRODUCCIÓN	3
1.1 IMAGEN DIGITAL	3
1.2 IMAGEN DIGITAL MULTIBANDA.....	5
1.3 IMÁGENES TOMOGRÁFICAS	6
1.4 LA SEGMENTACIÓN DE IMÁGENES.....	8
2 MÉTODOS DE AGRUPAMIENTO (CLUSTERING)	13
2.1 REPRESENTACIÓN DE LOS DATOS	14
2.1.1 Representación vectorial	15
2.1.2 Representación basada en proximidades.....	16
2.2 ENFOQUES CLÁSICOS.....	20
2.2.1 Agrupamiento particional.....	21
2.2.2 Agrupamiento jerárquico.....	29
2.2.3 Estrategias de implementación basadas en los vecinos más próximos para el agrupamiento jerárquico aglomerativo sin restricciones	35
2.3 OTROS ENFOQUES.....	38
3 LOS MÉTODOS DE SEGMENTACIÓN DE IMÁGENES	41
3.1 MÉTODOS BASADOS EN BORDES.....	43
3.2 MÉTODOS BASADOS EN REGIONES	50
3.2.1 Umbralización.....	51
3.2.2 Crecimiento de regiones	56
3.2.3 División y fusión	60
PARTE II	65
4 SOBRE LA AUTOCORRELACIÓN ESPACIAL.....	67
4.1 INTRODUCCIÓN.....	67
4.2 LA AUTOCORRELACIÓN ESPACIAL	67
4.2.1 Definiciones y propiedades estadísticas previas	70
4.2.2 El índice modificado de Moran	85
4.2.2.1 El gráfico de Moran	90
4.2.2.2 Índices locales de Moran.....	91
4.2.3 Índice modificado de Geary	93
4.2.3.1 El gráfico de Geary	96
4.2.3.2 Índices locales de Geary.....	97
4.2.4 Nuevos índices de autocorrelación espacial	101
4.2.4.1 Índices de vecindad η	101
4.2.4.2 Índices de vecindad δ	107
4.2.5 Relaciones entre los diversos tipos de índices.....	109
4.2.5.1 Relaciones entre índices globales	109
4.2.5.2 Relaciones entre índices locales.....	111
4.2.6 Casos particulares de estructura espacial	112
4.2.6.1 Caso a).....	112
4.2.6.2 Caso b).....	115
4.2.6.3 Caso c).....	116
4.2.6.4 Caso d).....	119

4.2.7	<i>Aplicación del índice modificado de Geary a particiones</i>	124
4.2.7.1	Introducción	124
4.2.7.2	Notaciones previas para una partición	125
4.2.7.3	Indices modificados asociados a una partición	128
4.2.7.4	Índice modificado de Geary asociado a una partición en un espacio estructurado.	132
4.2.7.4.1	Caso a).....	132
4.2.7.4.2	Caso b)	135
4.2.7.4.3	Caso c).....	135
4.3	CONCLUSIONES.....	136
PARTE III		137
5	CLUSTERING CON RESTRICCIONES	139
5.1	INTRODUCCIÓN	139
5.2	EL AGRUPAMIENTO CON RESTRICCIONES	139
5.3	PLANTEAMIENTO DEL PROBLEMA	146
5.4	NOTACIÓN Y DEFINICIONES.....	148
5.5	EL PROBLEMA DE LAS INVERSIONES.....	154
5.6	ESTRATEGIAS DE IMPLEMENTACIÓN BASADAS EN LOS VECINOS MÁS PRÓXIMOS PARA EL CASO DE RESTRICCIONES DE CONTIGÜIDAD	157
5.7	ANÁLISIS DE LA COMPLEJIDAD COMPUTACIONAL. EL CASO DE LAS IMÁGENES DIGITALES	161
5.8	COMPORTAMIENTO DEL ALGORITMO JERÁRQUICO AGLOMERATIVO CON RESTRICCIONES DE CONTIGÜIDAD.....	162
5.9	EJEMPLO ILUSTRATIVO	168
5.10	CONCLUSIONES	170
6	EXPERIMENTACIÓN	171
6.1	LAS IMÁGENES DE RESONANCIA MAGNÉTICA.....	171
6.2	EL PROBLEMA DE LA SEGMENTACIÓN DE IMÁGENES DE RESONANCIA MAGNÉTICA.....	172
6.3	FUENTES PARA LA EVALUACIÓN DE LOS MÉTODOS DE SEGMENTACIÓN	174
6.4	ÍNDICES DE CALIDAD DE LAS SEGMENTACIONES	174
6.5	SOBRE LAS BASES DE DATOS DE IMÁGENES	178
6.5.1	<i>Base de datos de imágenes simuladas (BrainWeb)</i>	179
6.5.2	<i>Base de datos de imágenes reales (IBSR)</i>	179
6.5.3	<i>Comentarios generales</i>	180
6.6	EXPERIMENTOS CON IMÁGENES SIMULADAS DE BRAINWEB. RESULTADOS	181
6.6.1	<i>Objetivo</i>	181
6.6.2	<i>Requisitos para abordar el objetivo</i>	181
6.6.3	<i>Elección de la región de interés</i>	182
6.6.4	<i>Selección del radio de la vecindad</i>	183
6.6.5	<i>Efecto del radio de vecindad utilizado en los resultados</i>	190
6.6.6	<i>Resultados para diferentes cortes en el volumen estudiado</i>	191
6.6.7	<i>Conclusiones</i>	193
6.7	EXPERIMENTOS CON IMÁGENES REALES DE IBSR. RESULTADOS	194
6.7.1	<i>Objetivo</i>	194
6.7.2	<i>Requisitos para abordar el objetivo</i>	195
6.7.3	<i>Elección de la región de interés</i>	195
6.7.4	<i>Selección del radio de la vecindad</i>	196
6.7.5	<i>Análisis de los resultados en un ejemplo</i>	197
6.7.6	<i>Resultados generales</i>	201
6.7.7	<i>Conclusiones</i>	203
6.8	EXPERIMENTOS PARA ACELERAR LA EJECUCIÓN DEL ALGORITMO. RESULTADOS	204

6.8.1	<i>Objetivo</i>	204
6.8.2	<i>Requisitos para abordar el objetivo</i>	204
6.8.3	<i>Descripción del heurístico</i>	205
6.8.4	<i>Evaluación experimental del heurístico</i>	206
6.8.4.1	Escenarios experimentales.....	208
6.8.4.2	Análisis de las particiones obtenidas.....	209
6.8.4.2.1	Escenario (a)	209
6.8.4.2.2	Escenario (b)	210
6.8.4.2.3	Escenario (c).....	213
6.8.4.2.4	Escenario (d)	216
6.8.4.2.5	Escenario (e)	219
6.8.4.2.6	Resultados.....	220
6.8.5	<i>Conclusiones</i>	222
PARTE IV		223
7	CONCLUSIONES Y FUTURAS INVESTIGACIONES	225
7.1	CONCLUSIONES	225
7.2	FUTURAS INVESTIGACIONES	227
7.2.1	<i>Capítulo 4</i>	227
7.2.2	<i>Capítulo 5</i>	228
7.2.3	<i>Capítulo 6</i>	228
REFERENCIAS		229

LISTA DE FIGURAS

FIGURA 1-1 REPRESENTACIÓN MATRICIAL DE UNA IMAGEN DIGITAL MONOCROMÁTICA.	3
FIGURA 1-2 REPRESENTACIÓN VECTORIAL DE UNA IMAGEN DIGITAL MONOCROMÁTICA.	4
FIGURA 1-3 (A) UNA IMAGEN CON b BYTES POR PÍXEL; (B) b BANDAS CON UN BYTE POR PÍXEL.	5
FIGURA 1-4 CORTE ASOCIADO A UNA IMAGEN TOMOGRÁFICA.	6
FIGURA 1-5 REPRESENTACIÓN DE UNA IMAGEN MEDIANTE VÓXELES.	7
FIGURA 1-6 IMÁGENES DE RESONANCIA MAGNÉTICA.	7
FIGURA 1-7 IMAGEN DE RESONANCIA MAGNÉTICA (IZQUIERDA) Y REGIONES DE INTERÉS SEGMENTADAS: SUSTANCIA GRIS, SUSTANCIA BLANCA Y LÍQUIDO CEFALORRAQUÍDEO (DERECHA).....	8
FIGURA 1-8 TIPOS DE ADYACENCIA. (A) - 4 VECINOS; (B) - 8 VECINOS	9
FIGURA 2-1 REPRESENTACIÓN VECTORIAL.	15
FIGURA 2-2 REPRESENTACIÓN BASADA EN PROXIMIDADES.....	17
FIGURA 2-3 DESCOMPOSICIÓN DE UNA MATRIZ DE CONEXIONES.	21
FIGURA 2-4 COMPORTAMIENTO DEL ESPACIO DE BÚSQUEDA DE LAS PARTICIONES.....	24
FIGURA 2-5 ÁRBOL MST	33
FIGURA 2-6 EJEMPLOS DONDE EL CRITERIO DE COMPACIDAD NO ES SUFICIENTE.	39
FIGURA 2-7 AGRUPAMIENTO DONDE EL CRITERIO DE CONECTIVIDAD ES APROPIADO.	39
FIGURA 3-1 RECTÁNGULO DE KANIZSA.	43
FIGURA 3-2 MÁSCARA DE 3X3 PÍXELES.	44
FIGURA 3-3 EJEMPLO DE APLICACIÓN DEL OPERADOR DE SOBEL.	47
FIGURA 3-4 MÁSCARA DE 5X5 DEL LOG.	48
FIGURA 3-5 APLICACIÓN DEL MÉTODO DE CANNY A UNA IMAGEN DE RESONANCIA MAGNÉTICA.....	48
FIGURA 3-6 OBTENCIÓN DE UN CAMINO ÓPTIMO SEGÚN EL MÉTODO <i>LIVE-WIRE</i>	49
FIGURA 3-7 IMAGEN DE RM, SU HISTOGRAMA, SEGMENTACIÓN MEDIANTE UMBRALIZACIÓN SIMPLE Y MEDIANTE UMBRALIZACIÓN MÚLTIPLE.....	52
FIGURA 3-8 EJEMPLO DE UNA IMAGEN CON 3 NIVELES DE GRISES Y SU MATRIZ DE CO-OCURRENCIAS.	54
FIGURA 3-9 SEGMENTACIÓN MEDIANTE LA TRANSFORMADA <i>WATERSHED</i>	59
FIGURA 3-10 SEGMENTACIÓN MEDIANTE DIVISIÓN Y FUSIÓN. (A) REPRESENTA LA IMAGEN ORIGINAL; (B) LA DIVISIÓN INICIAL EN CUATRO BLOQUES; (C) LA DIVISIÓN FINAL EN BLOQUES HOMOGÉNEOS; Y (D) LA SEGMENTACIÓN RESULTANTE DEL PROCESO DE FUSIÓN.....	61
FIGURA 3-11 DESCOMPOSICIÓN DE UNA IMAGEN MEDIANTE ÁRBOLES CUATERNARIOS.	62
FIGURA 4-1 TIPOS DE ADYACENCIA. (A) 4 – VECINOS; (B) 8 – VECINOS.	69
FIGURA 4-2 ANALOGÍA EN LA REPRESENTACIÓN TABULAR DE LOS DATOS: (A) HABITUAL, Y (B) DATOS ESPACIALES.	70
FIGURA 4-3 IMAGEN DE EJEMPLO.....	72
FIGURA 4-4 GRÁFICO DE LOS PRODUCTOS CRUZADOS PARA EJEMPLO DE LA FIGURA 4-3.....	85
FIGURA 4-5 IMAGEN DE EJEMPLO.....	87
FIGURA 4-6 EJEMPLOS DE REGIONES CON LOS VALORES ASOCIADOS DEL ÍNDICE MODIFICADO DE AUTOCORRELACIÓN ESPACIAL DE MORAN (USANDO 4 – ADYACENCIA).....	89
FIGURA 4-7 GRÁFICO DE MORAN.....	90
FIGURA 4-8 VALORES DE LOS ÍNDICES LOCALES DE MORAN	93
FIGURA 4-9 GRÁFICO DE GEARY.....	97
FIGURA 4-10 VALORES DE LO ÍNDICES LOCALES DE GEARY.....	99
FIGURA 4-11 VALORES DE LOS ÍNDICES LOCALES DE LEBART.....	100
FIGURA 4-12 GRÁFICO DE LOS ÍNDICES LOCALES η_{1i}^*	103
FIGURA 4-13 GRÁFICO DE ÍNDICE η_1^* Y η_2^*	104
FIGURA 4-14 GRÁFICO DE LOS ÍNDICES LOCALES η_{2i}^*	106

FIGURA 4-15 GRÁFICO DE ÍNDICE δ_1^* Y δ_2^*	108
FIGURA 4-16 EJEMPLO DE GRAFO REDUCIDO DE UNA IMAGEN	131
FIGURA 5-1 EJEMPLO DE INCLUSIÓN DE LA INFORMACIÓN ESPACIAL. LOS DATOS ORDENADOS (A), PARTICIÓN SIN CONSIDERAR EL ORDEN (B), PARTICIÓN USANDO EL NÚMERO DE ORDEN (C).....	141
FIGURA 5-2 EJEMPLO CON RESTRICCIONES <i>CANNOT-LINK</i>	145
FIGURA 5-3 MODIFICACIÓN DE LA VECINDADES DE LOS CLUSTERS ANTE UNA FUSIÓN. ANTES DE LA FUSIÓN (A), DESPUÉS DE LA FUSIÓN (B).....	164
FIGURA 5-4 COMPORTAMIENTO DE LOS TIEMPOS DE EJECUCIÓN DE LA IMPLEMENTACIÓN REALIZADA PARA DIFERENTES TAMAÑOS DE IMÁGENES UTILIZANDO DIFERENTES RADIOS DE VECINDAD.	167
FIGURA 5-5 COMPORTAMIENTO DEL TIEMPO DE EJECUCIÓN DEL MÉTODO PROPUESTO	168
FIGURA 5-6 EJEMPLO ARTIFICIAL: (A) IMAGEN ORIGINAL CON DOS PATRONES SIMILARES; (B) DENDROGRAMA OBTENIDO CON EL MÉTODO DE ENLACE PROMEDIO; (C)-(D) PARTICIONES DE DOS Y TRES CLUSTERS OBTENIDAS CON EL MÉTODO DE ENLACE PROMEDIO; (E)-(F) UNA PARTICIÓN DE TRES CLUSTERS OBTENIDA AL UTILIZAR LA 4 – VECINDAD; (G)-(H) UNA PARTICIÓN DE TRES CLUSTERS OBTENIDA AL UTILIZAR UNA VECINDAD DE RADIO 3.	169
FIGURA 6-1 CORTE CENTRAL SIMULADO OBTENIDO DE BRAINWEB CON 3% DE RUIDO Y 20% DE HETEROGENEIDADES DE INTENSIDAD (A), Y EL MODELO ANATÓMICO CORRESPONDIENTE (B).	182
FIGURA 6-2 REGIÓN DE INTERÉS DE DIMENSIÓN 75x70 PÍXELES EN EL CORTE SELECCIONADO.	183
FIGURA 6-3 REGIÓN DE INTERÉS AMPLIADA EN EL CORTE SELECCIONADO (A), Y EN EL MODELO ANATÓMICO (B).	183
FIGURA 6-4 DENDROGRAMA OBTENIDO AL SEGMENTAR LA REGIÓN DE INTERÉS DE LA FIG. 3 UTILIZANDO UN RADIO = $\sqrt{2}$ (8-VECINDAD) (A); PARTICIÓN DE 4 CLUSTERS (B); Y DE 16 CLUSTERS (C).....	184
FIGURA 6-5 DENDROGRAMA OBTENIDO AL SEGMENTAR LA REGIÓN DE INTERÉS DE LA FIG. 3 UTILIZANDO UNA VECINDAD CON RADIO DE 20 PÍXELES (A); PARTICIÓN DE 3 CLUSTERS.	185
FIGURA 6-6 ESQUEMA DE BARRIDO DE UNA IMAGEN UTILIZANDO UNA VENTANA DESLIZANTE.	186
FIGURA 6-7 VENTANA DESLIZANTE DE RADIO 2 SEGÚN LA DISTANCIA EUCLIDIANA.	187
FIGURA 6-8 GRÁFICO QUE RELACIONA EL RADIO DE LA VECINDAD DESLIZANTE Y LA DIVERSIDAD LOCAL EN EL EJEMPLO DE LA FIGURA 6-3A.	187
FIGURA 6-9 DENDROGRAMA OBTENIDO AL SEGMENTAR LA REGIÓN DE INTERÉS DE LA FIG. 3 UTILIZANDO UNA VECINDAD CON RADIO DE 15 PÍXELES (A); PARTICIÓN DE 3 CLUSTERS (B); MODELO ANATÓMICO (C).	188
FIGURA 6-10 PARTICIÓN DE 3 CLUSTERS OBTENIDA UTILIZANDO EL ÍNDICE DE CALINSKI-HARABZS Y EL ÍNDICE DE HARTIGAN PARA REALIZAR EL CORTE DEL DENDROGRAMA DE LA FIG. 9 (A); COMPORTAMIENTO DEL ÍNDICE C_B Y VALOR SUGERIDO DE 3 CLUSTERS (B).....	189
FIGURA 6-11 VOLUMEN COMPUESTO POR 21 CORTES EN LA ZONA CENTRAL SELECCIONADA DEL CEREBRO....	191
FIGURA 6-12 ÍNDICE AJUSTADO DE RAND PARA LAS PARTICIONES OBTENIDAS EN CADA UNO DE LOS 21 CORTES SELECCIONADOS.	193
FIGURA 6-13 CORTE DE EJEMPLO (A); LA SEGMENTACIÓN MANUAL CORRESPONDIENTE (B); LA REGIÓN DE INTERÉS SELECCIONADA (C); Y LA REGIÓN DE INTERÉS SEGMENTADA MANUALMENTE.	196
FIGURA 6-14 VALORES DE DIVERSIDAD LOCAL PARA DIFERENTES RADIOS DE VECINDAD EN EL EJEMPLO DE LA FIGURA 6-13.....	197
FIGURA 6-15 DENDROGRAMA OBTENIDO PARA EL CASO DE EJEMPLO DE LA FIGURA 6-13.....	198
FIGURA 6-16 COMPORTAMIENTO DE ÍNDICE C_B EL EJEMPLO DE LA FIGURA 6-13.....	198
FIGURA 6-17 SEGMENTACIÓN OBTENIDA (A); ORIGINAL (B); SEGMENTACIÓN MANUAL (C).	199
FIGURA 6-18 SEGMENTACIÓN DE 2 CLUSTERS (A), DE 3 CLUSTERS (B), DE 4 CLUSTERS (C), Y DE 5 CLUSTERS (D).	200
FIGURA 6-19 ESQUEMA DE LAS ESTRUCTURAS QUE CUBREN EL ENCÉFALO DEL HOMBRE.	201
FIGURA 6-20 ÍNDICE PROMEDIO PARA LA SEGMENTACIÓN DE LA SUSTANCIA GRIS Y DE LA SUSTANCIA BLANCA.	203
FIGURA 6-21 IMÁGENES DE PACIENTES DE EM.	205
FIGURA 6-22 REGIÓN SELECCIONADA EN UNA IMAGEN DE UN PACIENTE DE EM.	207
FIGURA 6-23 RADIO DE VECINDAD SUGERIDO PARA LA REGIÓN DE INTERÉS SELECCIONADA.....	207

FIGURA 6-24 DENDROGRAMA OBTENIDO PARA EL ESCENARIO (B).....	209
FIGURA 6-25 PARTICIONES OBTENIDAS EN EL ESCENARIO (A).....	210
FIGURA 6-26 DENDROGRAMA OBTENIDO PARA EL ESCENARIO (B).....	211
FIGURA 6-27 COMPORTAMIENTO DEL ÍNDICE C_B EN EL ESCENARIO (B).....	211
FIGURA 6-28 PARTICIONES OBTENIDAS EN EL ESCENARIO (B).....	212
FIGURA 6-29 DIAGRAMA DE BARRAS DE DISTANCIAS ENTRE PÍXELES.....	213
FIGURA 6-30 DENDROGRAMA OBTENIDO PARA EL ESCENARIO (C).....	214
FIGURA 6-31 COMPORTAMIENTO DEL ÍNDICE C_B EN EL ESCENARIO (C).....	215
FIGURA 6-32 PARTICIONES OBTENIDAS EN EL ESCENARIO (C).....	215
FIGURA 6-33 IMAGEN ORIGINAL (A); IMAGEN DESPUÉS DE LA REDUCCIÓN (B).....	217
FIGURA 6-34 DENDROGRAMA OBTENIDO PARA EL ESCENARIO (D).....	218
FIGURA 6-35 COMPORTAMIENTO DEL ÍNDICE C_B EN EL ESCENARIO (D).....	218
FIGURA 6-36 PARTICIONES OBTENIDAS EN EL ESCENARIO (E).....	219
FIGURA 6-37 PARTICIONES OBTENIDAS EN EL ESCENARIO (A).....	220

LISTA DE TABLAS

TABLA 2-1 ALGORITMO <i>K-MEANS</i> BÁSICO	28
TABLA 2-2 ALGORITMO JERÁRQUICO AGLOMERATIVO BÁSICO.....	30
TABLA 2-3 ALGORITMO JERÁRQUICO DIVISIVO BASADO EN EL ÁRBOL MST	33
TABLA 2-4 ALGORITMO BISECTING K-MEANS	34
TABLA 3-1 CLASIFICACIÓN DE MÉTODOS DE SEGMENTACIÓN [68].....	42
TABLA 3-2 ALGORITMO ITERATIVO SEGÚN LA PROPUESTA DE [90].....	53
TABLA 3-3 ALGORITMO BÁSICO DE CRECIMIENTO DE REGIONES SIN COMPETENCIA	57
TABLA 3-4 ESQUEMA DEL MÉTODO DE DIVISIÓN Y FUSIÓN	60
TABLA 4-1 RELACIONES ENTRE LAS VARIABLES ANALIZADAS	83
TABLA 4-2 VALORES DE LAS VARIABLES ANALIZADAS PARA EL EJEMPLO DE LA FIGURA 4-3	84
TABLA 4-3 RELACIÓN ENTRE LOS DIFERENTES ÍNDICES	110
TABLA 4-4 VALORES DE LOS ÍNDICES CALCULADOS PARA EL EJEMPLO DE LA FIGURA 4-3	110
TABLA 4-5 VALORES DE LOS ÍNDICES LOCALES CALCULADOS PARA EL EJEMPLO DE LA FIGURA 4-3	112
TABLA 4-6 MATRIZ DE CORRELACIONES ENTRE LOS ÍNDICES PARA EL EJEMPLO DE LA FIGURA 4-3	112
TABLA 4-7 NOTACIONES EMPLEADAS	127
TABLA 6-1 PARTICIONES OBTENIDAS CON DIFERENTES RADIOS DE VECINDAD.....	190
TABLA 6-2 COMPORTAMIENTO DE LOS RADIOS DE VECINDAD EN LOS 21 CORTES.....	192
TABLA 6-3 VALORES DE ÍNDICE DE JACCARD PARA CADA SUJETO POR TIPO DE TEJIDO.....	202
TABLA 6-4 TIEMPOS DEL ALGORITMO <i>K-MEANS</i>	220
TABLA 6-5 TIEMPOS DE EJECUCIÓN PARA CADA UNO DE LOS ESCENARIOS ANALIZADOS.....	221
TABLA 6-6 PARTICIONES SELECCIONADAS EN CADA UNO DE LOS ESCENARIOS ANALIZADOS.	222

PARTE I

1 Introducción

Antes de presentar un análisis de los métodos de segmentación de imágenes digitales es preciso definir qué es una imagen digital, cuáles son sus componentes, y las principales vías de obtención de las imágenes médicas.

1.1 Imagen digital

Una imagen digital de tonalidades de grises, como es el caso de la mayoría de las imágenes médicas, puede ser vista como una matriz de dimensiones $n_F \times n_C$ donde n_F y n_C representan la cantidad de filas y de columnas de la matriz, respectivamente (Figura 1-1). Los índices de las filas y las columnas sirven para identificar un punto de la imagen, mientras que el valor numérico de un elemento de la matriz está asociado a alguna magnitud física medida. Cada punto de la imagen se denomina también *píxel* (*picture element*).

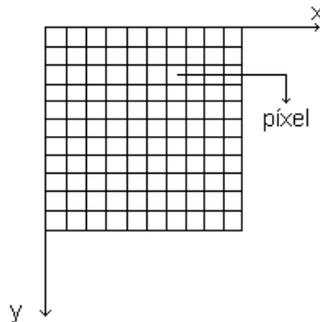


Figura 1-1 Representación matricial de una imagen digital monocromática.

Para generar una imagen digital se realiza un muestreo a lo largo de los ejes de coordenadas, por conveniencia orientados según se muestra en la Figura 1-1. De esta manera se obtiene una matriz de $n_F \times n_C$ muestras, cada una de las cuales representa un píxel. Se dice entonces que la imagen tiene una dimensión de n_F píxeles por filas, y n_C píxeles por columnas.

El conjunto de píxeles de una imagen se representará en forma de un vector \mathbf{x} de dimensión $n = n_F \times n_C$ para facilitar la realización de ciertas operaciones. En este caso se utilizará un vector columna $\mathbf{x} = (x_1, x_2, \dots, x_l, \dots, x_n)^T$, donde los componentes se corresponden con un

recorrido de la matriz de arriba hacia abajo y de izquierda a derecha. Es decir, un píxel que se encuentra en fila i y en la columna j aparecerá en la componente $l = (j-1)n_F + i$ del vector \mathbf{x} (ver Figura 1-2).

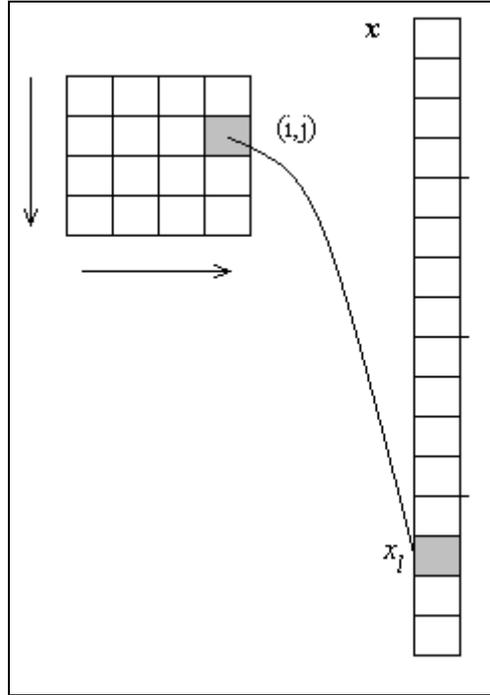


Figura 1-2 Representación vectorial de una imagen digital monocromática.

De la misma manera es posible realizar la transformación inversa, es decir, la componente x_m del vector \mathbf{x} se corresponde con la posición (i, j) , donde los índices son obtenidos según las siguientes expresiones

$$i = (m-1) \bmod n_F + 1$$

$$j = \lfloor m/n_F \rfloor + 1$$

donde **mod** representa el resto de la división y $\lfloor \ \rfloor$ la parte entera del argumento.

Una vez realizado el muestreo se procede a cuantificar el valor de la magnitud física de interés en cada una de las regiones, utilizando para ello G niveles enteros de tonalidades de grises. Se puede plantear entonces que cada píxel tiene asociado un valor escalar $x_i \in [G_{\min}, G_{\max}]$, donde $i = 1, \dots, n_F$, $j = 1, \dots, n_C$, y $G = G_{\max} - G_{\min} + 1$. Normalmente, y sin pérdida de generalidad, se asume que $G_{\min} = 0$.

La cantidad de niveles de cuantificación G es normalmente una potencia de 2, por ejemplo, 256, 1024, 4096. Un valor comúnmente utilizado es $G = 256$. Esto significa que cada píxel puede tener asociado un valor entre 0 y 255, donde el valor 0 representa el color negro, el valor 255 representa el blanco, y los valores intermedios representan tonalidades o niveles de gris. En este caso es suficiente utilizar un *byte* para almacenar cada píxel, y se dice que la *profundidad de la imagen* es de 8 bits por píxel, ó 1 *byte* por píxel. Se debe tener en cuenta qué parámetro, físico o químico, está siendo representado como intensidad luminosa para realizar una interpretación adecuada de lo que se aprecia en la imagen.

1.2 Imagen digital multibanda

En una imagen digital monocromática, el valor de un píxel es un escalar entre 0 y $G - 1$. Las imágenes digitales multibandas se pueden ver como la imagen de una misma región, por ejemplo geográfica ó anatómica, en la cual cada píxel tiene asociado un vector $x_i = (x_{i1}, \dots, x_{ib})$, donde $0 \leq x_{ik} < G_k$; $k = 1, \dots, b$, y el valor b define la cantidad de bandas. En general, los valores G_k pueden representar diferentes magnitudes, con diferentes intervalos de variación. Si se tiene, por ejemplo, que $G_k = 256$, para todo $k = 1, \dots, b$, se tendría una imagen con una profundidad de b bytes por píxel (ver Figura 1-3a).

Otra manera de visualizar una imagen multibanda es mediante una secuencia de imágenes monocromáticas (ver Figura 1-3b).

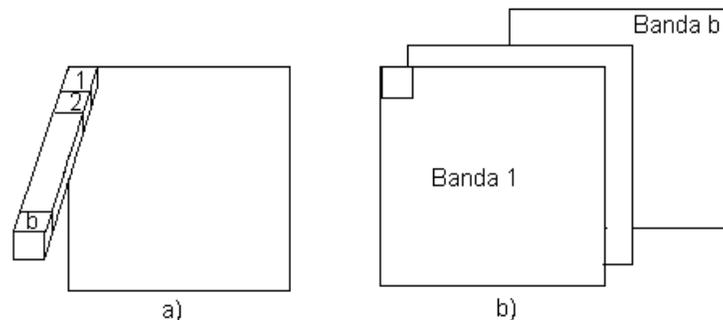


Figura 1-3 (a) Una imagen con b bytes por píxel; (b) b bandas con un byte por píxel.

Un ejemplo de imagen multibanda lo constituye una imagen a color. Una forma común de representar una imagen a color es mediante la utilización de 3 bandas R (*Red*), G (*Green*), B (*Blue*), con profundidad de un byte por píxel cada una. Si se considera la primera forma de representación tenemos una imagen con una profundidad de 24 bits por píxel.

Hasta el momento se ha tratado el caso del muestreo y cuantificación en el plano, por lo que las imágenes analizadas son imágenes digitales bidimensionales. Es posible realizar el muestreo y la cuantificación en una tercera dimensión, la cual puede ser el espacio o el

tiempo. En estos casos estaríamos en presencia de imágenes tridimensionales (3D), donde la tercera dimensión puede ser el eje espacial z , o el eje temporal t .

Las imágenes tridimensionales pueden ser vistas como una secuencia de imágenes monocromáticas o multibandas a lo largo del eje espacial z , o del eje temporal t .

1.3 Imágenes tomográficas

Los equipos tomográficos (del griego *tome*, que significa corte), tales como los tomógrafos de rayos X, o tomógrafos de resonancia magnética, permiten obtener imágenes de cortes o secciones, normalmente paralelos y uniformemente espaciados, de una región 3D.

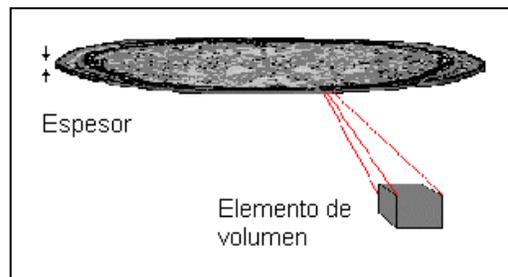


Figura 1-4 Corte asociado a una imagen tomográfica.

En medicina, por ejemplo, estos cortes cubren una determinada región del cuerpo humano la cual contiene un órgano determinado o está asociada a un determinado fenómeno o proceso bajo estudio (ver Figura 1-4). Según la orientación del corte las imágenes pueden ser axiales, sagitales o coronales. En una imagen axial el corte divide al cuerpo en una parte superior y otra inferior y la vista se obtiene como si se estuviera mirando desde la parte inferior hacia arriba (desde los pies a la cabeza). En una imagen sagital el corte divide al cuerpo en una parte izquierda y una derecha. Por último en una imagen coronal el corte divide al cuerpo en una parte frontal y otra posterior.

Si se consideran las dimensiones en el plano de un píxel en estas imágenes y el espesor del corte, tendríamos una extensión del píxel para el caso 3D en forma de un pequeño paralelepípedo con tres dimensiones (largo, ancho y altura), el cual es llamado *vóxel* (*volume element*). El centro de cada vóxel coincide con el centro del píxel en la imagen (ver Figura 1-5b).

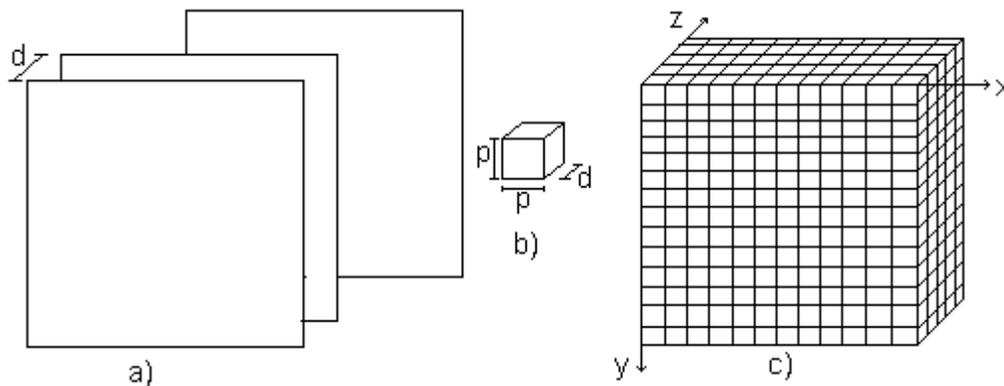


Figura 1-5 Representación de una imagen mediante vóxeles.

Cada vóxel representa un punto de muestreo y cuantificación de alguna magnitud medible, siendo la intensidad resultante proporcional a esta magnitud en todo el vóxel. En este caso la imagen puede ser considerada como una matriz 3D como se ilustra en la Figura 1-5c.

Las imágenes tomográficas poseen normalmente 256×256 ó 512×512 píxeles y una profundidad de 1 ó 2 bytes. En la Figura 1-6 se muestran tres imágenes de resonancia magnética de dimensiones 256×256 y 2 bytes de profundidad que representan un corte axial del cerebro (a), uno coronal (b), y otro sagital (c). La dimensión de cada píxel en el plano es de $0.859 \text{ mm} \times 0.859 \text{ mm}$ en todos los casos. El espesor del corte es de 5 mm . para las imágenes (a) y (c), y de 3 mm en el caso (b).

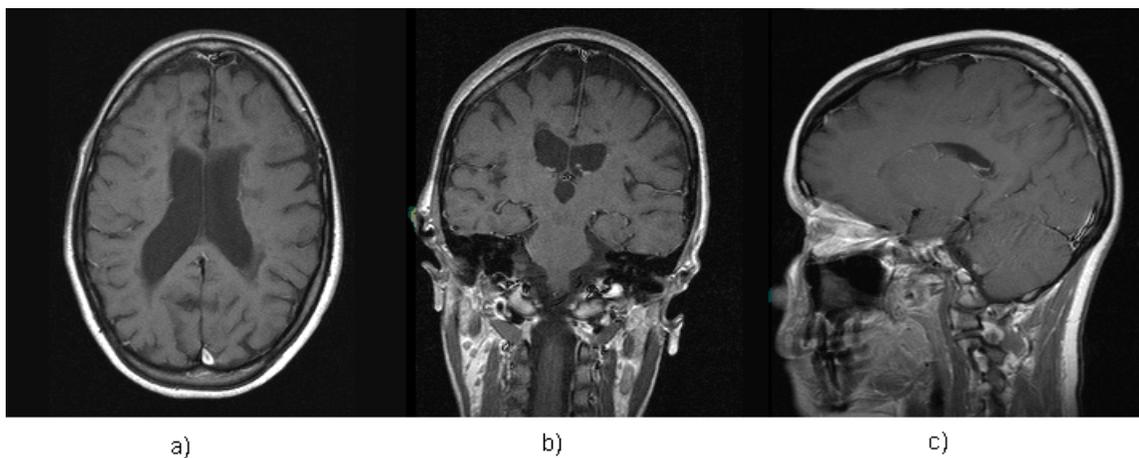


Figura 1-6 Imágenes de resonancia magnética.

Los tomógrafos permiten obtener cortes que no tienen por qué ser consecutivos. Esto genera un espacio entre cortes en el cual no se recupera información. Por ejemplo, en los casos de la Figura 1-6, los cortes (a) y (c) fueron tomados de una secuencia donde la separación entre cortes es de 1.5 mm . Por su parte en la secuencia del caso (b) la separación entre cortes es de 0.9 mm . Esto debe ser tenido en cuenta a la hora de realizar un procesamiento volumétrico que incluya todos los cortes.

Por último, consideraremos las imágenes de más de tres dimensiones. Los tomógrafos son capaces de generar imágenes en cuatro dimensiones (4D), realizando un muestreo en (x, y, z, t) . Este es el caso, por ejemplo, de imágenes tomográficas de un corazón latiendo, donde las tres primeras coordenadas son espaciales, y la cuarta es temporal.

1.4 La segmentación de imágenes

La división de una imagen en varios componentes o regiones con algún significado asociado, y la asignación a cada píxel de la imagen de una etiqueta correspondiente al componente al que pertenece, se denomina segmentación de imágenes. En la Figura 1-7 se muestra una imagen del cerebro y las regiones de interés segmentadas.

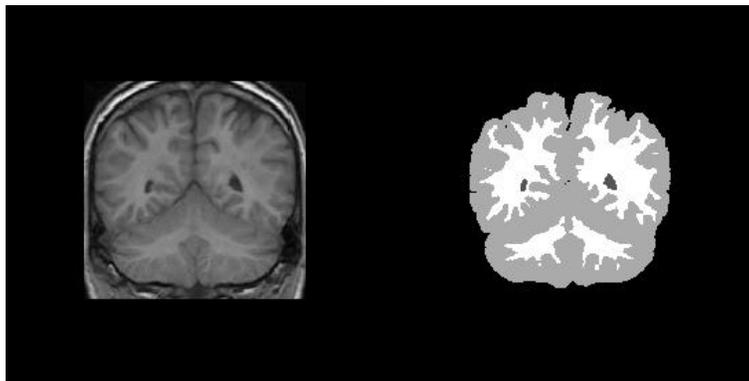


Figura 1-7 Imagen de resonancia magnética (izquierda) y regiones de interés segmentadas: sustancia gris, sustancia blanca y líquido cefalorraquídeo (derecha).

Se puede plantear que la segmentación es el proceso en el que se extraen los objetos de interés presentes en las imágenes. El resultado de la segmentación puede ser la frontera de un objeto respecto a su exterior, o los puntos que forman su interior. La forma de representación del resultado de la segmentación está asociada a la manera de representar ya sea la frontera, o ya sea el interior del objeto.

Las fronteras de los objetos delimitan las regiones que estos ocupan. Una vez que se hayan obtenido las fronteras es posible, en principio, reconstruir las regiones. De manera similar, si se tiene una partición de una imagen en regiones es posible hacer un seguimiento de la frontera de cada región.

La segmentación de imágenes médicas, en particular las del cerebro, es un proceso de alta complejidad, debido a su compleja estructura como objeto tridimensional. A lo anterior se adicionan las ambigüedades inducidas por el proceso de adquisición, entre las que se encuentran: el ruido, los bajos contrastes, las heterogeneidades de intensidad y el efecto de volumen parcial [1], todo lo cual hace difícil la interpretación aún para los especialistas.

Estos elementos sugieren un enfoque que tenga en cuenta todos estos factores a la hora de obtener una representación de los objetos de interés.

La calidad de una segmentación está dada por las características de sus componentes: deben ser coherentes desde el punto de vista espacial, mientras que componentes diferentes no deben ser semejantes [2, 3].

En el caso de imágenes digitales monocromáticas, como la mostrada en la Figura 1-7, los algoritmos de segmentación se basan fundamentalmente en dos propiedades básicas de los niveles de gris: la similitud, y la discontinuidad. La similitud está relacionada con la homogeneidad y se toma en calidad de criterio para la formación de regiones. Por su parte la discontinuidad es la medida de contraste entre el objeto y su exterior y se utiliza normalmente como base para la detección de puntos aislados, de líneas, y de bordes. En adelante se denominará borde a aquel conjunto conexo de píxeles que se encuentran en la frontera entre dos regiones. En este sentido el borde es concepto local, mientras la frontera de una región representa una idea más global.

La conectividad es de gran importancia en la segmentación pues simplifica la definición de varios conceptos asociados con las imágenes digitales. Para establecer si dos píxeles son conexos es preciso determinar si son vecinos o adyacentes y si sus niveles de intensidad cumplen cierta propiedad, por ejemplo, si son iguales o parecidos en un cierto sentido que habría que precisar. Normalmente se consideran los 4 vecinos, ó los 8 vecinos de un píxel (Figura 1-8).

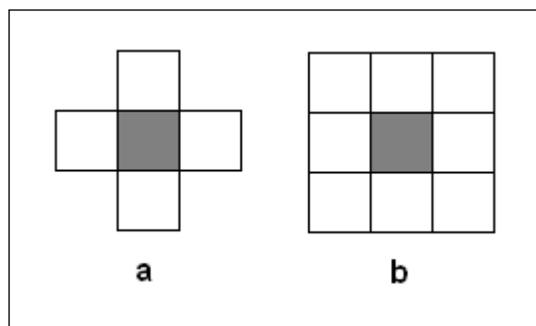


Figura 1-8 Tipos de adyacencia. (a) - 4 vecinos; (b) - 8 vecinos

Esta definición se puede generalizar a una vecindad alrededor de un píxel compuesta por el conjunto de píxeles que se encuentran a una distancia, no necesariamente euclidiana) menor o igual que un radio r dado.

Una vez establecida la relación de adyacencia es posible entonces representar la imagen mediante un grafo no dirigido. Cada píxel se hace corresponder con un vértice, y se establece una arista entre aquellos vértices que se consideran adyacentes. Más adelante se abunda en estos temas.

Tradicionalmente la segmentación de una imagen se interpreta como una partición de ésta en regiones disjuntas, las cuales son homogéneas en cierto sentido. Una posible definición sería la siguiente

Definición 1 Un conjunto de regiones $P = \{P_1, P_2, \dots, P_K\}$ es una segmentación de una imagen I en K regiones si se cumple que [4]

1. $\bigcup_{k=1}^K P_k = I$
2. $P_k \cap P_l = \emptyset, k \neq l$
3. P_k es conexa, $k = 1, 2, \dots, K$
4. $P_k, k = 1, 2, \dots, K$, es homogénea en cierto sentido, mientras $P_k \cup P_l, k \neq l$ deja de serlo, y P_k es adyacente a P_l .

Idealmente, aunque no necesariamente, cada P_k se debe corresponder con una región o estructura anatómica de interés en el caso de la segmentación de imágenes médicas.

Si se permite que las regiones no sean disjuntas (condición 2), es decir, que se solapen, estaríamos en presencia de lo que se conoce como segmentación flexible (*soft segmentation*). Este tipo de segmentación puede ser importante en el caso de las imágenes médicas como vía para tratar con las ambigüedades presentes en las imágenes.

La noción de segmentación flexible aparece a partir de una generalización del concepto de función característica de un conjunto la cual indica la pertenencia, o no, de un elemento a un conjunto. Si las regiones obtenidas no son disjuntas, es decir forman un recubrimiento, entonces un píxel de la imagen puede pertenecer a más de una región. Las funciones características pueden ser generalizadas mediante funciones de pertenencia que no necesariamente deben tomar los valores 0 ó 1. Las funciones de pertenencia pueden ser derivadas mediante técnicas difusas [5], o mediante técnicas estadísticas. En ambos casos la pertenencia de un elemento a un conjunto se especifica mediante un valor entre 0 y 1. Las funciones de pertenencia difusas pudieran ser sustituidas por densidades de probabilidad.

La tercera condición impone una restricción de contigüidad a las regiones que se obtengan. Esto no carece de sentido, aunque puede suceder que un mismo órgano o tejido aparezca en varios lugares no contiguos de la imagen, lo cual debe ser tenido en cuenta en etapas posteriores a la segmentación.

Por último, la cuarta condición está relacionada con el tipo de propiedades que deben poseer las regiones, por ejemplo, la uniformidad de los niveles de grises, y expresa además la maximalidad de las regiones.

En la definición anterior no se incluye ningún criterio de calidad de la partición obtenida y se acepta cualquiera que satisfaga las condiciones especificadas en la definición. Si a cada

miembro P_k de la partición P se le asocia cierto costo $J(P_k)$, el costo total de la partición $J_T(P)$ pudiera ser definido de manera aditiva como

$$J_T(P) = \sum_{k=1}^K J(P_k).$$

El problema de la segmentación puede ser entonces planteado como un problema de optimización,

$$J_T(P^*) = \min_{P \in \Psi_K} J_T(P),$$

donde Ψ_K representa el conjunto de todas las particiones factibles de cardinalidad K , y P^* la partición óptima en el sentido de menor costo, es decir,

$$P^* = \arg \min_{P \in \Psi_K} J_T(P)$$

La solución de este problema está asociada a altos costos computacionales y normalmente se recurre a métodos heurísticos como alternativa viable.

2 Métodos de agrupamiento (*clustering*)

La clasificación no supervisada, clasificación automática, o análisis de conglomerados (*cluster analysis*) ha sido utilizada en diversos contextos y disciplinas. Dentro de las principales aplicaciones se encuentra la segmentación de imágenes [6]. Aunque por el peso dentro del trabajo se le dedica a este tópico un capítulo aparte, este epígrafe se hace necesario para introducir los conceptos y enfoques clásicos.

El principal propósito de la clasificación no supervisada consiste formar grupos de objetos tomando como base la semejanza. Se desea descubrir clases “significativas” (grupos, conglomerados, *clusters*); pero no se posee información sobre las características de estas clases, o incluso cuál es su cantidad. Por tanto, en este caso las cuestiones de qué y cómo hacerlo resultan de interés. Esto apunta al cuidado que se debe tener en cuanto a la interpretación de los resultados que se obtienen a luz de las suposiciones hechas y de las limitaciones de las técnicas que se utilicen [7].

Evidentemente la formulación dada antes acerca del agrupamiento es demasiado general y abstracta como para formalizarla en un marco matemático apropiado. Han existido varios intentos de formalizar el problema del agrupamiento mediante la introducción de sistemas axiomáticos válidos para cualquier algoritmo de agrupamiento [8-11]. La idea principal de estos axiomas consiste en restringir la clase de algoritmos de agrupamiento solamente a aquellos que produzcan soluciones consistentes con el sistema axiomático dado.

Por lo general los axiomas resultan bastante intuitivos incluyendo, por ejemplo, la independencia respecto al reordenamiento de los objetos. Se ha investigado un sistema de tres axiomas que incluye: (1) la invariabilidad respecto a la escala; (2) la riqueza (toda partición debe ser, en principio, posible); y (3) la consistencia (la contracción de las distancias entre pares de puntos que forman parte de un mismo grupo, y el aumento de las distancias entre puntos de grupos diferentes, no debe afectar el agrupamiento). Sin embargo, a pesar del carácter intuitivo de estos axiomas, se demuestra que el sistema en su conjunto es inconsistente, es decir, no existe ninguna función de agrupamiento que lo satisfaga a todos de manera simultánea [12].

El principal señalamiento que se le puede hacer a estos enfoques axiomáticos está asociado al hecho de que la selección de los axiomas resulta un poco arbitraria y carente de justificación. Estos elementos confirman la dificultad de formalizar el problema del análisis de conglomerados o agrupamiento.

Una conclusión a la que se puede arribar es que, en ausencia de información o supuestos adicionales, el análisis de conglomerados es por naturaleza una técnica experimental [13]. Esto conduce al hecho de que, en la práctica, el agrupamiento se realice bajo diferentes

enfoques, con variedad de objetivos diferentes y en muchas ocasiones con métodos más o menos heurísticos [14].

Como se ha comentado antes, tanto la discontinuidad como la similitud entre píxeles se utilizan para formar grupos de elementos semejantes en cierto sentido. Es por esta razón que usualmente la segmentación de imágenes se asocia con la formación de conglomerados (*cluster analysis*), conocida también como clasificación no supervisada o clasificación automática.

El problema de la segmentación de imágenes se puede ver como la obtención de una partición tal que los píxeles pertenecientes a una región sean semejantes entre sí y que, por otra parte, se diferencien de aquellos pertenecientes a otras regiones. Normalmente también se impone la restricción de que las regiones sean conexas, es decir, que estén formadas por píxeles vecinos, y que posean cierta homogeneidad de la intensidad [15].

Otra opinión sobre la segmentación vista como un problema de formación de conglomerados se manifiesta en [3]. Según los autores la principal diferencia entre la segmentación y la formación de conglomerados consiste en que la primera se realiza en el dominio espacial de la imagen, mientras que el segundo se lleva a cabo en el espacio de las mediciones, es decir, el espacio de los rasgos o características, aunque se debe señalar que esta dicotomía es relativa toda vez que estos elementos pueden combinarse en una forma de representación que involucre a ambos.

Comentando esta relación se llega incluso a plantear que, de la misma manera que no existe una teoría del agrupamiento o *clustering*, no existe tampoco una teoría de la segmentación de imágenes [3], y que la mayoría de las propuestas en el campo de la segmentación de imágenes son técnicas *ad hoc* por naturaleza [16]. Sin embargo, en esta tesis se hace un esfuerzo de estructuración de los métodos desde el punto de vista del análisis de conglomerados.

Aunque el análisis de los métodos para lograr estos objetivos es esencial, no menos importante resulta el tema de la representación de los datos. Una representación adecuada de los objetos del espacio de entrada debe estar basada en ciertos datos. El problema de la representación de los datos resulta un tema medular en el reconocimiento de patrones [17].

2.1 Representación de los datos

Cualquier tipo de representación de los objetos constituye, de hecho, una simplificación. A pesar de ello es de esperar que se retengan los rasgos más esenciales que permitan llevar a cabo la división del espacio de entrada en regiones de una manera satisfactoria. Un estudio sistemático sobre el problema de la representación no es un cuestión simple, y es muy dependiente del dominio de aplicación [18].

Varios tipos de representación pueden ser utilizados. Así se puede citar la representación basada en rasgos, basada en píxeles u otras muestras, los modelos de probabilidad, los modelos estructurales, la representación basada en desemejanzas, basada en semejanzas, y la representación conceptual [19].

Las representaciones más comúnmente utilizadas son la representación vectorial y la representación basada en datos de proximidad a pares (semejanzas o desemejanzas) [20]. En el primer caso un conjunto $\Omega = \{\omega_1, \omega_2, \dots, \omega_n\}$ de n objetos es representado como n puntos en un espacio vectorial p -dimensional de rasgos, mientras en el segundo se tiene una matriz de proximidades de dimensión $n \times n$.

2.1.1 Representación vectorial

Tradicionalmente los objetos de interés se han representado mediante vectores en cierto espacio de rasgos. Este tipo de representación se basa en características, codificadas numéricamente, las cuales en su conjunto forman un espacio vectorial, usualmente euclidiano, en el cual cada objeto es representado mediante un vector. A este enfoque se le denomina geométrico o estadístico.

	X_1		X_l		X_p
ω_1					
ω_i	x_{i1}		x_{il}		x_{ip}
ω_n					

Figura 2-1 Representación vectorial.

En la Figura 2-1 se muestra esquemáticamente esta representación. Cada uno de los n objetos es representado mediante un vector de características o rasgos x_i , $i = 1, \dots, n$, de dimensión p , igual a la cantidad de rasgos o características X_l , $l = 1, \dots, p$, es decir, de manera compacta $x_i = (x_{i1}, x_{i2}, \dots, x_{ip})$ para cada objeto $\omega_i \in \Omega$, $i = 1, \dots, n$.

En el caso de las imágenes médicas, los objetos son los píxeles o vóxeles de la imagen, mientras el espacio de rasgos se toma, por lo general, como el espacio de los valores medidos. Estas mediciones realizadas en un espacio discreto bidimensional (píxeles), o tridimensional (vóxeles), denominadas intensidades de la imagen, pueden representar, por ejemplo, absorción de radiación como en el caso de las imágenes de rayos X, o la amplitud de las señales de radiofrecuencia en las imágenes de resonancia magnética. También es

posible que se realice más de una medición en cada localización, o simplemente que se utilicen varias imágenes de la misma región del cuerpo sometidas a un proceso previo de registro (*registration*) [21]. Por cualquiera de estas vías es posible asociar a cada localización un vector de características o rasgos [22].

Es posible utilizar otro tipo de información asociada a cada elemento de la imagen. Aquí se pueden incluir diferentes medidas de texturas [2] u otras magnitudes calculadas a partir de los valores de intensidad [23].

Se debe señalar que no existe una recomendación universal que indique cuáles rasgos o características son los más apropiados para utilizar en un contexto dado. Una tarea importante en este sentido consiste en la recolección de hechos y conjeturas sobre los datos así como la selección y extracción de características. Las técnicas de selección de características escogen aquellos atributos o rasgos más descriptivos y discriminatorios que proporcionen los mejores resultados al aplicar un método dado. Por su parte la extracción de características consiste en la aplicación de transformaciones a los datos originales con el fin de obtener otras nuevas que pudieran ser más prometedoras [6].

La principal objeción que se le puede hacer a este tipo de representación consiste en que la posible estructura de los objetos se pierde. Por ejemplo, las relaciones de vecindad entre píxeles de una imagen no son tenidas en cuenta. Este problema esencial con el uso de representaciones vectoriales ha sido señalado en [24], donde se propone la utilización de representaciones estructurales en las cuales se preserve la organización original de los objetos.

Un elemento a favor del tipo de representación vectorial es la disponibilidad de una gran cantidad de métodos para enfrentar el problema de la clasificación no supervisada [6, 25].

2.1.2 Representación basada en proximidades

Una representación adecuada de los objetos puede ser diseñada si se incorpora el conocimiento disponible del dominio de aplicación. Este es el caso de las relaciones de proximidad las cuales se pueden modelar mediante funciones de semejanza o de desemejanzas (distancias) [20]. Si se tiene en cuenta el contexto, es posible enriquecer la representación y tener en cuenta las relaciones estructurales presentes entre los objetos.

Esta forma de representación lleva a una matriz de dimensión $n \times n$ donde cada elemento está asociado a una cierta medida de proximidad ρ_{ij} entre dos objetos ω_i y ω_j (ver Figura 2-2).

Una imagen es una colección de elementos (píxeles o vóxeles) ordenada espacialmente. Al utilizar una representación basada puramente en rasgos resulta problemático preservar la conectividad espacial entre píxeles vecinos. Estas relaciones espaciales pudieran ser

derivadas quizás a partir de las ciertas transformaciones de los rasgos; pero no estarían incluidas en la representación de manera explícita. Por ejemplo, las coordenadas de los píxeles pudieran incorporarse en el vector de rasgos y a partir de ellas y otros rasgos definir alguna medida de proximidad (semejanza).

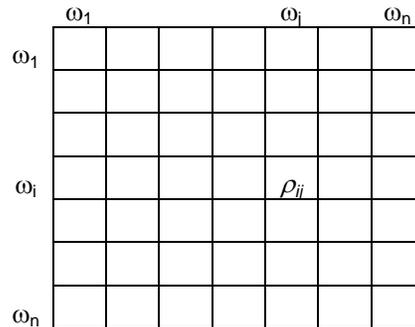


Figura 2-2 Representación basada en proximidades.

En la segmentación de imágenes las características numéricas aparecen antes de tomar en cuenta el concepto de proximidad y están esencialmente ligadas a los valores de intensidad asociados a los elementos de la imagen (por ejemplo, píxeles). Sin embargo, la distribución espacial impone ciertas restricciones toda vez que un elemento presente en una imagen se aproxima, usualmente, mediante un conjunto de píxeles vecinos. En una representación basada únicamente en características no resulta evidente como incluir esta información estructural. Lo mismo ocurre en el caso que se desee incorporar cierto conocimiento adicional acerca del dominio.

Una vía de solución a esta insuficiencia consiste en derivar medidas de proximidad que permitan tener en cuenta toda la información disponible. Esto sugiere la idea de utilizar la noción de proximidad como concepto primario asociado a la representación de los objetos.

Las representaciones de proximidad pueden dividirse en representaciones relativas y representaciones conceptuales [20]. En la representación relativa se comparan objetos relacionados y se calcula alguna medida de proximidad. Las medidas de proximidad pueden representar distancias entre vectores de rasgos en el caso en que los objetos tengan asociados vectores de rasgos. Sin embargo, otras representaciones pueden ser posibles: grafos, árboles, cadenas, u otros modelos estructurales [20, 26]. Por ejemplo, a un píxel se le puede asociar un histograma de frecuencias de los valores de intensidad en una cierta vecindad y la medición de proximidad ρ_{ij} entre dos píxeles cualesquiera estaría asociada a cierta semejanza entre dichos histogramas [27].

La representación conceptual, por su parte, extiende el análisis a la relación entre un objeto y otra cierta cantidad de ellos; o entre un modelo y un concepto; o entre dos conceptos. Tal

puede ser el caso de una consulta donde se tiene una imagen y se desean recuperar imágenes similares [28].

Se debe tener en cuenta que no existe una medida de proximidad (semejanza o desemejanza) universal que sea aplicable en todas las situaciones. Cualquier comparación entre objetos se realiza tomando como base un determinado contexto, o cierto punto de vista, y teniendo en cuenta determinadas características, las especificidades de los dominios, o cualquier información adicional disponible.

Generalmente en calidad de medida de proximidad se toma una medida d que refleje la noción de cercanía entre los objetos. Lo más común es utilizar una métrica o función de distancia.

Cualquier función de desemejanzas es una función $d : \Omega \times \Omega \rightarrow \mathbb{R}_+^0$, donde \mathbb{R}_+^0 representa el conjunto de los reales positivos con el cero incluido, y Ω es un conjunto de objetos. A continuación se presentan algunas propiedades que puede cumplir una función de desemejanzas

1. $d(\omega, \omega) = 0$
2. $d(\omega_i, \omega_j) \geq 0$ (no negatividad)
3. $d(\omega_i, \omega_j) = d(\omega_j, \omega_i)$ (Simetría)
4. $d(\omega_i, \omega_j) = 0 \Rightarrow \omega_i = \omega_j$
5. $d(\omega_i, \omega_j) \leq d(\omega_i, \omega_k) + d(\omega_k, \omega_j)$ (Desigualdad triangular)

Las propiedades (1) y (2) son muy intuitivas y representan la exigencia mínima para una función de desemejanzas.

La propiedad (3) aunque es común que se cumpla puede que no sea satisfecha por una función dada.

Por su parte la propiedad (4) se refiere a la invariabilidad dentro del espacio de los objetos. Diferentes objetos cuyas distancias entre sí sean cero no pueden ser discriminados y son tratados como miembros de una misma clase. Por lo tanto se debe tener cuidado con las funciones de desemejanzas que no satisfagan esta propiedad.

Las funciones que cumplan todas las cinco propiedades se denominan métricas o distancias. Si solamente se satisfacen las propiedades (1) – (3) y la (5) entonces se denominan semi-métricas.

En otros casos se le pueden exigir otras condiciones adicionales a la función de distancia. Así por ejemplo, si se impone la condición adicional de que

$$6. \quad \forall \omega_i, \omega_j, \omega_k \in \Omega, d(\omega_i, \omega_j) \leq \max \{d(\omega_i, \omega_k), d(\omega_k, \omega_j)\}$$

entonces estaríamos en presencia de una distancia ultramétrica, o simplemente de una ultramétrica, que encuentra su utilización en los métodos de agrupamiento jerárquicos que se comentan más adelante.

Algunos ejemplos de funciones de distancias para objetos con una representación vectorial son las siguientes.

Minkowski:
$$d_q(x_i, x_j) = \left(\sum_{k=1}^p |x_{ik} - x_{jk}|^q \right)^{\frac{1}{q}}, \quad q \geq 1.$$

Manhattan (*City-block*):
$$d_1(x_i, x_j) = \sum_{k=1}^p |x_{ik} - x_{jk}|$$

Euclidiana:
$$d_2(x_i, x_j) = d_E(x_i, x_j) = \sqrt{\sum_{k=1}^p (x_{ik} - x_{jk})^2} = \sqrt{(x_i - x_j)(x_i - x_j)^T}$$

Chebyshev:
$$d_\infty(x_i, x_j) = d_{\max}(x_i, x_j) = \max_{1 \leq k \leq p} |x_{ik} - x_{jk}|$$

Mahalanobis:
$$d_M(x_i, x_j) = \sqrt{(x_i - x_j) \Sigma^{-1} (x_i - x_j)^T}$$

donde $\Sigma = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x}_i)^T (x_i - \bar{x}_i)$ es la matriz de covarianzas y Σ^{-1} su inversa. Aquí se asume un conjunto de n vectores filas en un espacio de dimensión p .

En ocasiones se utilizan funciones o medidas de semejanzas o afinidades las cuales son de cierta manera contrapuestas a la noción de distancia o desemejanza en el sentido de que la semejanza entre dos objetos debe aumentar si la desemejanza disminuye. Particularmente la semejanza o afinidad debe aumentar a medida que los objetos sean más similares.

Esto hace posible utilizar ciertas transformaciones para pasar de desemejanzas a semejanzas y viceversa. Así si $d(x, y)$ es una función de desemejanzas, entonces cualquier función de $d(x, y)$ decreciente y no negativa se puede tomar como una función de semejanzas, por ejemplo, $s(x, y) = \exp(-d^2(x, y)/t)$, donde t sería un cierto parámetro a definir. De la misma manera si se tiene una función de semejanzas normalizada, es decir $0 \leq s(x, y) \leq 1$, entonces $d(x, y) = 1 - s(x, y)$ es una función de distancias [29].

La disponibilidad de diferentes medidas de proximidad, o de diferente conocimiento de expertos, sugiere la idea de su combinación [30]. La combinación de diferentes medidas

hace que la representación sea más universal y flexible, permitiendo tomar en cuenta diversos aspectos del problema que se analiza. Esta idea de combinar medidas de semejanzas está relacionada con la idea de combinar diferentes clasificadores [20].

Un ejemplo de medida de semejanza utilizada en aplicaciones de segmentación de imágenes es la siguiente [31].

$$s(i, j) = \begin{cases} e^{-\frac{|x_i - x_j|}{\sigma_1}} \cdot e^{-\frac{d(i, j)}{\sigma_2}}, & d(i, j) < r \\ 0 & , \quad d(i, j) \geq r \end{cases}$$

donde x_i y x_j representan los valores de intensidad de los píxeles i y j , respectivamente, $d(i, j)$ es la distancia euclidiana entre los píxeles i y j , r es un radio que delimita la vecindad que se analiza, y σ_1 , σ_2 son parámetros de control. En este caso la semejanza está asociada tanto a la cercanía espacial como a la de los valores de intensidad.

Una de las alternativas de representación de los datos de proximidad lo constituyen los grafos. Cada vértice del grafo representa un objeto, y cada arista se pondera con el valor de la medida de proximidad entre los objetos dados. Si la relación de proximidad no fuera simétrica se obtendría un grafo dirigido. La clasificación en este caso sería equivalente a un problema de partición del grafo. La utilización de grafos en este contexto permite hacer uso de todo el aparato teórico y algorítmico disponible en este campo.

Finalmente se puede plantear que en el problema de la segmentación de imágenes, y de las imágenes médicas en particular, es necesario incorporar toda la información disponible desde el momento mismo de la representación de los datos. Esto incluye también la cuestión de cómo medir la semejanza entre los elementos de la imagen, lo cual está muy asociado a la forma de representación, así como a los métodos aplicables en cada caso.

Una análisis más detallado sobre diferentes medidas de desemejanzas para diferentes tipos de representación de los datos puede ser consultado en [20].

2.2 Enfoques clásicos

Dos estrategias básicas se han desarrollado para el problema de formación de conglomerados: los métodos particionales y los métodos jerárquicos, cada uno de los cuales incluyen una gran variedad de algoritmos como vía de implementación de estas estrategias.

A diferencia del agrupamiento jerárquico, donde se obtiene un dendrograma de grupos o clusters anidados, el agrupamiento particional genera una única partición de los datos. Por otra parte, mientras el agrupamiento jerárquico exige, por lo general, solamente una matriz de proximidades (distancias o semejanzas), el agrupamiento particional opera usualmente

en un espacio de características, normalmente euclidiano. A continuación se presenta una revisión de las principales estrategias y métodos. Más detalles se pueden consultar, por ejemplo, en [6, 13, 15].

2.2.1 Agrupamiento particional

Formalmente el agrupamiento particional se puede plantear de la siguiente manera. Sean n objetos dados en un espacio métrico p -dimensional. El problema consiste en agrupar los n objetos en K clusters teniendo en cuenta un cierto criterio que favorezca el agrupamiento de objetos semejantes en un mismo cluster. El valor de K puede, o no, estar dado de antemano, aunque usualmente si lo está.

Es posible examinar esta problemática desde una perspectiva de copertenencia a una clase o adyacencia. Supongamos que se tiene una matriz $A = [a_{ij}]$, donde $a_{ij} = 1$ solamente en el caso en que los objetos i y j estén en la misma clase y $a_{ij} = 0$ si no. Si $K=1$ la matriz A está repleta de 1's ($A^{(1)}$), y si $K=n$ la matriz A es la matriz identidad. En general, $1 \leq K \leq n$, la matriz A representa la relación binaria de equivalencia asociada a una partición de K clases.

La cuestión de obtener una partición de los n objetos en K clases se puede interpretar como la descomposición de la matriz original $A^{(1)}$ cuando $K=1$ en otras dos matrices: $A^{(1)} = A^{(K)} + B^{(K)}$, donde $A^{(K)}$ describe las relaciones entre los objetos que se agrupan en cada clase, y $B^{(K)}$ señala las relaciones entre objetos de diferentes clases. En la Figura 2-3 se muestra un ejemplo donde se utilizan 3 clases y se asume alguna permutación de las filas y columnas de la matriz de conexiones original de manera que cada clase es descrita por un bloque de la diagonal principal de la primera matriz de la descomposición.

$$\begin{array}{|c|c|c|c|c|c|c|} \hline 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ \hline 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ \hline 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ \hline 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ \hline 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ \hline 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ \hline 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ \hline \end{array} = \begin{array}{|c|c|c|c|} \hline 1 & 1 & & \\ \hline 1 & 1 & & \\ \hline & 1 & 1 & \\ \hline & 1 & 1 & \\ \hline & & & 1 & 1 & 1 \\ \hline & & & 1 & 1 & 1 \\ \hline & & & 1 & 1 & 1 \\ \hline \end{array} + \begin{array}{|c|c|c|c|c|c|} \hline & & 1 & 1 & 1 & 1 & 1 \\ \hline & & 1 & 1 & 1 & 1 & 1 \\ \hline 1 & 1 & & & 1 & 1 & 1 \\ \hline 1 & 1 & & & 1 & 1 & 1 \\ \hline 1 & 1 & 1 & 1 & & & \\ \hline 1 & 1 & 1 & 1 & & & \\ \hline 1 & 1 & 1 & 1 & & & \\ \hline \end{array}$$

Figura 2-3 Descomposición de una matriz de conexiones.

Si se define alguna medida de desemejanza entre los objetos es posible entonces plantear la siguiente propiedad fundamental que relaciona la suma total de las desemejanzas y las desemejanzas internas (entre objetos de una misma clase, *Within*) y las cruzadas (entre objetos de diferentes clases, *Between*).

Sea una partición $P = \{P_1, P_2, \dots, P_K\}$ de tamaño K del conjunto de n objetos $i = 1, \dots, n$, y $d(\omega_i, \omega_j)$ una función de desemejanzas definida para cualquier de objetos ω_i y ω_j . Se tiene entonces que

$$\sum_{i=1}^n \sum_{j=1}^n d(\omega_i, \omega_j) = \sum_{k=1}^K \sum_{\omega_i \in P_k} \sum_{\omega_j \in P_k} d(\omega_i, \omega_j) + \sum_{k=1}^K \sum_{\substack{l=1 \\ l \neq k}}^K \sum_{\omega_i \in P_k} \sum_{\omega_j \in P_l} d(\omega_i, \omega_j)$$

De forma abreviada esta descomposición se puede expresar

$$S_T = S_W + S_B$$

Para un conjunto dado de objetos, el término de la izquierda es constante. A medida que el valor de K cambie así será la distribución de las sumas internas y cruzadas. En el caso de que todos los objetos formen una única clase ($K=1$) la suma cruzada será cero. De la misma manera si $K = n$, es decir, se tiene la partición más fina donde cada objeto es considerado una clase, y en este caso la suma interna sería igual a cero toda vez que $d(\omega_i, \omega_i) = 0, \forall i = 1, \dots, n$.

La suma interna permite caracterizar en cierto sentido cada una de las clases, es decir, a medida que la suma interna de una clase sea más pequeña, más compacta será dicha clase. Por su parte la suma cruzada nos brinda una idea de la diferencia entre las clases: mientras mayor sea la suma cruzada, más diferentes serán las clases. Es por ello que esta descomposición se utiliza como base para la formación de criterios de evaluación de las particiones.

La matriz $A = [a_{ij}]$ puede ser utilizada también como una vía para especificar ciertas restricciones asociadas en los datos, como puede ser la estructura espacial que caracteriza a una imagen digital. La expresión $S_T = S_W + S_B$ quedaría de la siguiente forma

$$\sum_{i=1}^n \sum_{j=1}^n a_{ij} d(\omega_i, \omega_j) = \sum_{k=1}^K \sum_{\omega_i \in P_k} \sum_{\omega_j \in P_k} a_{ij} d(\omega_i, \omega_j) + \sum_{k=1}^K \sum_{\substack{l=1 \\ l \neq k}}^K \sum_{\omega_i \in P_k} \sum_{\omega_j \in P_l} a_{ij} d(\omega_i, \omega_j)$$

Las utilidades de esta notación alternativa son estudiadas más adelante.

A continuación se presenta un análisis sobre los principales métodos de agrupamiento particional.

Los criterios que se utilizan para realizar las particiones pueden ser clasificados como globales o locales. Un criterio se dice que es global cuando cada cluster es representado

mediante uno o varios prototipos y la asignación de los objetos se realiza de acuerdo a la mayor semejanza con dichos prototipos. Los criterios locales utilizan la estructura local de los datos para efectuar el agrupamiento, por ejemplo, mediante la asignación de un objeto y sus vecinos más próximos al mismo cluster, o mediante la identificación de regiones de alta densidad en los datos.

De esta manera, si se denota por $J(P)$ el criterio del agrupamiento, donde P representa una partición de los objetos, el problema puede ser planteado como un problema de optimización discreta,

$$P^* = \arg \min_{P \in \Pi_k} J(P)$$

donde Π_k representa el conjunto de todas las particiones de cardinalidad k del conjunto dado de objetos.

Este planteamiento se ajusta a la definición de un problema de búsqueda. Según [32] un problema de búsqueda posee un conjunto de estados del mundo, un conjunto de operadores que cambian el estado, un estado inicial y un conjunto de estados metas, y la tarea consiste en hallar una secuencia de operadores que transformen el estado inicial en uno de los estados meta.

La solución ingenua de este problema, consiste en evaluar todas las posibles particiones de tamaño K de los n objetos y escoger aquella que optimice el criterio seleccionado. Sin embargo, este enfoque de búsqueda exhaustiva es impracticable, excepto para conjuntos de objetos muy pequeños.

La cuestión radica en que la cantidad de particiones de tamaño K de n objetos crece muy rápido en función de la cantidad de objetos y del tamaño de la partición. Para la cantidad de particiones de un tamaño K de n objetos, sin tener en cuenta el orden de los objetos dentro de cada grupo, ni el orden de los grupos dentro de la partición se verifica la ecuación en diferencias

$$S(n, K) = S(n-1, K-1) + K * S(n-1, K)$$

con las siguientes condiciones de frontera

$$S(n, 1) = 1, S(n, n) = 1, S(n, K) = 0 \text{ si } K > n.$$

Las soluciones de esta ecuación se denominan números de Stirling de segunda clase y se pueden calcular según la siguiente expresión [33]

$$S(n, K) = \frac{1}{K!} \sum_{i=1}^K (-1)^{K-i} \binom{K}{i} i^n$$

En la Figura 2-4 se ilustra este comportamiento para diferentes valores de n y de K . Así, por ejemplo, para 19 objetos la cantidad de particiones de cardinalidad 6 asciende a 6.931×10^{11} .

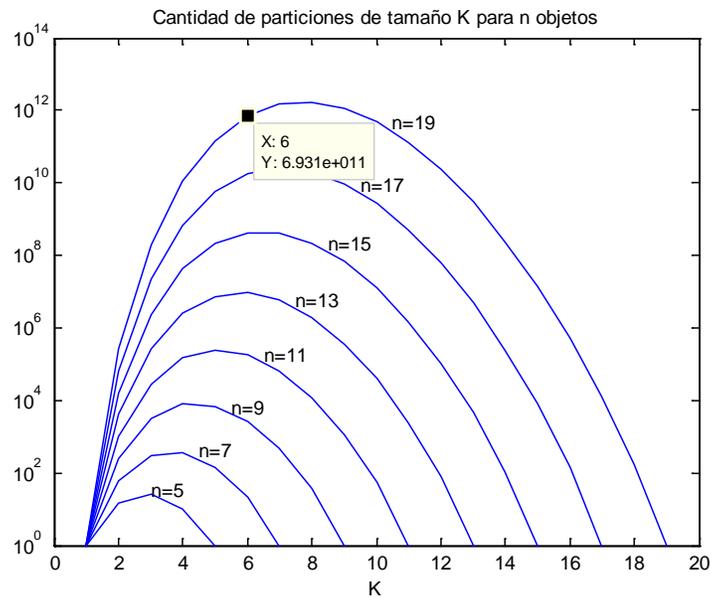


Figura 2-4 Comportamiento del espacio de búsqueda de las particiones

Otra problemática importante que surge con este enfoque está relacionada con la selección de un criterio que se ajuste a la noción intuitiva de cluster de manera que la partición que optimice el criterio sea la “mejor”. Evidentemente al no existir una definición precisa del concepto de cluster, no se puede hablar de un criterio universal para la obtención del mejor agrupamiento. Los clusters se pueden diferenciar por su tamaño, por su forma, incluso por la manera en que se define la semejanza entre sus miembros. Cada criterio asume o impone cierta estructura de los datos por lo que si estos supuestos no se cumplen los resultados a los que se llegan pueden no ser satisfactorios.

En la literatura se encuentran multitud de criterios para evaluar la calidad de un agrupamiento [34]. El más común es aquel que utiliza los centroides como prototipos de los clusters.

Sea una partición $P = \{P_1, P_2, \dots, P_K\}$ de un conjunto de n objetos ω_i , $i = 1, \dots, n$, cada uno representado por un vector x_i , $\bar{x}_k = \frac{1}{n_k} \sum_{i \in P_k} x_i$ el centroide del cluster P_k , y n_k su cardinalidad,

asumiendo que en el espacio de definición de los objetos sea euclidiano y tenga sentido el cálculo de la media.

La variación interna (*within cluster variation*) de un cluster P_k , denotada por $J_W(P_k)$ se puede definir de la siguiente manera

$$J_W(P_k) = \sum_{i \in P_k} d_E^2(x_i, \bar{x}_k)$$

donde $d_E^2(.,.)$ representa la distancia euclidiana cuadrática. También es posible utilizar la distancia de Mahalanobis $d_M^2(.,.)$ definida con anterioridad.

El valor $J_W(P_k)$ se conoce también como error cuadrático (*square-error*). El error cuadrático de la partición en su conjunto es la suma de los errores cuadráticos de cada uno de los clusters, es decir,

$$J_W(P) = \sum_{k=1}^K J_W(P_k) = \sum_{k=1}^K \sum_{i \in P_k} d_E^2(x_i, \bar{x}_k) = \sum_{k=1}^K n_k s_k^2$$

donde

$$s_k^2 = \frac{1}{n_k} \sum_{i \in P_k} d_E^2(x_i, \bar{x}_k)$$

representa la varianza del cluster P_k .

Es por esta razón que los agrupamientos definidos de esta manera se denominan también agrupamientos de mínima varianza. Este tipo de criterio es útil cuando los clusters poseen formas esféricas y se encuentran bien separados unos de otros.

Adicionalmente, se puede demostrar que

$$\sum_{i \in P_k} d_E^2(x_i, \bar{x}_k) = \frac{1}{2n_k} \sum_{i \in P_k} \sum_{j \in P_k} d_E^2(x_i, x_j)$$

Luego puede escribirse lo siguiente

$$J_W(P) = \sum_{k=1}^K J_W(P_k) = \sum_{k=1}^K \sum_{i \in P_k} d_E^2(x_i, \bar{x}_k) = \frac{1}{2} \sum_{k=1}^K \frac{1}{n_k} \sum_{i \in P_k} \sum_{j \in P_k} d_E^2(x_i, x_j)$$

Esta última expresión sugiere la posibilidad de introducir otros tipos de criterios [10, 25, 35]. Es decir, si se asume una función de desemejanzas arbitraria $d(\cdot, \cdot)$, entonces se pudiera escribir

$$J_W(P) = \sum_{k=1}^K J_W(P_k) = \frac{1}{2} \sum_{k=1}^K \frac{1}{n_k} \sum_{i, j \in P_k} d(x_i, x_j)$$

Otro criterio que se puede asumir es el de la variación entre clusters (*between cluster variation*) el cual se puede expresar mediante la distancia entre los centroides respectivos.

$$J_B(P) = \sum_{k=1}^K \sum_{\substack{l=1 \\ l \neq k}}^K d_E^2(\bar{x}_k, \bar{x}_l)$$

Al hacer la selección de un criterio o el otro se puede tener en cuenta su costo computacional. El cálculo de la variación interna, es decir, $J_W(P)$ requiere $O\left(\sum_{k=1}^K n_k\right) = O(n)$ operaciones. Por su parte la variación entre clusters, $J_B(P)$, puede ser calculada en $O(K^2)$ operaciones.

Otros criterios en espacios euclidianos pueden ser derivados a partir de la matriz de covarianzas. Para los objetos de un cluster P_k dado es posible definir la matriz de covarianzas interna

$$S_k = \sum_{i \in P_k} (x_i - \bar{x}_k)^T (x_i - \bar{x}_k)$$

que no es más que la matriz de covarianzas sin el factor $1/n_k$, de los objetos del cluster P_k . La matriz de covarianzas interna total sin el factor $1/n$ se define como la suma de las matrices particulares de cada cluster, es decir,

$$S_W = \sum_{k=1}^K S_k.$$

De la definición anterior se tiene que el criterio de la variación interna puede representarse como la suma de las trazas de las matrices S_k , es decir,

$$J_W(P) = \sum_{k=1}^K \text{tr}(S_k)$$

donde además se cumple que $tr(S_k) = J_w(P_k)$.

También es posible definir la matriz de covarianzas externa sin el factor $1/n$ de la siguiente manera

$$S_B = \sum_{k=1}^K n_k (\bar{x}_k - \bar{x})^T (\bar{x}_k - \bar{x}), \text{ donde } \bar{x} = \frac{1}{n} \sum_{k=1}^K n_k \bar{x}_k$$

La matriz de covarianzas total sin el factor $1/n$ se define como

$$S_T = \sum_{i=1}^n (x_i - \bar{x})^T (x_i - \bar{x})$$

y se puede demostrar [15] que

$$S_T = S_W + S_B$$

y que

$$tr(S_T) = tr(S_W) + tr(S_B).$$

Si se tiene en cuenta que la matriz de dispersión total no depende de un agrupamiento concreto, mientras que las matrices de dispersión interna y externa sí, se puede notar que minimizar la dispersión interna (la variación interna) es equivalente a maximizar la dispersión externa (la variación entre clusters).

Otras variantes de criterios a partir de las matrices de dispersión pueden ser consultados, por ejemplo, en [25, 34].

Una vez definido un criterio para el agrupamiento surge la problemática de la obtención de la partición óptima. Como ya se comentó antes, este problema es generalmente intratable desde el punto de vista computacional por lo que se recurre a métodos de búsqueda heurística. Estas son técnicas que brindan soluciones aceptables, pero que no garantizan llegar a la solución óptima (en ocasiones lo logran en términos de probabilidad).

Dentro de estas técnicas se pueden citar los métodos heurísticos iterativos de mejoras locales, o de optimización local, los cuales utilizan una heurística local glotona (*greedy heuristic*), y los métodos estocásticos de optimización, tales como los algoritmos genéticos [36] y el enfriamiento estadístico (*simulated annealing*) [37], entre otros.

Uno de los enfoques más utilizados al buscar la partición óptima es la optimización iterativa. La idea consiste en comenzar con una cierta partición razonable y realizar movimientos de los objetos de un cluster a otro siempre y cuando estos movimientos conduzcan a una partición que mejore el valor del criterio [38].

Este procedimiento, al igual que los escaladores de colinas (*hill-climbing*) sólo garantiza la obtención de un óptimo local y no uno global. Incluso, diferentes estados iniciales pueden conducir a diferentes soluciones. A pesar de estas limitaciones, la optimización iterativa es muy utilizada debido a que los costos computacionales son aceptables y mucho menores que en el caso de la búsqueda exhaustiva.

Otro algoritmo muy conocido de optimización iterativa lo constituye *k-means* [8, 25]. Este algoritmo asume que los n objetos están representados en un espacio euclidiano y se desea agruparlos en K clusters utilizando el criterio del error cuadrático para evaluar cada partición.

El algoritmo básico parte de un conjunto inicial de centroides los cuales sirven para obtener una partición inicial mediante la asignación de cada objeto al centroide más cercano, en el sentido de la distancia euclidiana. Una vez realizada la asignación se recalculan los centroides y se repite todo el proceso. El criterio de parada está dado por la ausencia de cambios en los centroides [39]. En la Tabla 2-1 se muestra este esquema básico.

Tabla 2-1 Algoritmo *k-means* básico

Algoritmo *k-means* básico

- 1. Seleccionar K puntos como centroides iniciales.**
- 2. Mientras los centros no cambien hacer**
- 3. Asignar cada objeto al centroide más cercano.**
- 4. Asegurarse que cada cluster contenga al menos un elemento.**
- 5. Reemplazar los centroides con las medias de cada uno de los clusters formados.**
- 6. Fin**

La complejidad computacional de este algoritmo es $O(KnIp)$, donde I representa la cantidad de iteraciones y p es la dimensión del espacio. En efecto, si se asumen los centroides conocidos entonces en una pasada por los datos es posible calcular todas las Kn distancias a los centroides, y para cada objeto seleccionar la menor de estas. El cálculo de los nuevos centroides también se puede realizar en $O(n)$ operaciones.

Se debe tener en cuenta que el algoritmo *k-means* puede conducir a clusters vacíos por lo que se deben tomar medidas apropiadas para garantizar la cantidad adecuada de estos, por ejemplo seleccionando un elemento del cluster de mayor error cuadrático. Por otra parte el algoritmo también es susceptible a la inicialización, es decir, diferentes conjuntos de centroides iniciales pueden conducir a particiones diferentes.

En ocasiones se utiliza un agrupamiento jerárquico (que será visto más adelante) para la determinación de los centroides iniciales. También se ensayan varias corridas y se selecciona la de menor error cuadrático. Un análisis de estas cuestiones puede ser consultado en [40].

Dentro de las diferentes variantes del algoritmo *k-means* se puede señalar el algoritmo ISODATA [41], el cual en un intento de encontrar la mejor partición permite la utilización de divisiones y fusiones de los clusters intermedios.

Cuando en calidad de representantes de los clusters se realiza otra selección se pueden obtener otras variantes. Así por ejemplo, se puede señalar el algoritmo *k-modes* que selecciona en calidad de centroide de un cluster la moda [42], o el *k-medois* que propone seleccionar en calidad de representante de un cluster a uno de sus miembros según un criterio dado. Ejemplos de estos últimos lo constituyen los algoritmos CLARA (*Clustering LARge Applications*) [43] y CLARANS (*Clustering Large Applications based upon RANdomized Search*) [44].

Dentro del agrupamiento particional se distingue también el enfoque probabilístico. En este enfoque se considera que los datos se han obtenido como muestras independientes de un modelo de mezclas de diferentes distribuciones de probabilidad [45]. La opción usual para la obtención de los estimadores de máxima verosimilitud los parámetros de la mezcla de distribuciones es el algoritmo EM (*Expectation Maximization*) [46]. Un análisis sobre diversos aspectos metodológicos de la utilización de los modelos de mezclas se puede consultar en [47].

2.2.2 Agrupamiento jerárquico

El agrupamiento jerárquico realiza de manera sucesiva una mezcla de grupos pequeños para formar otros más grandes; o una división de grupos grandes en otros más pequeños. Así se diferencian los algoritmos aglomerativos y los algoritmos divisivos.

El resultado final es una jerarquía, o árbol de grupos o conglomerados anidados; también denominado dendrograma. Si el árbol posee una medida o índice asociada a la fusión o la división de las clases, entonces una tala de este árbol a un nivel dado representa una partición (o *clustering*) de los objetos. Talas sucesivas según el nivel inducen una sucesión de particiones anidadas, desde la más fina (hojas del árbol o clases singulares) hasta la menos fina (la raíz del árbol que representa el conjunto completo).

La manera en que los clusters son mezclados (o divididos) depende de cómo esté definida la función de desemejanzas, que en este caso representa una función de desemejanzas conceptual. Los métodos divisivos se pueden basar en la construcción de grafos de

vecindad y la utilización de algún método para eliminar aristas en este grafo e ir creando los grupos o clases.

En el agrupamiento jerárquico aglomerativo a partir de un agrupamiento inicial se procede a unir aquellos clusters que estén más cercanos en cierto sentido disminuyendo así la cantidad de estos. Este proceso se repite hasta que se obtenga un único cluster formado por todos los objetos. Por lo general el punto inicial de este proceso lo constituye el caso donde cada objeto representa un cluster, aunque es posible que tenga sentido comenzar el proceso por una partición distinta a la más fina.

Sean n objetos $i = 1, \dots, n$ y una función $D(P_k, P_l)$ que mide la desemejanza entre dos cluster P_k y P_l cualesquiera. Un algoritmo aglomerativo básico puede ser descrito como se indica en la

Tabla 2-2. Esta propuesta posee una complejidad temporal y espacial $O(n^3)$.

Tabla 2-2 Algoritmo jerárquico aglomerativo básico

Algoritmo jerárquico aglomerativo básico

1. **Hacer** $P_i = \{i\}$, $i = 1, \dots, n$
2. **Mientras la cantidad de clusters sea mayor que 1 hacer**
3. **Sean** P_k y P_l **los clusters que minimizan la función** $D(P_k, P_l)$ **entre todos los clusters**
4. $P_k = P_k \cup P_l$
5. **Eliminar el cluster** P_l
6. **Actualizar las desemejanzas entre el cluster creado y el resto**
7. **Fin**

En este caso la función $D(P_k, P_l)$ puede ser utilizada en calidad de medida para caracterizar cada nivel del árbol que se forma.

Sean dos clusters P_k y P_l con cardinalidades n_k y n_l respectivamente y $d(x_i, x_j)$ una medida de desemejanza (o distancia) entre dos objetos cualesquiera. Algunos de los principales criterios utilizados en el agrupamiento jerárquico aglomerativo son los siguientes:

- Enlace simple (*single linkage*). En este caso la desemejanza entre dos clusters está determinada por los vecinos más próximos, es decir,

$$D(P_k, P_l) = \min_{i \in P_k} \min_{j \in P_l} d(x_i, x_j).$$

Esto enfatiza la conectividad en los clusters y resulta en clusters alargados.

- Enlace completo (*complete linkage*). Aquí la desemejanza entre clusters se define mediante los vecinos más lejanos, es decir,

$$D(P_k, P_l) = \max_{i \in P_k} \max_{j \in P_l} d(x_i, x_j)$$

Normalmente es útil cuando los objetos forman clusters separados ya que enfatiza la compacidad de estos. No resulta apropiado si los clusters presentan formas alargadas.

- Enlace basado en centroides (*centroid linkage*). Utiliza la distancia euclidiana entre los centroides de los clusters, es decir,

$$D(P_k, P_l) = d_E(\bar{x}_k, \bar{x}_l),$$

donde

$$\bar{x}_k = \frac{1}{n_k} \sum_{i \in P_k} x_i \quad \text{y} \quad \bar{x}_l = \frac{1}{n_l} \sum_{j \in P_l} x_j.$$

Se asume una representación vectorial de los datos. Este criterio puede producir dendrogramas no monótonos [15]. Esto ocurre cuando la distancia de un cluster obtenido por la unión de otros dos, r y s , a un tercero es menor que la distancia de cualquiera de los clusters componentes, r ó s a ese mismo tercero.

- Enlace basado en el criterio de Ward (*Ward's linkage*). Este criterio utiliza la suma de las distancias euclidianas cuadráticas de cada miembro de un cluster a su centroide. Este valor es conocido como el error cuadrático del cluster, o como inercia en la literatura francesa [48], y es utilizado también como criterio en el agrupamiento particional. El error cuadrático de una partición es la suma de los errores cuadráticos de cada uno de sus miembros. Al unir dos clusters cualquiera el error cuadrático de la partición aumenta.

El criterio de Ward propone unir en cada paso aquellos dos clusters para los cuales el incremento del error cuadrático, obtenido después de la unión, es menor. Para el caso de las distancias euclidianas se demuestra que el incremento en el valor del error cuadrático se puede expresar únicamente en función de los centroides. De esta manera el criterio queda finalmente así

$$D(P_k, P_l) = \frac{n_k n_l}{n_k + n_l} d_E^2(\bar{x}_k, \bar{x}_l),$$

donde \bar{x}_k y \bar{x}_l representan los centroides de los clusters P_k y P_l , respectivamente, tal y como en el caso del enlace basado en centroides.

- Enlace promedio (*average linkage*). La desemejanza utilizada es el promedio entre clusters, es decir,

$$D(P_k, P_l) = \frac{1}{n_k n_l} \sum_{i \in P_k} \sum_{j \in P_l} d(x_i, x_j)$$

Por lo general trabaja bien tanto en casos de clusters alargados como compactos. Tiende a producir clusters de varianza similar.

Cabe señalar que aunque en la mayoría de los criterios se exige únicamente una representación basada en desemejanzas, los criterios de Ward y el enlace basado en centroides presuponen una representación vectorial en un espacio euclidiano.

La selección de un criterio apropiado para el agrupamiento jerárquico es una cuestión esencial para el análisis de los datos. Sin embargo, existen muy pocas sugerencias teóricas para guiar esta selección.

Desde el punto de vista de implementación existe la fórmula general de Lance-Williams-Jambu [49, 50] que permite calcular las desemejanzas entre los nuevos clusters creados y el resto, en cualquier nivel de la jerarquía, a partir del agrupamiento actual.

Se debe destacar que en el agrupamiento aglomerativo se requieren tanto las desemejanzas entre los objetos (al inicio del proceso) así como entre clusters. Esta situación se puede considerar un aspecto favorable pues no impone la exigencia de que los objetos posean una representación vectorial (excepto en los criterios del centroide y de Ward mencionados antes); tan sólo se requiere que sea posible calcular la desemejanza entre ellos.

El agrupamiento jerárquico divisivo, por su parte, toma el camino inverso, es decir, se parte de un único cluster compuesto por todos los objetos y se procede a realizar divisiones para obtener sus componentes hasta que se obtenga una partición final donde cada cluster está compuesto por un único objeto. El principal problema que se presenta en este caso consiste en cómo restringir la búsqueda en el espacio de todas las posibles particiones con vistas a reducir su costo computacional.

Un ejemplo de algoritmo jerárquico divisivo es aquel que se basa en el árbol de expansión mínimo o MST (*Minimum Spanning Tree*). Este algoritmo comienza por la construcción del árbol MST. Para ello se parte de considerar cada objeto como miembro de un árbol. En los pasos subsiguientes se busca el par de objetos más cercanos entre sí con la condición de que uno de ellos, por ejemplo p , se encuentre en el árbol actual y el otro, q , no. Se procede entonces a agregar el objeto q al árbol actual y se coloca una arista entre ellos (ver Figura 2-5).

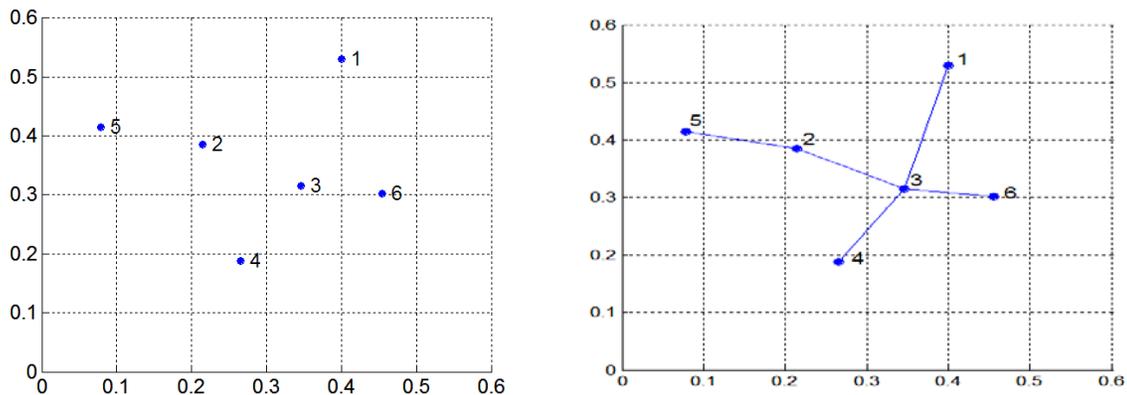


Figura 2-5 Árbol MST

Una vez construido al árbol MST el procedimiento de división procede a eliminar la arista correspondiente a la mayor distancia (o menor semejanza) entre dos objetos. De esa manera se va obteniendo una jerarquía. El algoritmo termina cuando cada objeto forme un cluster independiente (ver Tabla 2-3).

Tabla 2-3 Algoritmo jerárquico divisivo basado en el árbol MST

Algoritmo jerárquico divisivo MST

1. **Construir el árbol MST a partir de la matriz de proximidades.**
2. **Repetir**
3. **Crear un nuevo cluster eliminando la arista correspondiente a la mayor distancia (menor semejanza).**
4. **Hasta que cada objeto forme un cluster**

Un método híbrido consistente en la utilización del algoritmo *k-means* para la obtención de una jerarquía de clusters, o simplemente de una partición, es el conocido como *bisecting k-means* [51]. Esta variante es menos susceptible al problema de la inicialización. En la Tabla 2-4 se muestra el algoritmo básico. Se puede notar que durante el proceso de obtención de la partición final se va generando un árbol binario.

Se puede señalar que en el agrupamiento jerárquico resulta muy difícil separar el modelo del criterio que se utiliza, así como del método de búsqueda que se utiliza para encontrar el mejor agrupamiento en un momento dado. La naturaleza del modelo, es decir, la forma de los clusters está implícita en el algoritmo que se utiliza y no existe una función criterio global sino que se obtienen valoraciones locales con vistas a determinar los mejores candidatos para la aglomeración o para la división, según sea el caso. Se debe notar que diferentes funciones criterio pueden llevar a diferentes agrupamientos de los datos.

Tabla 2-4 Algoritmo bisecting k-means

Algoritmo *bisecting k-means* básico

1. **Inicializar la lista de clusters con un único cluster formado por todos los elementos**
2. **Mientras la lista no contenga K clusters hacer**
3. **Seleccionar un cluster de la lista.**
4. **for $i=1$ to Cantidad de Iteraciones do**
5. **Dividir en dos el cluster utilizando *k-means***
6. **end**
7. **Seleccionar la bipartición con menor error cuadrático del paso anterior.**
8. **Agregar la bipartición seleccionada a la lista de clusters**
9. **Fin**

La matriz de proximidades (desemejanzas o semejanzas) que se utiliza en el agrupamiento jerárquico es por definición densa, es decir, se analizan todos los pares de objetos. Algunas variantes de estos métodos proponen reducir la densidad de esta matriz reduciendo la cantidad de entradas en esta matriz. Esto se puede lograr de diversas maneras, por ejemplo, omitiendo aquellos valores que no superen un valor dado de proximidad, o analizando para cada punto una cierta cantidad de sus vecinos más próximos.

Una matriz construida de esta manera puede ser utilizada para representar los conceptos de cercanía o de conectividad la que puede ser vista además como un grafo. Este enfoque establece una relación entre el agrupamiento jerárquico y el particionamiento de grafos. En particular el algoritmo jerárquico divisivo MST (*Minimum Spanning Tree*) se basa en este enfoque [15].

Se debe tener en cuenta que la utilización de la distancia euclidiana para el agrupamiento jerárquico de datos espaciales tiende a formar clusters de forma convexa. Sin embargo, en muchas imágenes se aprecian objetos de formas arbitrarias. En este sentido se ha trabajado también con vistas a incorporar el concepto intuitivo de cluster como un componente conexo de forma arbitraria. Dos ejemplos lo constituyen los algoritmos jerárquicos aglomerativos CURE [52] y CHAMELEON [53].

Un elemento a resaltar en este punto es la similitud de las estrategias del agrupamiento jerárquico con las aplicadas en el caso de los métodos de segmentación orientados a regiones (que se verán más adelante). El análogo de la estrategia divisiva en el agrupamiento jerárquico lo constituye el proceso de división que se utiliza en los métodos de segmentación. Por su parte tanto en el agrupamiento jerárquico aglomerativo como en la segmentación se aplica el procedimiento de fusión (de regiones o de clusters).

La presencia de estas estrategias comunes, o su variantes híbridas (por ejemplo, división - fusión), constituyen un argumento a favor de la posibilidad de tratar el problema de la segmentación de imágenes como un problema de clustering.

2.2.3 Estrategias de implementación basadas en los vecinos más próximos para el agrupamiento jerárquico aglomerativo sin restricciones

El algoritmo general descrito en el apartado 2.2.2 presupone en cada paso la búsqueda del par de clusters que realizan la desemejanza mínima. Este proceso requiere un tiempo $O(n^2)$ si se tiene en cuenta que se trabaja con una matriz de desemejanzas de n objetos.

Por otra parte, en la construcción de la jerarquía se realizan $n-1$ fusiones de pares de clusters cada una de las cuales requiere un tiempo $O(n^2)$. Esto conlleva a que la complejidad temporal sea $O(n^3)$. La complejidad espacial, por su parte, es $O(n^2)$ debido a la necesidad de almacenar la matriz de desemejanzas.

Implementaciones más rápidas se pueden lograr considerando los vecinos más próximos para todos los clusters. Sea una partición $P = \{P_1, P_2, \dots, P_k\}$. El vecino más próximo de un cluster P_k dado es aquel cluster P_l que se encuentra más cercano a él desde el punto de vista de las desemejanzas y se denotará por $P_l = NN(P_k)$. Encontrar el vecino más próximo de un cluster requiere un tiempo $O(n)$.

Una vez que se tenga el vecino más próximo de cada cluster, la búsqueda del par de clusters más cercanos se puede realizar en un tiempo $O(n)$. Esto significa que al utilizar la estrategia de los vecinos más próximos la complejidad del algoritmo está determinada por el tiempo que se necesita en recalculer los vecinos más próximos después de cada fusión.

Los vecinos más próximos se deben recalculer para cada cluster que tenga como vecino más próximo a alguno de los dos clusters que se fusionan en la etapa dada. Esto puede conducir en general a $O(n)$ actualizaciones de vecinos más próximos en cada interacción [54, 55]. Si las filas de la matriz de desemejanzas se mantienen en un montículo (*heap*), estas actualizaciones se pueden realizar en un tiempo $O(n \log(n))$ lo que resulta en una implementación con un tiempo total $O(n^2 \log(n))$ y una complejidad espacial $O(n^2)$.

Otra estrategia rápida de implementación se basa en el hecho de que bajo ciertas condiciones es posible obtener la misma solución utilizando un orden diferente de aglomeraciones en el algoritmo jerárquico aglomerativo. La idea consiste en la obtención de los pares de vecinos más próximos mutuos o recíprocos (*reciprocal nearest neighbors*, *RNN*). Se dice que un par de clusters P_k y P_l son vecinos mutuos o recíprocos si se cumple que $P_l = NN(P_k)$ y $P_k = NN(P_l)$. Esta estrategia rápida presupone que en cada paso se fusionan pares de vecinos mutuos [56].

La estrategia de vecinos más próximos mutuos es aplicable si el criterio de aglomeración cumple con la propiedad de reducibilidad [57].

Se dice que un criterio de aglomeración posee con la propiedad de reducibilidad si se cumple que

$$d(P_k, P_l) \leq \min\{d(P_k, P_m), d(P_l, P_m)\} \Rightarrow \\ \min\{d(P_k, P_m), d(P_l, P_m)\} \leq d(P_k \cup P_l, P_m)$$

Según Diday una condición necesaria y suficiente para que no ocurran inversiones en un algoritmo jerárquico aglomerativo es que el criterio de aglomeración utilizado cumpla la propiedad de reducibilidad [58]. Los criterios de enlace simple, enlace completo, enlace promedio y de Ward para el caso del agrupamiento jerárquico sin restricciones cumplen con esta propiedad, no así el criterio del enlace basado en centroides para el que no se puede garantizar la ausencia de inversiones (ver apartado 2.2.2).

La búsqueda de los vecinos más próximos mutuos puede realizarse mediante el mantenimiento de una cadena completa de vecinos más próximos. Una secuencia $P = (P_{i_1}, P_{i_2}, \dots, P_{i_r})$ se denomina una cadena de vecinos más próximos respecto a una desemejanza dada si $P_{i_{s+1}} = NN(P_{i_s})$ para todo $1 \leq s < r$.

Una cadena de vecinos más próximos se denomina completa si contiene al menos dos clusters y además los dos últimos son vecinos mutuos o recíprocos. La fusión de los vecinos recíprocos de una cadena completa de vecinos más próximos no afecta la relación entre otros objetos o clusters. Normalmente se asume que no existen dos desemejanzas entre pares de objetos o de clusters que sean iguales y por tanto se considera que una cadena completa de vecinos más próximos cumple las siguientes propiedades [54].

Propiedad 1: Las desemejanzas entre objetos o clusters decrecen de manera monótona a medida que se avanza en la cadena de vecinos más próximos.

Propiedad 1: El enlace final de una cadena de vecinos más próximos siempre conecta a un par de vecinos más próximos mutuos.

Propiedad 3: Una cadena de vecinos más próximos no puede contener un ciclo de más de dos objetos.

A continuación se describe la estrategia basada en las cadenas de vecinos más próximos. El algoritmo comienza a construir una cadena de vecinos más próximos partir de un cluster arbitrario (singular o no) hasta que la cadena sea completa.

Dada una cadena completa $P = (P_{i_1}, P_{i_2}, \dots, P_{i_r})$ se mezclan los clusters $P_{i_{r-1}}$ y P_{i_r} que constituyen vecinos más próximos mutuos y se extraen de la cadena. Si la cadena resulta vacía después de la fusión, se comienza el proceso a partir de otro cluster cualquiera. El proceso termina cuando se queda un único cluster.

Se debe notar que si el criterio de agregación cumple con la propiedad de reducibilidad, es decir, es libre de inversiones, entonces la cadena resultante es una cadena válida de vecinos más próximos. De esta manera es posible entonces obtener una cadena de vecinos más próximos completa a partir de la anterior extendiéndola de manera iterativa.

La expansión de la cadena de vecinos más próximos y la comprobación de si es completa o no requiere un tiempo $O(n)$. Como la operación anterior se debe ejecutar no más de $2(n-1)$ veces, esto conduce a una complejidad temporal total $O(n^2)$ [54].

La complejidad espacial por su parte depende de si las desemejanzas entre clusters se pueden calcular en cada paso en tiempo constante, o si por el contrario están almacenadas. En el primer caso la complejidad espacial es $O(n)$, mientras en el segundo es $O(n^2)$.

Se debe tener en cuenta que para que esta estrategia sea exacta, es decir, que produzca los mismos resultados que el algoritmo estándar, el criterio de aglomeración debe satisfacer la propiedad de reducibilidad, es decir, debe ser libre de inversiones.

Las cadenas de vecinos más próximos y los vecinos más próximos mutuos permiten dos variantes para la construcción de una jerarquía:

1. Construir una cadena de vecinos más próximos. Al encontrar un par de vecinos más próximos mutuos proceder a su aglomeración y continuar la cadena a partir del cluster anterior al par de clusters recién fusionados.
2. Determinar los vecinos más próximos de todos los objetos; obtener los vecinos más próximos mutuos; fusionarlos, y continuar el proceso.

Respecto al cálculo de las desemejanzas se debe tener en cuenta si éstas se calculan en cada momento a partir de los datos originales, o si por el contrario se mantienen almacenadas y requieren actualizaciones en cada paso utilizando, por ejemplo, la fórmula de Lance – Williams – Jambu [49, 50].

Por último comentaremos una propiedad interesante del algoritmo estándar de agrupamiento jerárquico aglomerativo. Según este algoritmo en cada etapa se fusiona el par de clases cuyo valor de desemejanza es mínimo. Con esta premisa es válido lo siguiente [59].

Propiedad 4: En cada etapa del algoritmo estándar de agrupamiento jerárquico aglomerativo sin restricciones el par de clases que se fusiona está compuesto por un par de vecinos más próximos mutuos.

La validez de esta propiedad se demuestra fácilmente. Supongamos que en una etapa cualquiera se fusionan los clusters P_k y P_l . Eso significa que su desemejanza es la menor entre todos los pares posibles de clusters existentes en la etapa dada.

De lo anterior se deduce que P_l es un vecino más próximo de P_k . Si esto no fuera así entonces existiría otro cluster P_m que sería el vecino más próximo de P_k y por tanto $d(P_k, P_m) < d(P_k, P_l)$ lo que contradice la premisa de que P_k y P_l es el par de desemejanza mínima. De manera similar se demuestra que P_k es el vecino más próximo de P_l . Luego P_k y P_l son vecinos más próximos mutuos.

2.3 Otros enfoques

La categorización de los algoritmos de agrupamiento no es sencilla. En ocasiones existen propuestas híbridas que incluyen las más diversas etapas y que las hace difícil ubicar en una categoría pura. A continuación se comentan algunas propuestas recientes.

Un enfoque interesante para el agrupamiento general, tanto jerárquico como particional, basado en proximidades ha sido propuesto en [10]. Aquí el problema del agrupamiento es formulado como un problema combinatorio; se especifica una función objetivo, cuya idea es la formación de clusters compactos y que se basa en la suma normalizada de las desemejanzas intracluster (desemejanza promedio dentro de cada cluster)

$$J_W(P) = \sum_{k=1}^K J_W(P_k) = \frac{1}{2} \sum_{k=1}^K \frac{1}{n_k} \sum_{\omega_i, \omega_j \in P_k} d(\omega_i, \omega_j)$$

En el caso que las desemejanzas sean distancias euclidianas cuadráticas, la función objetivo coincide con la función de costo utilizada en el algoritmo *k-means*. Se proponen algunas heurísticas para la optimización. Esta propuesta no es adecuada para la detección de clusters alargados ya que en la función objetivo se suman todas las desemejanzas intracluster sin tener en cuenta las posibles relaciones topológicas.

La propuesta anterior se centra en la compacidad de los clusters como criterio primario. Sin embargo en ocasiones el énfasis en el criterio de la compacidad solamente no permite capturar las propiedades esenciales de los datos. En la Figura 2-6 se muestran dos ejemplos clásicos los cuales han sido procesados con el algoritmo *k-means*. Normalmente se esperaría que cada brazo de las espirales y cada círculo concéntrico formen un cluster.

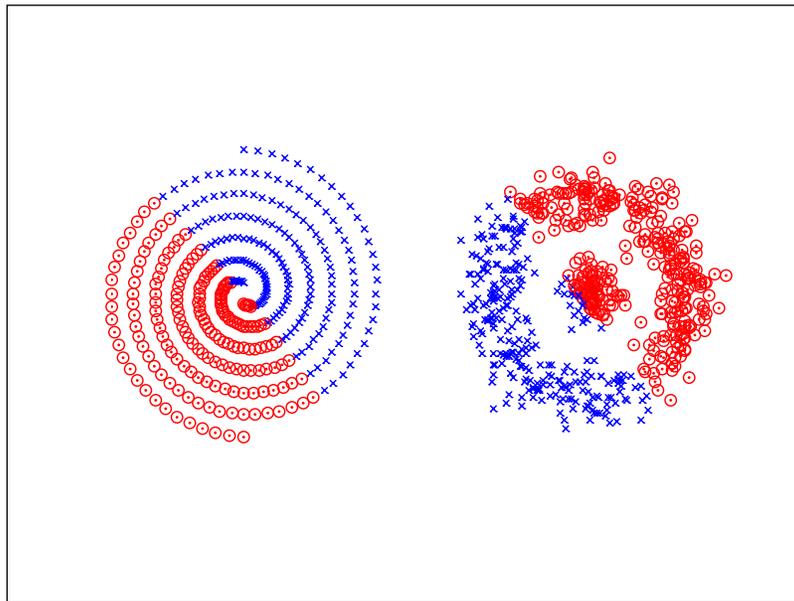


Figura 2-6 Ejemplos donde el criterio de compacidad no es suficiente.

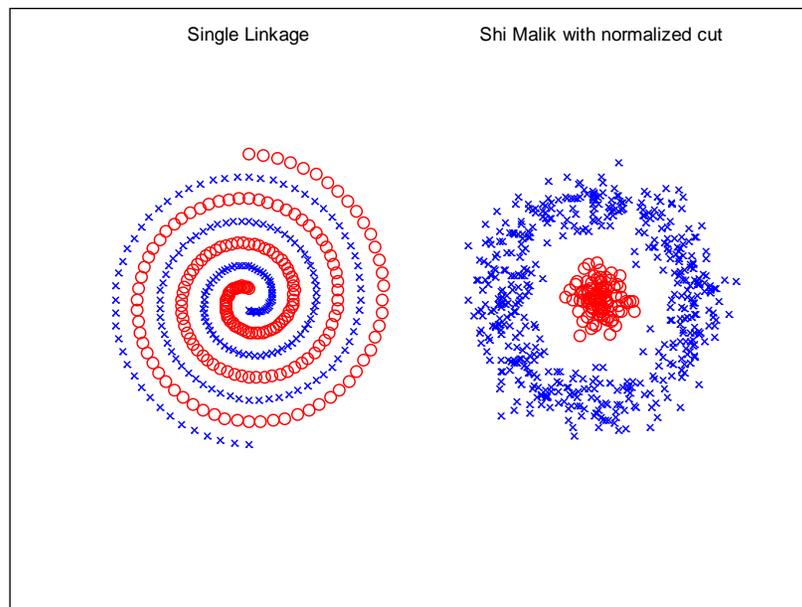


Figura 2-7 Agrupamiento donde el criterio de conectividad es apropiado.

En estos casos sería conveniente hacer una distinción entre la compacidad y la conectividad. Una propuesta en este sentido es el denominado agrupamiento a pares basado en caminos (*path-based pairwise clustering*) basado en métodos basados en la teoría de grafos [35]. En este caso dos objetos se consideran semejantes si existe un camino

dentro del cluster que no posea ningún arco con alta desemejanza. Aquí se mantiene la función objetivo, [10], pero las desemejanzas tienen en cuenta los posibles caminos entre dos objetos.

Por otra parte, la aplicación de la teoría espectral de grafos ha conducido a los algoritmos espectrales de agrupamiento (*spectral clustering algorithms*) los cuales se basan en la utilización de los valores propios de alguna matriz de semejanzas (u otra derivada de ésta) entre los objetos [31, 60]. Una propiedad interesante de estos algoritmos es su capacidad de extraer clusters no convexos o disjuntos. En su etapa final se utilizan métodos particionales. En esencia estos algoritmos determinan una inmersión o proyección, en un espacio euclidiano, de la matriz de semejanzas transformada de manera apropiada, y por último aplican métodos particionales en este espacio (ver Figura 2-7).

El concepto de proyección o inmersión es bastante general. Es aplicable tanto al problema de reducción de la dimensionalidad (encontrar representaciones de datos de grandes dimensiones en espacio de pequeñas dimensiones, usualmente de 2 ó 3), como al problema de encontrar representaciones vectoriales de datos de proximidad a pares. Ejemplos de estos métodos son el análisis de componentes principales (*Principal Component Analysis - PCA*) [61] y el escalado multidimensional (*Multidimensional Scaling - MDS*) [62]. Una panorámica de estos métodos puede ser consultada en [63].

3 Los métodos de segmentación de imágenes

Los métodos de segmentación de imágenes pueden agruparse en dos categorías: supervisados y no supervisados. En los métodos supervisados se dispone de un conjunto de muestras u objetos con sus correspondientes etiquetas que sirven con vía para la obtención de un clasificador, de manera que al presentarle una muestra desconocida éste le asigne la etiqueta correspondiente. Por su parte los métodos no supervisados persiguen como objetivo la búsqueda de los grupos “naturales” en los datos disponibles, para lo cual es preciso definir alguna noción de grupo que se ajuste a las necesidades concretas. Como el trabajo se centra en los métodos automáticos, en lo adelante se analizan fundamentalmente las técnicas no supervisadas.

Se han desarrollado diversas técnicas de segmentación de imágenes [64], y se han propuesto varios esquemas de clasificación de estas técnicas [3, 6, 65].

Independientemente del enfoque que se utilice, y atendiendo al grado de intervención del usuario, los métodos de segmentación pueden considerarse automáticos o interactivos [66]. Lógicamente, a mayor componente manual, más lenta y tediosa resulta la segmentación. Sin embargo, los métodos automáticos son incapaces, en general, de obtener resultados adecuados en todas las situaciones, independientemente del enorme esfuerzo de investigación realizado [3, 16, 22, 65, 67]. La principal causa de esta dificultad radica en la necesidad de utilizar conocimiento adicional a la información contenida en las imágenes. A pesar de todo, los métodos automatizados se deben convertir en complementos cruciales en el análisis de imágenes, particularmente de las imágenes médicas.

Una clasificación compacta se brinda en [68] que agrupa los métodos de segmentación en: métodos basados en bordes, métodos basados en regiones, y métodos híbridos (ver Tabla 3-1).

Tabla 3-1 Clasificación de métodos de segmentación [68]

ESTRATEGIA	MÉTODOS	DESCRIPCIÓN
Basada en bordes (Pretende delinear la interfase entre los objetos)	Detección de contornos óptimos	Plantean el problema del delimitado del borde como un problema de optimización. [69]
	Contornos activos o bordes deformables	Se plantea un sistema mecánico que deforma un contorno inicial bajo determinadas fuerzas. El contorno final es aquel que minimiza cierto funcional de energía [70]
	Detección guiada por el usuario (<i>live-wire</i>)	El usuario reconoce los bordes mediante la señalización de ciertos puntos y el proceso de delimitado se realiza de manera automática y cooperada. [71, 72]
	Formas y apariencias activas	Se imponen restricciones cerca de la forma del contorno, así como de los patrones de intensidad en su vecindad. [73, 74]
Basada en regiones	Umbralización	Consiste en la binarización de una imagen, es decir, en su conversión a una imagen con dos niveles de gris. Ver, por ejemplo, [75, 76]
	Clustering	Se plantea el problema como uno de clasificación no supervisada.
	Crecimiento de regiones	Se pretende extraer regiones conexas a partir de ciertos puntos iniciales o semillas.
	División y fusión	Están relacionados con el crecimiento de regiones; pero no necesitan semillas iniciales.
	Enfoques basados en grafos	La segmentación se plantea como un problema de grafos, y se utiliza el aparato teórico asociado. [77, 78]
	Métodos basados en modelos estadísticos	La clasificación de los objetos se realiza en términos de probabilidades.
	Watershed (línea de unión de aguas)	La imagen se ve como una superficie 3D. Se realiza un proceso de inundación de esa superficie hasta que las cuencas de agua se encuentran. En ese momento se forman las líneas de unión de aguas.
Híbrida (combina las fortalezas de las estrategias anteriores)		En muchas aplicaciones es necesario combinar varias de las técnicas descritas antes con el fin de obtener resultados satisfactorios.

Por otra parte, cualquier método de segmentación necesita una valoración cuantitativa de su desempeño. En este sentido se requieren experimentos de validación. La validación se realiza normalmente utilizando un cierto modelo de verdad [67].

El modelo más directo de validación consiste en la comparación de los resultados obtenidos por el método bajo análisis con los resultados obtenidos mediante la segmentación manual. Este enfoque se ve afectado por los problemas inherentes a la segmentación manual, la cual no garantiza un modelo de verdad perfecta toda vez que los resultados obtenidos por el operador pueden estar distorsionados. Independientemente de estos señalamientos, en muchas ocasiones se recurre a este enfoque de validación.

Otro enfoque muy utilizado es aquel que hace uso de modelos físicos, o modelos computacionales [79]. Los modelos físicos brindan una descripción exacta del proceso de

adquisición de la imagen, aunque no representan de manera fiel la anatomía. Por su parte los modelos computacionales pueden ser más realistas en este último aspecto; pero simulan el proceso de adquisición utilizando solamente modelos simplificados; por ejemplo, de generación de ruido, o de variación de fondo.

Una vez que se ha asumido un modelo de verdad, se precisa de la definición de un indicador para cuantificar la exactitud. La selección de este indicador es dependiente de la aplicación y puede estar basada, por ejemplo, en información sobre las regiones tal como la cantidad de píxeles clasificados erróneamente, ya sea porque pertenecen a la región de interés y no fueron incluidos (falsos negativos), o porque fueron incluidos y no pertenecían (falsos positivos). Un revisión de esta problemática puede ser consultada en [67].

A continuación se describen los métodos de segmentación tomando como base la clasificación presentada en [68].

3.1 Métodos basados en bordes

Estos métodos se basan en la detección de discontinuidades o transiciones de los niveles de grises, el color, la textura, o cualquier otra magnitud de interés en la imagen. Los bordes marcan, precisamente, los lugares donde ocurren estas discontinuidades, y se encuentran en las fronteras de los objetos. Una vez establecido el mecanismo o el criterio para la determinación de las discontinuidades o transiciones lo que ocurre es, en esencia, un proceso de clasificación de los píxeles de la imagen en dos clases: bordes y no bordes.

Existen situaciones que estos métodos no contemplan, como es el caso de los denominados bordes perceptuales. En la Figura 3-1 se muestra un ejemplo: el aparente rectángulo enmarcado en los cuatro círculos no presenta discontinuidades respecto al fondo por ser del mismo color.

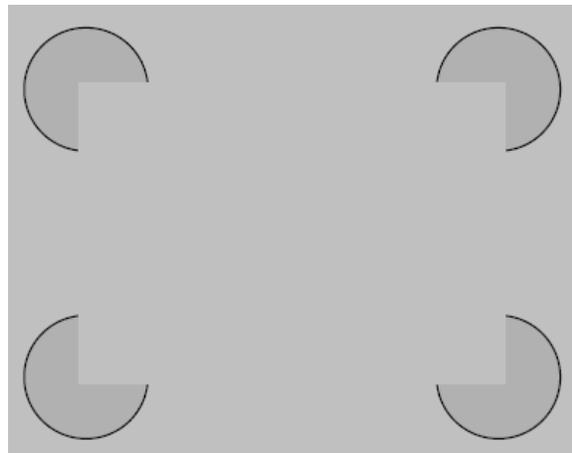


Figura 3-1 Rectángulo de Kanizsa.

La detección de los bordes se realiza mediante operadores de detección. Sin embargo; por lo general las imágenes que se obtienen no pueden ser utilizadas directamente como resultados de la segmentación. La segmentación basada en bordes se apoya en los resultados obtenidos por medio de dichos operadores. En este sentido se requieren etapas suplementarias que combinen los bordes obtenidos con vistas a la formación de cadenas que se ajusten lo más posible a las fronteras de los objetos presentes en las imágenes. El objetivo final es obtener al menos una segmentación parcial donde se retengan solamente aquellas cadenas de bordes que tengan alguna correspondencia con los objetos de la imagen.

Un método para el procesamiento local en las imágenes consiste en la utilización de máscaras espaciales o filtros que recorren por toda la imagen. Esto constituye un mecanismo de inclusión del contexto, en forma de píxeles vecinos, en las tareas de procesamiento de imágenes. Un ejemplo de máscara se muestra en la Figura 3-2.

$$\begin{bmatrix} w_1 & w_4 & w_7 \\ w_2 & w_5 & w_8 \\ w_3 & w_6 & w_9 \end{bmatrix}$$

Figura 3-2 Máscara de 3x3 píxeles.

Estos filtros pueden ser lineales o no. La manera más común de procesamiento consiste en obtener la suma de los productos de los coeficientes de la máscara y las intensidades de los píxeles que se encuentran bajo ésta en una localización específica de la imagen, es decir, $\sum_{i=1}^9 w_i x_i$, donde x_i representa la intensidad presente en la posición i .

Así, por ejemplo, si $w_i = 1/9$, $i = 1, \dots, 9$ el valor resultante de la aplicación del filtro sería la media de los valores de la vecindad incluyendo el píxel central. Si se toma $w_i = 1/8$, $\forall i = 1, \dots, 9, i \neq 5$ el valor resultante sería la media de la vecindad sin incluir el píxel central. En cualquier caso el valor resultante se le asocia al píxel correspondiente a la posición 5.

La utilización de este tipo de media ponderada en una ventana es muy común en varios operadores de procesamiento de imágenes. Una interesante reformulación de este proceso en el marco de un espacio vectorial, y el papel que juegan los diferentes subespacios se analiza en [80].

Si se tiene en cuenta que el resultado de esta operación se le asocia al píxel central, es posible entonces obtener una nueva imagen con estos valores. Mediante la definición de máscaras apropiadas el procesamiento se puede enfocar hacia la detección de discontinuidades, tales como puntos y líneas. Estos filtros lineales o suavizantes pueden ser

utilizados también para la reducción del ruido y pueden además producir cierto desdibujado de la imagen [81].

Aunque la detección de puntos y líneas tienen cierta importancia desde el punto de vista de la segmentación, es la detección de bordes el enfoque más común para la detección de discontinuidades. Los operadores de detección de bordes consideran que un borde es un conjunto de píxeles conexos que se encuentran en la frontera entre dos regiones que poseen distintas características de intensidad. Se asume que las regiones son lo suficientemente homogéneas como para que la transición se pueda determinar a partir de las discontinuidades de los niveles de grises. Sin embargo, ésta es también la característica del ruido que puede estar presente en una imagen. De ahí que cualquier operador que responda bien ante la presencia de un borde, lo hará también ante la presencia del ruido.

Existen dos enfoques locales básicos: la diferenciación de primer y segundo órdenes. Estos operadores se basan en el cálculo del gradiente en cada punto. El vector del gradiente indica la dirección de mayor cambio de la función en un punto dado y se obtiene de la siguiente manera

$$\nabla F = \begin{bmatrix} G_x \\ G_y \end{bmatrix} = \begin{bmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{bmatrix}$$

El módulo del vector, es decir la magnitud del gradiente, definido como $\nabla f = \sqrt{G_x^2 + G_y^2}$, representa la mayor tasa de crecimiento de f por unidad de distancia. Usualmente se toma de manera aproximada $\nabla f \approx |G_x| + |G_y|$.

Por su parte, la dirección de esa variación la brinda el ángulo del vector, definido como

$$\theta = \tan^{-1} \left(\frac{G_x}{G_y} \right).$$

En el caso de las imágenes digitales para el cálculo del gradiente se parte de las aproximaciones de las derivadas a partir del cociente incremental

$$\frac{\partial f(x, y)}{\partial x} = G_x \approx \frac{f(x + \Delta x, y) - f(x, y)}{\Delta x}, \text{ y}$$

$$\frac{\partial f(x, y)}{\partial y} = G_y \approx \frac{f(x, y + \Delta y) - f(x, y)}{\Delta y},$$

donde Δx y Δy representan el incremento en el eje x y en el eje y , respectivamente.

Así, en una imagen de dimensión 3×3 como la siguiente

$$\begin{bmatrix} x_1 & x_4 & x_7 \\ x_2 & x_5 & x_8 \\ x_3 & x_6 & x_9 \end{bmatrix}$$

la primera y segunda derivadas en la dirección del eje x en el punto x_5 se aproximan de la siguiente manera

$$\frac{\partial f}{\partial x} = x_8 - x_2$$
$$\frac{\partial^2 f}{\partial x^2} = x_7 + x_9 - 2x_5$$

Supongamos que se tiene el siguiente perfil de dos regiones separadas por un borde

a)	0	0	0	1	1
b)	0	0	1	0	0
c)	0	1	-1	0	0

En a) se tiene el perfil original; en b) la primera derivada; y en c) la segunda derivada. Como se puede apreciar, la primera derivada brinda una respuesta alta en el lugar de la transición de la zona oscura (0's) hacia la zona clara (1's). Por su parte la segunda derivada presenta un cruce por cero en la zona de transición, es decir, en el borde.

Entre los métodos que caen en esta categoría se encuentran los de Sobel (ver Figura 3-3), Roberts, Prewitt [81]. Estos métodos tienen en común la utilización de máscaras espaciales para la estimación de la magnitud del gradiente en diferentes direcciones.

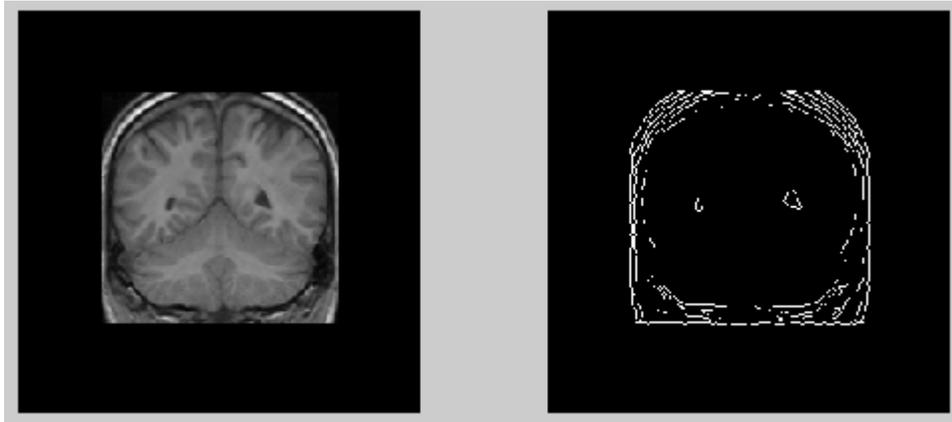


Figura 3-3 Ejemplo de aplicación del operador de Sobel.

Por su parte los métodos de segundo orden buscan los bordes en los sitios donde la segunda derivada tiene un cruce por cero. Para ello se basan en el operador Laplaciano definido como

$$\nabla^2 f = \frac{\partial^2 f}{\partial^2 x} + \frac{\partial^2 f}{\partial^2 y}.$$

Sin embargo, el operador Laplaciano no es muy utilizado debido a su alta sensibilidad al ruido y a su incapacidad de detectar la dirección del borde. Su utilidad radica en la detección de bordes a partir de su propiedad de resaltar los cruces por cero.

Con vistas a superar estas limitaciones se utiliza un filtro gaussiano de la forma

$$h(x, y) = -\exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right),$$

donde σ representa la desviación estándar del filtro.

El Laplaciano del filtro gaussiano (LoG - Laplacian of Gaussian) es

$$\nabla^2 h = -\left(\frac{r^2 - 2\sigma^2}{\sigma^4}\right)\exp\left(-\frac{r^2}{2\sigma^2}\right),$$

donde $r^2 = x^2 + y^2$, y σ determina el grado de emborronamiento que ocurre.

Una máscara espacial de 5×5 píxeles que aproxima al Laplaciano del filtro gaussiano se muestra en la Figura 3-4.

$$\begin{bmatrix} 0 & 0 & -1 & 0 & 0 \\ 0 & -1 & -2 & -1 & 0 \\ -1 & -2 & 16 & -2 & -1 \\ 0 & -1 & -2 & -1 & 0 \\ 0 & 0 & -1 & 0 & 0 \end{bmatrix}$$

Figura 3-4 Máscara de 5x5 del LoG.

En la línea de los métodos basados en el gradiente se puede señalar también el de Canny [82]. Para el cálculo del gradiente, este método utiliza la derivada de un filtro gaussiano. Con vistas a disminuir la influencia del ruido se utilizan dos umbrales que permiten detectar bordes fuertes y débiles. Los bordes débiles se incluyen en el resultado final sólo si están conectados con bordes fuertes. Este método es más robusto al ruido que los otros y más propenso a detectar los bordes reales presentes en la imagen.

En la Figura 3-5 se muestra la aplicación de este método a una imagen de resonancia magnética.

Otros métodos generales para la detección de bordes lo constituyen los basados en búsquedas en grafos. Con este fin se construye un grafo cuyos nodos se corresponden con los píxeles de la imagen. El grafo es ponderado asociando costos a los nodos y/o los arcos. Estos costos representan propiedades locales o globales de las fronteras que se pretenden obtener. La especificación de los costos es un elemento crucial en estos algoritmos. Un ejemplo de asignación de costos puede ser consultado en [81], págs. 592-593.

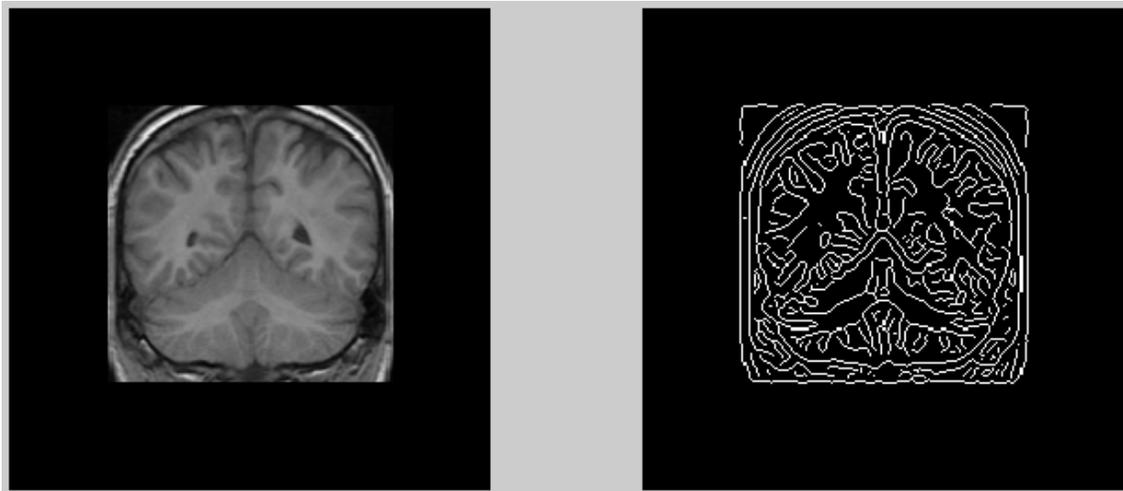


Figura 3-5 Aplicación del método de Canny a una imagen de resonancia magnética.

El problema de detección de bordes se transforma en uno de búsqueda de caminos óptimos (usualmente de costo mínimo) en el grafo. En este sentido las dos técnicas más utilizadas han sido la búsqueda heurística en grafos y la programación dinámica.

En esta categoría de búsqueda de caminos óptimos se puede señalar el método de delineado de bordes guiado por el usuario denominado *live-wire* [71]. Este método combina el delineado automático del borde con la definición manual por parte del usuario de un punto inicial y otro final. Entre esos dos puntos se encuentra el camino óptimo que identifica al borde. Este procedimiento se ilustra en la Figura 3-6.

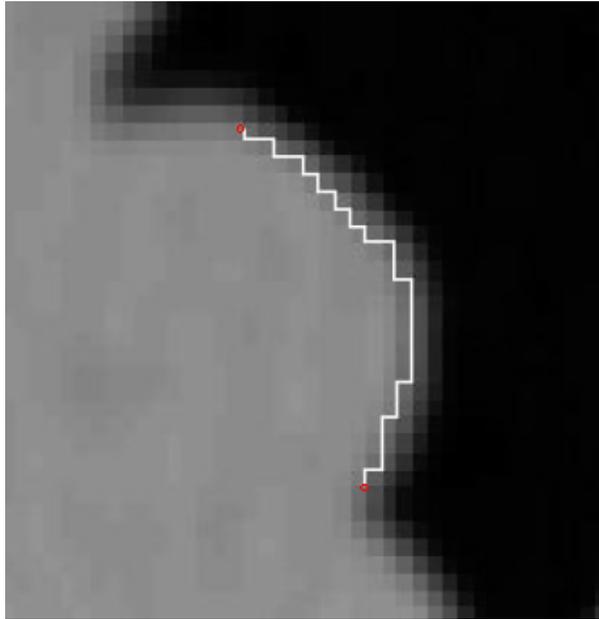


Figura 3-6 Obtención de un camino óptimo según el método *live-wire*.

Los métodos basados en grafos presentan el inconveniente de su complejidad y costo computacional. Su principal ventaja consiste en la robustez en presencia de ruido.

Otro enfoque razonable consiste en la búsqueda de contornos de una forma dada. Este es el caso de la transformada de Hough, utilizada inicialmente para detectar líneas rectas y posteriormente se extiende a la búsqueda de formas definidas analíticamente [83-85].

Dentro de la búsqueda de la frontera de un objeto se destaca también la aproximación basada en serpientes (*snakes*), y en general en contornos deformables y activos, cuyo objetivo es la construcción de una banda que de manera elástica e iterativa se va ajustando a la frontera de los objetos. Las *snakes* son curvas abiertas o cerradas que tienen asociado un funcional de energía y que se deforman bajo la acción de fuerzas internas y externas con el objetivo de minimizar el funcional. El contorno activo es un caso particular donde la curva es cerrada. Más detalles sobre los modelos deformables se pueden encontrar en [86].

Los bordes, tal y como se obtienen por los métodos descritos se encuentran en general afectados por el ruido y comúnmente ciertas partes de las fronteras reales están ausentes. Se requiere por tanto tomar en consideración el contexto en que aparecen los bordes resultantes con el fin de mejorar la imagen resultante.

Los bordes resultantes deben ser extendidos o eliminados de manera que se puedan obtener las fronteras cerradas de los objetos.

La relajación de bordes es un proceso iterativo que asocia a cada borde un valor de confianza en función de su tipo, es decir, de su vecindad. Este valor debe converger a cero (eliminación del borde), o hacia uno (el borde forma parte de la frontera) [87].

Otra idea para la formación de las fronteras lo constituye el seguimiento o rastreo de los bordes. En esencia el procedimiento consiste en hacer crecer la frontera a partir del punto de mayor intensidad en la imagen de bordes. Para ello se hace una búsqueda local y se incluye en el resultado el píxel más parecido al actual. El proceso continúa de manera iterativa. Este procedimiento funciona, en esencia, como el método de crecimiento de regiones que se comenta en el siguiente apartado.

Las fronteras de los objetos delimitan las regiones que estos ocupan. Una vez que se hayan obtenido las fronteras es posible, en principio, reconstruir la regiones. De manera similar, si se tiene una partición de una imagen en regiones es posible hacer un seguimiento de la frontera de cada región.

Se pueden diferenciar la frontera interna y la frontera externa de la región. La primera es un subconjunto de la región, mientras la segunda no lo es. Eso significa que dos regiones adyacentes poseen diferentes fronteras internas. Lo mismo ocurre con las externas.

En este sentido mejores propiedades poseen las fronteras extendidas. La frontera extendida representa una frontera común entre las regiones, posee la misma forma que la frontera inter-píxeles, y se especifica en coordenadas de píxeles.

Una presentación de diferentes algoritmos de relajación de bordes, de seguimientos de fronteras internas, externas y entendidas puede ser consultado en [87].

3.2 Métodos basados en regiones

Los métodos basados en regiones tradicionales asumen que píxeles conexos y de similares características deben pertenecer a la misma región, es decir, tratan de identificar zonas cuyos componentes posean alguna propiedad en común. En este apartado se incluyen además aquellos métodos que relajan la condición de contigüidad de los píxeles, como es el caso de la umbralización.

Estos métodos hacen uso de dos operaciones fundamentales que ya se han mencionado dentro de los métodos de agrupamiento (*clustering*): la fusión o agregación, y la división o separación. En el primer caso, las regiones vecinas se fusionan si cumplen con determinado criterio. Por su parte, la división de una región o clase ocurre cuando ésta no cumple un cierto criterio y genera regiones de menor tamaño que pueden cumplirlo.

También se encuentran algoritmos híbridos que combinan ambas operaciones. En cualquier caso se debe utilizar algún criterio para decidir cuándo dividir una región, o cuando fusionar o mezclar dos.

A continuación se presentan varios métodos que generan regiones a partir de una imagen digital. Aunque la revisión no es exhaustiva, si ilustra la manera en que las operaciones fundamentales están presentes en las distintas propuestas dentro del área del procesamiento de imágenes.

3.2.1 Umbralización

Uno de los métodos más simples, antiguos y aún bastante utilizado es conocido por umbralización. Este método se basa en el histograma de la imagen. El histograma de una imagen es una representación de la frecuencia de ocurrencia de los niveles de grises (ver Figura 3-7). Cada imagen posee un único histograma; aunque lo contrario no se puede afirmar de manera general. Por ejemplo, un tablero de ajedrez con casillas blancas y negras, y un cuadrado, de las mismas dimensiones, con una mitad blanca y la otra negra poseen el mismo histograma.

Cada elevación en el histograma, o moda, está asociada a una determinada región de la imagen. Los valles del histograma pudieran ser utilizados como posibles umbrales para separar esas regiones. La idea principal radica en que los píxeles con valores de niveles de grises por debajo del umbral se asocian a una región y el resto a otra. El mismo procedimiento se aplicaría para el caso de varios umbrales. La umbralización puede resultar un medio sencillo y efectivo de lograr la segmentación cuando las estructuras de interés presentan intensidades, u otras características cuantificables, contrastantes.

Cuando se utiliza la información de toda la imagen en su conjunto para obtener un umbral, el método se denomina umbralización global, que puede ser múltiple si se utiliza más de un umbral. Si el método se aplica en una vecindad de un píxel se dice que la umbralización es local, también denominada umbralización dinámica o adaptativa. Este es el caso cuando la imagen se divide en varias sub-regiones y se realiza la umbralización mediante la selección de uno o varios umbrales en cada una de ellas por separado.

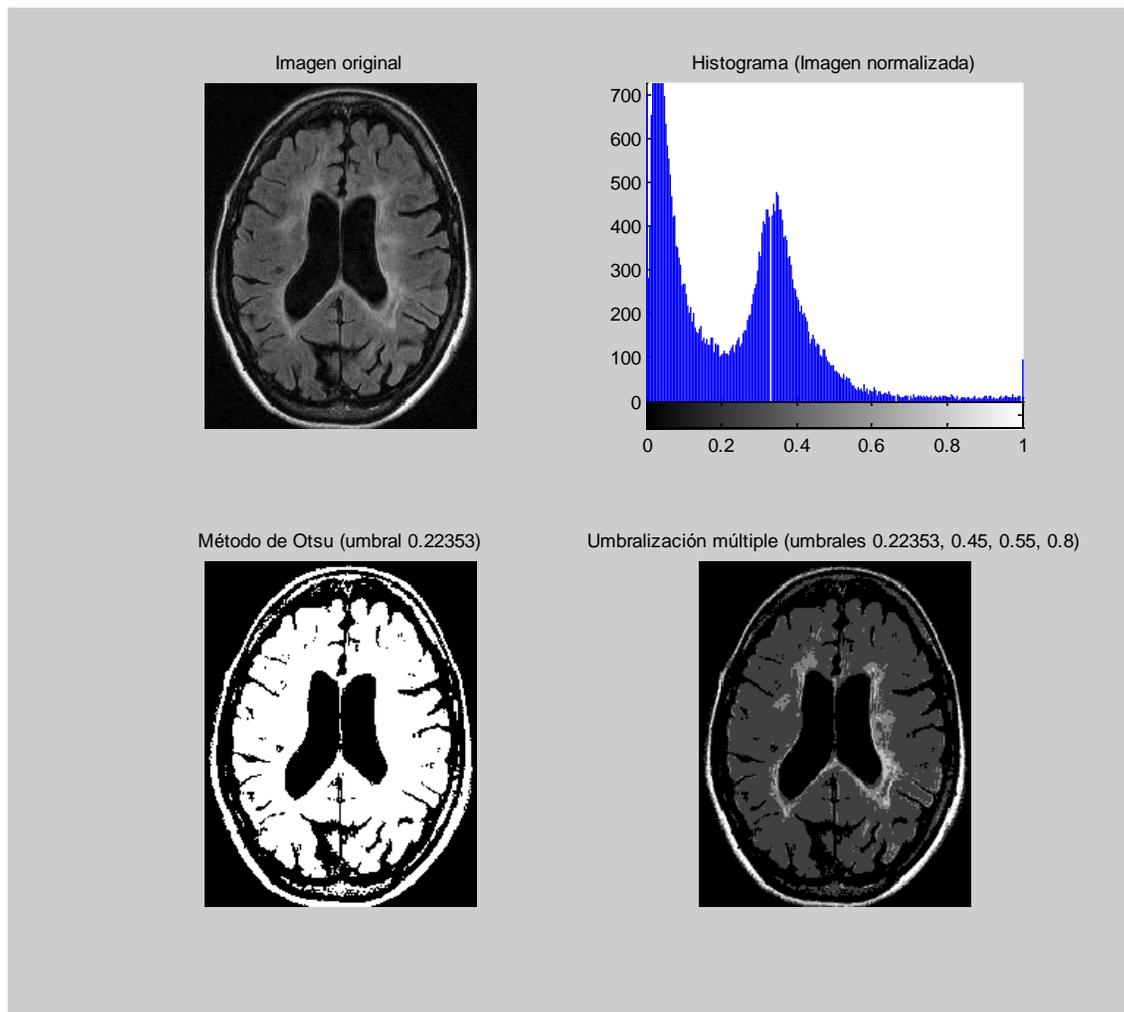


Figura 3-7 Imagen de RM, su histograma, segmentación mediante umbralización simple y mediante umbralización múltiple.

La umbralización también se puede clasificar como simple o múltiple. Si en la imagen están presentes dos regiones diferenciables, el objeto y el fondo, estos pueden ser segmentados mediante un único umbral, es decir, mediante umbralización simple. En el caso que existan más objetos diferenciables por sus niveles de grises se requieren varios umbrales para su separación. Este tipo se conoce como umbralización múltiple. En la Figura 3-7 se muestra un ejemplo de una imagen de resonancia magnética segmentada mediante umbralización simple y umbralización múltiple.

Normalmente la umbralización se aplica a los niveles de gris de la imagen, pero existe la posibilidad de utilizarla cuando los valores representan el gradiente, o cualquier otra propiedad asociada al contexto local.

Aunque como se señaló antes, los picos y valles del histograma de una imagen pueden servir de guía en la selección de un umbral, esto no resulta una tarea trivial.

Se han propuesto varios métodos para la detección automática de umbrales así como varias clasificaciones de estos métodos [65, 88, 89].

El problema de la umbralización se puede analizar desde la perspectiva de la clasificación automática. Por ejemplo, el caso de la umbralización simple es equivalente a asignar un píxel a una de dos clases posibles, objeto o fondo. Para la umbralización múltiple se utilizarían más de dos clases.

Un posible algoritmo que emana de la propuesta de Ridler y Calvard [90] se muestra en la Tabla 3-2.

Tabla 3-2 Algoritmo iterativo según la propuesta de [90]

<p>Algoritmo iterativo de umbralización para dos clases.</p> <ol style="list-style-type: none"> 1. En ausencia de información adicional considerar, como una primera aproximación, que los cuatro píxeles de las esquinas de la imagen pertenecen al fondo, y resto de los píxeles se corresponden con el objeto. 2. En el paso t calcular μ_f^t y μ_o^t como la media de las intensidades de los píxeles pertenecientes al fondo y al objeto, respectivamente. Para la segmentación en el paso t se utiliza el umbral T^t determinado en el paso anterior (ver ecuación del paso 3). 3. Calcular el nuevo valor del umbral según la ecuación $T^{t+1} = \frac{\mu_f^t + \mu_o^t}{2}$ 4. Si la magnitud $T^{t+1} - T^t$ es suficientemente pequeña entonces parar. De lo contrario ir al paso 2.
--

Entre los métodos más conocidos se encuentra el de Otsu [75] que utiliza la varianza como medida de separabilidad entre clases correspondientes al fondo y al objeto y propone maximizar la razón entre la varianza entre clases y una suma ponderada de las varianzas intraclase del fondo y del objeto con el fin de establecer un umbral óptimo. Este criterio guarda semejanza con lo visto en el apartado 2.2 acerca de la descomposición de las desemejanzas internas y cruzadas entre clusters.

Este método, uno de los más referenciados, brinda buenos resultados cuando las intensidades de los píxeles de cada clase están cercanas entre sí. Algunas limitaciones del método de Otsu son discutidas en [91]. En [92] se propone una generalización al caso 2D. En la Figura 3-7 se muestra un ejemplo de la segmentación de una imagen de RM en dos clases mediante un umbral obtenido por el método de Otsu.

También se han realizado propuestas en el marco de los conjuntos difusos [5]. Un ejemplo lo constituye la asignación a los píxeles de valores de pertenencia en dependencia de sus diferencias respecto a las medias de cada clase [93]. Esta propuesta en su caso extremo coincide con el método de agrupamiento *k-means*.

El histograma de la imagen no contiene información suficiente para caracterizar toda la imagen –en particular no toma en cuenta las características espaciales- por lo que los métodos que se basan solamente en él son sensibles, por ejemplo, al ruido y las heterogeneidades de intensidad que aparecen en las imágenes de resonancia magnética. Eso hace que no siempre logren obtener un umbral apropiado. Una manera de lograr mejoras consiste en utilizar información adicional.

Es por esa razón que se han propuesto métodos de umbralización que utilizan, además de la distribución de los niveles de intensidad, la información espacial en las vecindades de los píxeles. La información espacial se ha utilizado, por ejemplo, en forma de frecuencias de co-ocurrencia de valores de píxeles en cierta vecindad, mediante funciones correlación, o modelos de dependencia lineal entre los valores de intensidad de los píxeles. Uno de los primeros intentos en este sentido es el de [94] que tiene en cuenta el promedio de los niveles de intensidad en una vecindad local con vistas a la definición de un umbral.

También se ha ensayado el uso de la matriz de co-ocurrencias de niveles de grises para la umbralización [95]. La matriz de co-ocurrencias se define a través de la frecuencia de aparición de los niveles de grises a una distancia y en una dirección dadas [81]. De esta manera se obtiene un histograma bidimensional que captura la dependencia espacial de los niveles de gris en la imagen.

En la Figura 3-8 se muestra el caso de una imagen y la matriz de co-ocurrencias calculada al asociar un píxel dado y el que se encuentra “a la derecha y hacia abajo”. Se resalta el par 1–1 según el desplazamiento definido.

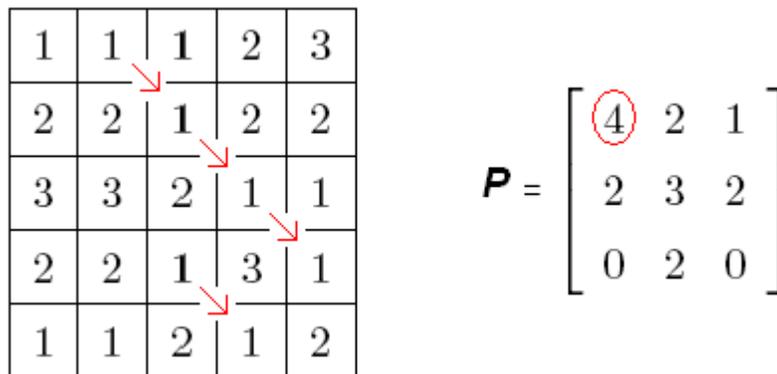


Figura 3-8 Ejemplo de una imagen con 3 niveles de grises y su matriz de co-ocurrencias.

Normalmente se utilizan valores derivados de esta matriz como características de textura las cuales pueden ser utilizadas en tareas de clasificación [2, 96]. Utilizando las matrices de co-ocurrencias se han propuesto varios algoritmos de umbralización [97]. Para más detalles se puede consultar la reseña de [89].

La umbralización se utiliza en ocasiones como una etapa inicial en los procesos de segmentación. También se han propuesto modificaciones a las variantes clásicas en el caso de la segmentación de imágenes médicas, las cuales incluyen información local sobre las intensidades, y sobre la conectividad [98].

Al utilizar la umbralización en el caso de la segmentación de imágenes médicas, el objetivo que se plantea consiste en determinar valores de intensidad, es decir umbrales, que permitan separar los diferentes tipos de tejidos.

Una posible vía de determinación de un umbral consiste en seleccionar de una manera iterativa cada posible valor y evaluar la segmentación resultante tomando en cuenta ciertos criterios que permitan evaluar la calidad del resultado final. Este enfoque ha sido utilizado para la segmentación de imágenes de cerebro de RM [98, 99].

Otra alternativa se basa en el hecho de que cada tipo de tejido posee una intensidad característica, y por tanto es de esperar que el histograma de intensidades exhiba un pico discernible para cada uno [100]. Luego, el problema de segmentar los tipos de tejidos se transforma en un problema, presumiblemente más simple, de encontrar esos picos.

En este línea, por ejemplo, [101] utilizan el algoritmo *k-means* para obtener una partición del histograma en K clusters (picos). En cada paso se ajusta el valor central de cada cluster de manera que se minimice la suma de las diferencias cuadráticas entre cada píxel y el valor central de intensidad de su cluster; con la subsiguiente posible modificación de la pertenencia de los píxeles en correspondencia con los nuevos valores de intensidad central que se obtienen. Este proceso se repite hasta que no ocurran cambios en los clusters.

Otros ejemplos de segmentación de imágenes cerebrales de RM basados en la detección de picos en el histograma de intensidades pueden ser consultados en [102, 103].

En lugar de la búsqueda directa de los picos se han utilizado otras ideas. Por ejemplo, [99] utilizan el percentil 70 de toda el área del histograma como cota superior del rango asociado al pico correspondiente al aire ya que el fondo negro asociado a este ocupa, por lo general, el 70 por ciento de toda el área de la imagen.

Cabe señalar que en el histograma de una imagen típica de RM de cráneo los picos en general no se encuentran tan bien definidos y puede resultar problemático la selección de umbrales apropiados.

Varios factores influyen en esta situación. Por ejemplo, el ruido provoca un ensanchamiento de los picos y una reducción de su altura. Por otra parte, otros tipos de

tejidos, como son el caso de los músculos y la grasa, aparecen con rangos de intensidad que se solapan con los tejidos cerebrales. Aún los mismos tejidos cerebrales exhiben un rango de valores amplio, toda vez que tanto la sustancia gris, como la sustancia blanca, están compuestos por núcleos celulares y por axones aunque en diferentes proporciones.

Por otra parte el efecto de volumen parcial puede también contribuir a una pobre definición de los picos. Esto puede ocurrir en zonas donde, producto a la resolución espacial de la imagen, concurren varios tipos de tejidos en un mismo píxel (o vóxel) con lo cual la intensidad resultante no se corresponde con la de un tipo de tejido en específico. Este efecto se puede mitigar si se opta por no incluir en el histograma aquellos píxeles que se encuentran en zonas de alta variación de intensidad, tal y como se propone por [103].

Estos factores provocan que, aún en el caso de que los picos sean identificables, la región de solapamiento entre ellos no muestre una clara división que permita establecer un umbral de manera única y definitiva. Por esta razón solamente en muy raras ocasiones los métodos de umbralización por sí solos brindan resultados aceptables, y por tanto se utilizan de manera combinada con otros.

3.2.2 Crecimiento de regiones

El crecimiento de regiones [104] es uno de los métodos más simples basados en regiones. Se inicia con la selección de uno o varios puntos iniciales o semillas. La forma más habitual consiste en la selección aleatoria, o siguiendo cierta dirección de recorrido de la imagen. También son posibles otras formas de especificación de las semillas.

Cada región crece mediante la unión de los píxeles vecinos que sean similares según cierto criterio de homogeneidad. En caso de no cumplirse el criterio para un píxel no etiquetado - lo que significa que el píxel pertenece a otro objeto-, entonces se detiene la búsqueda en esa dirección.

En calidad de criterios para el crecimiento de las regiones se puede seleccionar cierta medida de homogeneidad de la región, exigiendo que ésta no supere un umbral predefinido. También es posible exigir que la diferencia entre el valor de intensidad de un píxel dado y un prototipo de la región sea suficientemente pequeña.

La idea de estos métodos consiste en formar regiones conexas de píxeles adyacentes partiendo de las semillas iniciales y asumiendo cierto criterio de homogeneidad. Precisamente estos son los dos puntos clave de este enfoque: cómo y cuáles semillas escoger, y cómo formular el criterio de homogeneidad.

Supongamos que la imagen está compuesta por K regiones R_1, \dots, R_K , y que para cada región i se tiene un conjunto S_i de semillas iniciales escogidas mediante algún procedimiento automático, o seleccionadas manualmente por un operador. Denotemos por

$N(p)$ la vecindad de un píxel p dado, y sea $g(p, q)$ una función que sirve como criterio de aglomeración.

Un algoritmo simple para la formación de las regiones se ilustra en la Tabla 3-3. Como resultado se obtiene una imagen E etiquetada, donde $E(p)$ representa el índice de la región a la que pertenece el píxel p . El algoritmo utiliza como estructura de datos auxiliar un conjunto Q .

Tabla 3-3 Algoritmo básico de crecimiento de regiones sin competencia

<p>Algoritmo básico de crecimiento de regiones sin competencia.</p> <ol style="list-style-type: none"> 1. Para todo píxel p de la imagen, si $p \in S_i$, hacer $E(p) = i$ e insertar p en Q. 2. Mientras $Q \neq \emptyset$ hacer 3. Extraer p de Q. 4. Para todo $q \in N(p)$, tal que $E(q) = 0$, hacer 5. Si $g(p, q) = \text{verdadero}$, $E(q) \leftarrow E(p)$ e insertar q en Q. 6. Fin
--

Una cuestión a tener en cuenta en el algoritmo dado anteriormente consiste en el hecho de que es posible que un píxel cumpla el criterio de semejanza para más de una región. Sin embargo; una vez que un píxel p es etiquetado como perteneciente a una R_i esta decisión no puede ser revocada, aún en el caso que el píxel p posea propiedades que lo hagan más semejante a otra región R_j , $i \neq j$, cuyas semillas sean analizadas con posterioridad. Por otra parte, el criterio que se utiliza para decidir si un píxel pertenece a un objeto es local y depende sólo de las propiedades del píxel, o de una vecindad de este.

Una propuesta más interesante sería definir una medida de la fuerza de semejanza o de conectividad de un píxel p respecto a una región R_i . Esta medida puede estar dada, por ejemplo, por el costo de un camino óptimo desde S_i hasta el píxel p .

En esta línea se ha propuesto el enfoque de la conectividad difusa absoluta [105, 106] para el caso de un único objeto. Si la fuerza de conectividad es superior a cierto umbral T_i se dice entonces que el píxel p pertenece a la región R_i .

En el caso que se tengan en cuenta varios objetos de manera simultánea la idea sería entonces asignar un píxel p a aquel objeto cuya fuerza de conectividad con él sea mayor. En ese sentido se establece una especie de competencia entre los objetos por el píxel en cuestión. El enfoque de la conectividad difusa relativa [107, 108] se enmarca precisamente en este contexto. En esta propuesta hay que tener en cuenta el hecho de que para cada objeto puede existir su propia medida de conectividad. Esto plantea el problema de cómo

combinar estas medidas individuales toda vez que la validez de los resultados teóricos que se proponen requiere la definición de una medida única.

En la propuesta inicial se utiliza la unión difusa de las medidas, implementada como el máximo de éstas [107], aunque como se señala en el mismo artículo, este es un punto que merece ser investigado con el fin de obtener formas más sofisticadas de combinar las medidas individuales que permitan mejorar la efectividad de la segmentación.

Otro ejemplo de crecimiento de regiones con competencia entre las semillas lo constituye la transformada *watershed* (líneas de unión de aguas) [109]. Si vemos una imagen con una superficie 3D, donde la altura de cada píxel se corresponde con su valor de intensidad, entonces podemos esperar que los objetos de interés se encuentren a diferentes alturas y están delimitados por elevaciones más altas que las elevaciones internas y externas. Esto puede ocurrir, sobre todo, cuando la imagen que se analiza se corresponde con algún tipo de gradiente de la imagen original. Los “valles” –mínimos locales- que se forman pueden ser utilizados, en principio, como punto de partida o semillas.

No existe una definición única y precisa de la transformada *watershed* en la literatura [110]. De manera informal se describe como un proceso de inundación en la superficie de la imagen, comenzando con una fuente de agua en cada semilla y creando cuencas de inundación. Cuando el agua proveniente de diferentes cuencas se encuentra, se construye un dique para evitar que se mezclen cuencas diferentes. Estos diques forman las líneas de unión de agua (*watershed lines*), que se deben corresponder con los bordes de los objetos.

Para el caso del crecimiento de regiones en lugar de comenzar la inundación por los mínimos locales, lo cual genera un fenómeno de sobre-segmentación (*oversegmentation*), es más conveniente utilizar marcadores, obteniéndose una cantidad dada de objetos solamente.

En la Figura 3-9 se muestra un ejemplo de segmentación utilizando este enfoque. Ahí se aprecia el fenómeno de sobre-segmentación y las posibles mejoras que se pueden obtener mediante la utilización de marcadores. Se debe señalar que esto es solamente un ejemplo ilustrativo.

Los enfoques anteriormente descritos se pueden implementar mediante la denominada Transformada Imagen – Foresta (*Image Foresting Transform, IFT*) [111]. Esta propuesta se basa en un enfoque grafo – teórico que utiliza la conectividad entre píxeles. La *IFT* define una floresta de caminos de costo mínimo sobre un grafo cuyos nodos son los píxeles de la imagen y cuyos arcos se definen mediante una relación de adyacencia entre píxeles. El costo de un camino se define mediante una función de costo dependiente de la aplicación, y por lo general tiene en cuenta propiedades locales de la imagen a lo largo de un camino – tales como, la intensidad, el gradiente, la posición de los píxeles.

Las raíces de este bosque comienzan en un conjunto inicial de semillas. Para funciones de costo adecuadas la *IFT* asigna a un camino de costo mínimo desde el conjunto de semillas

hasta cada píxel de la imagen, de manera tal que la unión de estos caminos es un bosque orientado que cubre toda la imagen. Para cada píxel de la imagen se obtiene tres atributos: su antecesor en el camino óptimo, el costo del camino óptimo, y la raíz –o alguna etiqueta asociada. Para más detalles y algunas aplicaciones se puede consultar a [111].

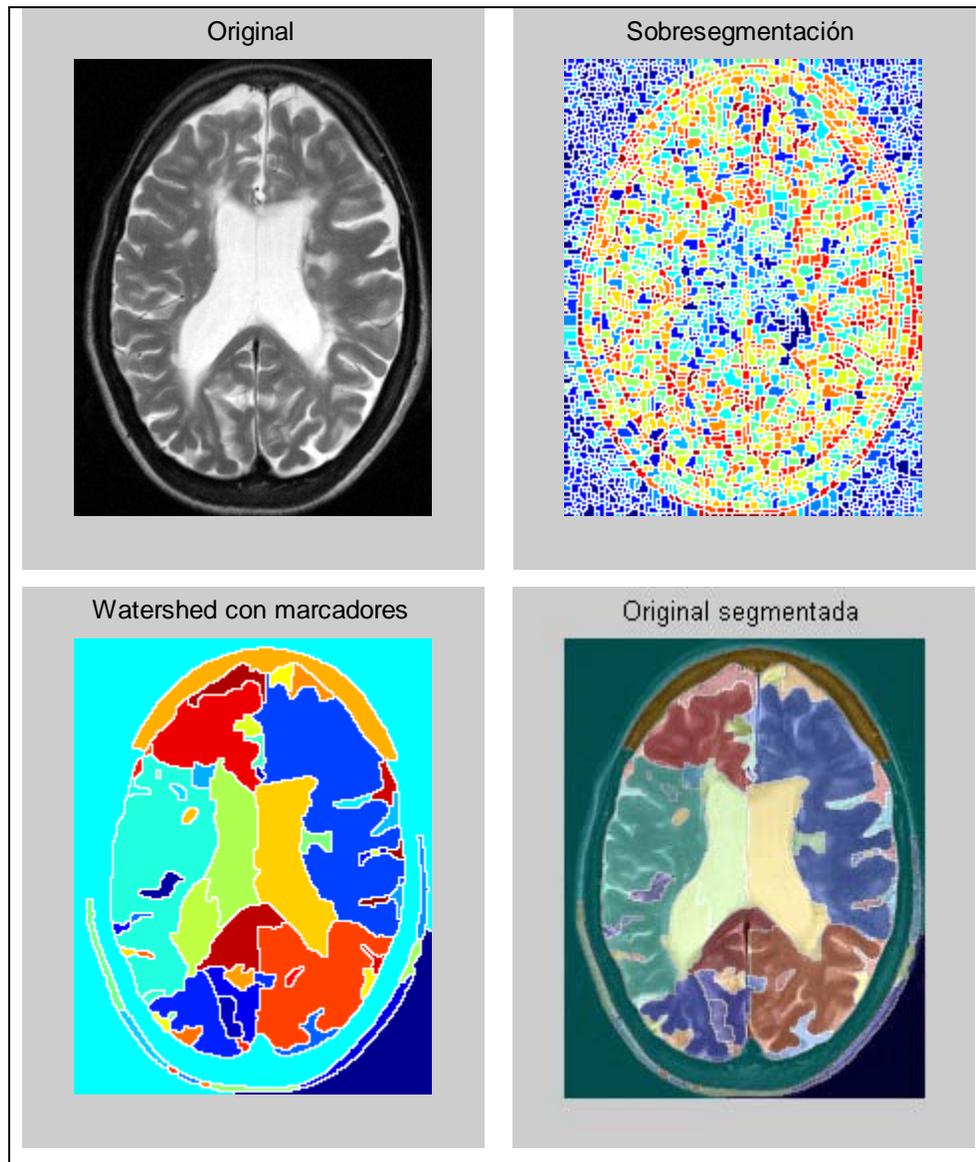


Figura 3-9 Segmentación mediante la transformada *watershed*.

3.2.3 División y fusión

La estrategia de crecimiento de regiones parte de un conjunto dado de semillas. Sin embargo, son posibles otros dos enfoques que prescinden de la utilización de semillas iniciales: uno ascendente y otro descendente. El primero consiste en partir de una serie de regiones que cubran toda la imagen y que se consideren homogéneas en cierto sentido y comenzar a unir aquellas que posean fronteras comunes con la condición de que la nueva región que se forme sea también homogénea. El segundo enfoque parte de la imagen como un todo y prosigue con la división sucesiva de aquellas regiones que no sean homogéneas mientras exista alguna.

También es posible utilizar una estrategia híbrida es decir de división y fusión. En esta línea se encuentra el método propuesto por [112]. La idea consiste en ir en sentido opuesto al crecimiento de regiones y consta de dos fases. En la primera, se parte de una única región que representa toda la imagen; si dicha región no satisface un cierto criterio o predicado de homogeneidad, entonces la región se subdivide de manera secuencial en subregiones de las cuales se analiza su homogeneidad. Si una subregión está compuesta de puntos homogéneos, no se subdivide. En una segunda fase, las subregiones obtenidas que sean homogéneas son agrupadas. Por esta razón este procedimiento se conoce como método de división y fusión (*split and merge*).

Usualmente se utiliza la estructura de árboles cuaternarios (*quad-trees*) en la fase de división. La idea consiste en dividir cada región que no satisface el criterio de homogeneidad en cuatro subregiones, comenzando por la imagen en su conjunto, y continuando de manera recursiva.

Si se utiliza solamente la división se obtendrían regiones adyacentes de iguales características. Para evitar esto se puede utilizar entonces la fusión. La etapa de fusión consiste en la unión de los bloques adyacentes que fueron obtenidos mediante la descomposición anterior.

El procedimiento se puede resumir de manera esquemática de la siguiente manera (ver Tabla 3-4).

Tabla 3-4 Esquema del método de división y fusión

Algoritmo básico del método de división y fusión. 1. Comenzar con toda la imagen como un bloque. 2. Mientras existan regiones R_i no homogéneas hacer 3. Dividir en cuatro cuadrantes la región R_i . 4. Fusionar aquellas regiones R_k y R_l que sean adyacentes y cumplan en su conjunto con el criterio. 5. Fin
--

La Figura 3-10 muestra un ejemplo de este método. Se trata de un objeto semioscuro único con intensidad constante sobre un fondo blanco. El predicado criterio que se asume plantea que una región es homogénea si todos sus píxeles tienen la misma intensidad.

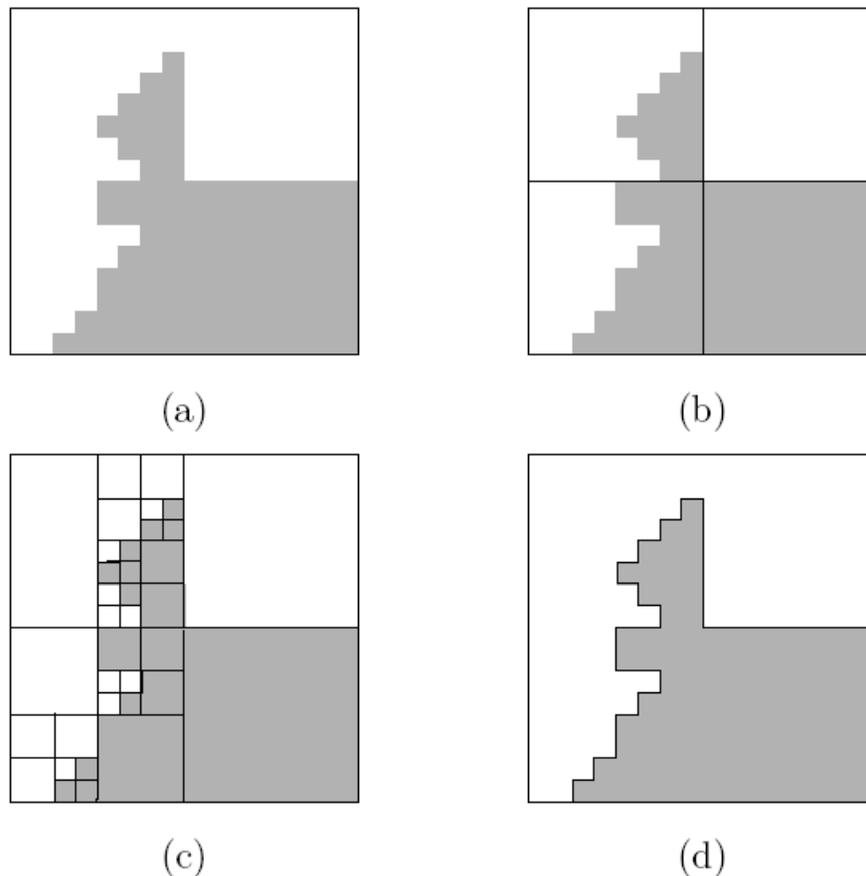


Figura 3-10 Segmentación mediante división y fusión. (a) representa la imagen original; (b) la división inicial en cuatro bloques; (c) la división final en bloques homogéneos; y (d) la segmentación resultante del proceso de fusión.

En la primera división solamente la región inferior derecha cumple el predicado, por lo que no se divide. El resto de las regiones se dividen en cuatro cuadrantes como se muestra en la Figura 3-10b. Siguiendo el procedimiento se llega a la división final en regiones homogéneas Figura 3-10c, y por último se realiza la fusión de regiones adyacentes que cumplan con el criterio de homogeneidad (Figura 3-10d), aunque estas no sean de igual tamaño. Se debe señalar que es posible intercalar los procesos de división y fusión según se especifica en el esquema de la Figura 3-11.

En la Figura 3-11 se muestra un ejemplo de este proceso de descomposición. En el ejemplo se parte de un bloque inicial de tamaño 256×256 . Las subsiguientes divisiones forman 4 bloques de 128×128 . Si alguno de estos no es homogéneo entonces se divide en 4 bloques de tamaño 64×64 , y así sucesivamente hasta llegar a bloques de tamaño 1×1 . Aquí un

bloque se considera homogéneo cuando el valor absoluto de la diferencia entre la mayor y la menor intensidad de un bloque no supera un umbral dado.

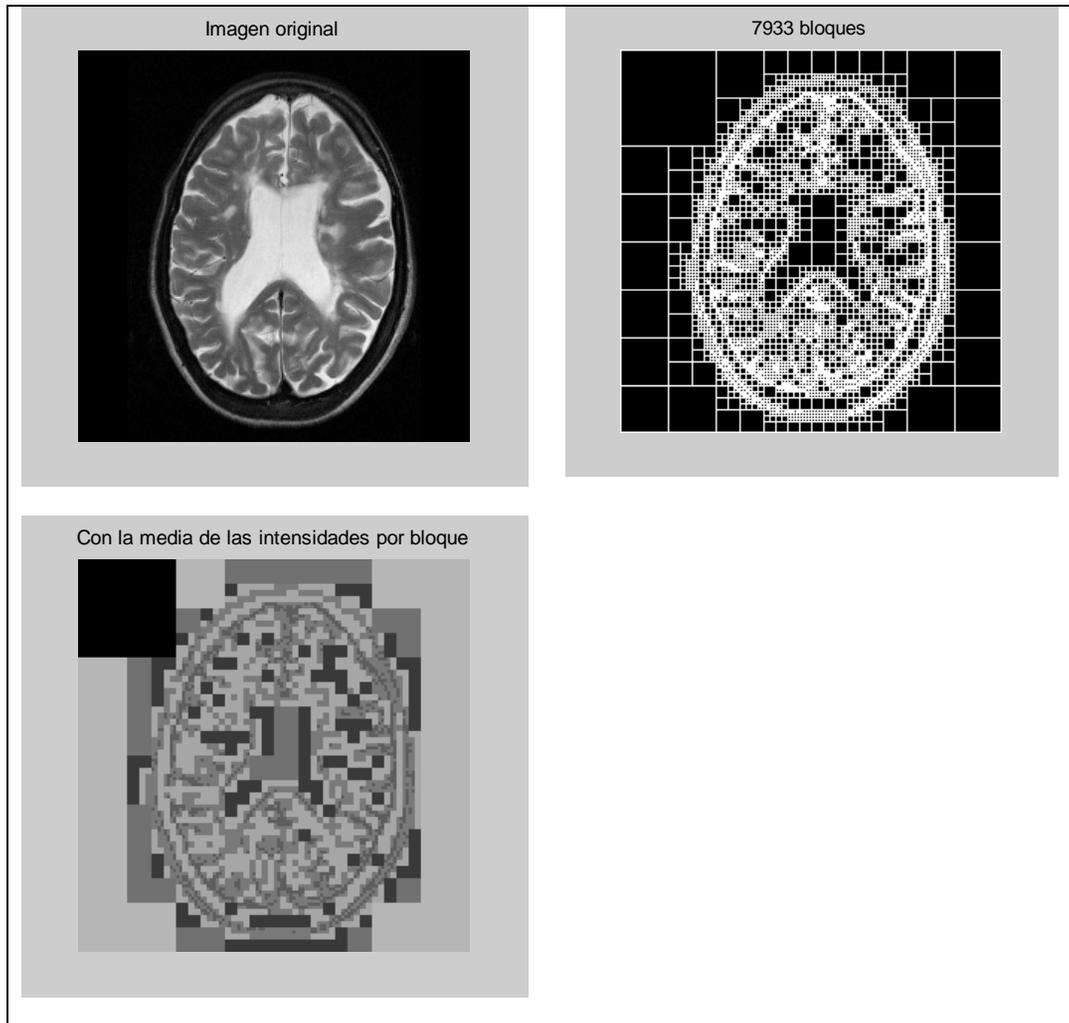


Figura 3-11 Descomposición de una imagen mediante árboles cuaternarios.

Los métodos de crecimiento de regiones y los de división y fusión se han utilizado ampliamente en la segmentación de imágenes, tanto de manera independiente como combinados con otros métodos [113, 114].

El algoritmo de segmentación rápida adaptativa (*Fast Adaptive Segmentation*) parte de un conjunto de regiones primitivas pequeñas que se asumen homogéneas. Estas regiones primitivas se comienzan a mezclar hasta que no sea posible ninguna otra fusión. Dos regiones son fusionadas si la nueva región que se obtiene cumple el criterio de homogeneidad y además están separadas por un borde débil [115]. En el trabajo se investiga el efecto que tienen diferentes criterios de fusión en la calidad del resultado final, así como en el tiempo de procesamiento.

También se ha analizado el efecto de la selección manual de las semillas para la segmentación de imágenes de tonos de grises en una propuesta conocida como *Seeded Region Growing (SRG)* [116]. Cada semilla genera una región que no es más que un componente conexo compuesto por uno o más píxeles. Para un píxel dado se obtiene el conjunto de píxeles vecinos. Cada píxel vecino se analiza con vistas a determinar si es adyacente a alguna región. En caso de ser así se calcula una medida que representa una diferencia entre el píxel y la región. En caso de que el píxel dado sea adyacente a más de una región, entonces se calcula el menor valor de la medida y el píxel se adjudica a dicha región. El proceso continúa hasta que todos los píxeles sean asimilados por alguna región. Este algoritmo es dependiente del orden de exploración de la imagen, es decir, diferentes órdenes de exploración pueden llevar a resultados diferentes.

Con vistas a superar el efecto de orden de exploración sobre los resultados que se obtienen, se ha propuesto una mejora al algoritmo anterior conocida como *Improved Seeded Region Growing (ISRG)* que posee además un mayor grado de paralelismo [117]. Se propone un esquema de desempate cuando más de un píxel vecino posea el mismo valor mínimo de la medida en cuyo caso todos ellos se procesan en paralelo. No se realiza ninguna actualización hasta que todos los píxeles con la misma prioridad hayan sido examinados. Si un píxel dado no puede ser etiquetado debido a que pudiera ser asignado más de una región debido entonces se clasifica como “empatado” y no toma parte en el proceso de crecimiento. Una vez analizados todos los píxeles se examinan entonces los “empatados” con vistas a resolver el conflicto. Para ello se utiliza un criterio adicional que puede ser, por ejemplo, asignarlo a la región vecina de mayor tamaño. En los experimentos realizados los autores concluyen que el algoritmo produce segmentaciones consistentes.

La propuesta de [118] consiste en la obtención de una descomposición jerárquica de la imagen para la que se parte de una partición inicial de la imagen y se aplica una estrategia de optimización paso a paso de manera que en cada iteración se mezclan aquellos segmentos que minimizan un criterio de mezcla. Eso significa que en cada iteración se fusionan los segmentos más semejantes. De esta manera el algoritmo va produciendo una secuencia de particiones que refleja la estructura jerárquica de la imagen. Este enfoque coincide en esencia con la estrategia de agrupamiento jerárquico aglomerativo que se comentó antes en los métodos de *clustering*.

Al igual que en el caso de la umbralización, los métodos basados en regiones por sí solos por lo general no brindan resultados satisfactorios. Es por esa razón que también aquí se han ensayado estrategias híbridas [119].

Un análisis acerca de diferentes estrategias que integran información de los bordes y de las regiones con vistas a mejorar los resultados de la segmentación puede ser consultada en [120].

PARTE II

4 Sobre la autocorrelación espacial

4.1 Introducción

El análisis de datos espaciales introduce en sus técnicas las propiedades topológicas o geométricas que posee el conjunto de los objetos del análisis. Estas técnicas han empezado a desarrollarse en la segunda mitad del siglo XX, y están en auge debido a la cantidad de nuevos problemas que plantean las relaciones espaciales, en relación al punto de vista clásico de ausencia de relación espacial entre los objetos, y la gran variedad de campos de aplicación en las que se plantean.

Las interacciones espaciales pueden ser interpretadas a luz del concepto de autocorrelación, entendida ésta como la correlación entre los valores de la misma variable. La autocorrelación es una característica común a muchos conjuntos de datos. Por ejemplo, los valores de intensidad de los píxeles de una imagen digital están relacionados, en general, en función de su proximidad en la imagen misma. También las páginas web enlazadas mediante hipervínculos son más propensas a compartir un mismo tema que aquéllas que sean seleccionadas al azar [121]. De la misma manera, es más probable que las proteínas que se encuentran en las mismas localizaciones de una célula compartan la misma función que proteínas seleccionadas al azar [122].

La cercanía espacial frecuentemente conduce a observaciones autocorrelacionadas debido, por ejemplo, al efecto de algún fenómeno o variable [123].

La interacción temporal puede considerarse como el caso particular de la interacción espacial de una única dimensión.

En este capítulo se introducen modificaciones de índices clásicos de autocorrelación espacial con el fin de equipararlos mejor. Se presentan sus propiedades fundamentales, se proponen otros índices, y se analizan algunos casos particulares de datos estructurados espacialmente.

4.2 La autocorrelación espacial

Dado un conjunto de objetos, $\Omega = \{\omega_i / i = 1, \dots, n\}$, que están situados en un mapa, es decir, están estructurados espacialmente, se les observan una o varias características, y respecto a éstas, se analiza una propiedad o un patrón de organización en el mapa. La autocorrelación espacial trata, precisamente, de delatar ese “patrón de organización”. Este término se presta

a diversas interpretaciones. Es por eso que los autores refinan un poco la definición, y la autocorrelación espacial se puede definir de varias maneras.

En una interpretación, que podemos calificarla de descriptiva, se plantea que si los valores relativamente altos (bajos) de la magnitud analizada en una determinada localización están acompañados de valores relativamente altos (bajos) en las localizaciones vecinas, entonces puede hablarse de la existencia de autocorrelación positiva. Si por el contrario, los valores relativamente altos (bajos) se alternan con valores relativamente bajos (altos) en las localizaciones vecinas, la autocorrelación será negativa [124].

Una interpretación probabilista plantea que la autocorrelación espacial es una característica de los datos según la cual la presencia de una determinada calidad o cantidad de la variable analizada en las zonas vecinas hace más o menos probable su presencia en una región dada [125].

Otra interpretación de este concepto está asociada a la independencia, es decir, en qué medida el valor observado de una variable en una localización es independiente de los valores de la misma variable en las localizaciones vecinas [126].

Estas dos últimas interpretaciones no son las que interesan en el desarrollo de este trabajo, sino la primera, la calificada como descriptiva.

La propensión de una variable a tomar valores semejantes (o diferentes) a sus vecinos se denomina dependencia espacial. Los índices dedicados a la medición de la autocorrelación espacial parten de una raíz común, materializada en una matriz, $P = [\rho_{ij}]$, cuyos valores dependen de los valores que tome(n) la(s) variable(s) en cuestión para cada par de objetos (ω_i, ω_j) , y de un estadístico basado en el mismo. Por ejemplo [127]:

$$\Gamma = \sum_i \sum_j a_{ij} \rho_{ij}$$

Aquí los valores de la matriz $A = [a_{ij}]$ representan una forma de medición de la adyacencia o contigüidad en los datos originales. La matriz A se denomina de manera equivalente como matriz de adyacencias, de conexión, de contigüidad o de pesos espaciales. Por su parte, los valores de la matriz $P = [\rho_{ij}]$ miden la proximidad entre los objetos ω_i y ω_j en otra dimensión distinta a la propiamente espacial, es decir, la de la(s) variable(s) observada(s). La medida de proximidad puede estar asociada a una función de desemejanza o de semejanza entre los objetos.

Con vistas a representar la conexión de los datos, A , es posible utilizar la contigüidad de los objetos en el mapa (1 , si ω_i y ω_j son contiguos, 0 , si no lo son), o pesos basados en las distancias en el mapa entre cada par de objetos.

La contigüidad puede ser considerada a partir del concepto de vecindad. Por ejemplo, en un retículo regular es común utilizar las 4- u 8-vecindades (ver Figura 4-1).

En general, estas vecindades pueden caracterizarse a partir de un radio y la distancia habitual euclidiana. Se consideran vecinos aquellos objetos que se encuentren a una distancia menor o igual que un r arbitrario estipulado. Así, la 4-vecindad se corresponde con $r=1$, y la 8-vecindad a $r=\sqrt{2}$. En este trabajo se consideran también vecindades de radios mayores.

En los casos de regiones irregulares se pueden considerar contiguos aquellos objetos que compartan alguna frontera en común. Si se tratara de puntos esparcidos en un cierto espacio se pueden considerar vecinos los k puntos más próximos, con k arbitrario.

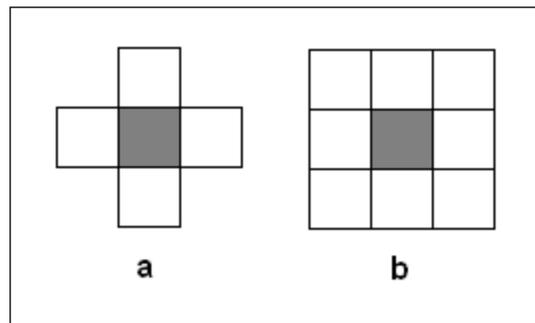


Figura 4-1 Tipos de adyacencia. (a) 4 – vecinos; (b) 8 – vecinos.

En cualquiera de los casos anteriores se suele definir $a_{ij} = 1$ siempre y cuando los objetos ω_i y ω_j se consideren contiguos o vecinos, y $a_{ij} = 0$, si no. Es arbitrario considerar si un objeto es o no adyacente a sí mismo (propiedad reflexiva), o si es elemento de su propia vecindad. Habitualmente se considera que no, es decir, que $a_{ii} = 0$. En el desarrollo de este capítulo esta cuestión resulta irrelevante, y la decisión dependerá del contexto práctico en se aplique la noción de autocorrelación.

Otra alternativa para la definición de la matriz de adyacencias, A , es a través de alguna noción de distancia espacial, d_{ij} , entre los objetos ω_i y ω_j , que puede ser lineal o a través de cierta red. Dada la distancia d_{ij} se puede considerar, por ejemplo, $a_{ij} = 1/d_{ij}$, $a_{ij} = 1/d_{ij}^2$, o sus variantes exponenciales, $a_{ij} = e^{-d_{ij}}$ y $a_{ij} = e^{-d_{ij}^2}$. En todos los casos se debe tener en cuenta la posible indefinición en el caso de que algún $d_{ij} = 0$, es decir, el caso en que dos objetos se ubiquen igualmente en el mapa, o en el caso de la propiedad reflexiva, pues $d_{ii} = 0$.

Es importante tener en cuenta que al utilizar diferentes matrices de pesos espaciales se deben esperar diferentes valores para las medidas de autocorrelación. Las matrices de pesos espaciales como vía para representar las relaciones entre los objetos dan pie a la utilización de diferentes técnicas de análisis.

A continuación se presenta una versión modificada de los índices clásicos de Moran y Geary, que toma en cuenta las relaciones de contigüidad en la definición de cualquier estadístico previo usado en su definición, y que permite tratarlos de una manera unificada, además de lograr otras propiedades interesantes. A partir de las ideas de Lebart es posible derivar un índice en este mismo contexto, y se presentará una versión modificada. Existen otros índices para evaluar la autocorrelación espacial, pero quedan fuera de este contexto [128].

4.2.1 Definiciones y propiedades estadísticas previas

En este apartado se presentan las definiciones de la media y la varianza en las que se basan los índices modificados de autocorrelación propuestos. Luego, se definen nuevas variables en función de las vecindades, y se calculan sus características propias, así como sus covarianzas respecto a las demás variables, con el fin de ayudar a interpretar los índices de autocorrelación.

Sea Ω el conjunto de los n objetos, ω_i , $i=1,2,\dots,n$, cada uno con un valor x_i asociado, y relacionados en una estructura de adyacencia expresada mediante una matriz de pesos espaciales $A=[a_{ij}]$, donde $a_{ij}=1 \forall i,j$, con $i,j=1,\dots,n$, si los objetos ω_i y ω_j son adyacentes, y $a_{ij}=0$ en otro caso. La matriz de pesos espaciales puede ser vista como la matriz de adyacencias de un grafo donde cada objeto representa un vértice. En general, se suele considerar que $a_{ii}=0$, pero no es necesario que sea así para las definiciones y propiedades que siguen.

Una representación conjunta de los valores y los pesos consiste en desplegar en una fila las filas de la matriz A , en otra fila los valores correspondientes al objeto del primer subíndice de los pesos, y en otra los del objeto del segundo subíndice. Es análoga a la que habitualmente se hace para describir los datos brutos de dos variables sobre un conjunto de objetos. Esta representación tabular es la siguiente

<table style="border-collapse: collapse; margin: auto;"> <tr> <td style="padding: 5px;">x_1</td> <td style="padding: 5px;">x_2</td> <td style="padding: 5px;">\dots</td> <td style="padding: 5px;">x_n</td> </tr> <tr> <td style="padding: 5px;">y_1</td> <td style="padding: 5px;">y_2</td> <td style="padding: 5px;">\dots</td> <td style="padding: 5px;">y_n</td> </tr> </table>	x_1	x_2	\dots	x_n	y_1	y_2	\dots	y_n	<table style="border-collapse: collapse; margin: auto;"> <tr> <td style="padding: 5px;">\dots</td> <td style="padding: 5px;">a_{i1}</td> <td style="padding: 5px;">a_{i2}</td> <td style="padding: 5px;">\dots</td> <td style="padding: 5px;">a_{in}</td> <td style="padding: 5px;">\dots</td> </tr> <tr> <td style="padding: 5px;">\dots</td> <td style="padding: 5px;">x_i</td> <td style="padding: 5px;">x_i</td> <td style="padding: 5px;">\dots</td> <td style="padding: 5px;">x_i</td> <td style="padding: 5px;">\dots</td> </tr> <tr> <td style="padding: 5px;">\dots</td> <td style="padding: 5px;">x_1</td> <td style="padding: 5px;">x_2</td> <td style="padding: 5px;">\dots</td> <td style="padding: 5px;">x_n</td> <td style="padding: 5px;">\dots</td> </tr> </table>	\dots	a_{i1}	a_{i2}	\dots	a_{in}	\dots	\dots	x_i	x_i	\dots	x_i	\dots	\dots	x_1	x_2	\dots	x_n	\dots
x_1	x_2	\dots	x_n																								
y_1	y_2	\dots	y_n																								
\dots	a_{i1}	a_{i2}	\dots	a_{in}	\dots																						
\dots	x_i	x_i	\dots	x_i	\dots																						
\dots	x_1	x_2	\dots	x_n	\dots																						
(a)	(b)																										

Figura 4-2 Analogía en la representación tabular de los datos: (a) habitual, y (b) datos espaciales.

Conviene remarcar este tipo de representación tabular de un grafo. Desde el punto de vista estadístico se suele definir una población o universo, Ω , y cada elemento del mismo, ω , es un registro, individuo, objeto, o unidad estadística. Las estadísticas se hacen sobre los valores que se observan en esos objetos.

Cuando el conjunto de objetos está estructurado es necesario incorporar esta información en las estadísticas. En la representación tabular recién presentada, que es la idea básica de todo lo que se desarrolla a continuación, se considera como registro o unidad estadística, no un objeto, sino un elemento de adyacencia al que le corresponden dos valores, los de los objetos adyacentes. Un objeto queda representado por su conjunto de adyacencias, y el universo Ω por el conjunto de adyacencias definido por la matriz A sobre Ω .

Por ejemplo, una imagen se puede considerar como un retículo regular, y la intensidad en cada píxel como la variable en cuestión (ver Figura 4-3). En el ejemplo la matriz de adyacencias A se define mediante la 4-vecindad, según la cual para cada píxel ω_i son adyacentes los píxeles que se encuentran encima, debajo, a la derecha y a la izquierda.

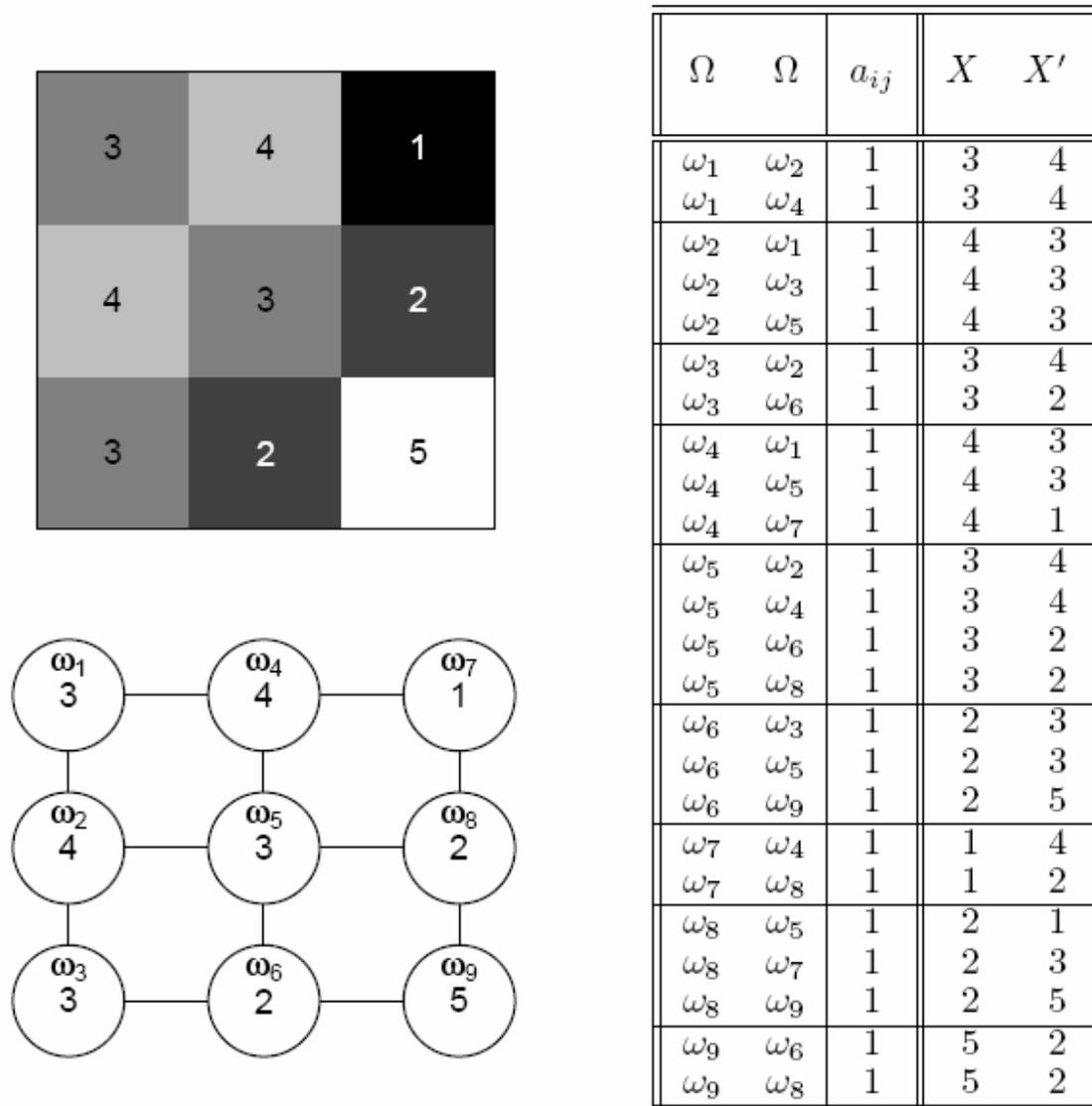


Figura 4-3 Imagen de ejemplo

Si se consideran los 4-vecinos de los objetos ω_i , y se observan sus valores, se puede apreciar que ω_1 tiene una vecindad homogénea no muy diferente del valor que se le ha asignado. Lo mismo puede decirse de ω_2 .

Por su parte ω_9 tiene una vecindad totalmente homogénea, pero muy diferente a su valor. En el caso de ω_4 y ω_6 ambos tienen la vecindad un poco más homogénea, pero más diferente del valor asignado.

El píxel ω_3 , aún no teniendo una vecindad muy diferente a su valor, tiene una vecindad más heterogénea. Lo mismo puede decirse de ω_5 , aún siendo su vecindad más amplia. Para ω_7 se aprecia que tiene una vecindad del mismo tipo, pero él difiere un tanto de sus vecinos. Finalmente, ω_8 tiene una vecindad muy heterogénea.

Las definiciones que siguen, y sus propiedades, pretenden poner de manifiesto numéricamente este tipo de apreciaciones mediante índices convenientes. En los comentarios se supone que los valores a_{ij} son 0's ó 1's (caso habitual de una imagen), pero no se pierde la generalidad de las fórmulas para cualesquiera que sean los valores de adyacencia comentados anteriormente.

El peso espacial de cada objeto ω_i se define de la siguiente manera

$$A_i = \sum_{j=1}^n a_{ij}$$

y coincide con el grado del vértice asociado al objeto ω_i en la representación basada en grafos.

El peso espacial total se define como

$$A_{Total} = \sum_{i=1}^n A_i = \sum_{i=1}^n \sum_{j=1}^n a_{ij}$$

El valor A_{Total} representa la cantidad total de conexiones presente en los datos, contadas doblemente si $a_{ii} = 0$, toda vez que $a_{ij} = a_{ji}$. Es decir, $A_{Total}/2$ se corresponde la cantidad de aristas del grafo no orientado definido tomando como vértices los objetos y con matriz de adyacencias A .

Considerando los pesos de cada objeto, la media de los valores se define como

$$\bar{x}_A = \frac{1}{A_{Total}} \sum_{i=1}^n A_i x_i$$

es decir, la contribución de cada objeto a la media es ponderada según su peso espacial.

En el caso de un retículo regular, el efecto de esta ponderación provoca que la contribución de los objetos con menos (más) vecinos sea menor (mayor). Es decir, los bordes o márgenes del retículo contribuyen en menor medida que los puntos interiores, y las esquinas menos aún.

La media, así definida, será denominada como *media modificada*. En el cálculo de la media habitual no se consideran pesos espaciales

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i,$$

o lo que es equivalente, se considera que todos los objetos tienen el mismo peso espacial.

La definición de la media modificada, al tratarse de un sistema de pesos o ponderaciones, mantiene la propiedad habitual de la media, a saber

$$\frac{1}{A_{Total}} \sum_{i=1}^n A_i (x_i - \bar{x}_A) = \frac{1}{A_{Total}} \sum_{i=1}^n A_i x_i - \frac{1}{A_{Total}} A_{Total} \bar{x}_A = 0$$

La varianza de los valores se define, en consonancia con lo anterior, de la siguiente manera

$$s_A^2 = \frac{1}{A_{Total}} \sum_{i=1}^n A_i (x_i - \bar{x}_A)^2$$

a diferencia de la varianza habitual,

$$s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2.$$

La varianza, así definida, es denominada también como *varianza total modificada*.

Cuando los pesos A_i de cada objeto son iguales, la media y la varianza modificadas coinciden con las habituales. En el caso de retículos regulares (imágenes) y 4-vecindad, los puntos interiores tienen peso proporcional a 4, los márgenes a 3 y las esquinas a 2. Si el retículo es grande (resolución alta de imágenes), la influencia de los márgenes es pequeña, y, por tanto, los valores de la media y de la varianza tienden a ser iguales en los sistemas habitual y modificado.

Por supuesto, al tratarse simplemente de un sistema de pesos, se verifica la propiedad fundamental que relaciona la media y la varianza

$$s_A^2 = \min_x \frac{1}{A_{Total}} \sum_{i=1}^n A_i (x_i - x)^2$$

La relación del mínimo con otro tipo de desviaciones es la siguiente

$$s_A^2 = \frac{1}{A_{Total}} \sum_{i=1}^n A_i (x_i - \bar{x})^2 + (x - \bar{x}_A)^2$$

Para ayudar a su interpretación, conviene recordar la propiedad de la varianza que la relaciona con el valor medio de las diferencias entre los valores de todos los objetos tomados de dos en dos donde cada objeto ω_i actúa con su respectivo peso espacial, A_i . El resultado coincide con la varianza s_A^2 (multiplicada por el factor 2, debido al hecho que cada comparación se hace dos veces; también se contabilizan las diferencias nulas de un objeto respecto a sí mismo si $a_{ii} \neq 0$). A continuación se presenta dicha relación.

$$\begin{aligned} & \frac{1}{A_{Total}^2} \sum_{i=1}^n \sum_{j=1}^n A_i A_j (x_i - x_j)^2 = \frac{1}{A_{Total}^2} \sum_{i=1}^n \sum_{j=1}^n A_i A_j ((x_i - \bar{x}_A)(x_j - \bar{x}_A))^2 = \\ & = \frac{1}{A_{Total}^2} \sum_{i=1}^n \sum_{j=1}^n A_i A_j (x_i - \bar{x}_A)^2 + \frac{1}{A_{Total}^2} \sum_{i=1}^n \sum_{j=1}^n A_i A_j (x_j - \bar{x}_A)^2 \\ & \quad - 2 \frac{1}{A_{Total}^2} \sum_{i=1}^n \sum_{j=1}^n A_i A_j (x_i - \bar{x}_A)(x_j - \bar{x}_A) = \\ & = \frac{1}{A_{Total}^2} \sum_{j=1}^n A_j \sum_{i=1}^n A_i (x_i - \bar{x}_A)^2 + \frac{1}{A_{Total}^2} \sum_{i=1}^n A_i \sum_{j=1}^n A_j (x_j - \bar{x}_A)^2 - \\ & \quad - 2 \frac{1}{A_{Total}^2} \sum_{i=1}^n A_i (x_i - \bar{x}_A) \sum_{j=1}^n A_j (x_j - \bar{x}_A) = \\ & = \frac{1}{A_{Total}^2} \sum_{j=1}^n A_j A_{Total} s_A^2 + \frac{1}{A_{Total}^2} \sum_{i=1}^n A_i A_{Total} s_A^2 = 2s_A^2 \end{aligned}$$

El núcleo de todas las definiciones y propiedades que se desarrollan a continuación se basan precisamente en la consideración del valor que toman estas diferencias cuadráticas medias con el sistema de pesos a_{ij}/A_{Total} en lugar de las recién mostradas $A_i A_j/A_{Total}^2$, es decir, en la siguiente expresión

$$\frac{1}{A_{Total}} \sum_{i=1}^n \sum_{j=1}^n a_{ij} (x_i - x_j)^2.$$

Los valores de una variable se tipifican o estandarizan como es habitual, pero respecto a la media y varianza modificadas

$$z_i = \frac{x_i - \bar{x}_A}{s_A}$$

Este conjunto de valores z_i , $i=1,2,\dots,n$, con el sistema de pesos espaciales correspondiente A_i , $i=1,2,\dots,n$, tiene media 0 y varianza 1.

Dada la matriz de adyacencias, A , entre los objetos, se define la covarianza de una variable X consigo misma (autocovarianza) respecto al sistema de pesos a_{ij}/A_{Total} como la media de los productos cruzados de las desviaciones respecto a la media modificada

$$\text{cov}(X, X') = \frac{1}{A_{Total}} \sum_{i=1}^n \sum_{j=1}^n a_{ij} (x_i - \bar{x}_A)(x_j - \bar{x}_A)$$

La variable X' es una reordenación de la variable original X , que expresa los valores en las vecindades de los objetos (ver Figura 4-3). Sus características principales (media y varianza) son las mismas que las de la variable X .

La vecindad de un objeto puede ser caracterizada por la media de sus valores. Este valor puede interpretarse como una suavización o filtrado cuando cada valor es sustituido por la media de su vecindad (ver apartado 3.1)

$$y_i = \frac{1}{A_i} \sum_{j=1}^n a_{ij} x_j.$$

Este conjunto de valores tiene las siguientes características

$$\bar{y} = \frac{1}{A_{Total}} \sum_{i=1}^n A_i y_i = \frac{1}{A_{Total}} \sum_{i=1}^n A_i \frac{1}{A_i} \sum_{j=1}^n a_{ij} x_j = \frac{1}{A_{Total}} \sum_{i=1}^n \sum_{j=1}^n a_{ij} x_j = \frac{1}{A_{Total}} \sum_{j=1}^n A_j x_j = \bar{x}_A,$$

y

$$s_y^2 = \frac{1}{A_{Total}} \sum_{i=1}^n A_i (y_i - \bar{x}_A)^2$$

La covarianza de la variable Y con la variable X , que en adelante será denominada s_{xy} , coincide con la autocovarianza de X con X' , y la covarianza de la variable Y respecto a X' con la varianza de Y .

$$\begin{aligned} \text{cov}(X, Y) &= s_{xy} = \frac{1}{A_{Total}} \sum_{i=1}^n A_i (x_i - \bar{x}_A)(y_i - \bar{x}_A) = \\ &= \frac{1}{A_{Total}} \sum_{i=1}^n A_i (x_i - \bar{x}_A) \frac{1}{A_i} \sum_{j=1}^n a_{ij} (x_j - \bar{x}_A) = \text{cov}(X, X') \end{aligned}$$

$$\begin{aligned}
\text{cov}(Y, X') &= \frac{1}{A_{Total}} \sum_{i=1}^n \sum_{j=1}^n a_{ij} (y_i - \bar{x}_A) (x_j - \bar{x}_A) = \\
&= \frac{1}{A_{Total}} \sum_{i=1}^n (y_i - \bar{x}_A) \sum_{j=1}^n a_{ij} (x_j - \bar{x}_A) = \\
&= \frac{1}{A_{Total}} \sum_{i=1}^n (y_i - \bar{x}_A) A_i (y_i - \bar{x}_A) = s_y^2
\end{aligned}$$

Por otra parte, se pueden considerar los “errores” entre cada valor x_i y su correspondiente valor suavizado según la vecindad, y_i

$$v_i = x_i - y_i.$$

Las características de este conjunto de valores son

$$\bar{v} = \frac{1}{A_{Total}} \sum_{i=1}^n A_i v_i = \frac{1}{A_{Total}} \sum_{i=1}^n A_i (x_i - y_i) = \bar{x}_A - \bar{y} = 0,$$

y

$$s_v^2 = \frac{1}{A_{Total}} \sum_{i=1}^n A_i (x_i - y_i)^2 = \frac{1}{A_{Total}} \sum_{i=1}^n A_i ((x_i - \bar{x}_A) - (y_i - \bar{x}_A))^2 = s_A^2 + s_y^2 - 2s_{xy}$$

Así pues, inversamente, la autocovarianza de la variable X , que coincide con s_{xy} , puede calcularse a partir de las varianzas de estas tres variables

$$s_{xy} = \frac{1}{2} (s_A^2 + s_y^2 - s_v^2)$$

Si la variable Y es un suavizado de la variable X , s_v^2 se puede interpretar como una distancia cuadrática entre la estructura original y su suavizado.

Las relaciones entre esta variable V con las anteriores X , X' e Y , en términos de covarianzas o productos cruzados, son las siguientes

$$\begin{aligned}
\text{cov}(X, V) &= \frac{1}{A_{Total}} \sum_{i=1}^n A_i (x_i - \bar{x}_A) (x_i - y_i) = \\
&= \frac{1}{A_{Total}} \sum_{i=1}^n A_i (x_i - \bar{x}_A) ((x_i - \bar{x}_A) - (y_i - \bar{x}_A)) = \\
&= s_A^2 - s_{xy} = \frac{1}{2} (s_A^2 - s_y^2 + s_v^2)
\end{aligned}$$

$$\begin{aligned}
\text{cov}(X', V) &= \frac{1}{A_{Total}} \sum_{i=1}^n \sum_{j=1}^n a_{ij} (x_j - \bar{x}_A) (x_i - y_i) = \\
&= \frac{1}{A_{Total}} \sum_{i=1}^n (x_i - y_i) \sum_{j=1}^n a_{ij} (x_j - \bar{x}_A) = \\
&= \frac{1}{A_{Total}} \sum_{i=1}^n (x_i - y_i) A_i (y_i - \bar{x}_A) = \\
&= \frac{1}{A_{Total}} \sum_{i=1}^n A_i ((x_i - \bar{x}_A) - (y_i - \bar{x}_A)) (y_i - \bar{x}_A) = \\
&= s_{xy} - s_y^2 = \frac{1}{2} (s_A^2 - s_y^2 - s_v^2)
\end{aligned}$$

$$\begin{aligned}
\text{cov}(Y, V) &= \frac{1}{A_{Total}} \sum_{i=1}^n A_i (y_i - \bar{x}_A) (x_i - y_i) = \text{cov}(X', V) = \\
&= s_{xy} - s_y^2 = \frac{1}{2} (s_A^2 - s_y^2 - s_v^2)
\end{aligned}$$

Consideremos la diferencia entre dos valores $(x_i - x_j)$ para aquellos objetos contiguos, es decir, las diferencias en las vecindades, en lugar de calcularlas respecto a los valores medios

$$w_{ij} = x_i - x_j.$$

Estas diferencias son afectadas por un sistema de pesos para el cálculo de estadísticos. Lebart [129, 130] y Ollier [131] consideran el sistema $A_i A_j / A_{Total}^2$ y $w'_{ij} = A_i A_j (x_i - x_j)$. Aquí, en cambio, se utiliza el sistema de pesos a_{ij} / A_{Total} . Se tienen las siguientes características de este conjunto de valores w_{ij}

$$\bar{w} = \frac{1}{A_{Total}} \sum_{i=1}^n \sum_{j=1}^n a_{ij} w_{ij} = \frac{1}{A_{Total}} \sum_{i=1}^n \sum_{j=1}^n a_{ij} (x_i - x_j) = \bar{x}_A - \bar{x}_A = 0, \text{ y}$$

$$\begin{aligned}
s_w^2 &= \frac{1}{A_{Total}} \sum_{i=1}^n \sum_{j=1}^n a_{ij} (x_i - x_j)^2 = \frac{1}{A_{Total}} \sum_{i=1}^n \sum_{j=1}^n a_{ij} \left((x_i - \bar{x}_A) - (x_j - \bar{x}_A) \right)^2 = \\
&= 2s_A^2 - 2s_{xy} = s_A^2 - s_y^2 + s_v^2
\end{aligned}$$

De esta expresión se deducen otras para las covarianzas anteriores

$$\text{cov}(X, X') = \text{cov}(X, Y) = s_{xy} = s_A^2 - \frac{1}{2}s_w^2$$

$$\text{cov}(X, V) = \frac{1}{2}s_w^2$$

$$\text{cov}(X', Y) = s_A^2 - s_w^2 + s_v^2$$

$$\text{cov}(X', V) = \text{cov}(Y, V) = \frac{1}{2}s_w^2 - s_v^2$$

Reagrupando de manera conveniente se obtiene

$$s_A^2 = \text{cov}(X, X') + \text{cov}(X, V) = \text{cov}(X, Y) + \text{cov}(X, V)$$

$$s_y^2 = \text{cov}(X, X') - \text{cov}(X', V) = \text{cov}(X, Y) - \text{cov}(X', V) = \text{cov}(X, Y) - \text{cov}(Y, V)$$

$$s_v^2 = \text{cov}(X, V) - \text{cov}(X', V) = \text{cov}(X, V) - \text{cov}(Y, V)$$

$$s_w^2 = 2\text{cov}(X, V)$$

Las covarianzas de la variable W con las variables X , X' , Y y V son

$$\begin{aligned}
\text{cov}(X, W) &= \frac{1}{A_{Total}} \sum_{i=1}^n \sum_{j=1}^n a_{ij} (x_i - \bar{x}_A)(x_i - x_j) = \\
&= \frac{1}{A_{Total}} \sum_{i=1}^n (x_i - \bar{x}_A) \sum_{j=1}^n a_{ij} (x_i - x_j) = \\
&= \frac{1}{A_{Total}} \sum_{i=1}^n (x_i - \bar{x}_A) A_i (x_i - y_i) = \text{cov}(X, V) = s_A^2 - s_{xy} = \\
&= \frac{1}{2}s_w^2 = \frac{1}{2}(s_A^2 - s_y^2 + s_v^2)
\end{aligned}$$

$$\begin{aligned}
\text{cov}(X', W) &= \frac{1}{A_{Total}} \sum_{i=1}^n \sum_{j=1}^n a_{ij} (x_i - x_j)(x_i - \bar{x}_A) = \\
&= \frac{1}{A_{Total}} \sum_{i=1}^n \sum_{j=1}^n a_{ij} ((x_i - \bar{x}_A) - (x_j - \bar{x}_A))(x_i - \bar{x}_A) = \\
&= s_{xy} - s_A^2 = -\text{cov}(X, V) = -\text{cov}(X, W) = -\frac{1}{2} s_w^2 = -\frac{1}{2} (s_A^2 - s_y^2 + s_v^2)
\end{aligned}$$

$$\begin{aligned}
\text{cov}(Y, W) &= \frac{1}{A_{Total}} \sum_{i=1}^n \sum_{j=1}^n a_{ij} (y_i - \bar{x}_A)(x_i - x_j) = \\
&= \frac{1}{A_{Total}} \sum_{i=1}^n (y_i - \bar{x}_A) \sum_{j=1}^n a_{ij} (x_i - x_j) = \frac{1}{A_{Total}} \sum_{i=1}^n (y_i - \bar{x}_A) A_i (x_i - y_i) = \\
&= \text{cov}(Y, V) = s_{xy} - s_y^2 = \frac{1}{2} (s_A^2 - s_y^2 - s_v^2)
\end{aligned}$$

$$\begin{aligned}
\text{cov}(V, W) &= \frac{1}{A_{Total}} \sum_{i=1}^n \sum_{j=1}^n a_{ij} (x_i - y_i)(x_i - x_j) = \frac{1}{A_{Total}} \sum_{i=1}^n (x_i - y_i) \sum_{j=1}^n a_{ij} (x_i - x_j) = \\
&= \frac{1}{A_{Total}} \sum_{i=1}^n (x_i - y_i) A_i (x_i - y_i) = s_v^2 = s_A^2 + s_y^2 - 2s_{xy}
\end{aligned}$$

Para finalizar con las definiciones previas se considera la variabilidad en las vecindades de cada objeto ω_i , denotada por T , no en relación al valor x_i , sino respecto al valor suavizado en la vecindad, es decir, y_i

$$t_{ij} = y_i - x_j$$

Sus características principales son

$$\begin{aligned}
\bar{t} &= \frac{1}{A_{Total}} \sum_{i=1}^n \sum_{j=1}^n a_{ij} t_{ij} = \frac{1}{A_{Total}} \sum_{i=1}^n \sum_{j=1}^n a_{ij} (y_i - x_j) = \frac{1}{A_{Total}} \sum_{i=1}^n \left(A_i y_i - \sum_{j=1}^n a_{ij} x_j \right) = 0, \text{ y} \\
s_t^2 &= \frac{1}{A_{Total}} \sum_{i=1}^n \sum_{j=1}^n a_{ij} (y_i - x_j)^2 = \frac{1}{A_{Total}} \sum_{i=1}^n \sum_{j=1}^n a_{ij} ((y_i - \bar{x}_A) - (x_j - \bar{x}_A))^2 = \\
&= s_A^2 + s_y^2 - 2 \frac{1}{A_{Total}} \sum_{i=1}^n \sum_{j=1}^n a_{ij} (y_i - \bar{x}_A)(x_j - \bar{x}_A) = s_A^2 + s_y^2 - 2s_y^2 = s_A^2 - s_y^2
\end{aligned}$$

Desde el punto de vista interpretativo, esta varianza es la media de las diferencias cuadráticas medias en las vecindades.

Así, se tiene la siguiente descomposición de la varianza modificada total o global

$$s_A^2 = s_y^2 + s_t^2.$$

Esta relación expresa que la varianza global se descompone como suma de la varianza de los valores medios o suavizados de las vecindades (y_i), y la suma ponderada de las varianzas locales de las vecindades respecto a su valor medio o suavizado. Es decir, es la descomposición clásica de la varianza entre (*between*) las vecindades y dentro (*within*) de las mismas.

Siempre que la matriz A sea simétrica, la varianza de la variable T tiene también otro sentido (equivalente). Con la reorganización de los subíndices y respectivos sumandos se tiene

$$s_t^2 = \frac{1}{A_{Total}} \sum_{i=1}^n \sum_{j=1}^n a_{ij} (x_i - y_j)^2$$

Se observa que se miden las diferencias de cada valor con los de su vecindad “alisada” o promediada.

De la descomposición $s_A^2 = s_y^2 + s_t^2$ y de igualdades anteriores se deducen las siguientes

$$\begin{aligned} s_t^2 &= s_A^2 - s_y^2 = (\text{cov}(X, Y) + \text{cov}(X, V)) - (\text{cov}(X, Y) - \text{cov}(Y, V)) = \\ &= \frac{1}{2} s_w^2 + \left(\frac{1}{2} s_w^2 - s_v^2 \right) = s_w^2 - s_v^2 \end{aligned}$$

Así, se tiene la descomposición de la variabilidad entre los objetos, como suma de la variabilidad respecto a los valores medios de las vecindades y la variabilidad de las respectivas vecindades.

$$s_w^2 = s_v^2 + s_t^2$$

Además, en este caso, para cada objeto ω_i , en particular, se cumple esa misma relación, en virtud de la propiedad de mínimo de la varianza, mencionada al principio de este apartado

$$\frac{1}{A_i} \sum_{j=1}^n a_{ij} (x_i - x_j)^2 = (x_i - y_i)^2 + \frac{1}{A_i} \sum_{j=1}^n a_{ij} (y_i - x_j)^2$$

pues basta ver para ello el desarrollo de

$$\frac{1}{A_i} \sum_{j=1}^n a_{ij} (x_i - y_i + y_i - x_j)^2$$

Es decir, la variabilidad de un objeto ω_i respecto a sus vecinos, se descompone en la suma de la variabilidad dentro de la vecindad y la variabilidad entre el objeto dado y la media de su vecindad (valor suavizado).

Las covarianzas de la variable T respecto de las cuatro variables anteriores son

$$\text{cov}(X, T) = 0$$

$$\text{cov}(Y, T) = 0$$

$$\text{cov}(V, T) = 0$$

$$\begin{aligned} \text{cov}(X', T) &= \frac{1}{A_{Total}} \sum_{i=1}^n \sum_{j=1}^n a_{ij} (x_j - \bar{x}_A) (y_i - x_j) = \\ &= \frac{1}{A_{Total}} \sum_{i=1}^n \sum_{j=1}^n a_{ij} (x_j - \bar{x}_A) ((y_i - \bar{x}_A) - (x_j - \bar{x}_A)) = \\ &= \frac{1}{A_{Total}} \sum_{i=1}^n \sum_{j=1}^n a_{ij} (x_j - \bar{x}_A) (y_i - \bar{x}_A) - \frac{1}{A_{Total}} \sum_{i=1}^n \sum_{j=1}^n a_{ij} (x_j - \bar{x}_A) (x_j - \bar{x}_A) = \\ &= \frac{1}{A_{Total}} \sum_{i=1}^n (y_i - \bar{x}_A) \sum_{j=1}^n a_{ij} (x_j - \bar{x}_A) - s_A^2 = \\ &= \frac{1}{A_{Total}} \sum_{i=1}^n (y_i - \bar{x}_A) A_i (y_i - \bar{x}_A) - s_A^2 = s_y^2 - s_A^2 = -s_t^2 \end{aligned}$$

$$\begin{aligned} \text{cov}(W, T) &= \frac{1}{A_{Total}} \sum_{i=1}^n \sum_{j=1}^n a_{ij} (x_i - x_j) (y_i - x_j) = \\ &= \frac{1}{A_{Total}} \sum_{i=1}^n \sum_{j=1}^n a_{ij} ((x_i - y_i) + (y_i - x_j)) (y_i - x_j) = \\ &= \frac{1}{A_{Total}} \sum_{i=1}^n \sum_{j=1}^n a_{ij} (y_i - x_j) (y_i - x_j) = \\ &= s_A^2 - s_y^2 = s_t^2 \end{aligned}$$

Las relaciones encontradas pueden resumirse, de forma sintética, utilizando únicamente los valores de s_A^2 , s_y^2 y s_{xy} (ver Tabla **4-1**).

Tabla 4-1 Relaciones entre las variables analizadas

	X	X'	Y	V	W	T
				X - Y	X - X'	Y - X'
<i>Media</i>	\bar{x}_A	\bar{x}_A	\bar{x}_A	0	0	0
<i>Varianza</i>	s_A^2	s_A^2	s_y^2	$s_A^2 + s_y^2 - 2 \cdot s_{xy}$	$s_A^2 + s_A^2 - 2 \cdot s_{xy}$	$s_A^2 - s_y^2$
<i>Covarianza</i>	X	X'	Y	V	W	T
X	s_A^2	s_{xy}	s_{xy}	$+s_A^2 - s_{xy}$	$+s_A^2 - s_{xy}$	0
X'		s_A^2	s_y^2	$-s_y^2 + s_{xy}$	$-s_A^2 + s_{xy}$	$-s_A^2 + s_y^2$
Y			s_y^2	$-s_y^2 + s_{xy}$	$-s_y^2 + s_{xy}$	0
V				$s_A^2 + s_y^2 - 2 \cdot s_{xy}$	$s_A^2 + s_y^2 - 2 \cdot s_{xy}$	0
W					$s_A^2 + s_A^2 - 2 \cdot s_{xy}$	$+s_A^2 - s_y^2$
T						$+s_A^2 - s_y^2$

En el ejemplo presentado anteriormente (ver Figura 4-3) los valores de cada variable así como sus covarianzas son los siguientes (ver Tabla 4-2).

Tabla 4-2 Valores de las variables analizadas para el ejemplo de la Figura 4-3

Ω	Ω	a_{ij}	X	X'	Y	V $X - Y$	W $X - X'$	T $Y - X'$
ω_1	ω_2	1	3	4	4.00	-1.00	-1	0.00
ω_1	ω_4	1	3	4	4.00	-1.00	-1	0.00
ω_2	ω_1	1	4	3	3.00	1.00	1	0.00
ω_2	ω_3	1	4	3	3.00	1.00	1	0.00
ω_2	ω_5	1	4	3	3.00	1.00	1	0.00
ω_3	ω_2	1	3	4	3.00	0.00	-1	-1.00
ω_3	ω_6	1	3	2	3.00	0.00	1	1.00
ω_4	ω_1	1	4	3	2.33	1.67	1	-0.67
ω_4	ω_5	1	4	3	2.33	1.67	1	-0.67
ω_4	ω_7	1	4	1	2.33	1.67	3	1.33
ω_5	ω_2	1	3	4	3.00	0.00	-1	-1.00
ω_5	ω_4	1	3	4	3.00	0.00	-1	-1.00
ω_5	ω_6	1	3	2	3.00	0.00	1	1.00
ω_5	ω_8	1	3	2	3.00	0.00	1	1.00
ω_6	ω_3	1	2	3	3.67	-1.67	-1	0.67
ω_6	ω_5	1	2	3	3.67	-1.67	-1	0.67
ω_6	ω_9	1	2	5	3.67	-1.67	-3	-1.33
ω_7	ω_4	1	1	4	3.00	-2.00	-3	-1.00
ω_7	ω_8	1	1	2	3.00	-2.00	-1	1.00
ω_8	ω_5	1	2	1	3.00	-1.00	1	2.00
ω_8	ω_7	1	2	3	3.00	-1.00	-1	0.00
ω_8	ω_9	1	2	5	3.00	-1.00	-3	-2.00
ω_9	ω_6	1	5	2	2.00	3.00	3	0.00
ω_9	ω_8	1	5	2	2.00	3.00	3	0.00

	X	X'	Y	V $X - Y$	W $X - X'$	T $Y - X'$
<i>Media</i>	3.0000	3.0000	3.0000	0.0000	0.0000	0.0000
<i>Varianza</i>	0.5833	0.5833	0.1667	0.2500	0.6667	0.4167
<i>Covarianza</i>	X	X'	Y	V	W	T
X	1.1667	-0.3333	-0.3333	1.5000	1.5000	0.0000
X'		1.1667	0.2778	-0.6111	-1.5000	-0.8889
Y			0.2778	-0.6111	0.0833	0.0000
V				2.1111	2.1111	0.0000
W					3.0000	0.8889
T						0.8889

4.2.2 El índice modificado de Moran

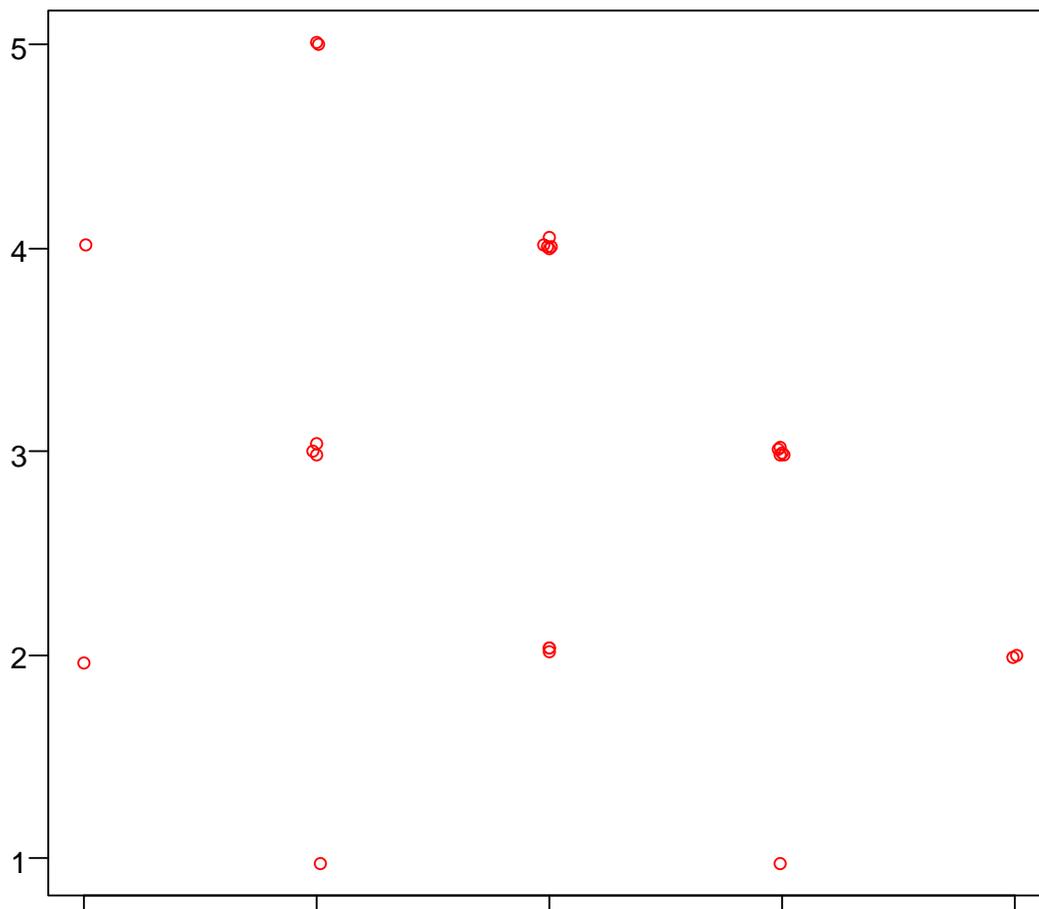
El índice modificado de Moran, I^* , es un caso especial de la fórmula general de medición de autocorrelación anteriormente mencionada

$$\Gamma = \sum_i \sum_j a_{ij} \rho_{ij}$$

en la que los pesos espaciales a_{ij} son relativizados respecto al peso total A_{Total} , y los ρ_{ij} son los productos cruzados de los valores típicos o estándares, z_i y z_j , es decir,

$$\rho_{ij} = z_i z_j$$

A continuación en la Figura 4-4 se muestran las ubicaciones de los valores de los productos cruzados. Las ubicaciones han sido alteradas aleatoria y levemente para mostrar la cantidad de coincidencias espaciales.



El índice de autocorrelación espacial de Moran se define de la siguiente manera.

$$\begin{aligned}
 I^* &= \frac{1}{A_{Total}} \sum_{i=1}^n \sum_{j=1}^n a_{ij} \left(\frac{x_i - \bar{x}_A}{s_A} \right) \left(\frac{x_j - \bar{x}_A}{s_A} \right) = \\
 &= \frac{1}{A_{Total}} \sum_{i=1}^n \sum_{j=1}^n a_{ij} (x_i - \bar{x}_A)(x_j - \bar{x}_A) = \frac{\sum_{i=1}^n \sum_{j=1}^n a_{ij} (x_i - \bar{x}_A)(x_j - \bar{x}_A)}{s_A^2} = \frac{\sum_{i=1}^n A_i (x_i - \bar{x}_A)^2}{\sum_{i=1}^n A_i (x_i - \bar{x}_A)^2}
 \end{aligned}$$

Este índice es una modificación del índice de Moran [132], I , en el que se consideran la media y la varianza habituales, $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ y $s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$, respectivamente.

Por tanto, cuando los pesos espaciales A_i son iguales el índice modificado de Moran, I^* , coincide con el clásico, I . Por otra parte, en el caso de retículos regulares y 4-vecindades, como los puntos interiores tienen peso igual a 4, los márgenes a 3 y los vértices a 2, si el retículo es amplio, la influencia de los márgenes es pequeña, y como los valores de la media, y de la varianza tienden a ser iguales en los sistemas clásico y modificado, los de ambos índices de autocorrelación también tienden a ser muy parecidos.

El índice de Moran, por su parte, se puede considerar como un caso especial del coeficiente de correlación lineal de Pearson, r , que mide la relación lineal entre dos variables

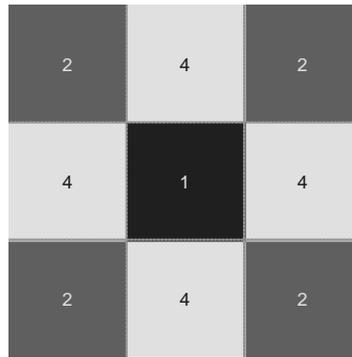
$$\begin{aligned}
 r &= \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{j=1}^n (y_j - \bar{y})^2}} = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \frac{1}{n} \sum_{j=1}^n (y_j - \bar{y})^2}} = \\
 &= \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{s_x s_y} = \frac{1}{n} \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{s_x} \right) \left(\frac{y_i - \bar{y}}{s_y} \right)
 \end{aligned}$$

donde s_x , s_y representan las desviaciones estándares de las variables x e y , respectivamente.

En efecto, si se considera la relación de una variable consigo mismo atendiendo a la adyacencia de los objetos, A (a_{ij} toma el valor uno si se considera el par (x_i, x_j) , y el valor cero no se considera), y se calcula r , se tiene el índice de Moran. Es decir, en esencia,

el índice de Moran mide la correlación lineal entre la variable y sus vecinos espaciales obtenidos según la matriz de conexiones A .

En el caso del índice clásico no se trata exactamente de la correlación lineal de Pearson. De hecho el coeficiente de Pearson se ubica siempre entre -1 y +1, y sin embargo, el índice clásico de Moran puede rebasar estas cotas. Esto se ilustra en el siguiente ejemplo (ver Figura 4-5).



$$I = -1.0577, I^* = -0.9245$$

Figura 4-5 Imagen de ejemplo

La razón de esta diferencia estriba en que en el cálculo de la media y la varianza para el índice clásico no se tiene en cuenta la relación de contigüidad. Es decir, para el cálculo de estos dos estadísticos los pesos de todos los objetos son iguales, y para el cálculo de la covarianza, en el que los objetos toman parte en función de su adyacencia, no.

En el caso del índice modificado de Moran, I^* , sin embargo, al considerarse los pesos espaciales en todos los cálculos embebidos en la fórmula del coeficiente de Pearson, se cumple, evidentemente, la propiedad de r de ubicarse entre los límites -1 y +1.

$$-1 \leq I^* \leq 1$$

En el ejemplo anteriormente detallado (ver Figura 4-3) $I^* = -0.2857$, lo cual denota que hay cierta autocorrelación negativa que se aprecia en la imagen misma pues los valores altos tienden a tener en su vecindad valores bajos, y viceversa.

Una expresión de I^* usando la variable de vecindad Y , que resulta interesante desde el punto de vista interpretativo, es la siguiente.

$$I^* = \frac{1}{A_{Total}} \frac{\sum_{i=1}^n \sum_{j=1}^n a_{ij} (x_i - \bar{x}_A)(x_j - \bar{x}_A)}{s_A^2} = \frac{\sum_{i=1}^n \frac{A_i}{A_{Total}} (x_i - \bar{x}_A)(y_i - \bar{x}_A)}{s_A^2} = \frac{s_y}{s_A} r_{xy}$$

donde r_{xy} es el coeficiente de correlación lineal entre las variables X e Y , y, por tanto, está limitado por -1 y $+1$. Como $s_y \leq s_A$, podría ser que aun estando X e Y totalmente correladas ($r_{xy}=1$, por ejemplo), I^* no alcanzara la cota ($I^* < 1$, si $s_y < s_A$).

El hecho de que $+1$ y -1 sean, en general, las cotas máxima y mínima del índice modificado de Moran, no significa que se puedan alcanzar para cualquier estructura de adyacencia o vecindad. En efecto, la estructura espacial definida por la matriz de adyacencia, A , permite precisar más estas cotas [130].

Respecto al rango de los posibles valores del índice de Moran se puede consultar el trabajo [133], citado también por [134]. Allí se demuestra que para una matriz de pesos espaciales no negativos $A = [a_{ij}]$ de dimensión n la cota inferior y superior de los valores del índice de Moran vienen dadas por $(n/1_n^t A 1_n) \lambda_{\min}$ y $(n/1_n^t A 1_n) \lambda_{\max}$, respectivamente; donde λ_{\min} y λ_{\max} son los autovalores extremos de la matriz QAQ , también llamada matriz centralizada, donde la matriz $Q = I - 1_n 1_n^t / n$ siendo $1_n = (1, 1, \dots, 1)^T$. Si uno de los autovalores extremos obtenidos es cero, entonces el segundo menor (o mayor, según sea al caso) es la cota mínima (o máxima).

A continuación, en la Figura 4-6, se muestran algunos ejemplos ilustrativos del índice modificado de Moran, según la distribución de los datos en un retículo 3x3 con adyacencias definidas a partir de la 4-vecindad. Se obtienen diferentes valores del índice de autocorrelación espacial que van desde valores positivos en el caso (a) hasta valores de autocorrelación espacial negativa en el caso (d).

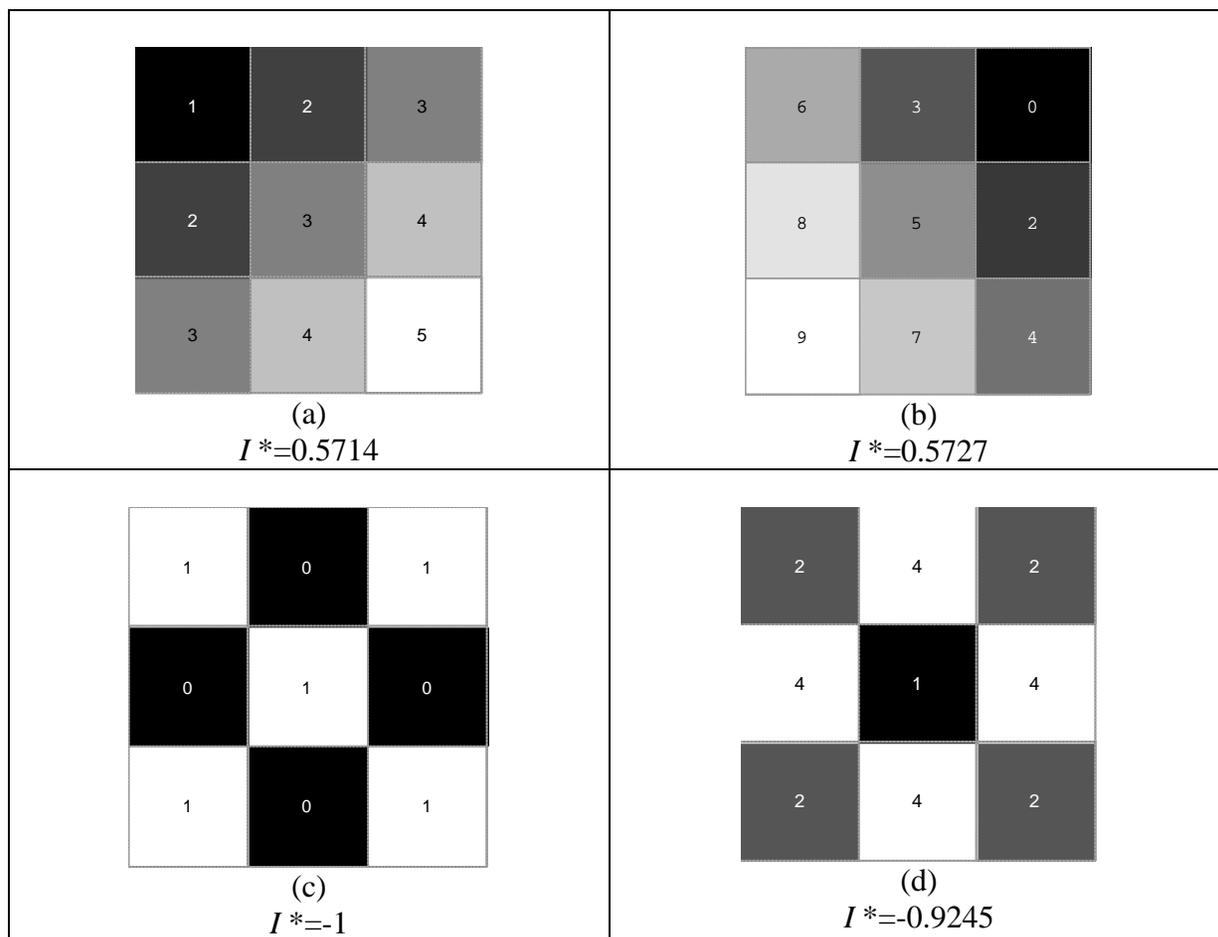


Figura 4-6 Ejemplos de regiones con los valores asociados del índice modificado de autocorrelación espacial de Moran (usando 4 – adyacencia)

También es posible comparar el índice modificado de Moran para diferentes niveles de contigüidad, es decir, utilizando diferentes tipos de vecindades. En ocasiones los datos pueden estar altamente autocorrelacionados de manera local, es decir, en vecindades cercanas, pero pueden dejar de serlo al tomar vecindades más amplias.

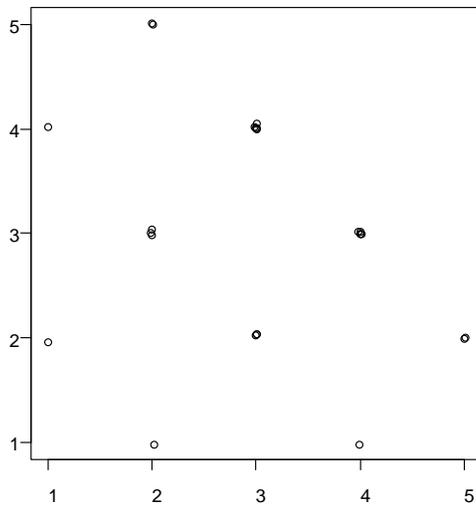
El estudio del índice modificado de Moran puede apoyarse, además de la interpretación anteriormente dada de la correlación lineal de Pearson, en dos instrumentos o herramientas: el denominado *gráfico de Moran* [135] y los *indicadores locales* de asociación espacial [136].

4.2.2.1 El gráfico de Moran

El gráfico de dispersión de los pares de valores (x_i, x_j) para aquellos cuyo peso espacial, a_{ij} , es distinto de cero, proporciona una interpretación añadida del índice modificado de Moran. En efecto, ya que el índice modificado de Moran se puede expresar como el cociente de una covarianza y la varianza de una de las variables implicadas, se puede pensar en la pendiente de una recta de regresión obtenida por el método de los mínimos cuadrados

$$I^* = \frac{\text{cov}(X, X')}{s_A^2} = \frac{\text{cov}(X, Y)}{s_A^2} = \frac{s_{xy}}{s_A^2}$$

Como se tiene la expresión de la covarianza con X para dos variables, X' e Y , cuyos valores medios coinciden (\bar{x}_A), se tienen la misma recta de regresión para ambos gráficos de dispersión (X vs. X' , X vs. Y). En el ejemplo de la Figura 4-3, para entender conjuntamente ambos gráficos, hay que reparar en que ambos ejes de ordenadas están dimensionados con distintas escalas de valores típicos, pues, en general, s_A^2 y s_y^2 son distintos (ver Figura 4-7).



Esta segunda interpretación es la que propone Anselin [135] usando las medias y varianzas habituales, y una normalización sobre las filas de la matriz de adyacencias, A , para que todos los objetos tengan el mismo peso y lo denomina gráfico de dispersión de Moran.

El índice modificado de Moran, aún brindando una visión global acerca de la autocorrelación espacial, no ofrece ninguna información sobre la estructura subyacente. Por ejemplo, es interesante conocer qué regiones o zonas contribuyen en mayor medida a la autocorrelación espacial global, o si existen concentraciones locales de valores semejantes, o hasta qué punto la autocorrelación global obtenida enmascara localizaciones o zonas atípicas.

Más aún, desde el punto de vista de la utilidad predictiva de la regresión, y considerando la interpretación mediante los valores medios de las vecindades (y_i), parecería más útil que la variable independiente fuera precisamente Y , y la dependiente X ; es decir, que conocido el valor medio de la vecindad de un objeto, si el valor del índice I^* fuera extremo (cercano a $+1$ ó -1) se pudiera predecir en buen grado cuál sería el valor del objeto en cuestión.

El gráfico de Moran ofrece una ventaja sobre el índice, y es que descompone la asociación espacial en cuatro zonas o cuadrantes, y permite detectar objetos extraños, así como los objetos que ejercen más influencia en el valor de I^* .

Cada uno de los cuatro cuadrantes que determina el valor de la media \bar{x}_A se corresponde con diferentes formas de autocorrelación espacial, de manera análoga a como ha quedado reflejado en la definición misma del índice. Allí cada punto se refería a un par de objetos vecinos, y aquí cada punto designa a un objeto junto a sus vecinos o junto a su vecino medio. El primer y tercer cuadrantes están asociados a formas positivas, mientras el segundo y cuarto representan formas negativas de autocorrelación espacial.

Todo ello está ligado a la descomposición de I^* en suma de valores relacionados con los objetos.

4.2.2.2 Índices locales de Moran

El índice modificado de Moran (y, en general, cualquier índice de autocorrelación) refleja si hay o no agrupamientos (concentraciones, clusters) espaciales de valores altos (bajos), pero no revela dónde esto ocurre, o incluso si los valores afectados son los altos, los bajos, o ambos, pues consiste en promediar las variaciones locales de la autocorrelación espacial.

Esto ha llevado al desarrollo de índices locales de autocorrelación espacial. Para el caso del índice de Moran, observando la posibilidad de descomposición aditiva de I^*

$$I^* = \frac{1}{A_{Total}} \sum_{i=1}^n A_i \frac{1}{A_i} \sum_{j=1}^n a_{ij} \left(\frac{x_i - \bar{x}_A}{s_A} \right) \left(\frac{x_j - \bar{x}_A}{s} \right)$$

$$I_i^* = \frac{1}{A_i} \sum_{j=1}^n a_{ij} \left(\frac{x_i - \bar{x}_A}{s_A} \right) \left(\frac{x_j - \bar{x}_A}{s_A} \right) = \left(\frac{x_i - \bar{x}_A}{s_A} \right) \frac{1}{A_i} \sum_{j=1}^n a_{ij} \left(\frac{x_j - \bar{x}_A}{s_A} \right) =$$

$$= \frac{(x_i - \bar{x}_A)(y_i - \bar{x}_A)}{s_A^2}$$

Así pues, el índice modificado de Moran es el valor medio de los índices locales

$$I^* = \frac{1}{A_{Total}} \sum_{i=1}^n A_i I_i^*$$

El cálculo de I_i^* equivale, en cierto sentido, a la ubicación del objeto ω_i en el gráfico de Moran con coordenadas (x_i, y_i) , ya que el denominador s_A^2 es común a los I_i^* de todos los objetos.

El punto de vista adoptado por los índices LISA (*Local Indicators of Spatial Autocorrelation*) de Anselin [136], es más adecuado para detectar la influencia de un objeto en el valor global del índice (I en su caso, en lugar de I^*), pues considera los valores $A_i I_i$ (correspondientes a los $A_i I_i^*$ de aquí), y descompone el valor del índice global como suma (no ponderada) de los locales.

En el caso del ejemplo que se viene analizando (ver Figura 4-8), el objeto ω_3 , que se ubica en la esquina inferior derecha de la imagen, es el que ofrece el valor mínimo, con gran diferencia respecto a los demás, debido a la gran disparidad entre el valor medio de la vecindad respecto al valor medio (2-3) y el suyo propio (5-2). Los objetos ω_4 y ω_6 , que ya habían sido detectados como equivalentes, tienen también un valor negativo, pues respecto al valor medio (3) sus valores, el asignado y el medio de la vecindad, están opuestos.

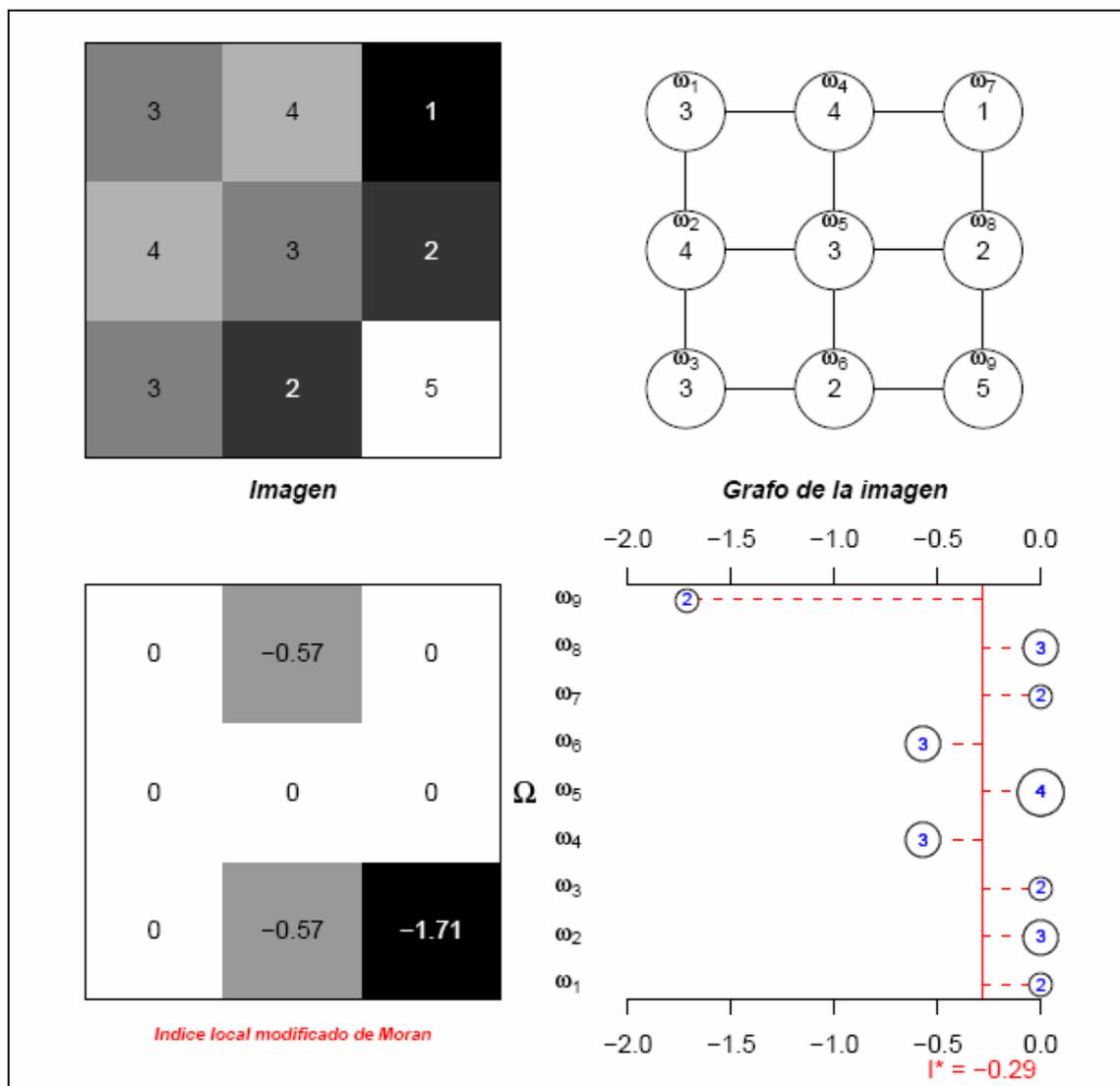


Figura 4-8 Valores de los índices locales de Moran

4.2.3 Índice modificado de Geary

Si en la raíz común general para la medición de autocorrelación espacial formulada anteriormente

$$\Gamma = \sum_i \sum_j a_{ij} \rho_{ij},$$

además de la matriz de contigüidad o pesos espaciales, $A = [a_{ij}]$, se considera $P = [\rho_{ij}]$ como una matriz de desemejanza entre los objetos ω_i y ω_j , se pueden obtener otros índices. Más concretamente, en el caso de una variable, se puede definir P así

$$\rho_{ij} = (z_i - z_j)^2 = \left(\frac{x_i - \bar{x}_A}{s_A} - \frac{x_j - \bar{x}_A}{s_A} \right)^2 = \frac{(x_i - x_j)^2}{s_A^2}$$

y el índice modificado de autocorrelación espacial de Geary [137] de la siguiente manera

$$\begin{aligned} C^* &= \frac{1}{2} \frac{1}{A_{Total}} \sum_{i=1}^n \sum_{j=1}^n a_{ij} (z_i - z_j)^2 = \frac{1}{2} \frac{1}{A_{Total}} \sum_{i=1}^n \sum_{j=1}^n a_{ij} \left(\frac{x_i - \bar{x}_A}{s_A} - \frac{x_j - \bar{x}_A}{s_A} \right)^2 = \\ &= \frac{1}{2} \frac{\sum_{i=1}^n \sum_{j=1}^n a_{ij} (x_i - x_j)^2}{s_A^2} = \frac{1}{2} \frac{s_w^2}{s_A^2} = \frac{1}{2} \frac{\sum_{i=1}^n \sum_{j=1}^n a_{ij} (x_i - x_j)^2}{\sum_{i=1}^n A_i (x_i - \bar{x}_A)^2} \end{aligned}$$

Puede pensarse que el factor 1/2 que precede a la ponderación se aplica para compensar, en el caso habitual en que la matriz A es simétrica, el hecho de que los sumandos aparecen dos veces. Lo mismo podría haberse pensado para el índice de Moran cuando fuera $a_{ii} = 0$, pero en aquel caso no aparece este factor.

Este índice, así definido, es una modificación del índice de Geary [137], en el que se consideran la media y varianzas habituales. Además, esta última es considerada desde el punto de vista de la estimación insesgada, es decir, promediando las diferencias cuadráticas a la media con $1/(n-1)$ en lugar de hacerlo $1/n$.

$$C = \frac{(n-1) \sum_{i=1}^n \sum_{j=1}^n a_{ij} (x_i - x_j)^2}{2 \left(\sum_{i=1}^n \sum_{j=1}^n a_{ij} \right) \sum_{i=1}^n (x_i - \bar{x})^2} = \frac{\sum_{i=1}^n \sum_{j=1}^n a_{ij} (x_i - x_j)^2}{2 \left(\sum_{i=1}^n \sum_{j=1}^n a_{ij} \right) \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}} = \frac{\sum_{i=1}^n \sum_{j=1}^n a_{ij} \left(\frac{x_i - \bar{x}}{s} - \frac{x_j - \bar{x}}{s} \right)^2}{2 \left(\sum_{i=1}^n \sum_{j=1}^n a_{ij} \right)}$$

Los valores del índice de Geary se encuentran normalmente entre 0 y 2 (pero pueden superar este último [138]). La relación que hay entre los índices de Moran y de Geary es la siguiente [139]

$$C = \frac{n-1}{n} \left[\frac{n \sum_{i=1}^n A_i (x_i - \bar{x})^2}{\left(\sum_{i=1}^n \sum_{j=1}^n a_{ij} \right) \sum_{i=1}^n (x_i - \bar{x})^2} - I \right]$$

Se observa que esta relación de complementariedad entre los índices se ve afectada por el número de objetos, n . Si los pesos de la matriz de adyacencia A fueran normalizados proporcionalmente por las filas, es decir, si todos los objetos tuvieran el mismo peso $A_i=1$, y $\sum_{i=1}^n \sum_{j=1}^n a_{ij} = 1$, y se considerase el índice Geary con la varianza habitual en lugar de hacerlo con el factor $1/(n-1)$, ambos índices serían complementarios respecto a la unidad [131].

Sin embargo, esta misma propiedad de complementariedad la cumplen los índices modificados sin hacer ninguna consideración. En efecto, $I^* + C^* = 1$. Para demostrar esta relación de complementariedad respecto a la unidad, basta con tomar en cuenta que

$$I^* = \frac{\text{cov}(X, Y)}{s_A^2}, \quad C^* = \frac{\text{cov}(X, V)}{s_A^2}, \text{ y}$$

$$\text{cov}(X, Y) = s_A^2 - \frac{1}{2} s_w^2, \quad \text{cov}(X, V) = \frac{1}{2} s_w^2$$

Una consecuencia inmediata de esta relación de complementariedad, junto con la anterior desigualdad $-1 \leq I^* \leq 1$, es que los valores del índice modificado de Geary están estrictamente entre 0 y 2

$$0 \leq C^* \leq 2.$$

Al igual que en el caso del índice de Moran, estos límites pueden ser más precisos, dependiendo de la estructura de los datos reflejada en la matriz de adyacencia A .

En relación con este índice, Lebart introduce una modificación [130], calculando la diferencias cuadráticas de cada valor respecto al valor promedio de su vecindad, tomando en cuenta los pesos espaciales, es decir, la media modificada, en lugar de calcularlos únicamente respecto a todos los valores vecinos. Sin embargo, no toma en cuenta los pesos espaciales para calcular esa nueva varianza “local” (tal y como él la denomina) y la compara con la varianza habitual, tal y como propuso Geary (prescindiendo del factor multiplicativo $1/2$).

En realidad, desde el punto de vista de este trabajo, teniendo como base la igualdad $s_w^2 = s_v^2 + s_t^2$, lo que se propone es tomar una parte de las diferencias por pares adyacentes (s_v^2).

Si se usasen los pesos y la varianza modificada se tendría el *índice modificado de Lebart* [129]

$$L^* = \frac{\frac{1}{A_{Total}} \sum_{i=1}^n A_i (x_i - y_i)^2}{s_A^2} = \frac{s_v^2}{s_A^2} = \frac{\sum_{i=1}^n A_i (x_i - y_i)^2}{\sum_{i=1}^n A_i (x_i - \bar{x}_A)^2}$$

Los límites generales de este índice son: $0 \leq L^* \leq 4$. En efecto, por la definición misma se tiene que $L^* \geq 0$, y, desarrollando su relación directa con el índice modificado de Geary se tiene

$$\begin{aligned} L^* &= \frac{\frac{1}{A_{Total}} \sum_{i=1}^n \sum_{j=1}^n a_{ij} ((x_i - x_j) - (y_i - x_j))^2}{s_A^2} = \\ &= \frac{\frac{1}{A_{Total}} \sum_{i=1}^n \sum_{j=1}^n a_{ij} (x_i - x_j)^2}{s_A^2} - 2 \frac{\frac{1}{A_{Total}} \sum_{i=1}^n \sum_{j=1}^n a_{ij} (x_i - x_j)(y_i - x_j)}{s_A^2} + \frac{\frac{1}{A_{Total}} \sum_{i=1}^n \sum_{j=1}^n a_{ij} (y_i - x_j)^2}{s_A^2} = \\ &= 2C^* - 2 \frac{\text{cov}(W, T)}{s_A^2} + \frac{s_t^2}{s_A^2} = 2C^* - \frac{s_t^2}{s_A^2} \end{aligned}$$

Como $s_A^2 = s_v^2 + s_t^2$ entonces $L^* \leq 2C^* \leq 4$.

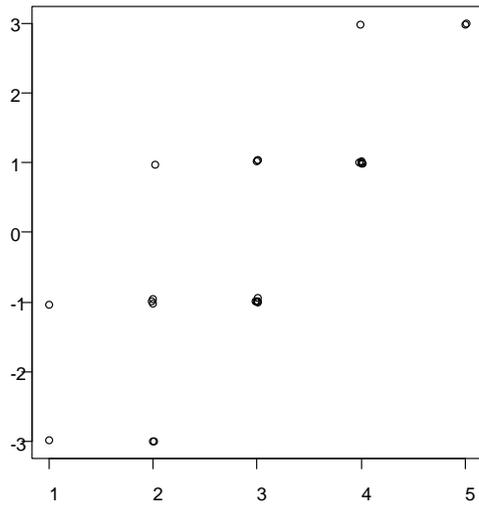
4.2.3.1 El gráfico de Geary

Teniendo en cuenta la expresión del índice de Geary en términos de cociente de una covarianza y una varianza, se puede buscar su interpretación como si fuera la pendiente de una regresión lineal.

$$C^* = \frac{1}{2} \frac{s_w^2}{s_A^2} = \frac{\text{cov}(X, W)}{s_A^2} = \frac{\text{cov}(X, V)}{s_A^2}$$

En efecto, si hace el diagrama de dispersión de las variables X vs. W y X vs. V , las regresiones lineales de W y V sobre X tienen por pendiente el índice modificado C^* .

Teniendo en cuenta que $W = X - X'$ y $V = X - Y$, se tienen los siguientes gráficos para el ejemplo analizado (ver Figura. 4-9).



El cálculo de C_i^* equivale, en cierto sentido, como sucedía con I_i^* , a la ubicación del objeto ω_i en el gráfico de Geary con coordenadas (x_i, w_i) , ya que el denominador s_A^2 es común a los C_i^* de todos los objetos. Se entiende que cuanto más alto sea su valor más diferente es la vecindad de un objeto respecto al valor del mismo.

Al ser no negativos todos los valores del índice local, y ser la suma de todos el valor del índice global modificado, cada sumando $A_i C_i^*$ expresa su contribución al valor definitivo (éste es el enfoque original de LISA). Más aún, se pueden explicar estas contribuciones en términos porcentuales, lo que ayuda a detectar más fácilmente los objetos más contribuyentes al valor de C^*

$$\frac{A_i C_i^*}{A_{Total} C^*} = \frac{A_i}{A_{Total}} \frac{C_i^*}{C^*}$$

En la Figura 4-10 se observa la disposición de los objetos en la que se refleja la disparidad de ω_9 y su vecindad, y la de ω_7 .

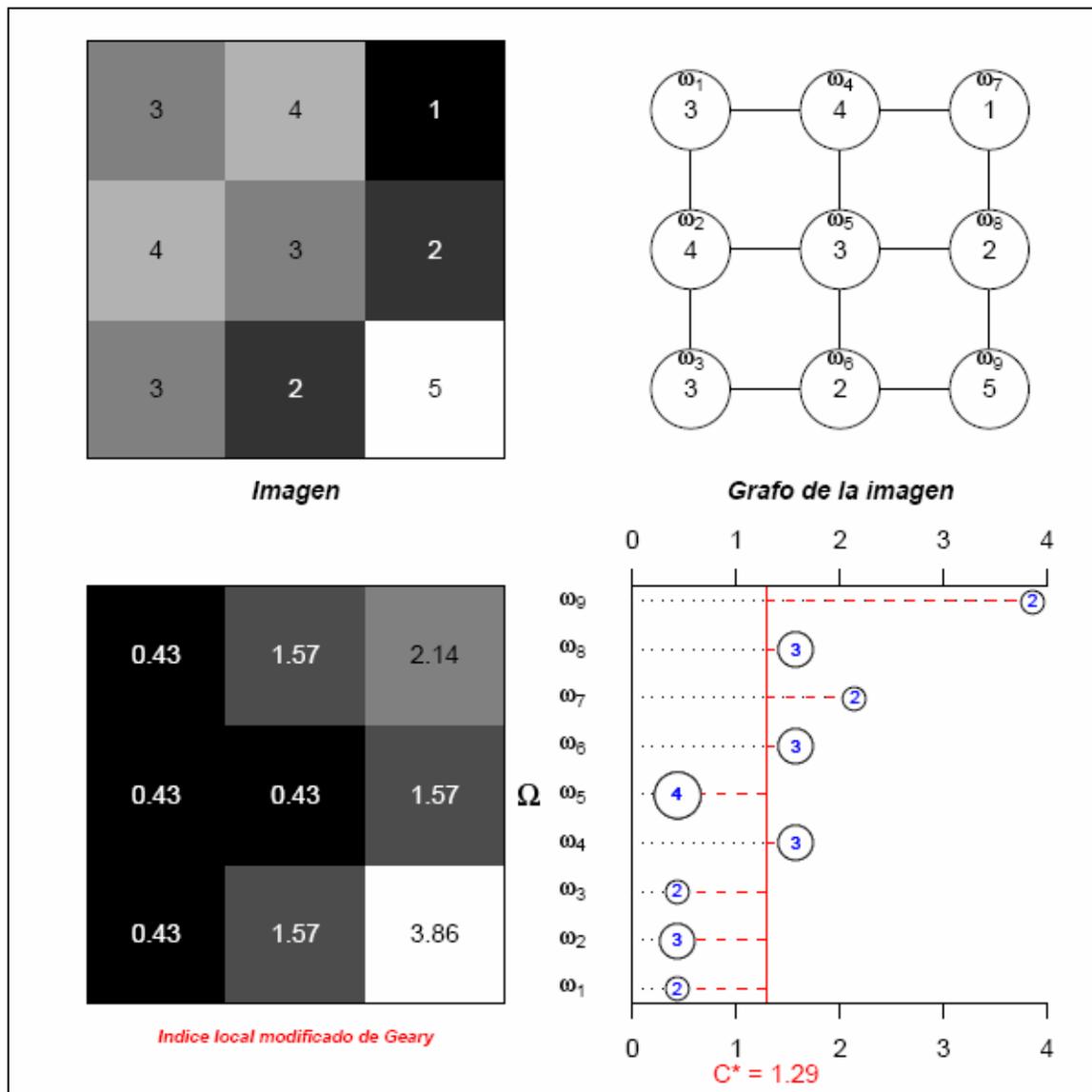


Figura 4-10 Valores de lo índices locales de Geary

Por su parte, el índice modificado de Lebart también ofrece la posibilidad de definir un índice local que lo descomponga aditivamente en valores no negativos, tal que L^* resulte ser el valor medio de los valores del índice local. En efecto, si se considera

$$L_i^* = \frac{(x_i - y_i)^2}{s_A^2}$$

se tiene que

$$L^* = \frac{1}{A_{Total}} \sum_{i=1}^n A_i L_i^*$$

La Figura 4-11 muestra que los objetos ω_9 y ω_7 , situados en dos vértices, son los más dispares en relación a sus respectivas vecindades, consideradas éstas con valores medios, y ω_3 y ω_5 los más similares (de hecho, son iguales).

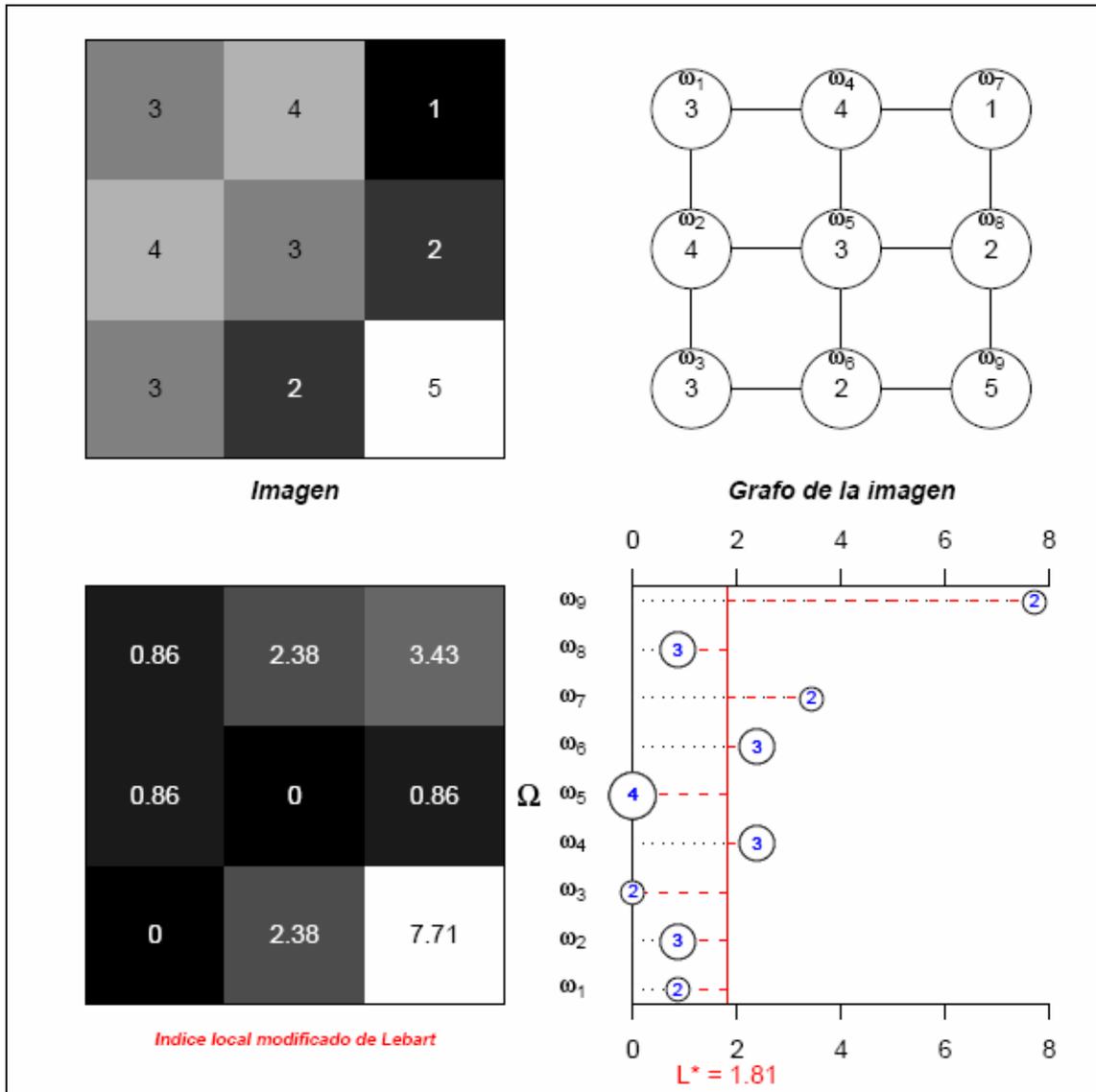


Figura 4-11 Valores de los índices locales de Lebart

4.2.4 Nuevos índices de autocorrelación espacial

La relación entre las varianzas de las distintas variables definidas al principio sugiere la definición de algunos otros índices. Dichas relaciones son las siguientes

$$s_A^2 = s_y^2 + s_t^2 \text{ y}$$

$$s_w^2 = s_v^2 + s_t^2$$

Cada una ellas conlleva la definición de los correspondientes índices. Aquí también se determinan sus relaciones con los índices de Moran, Geary, y con el índice de Lebart.

4.2.4.1 Índices de vecindad η

La relación $s_A^2 = s_y^2 + s_t^2$ con

$$s_A^2 = \frac{1}{A_{Total}} \sum_{i=1}^n A_i (x_i - \bar{x}_A)^2, \quad s_y^2 = \frac{1}{A_{Total}} \sum_{i=1}^n A_i (y_i - \bar{x}_A)^2, \quad \text{y} \quad s_t^2 = \frac{1}{A_{Total}} \sum_{i=1}^n \sum_{j=1}^n a_{ij} (y_i - x_j)^2$$

expresa la descomposición de la varianza modificada global como suma de la varianza entre (*between*) las vecindades, consideradas éstas como valores promediados, y la varianza dentro (*within*) de las mismas.

Este segundo sumando tiene otro sentido (equivalente) con la reorganización de los sumandos, tal y como ha quedado señalado anteriormente

$$s_t^2 = \frac{1}{A_{Total}} \sum_{i=1}^n \sum_{j=1}^n a_{ij} (x_i - y_j)^2$$

En efecto, de esta forma se observa que se miden las diferencias de cada valor con los de su vecindad, suavizada o promediada.

La igualdad de varianzas puede expresarse más adecuadamente para los fines perseguidos como

$$\frac{s_y^2}{s_A^2} + \frac{s_t^2}{s_A^2} = 1$$

El cociente s_y^2/s_A^2 da la proporción de varianza global explicada por la variabilidad entre (*between*) las vecindades, tomadas en términos medios, y el cociente s_i^2/s_A^2 , la proporción explicada por la variabilidades dentro (*within*) de las vecindades.

Cada uno de los cocientes sugiere una medida de autocorrelación espacial, y se tiene, además, la complementariedad de ambos respecto a 1.

Si se tomara como índice de autocorrelación s_y^2/s_A^2 se tendría

$$\eta_1^* = \frac{\sum_{i=1}^n A_i (y_i - \bar{x}_A)^2}{\sum_{i=1}^n A_i (x_i - \bar{x}_A)^2}$$

Este es un índice parecido al índice de determinación de un modelo, que en este caso se trata de la media de la vecindad.

Su rango va desde de 0 a 1. Si se acerca a 0 denotará una ausencia de autocorrelación, y si se acerca a 1 una gran autocorrelación, sea ésta positiva o negativa.

Teniendo en cuenta que s_y^2 es precisamente $\text{cov}(X', Y)$, y la varianza de X' es s_A^2 , η_1^* se puede interpretar también como la pendiente de la recta de regresión de la variable Y respecto de X' (ver Figura 4-13).

Junto a este índice podrían considerarse también los índices locales para cada objeto ω_i , siguiendo la metodología LISA

$$\eta_{li}^* = \frac{(y_i - \bar{x}_A)^2}{s_A^2}$$

Hay que hacer notar que no se trata de valores tipificados, pues, en general, aunque \bar{x}_A sea la media de la variable Y , s_y^2 no es igual a s_A^2 .

En la Figura 4-12 se muestran los valores de los índices locales calculados en un ejemplo.

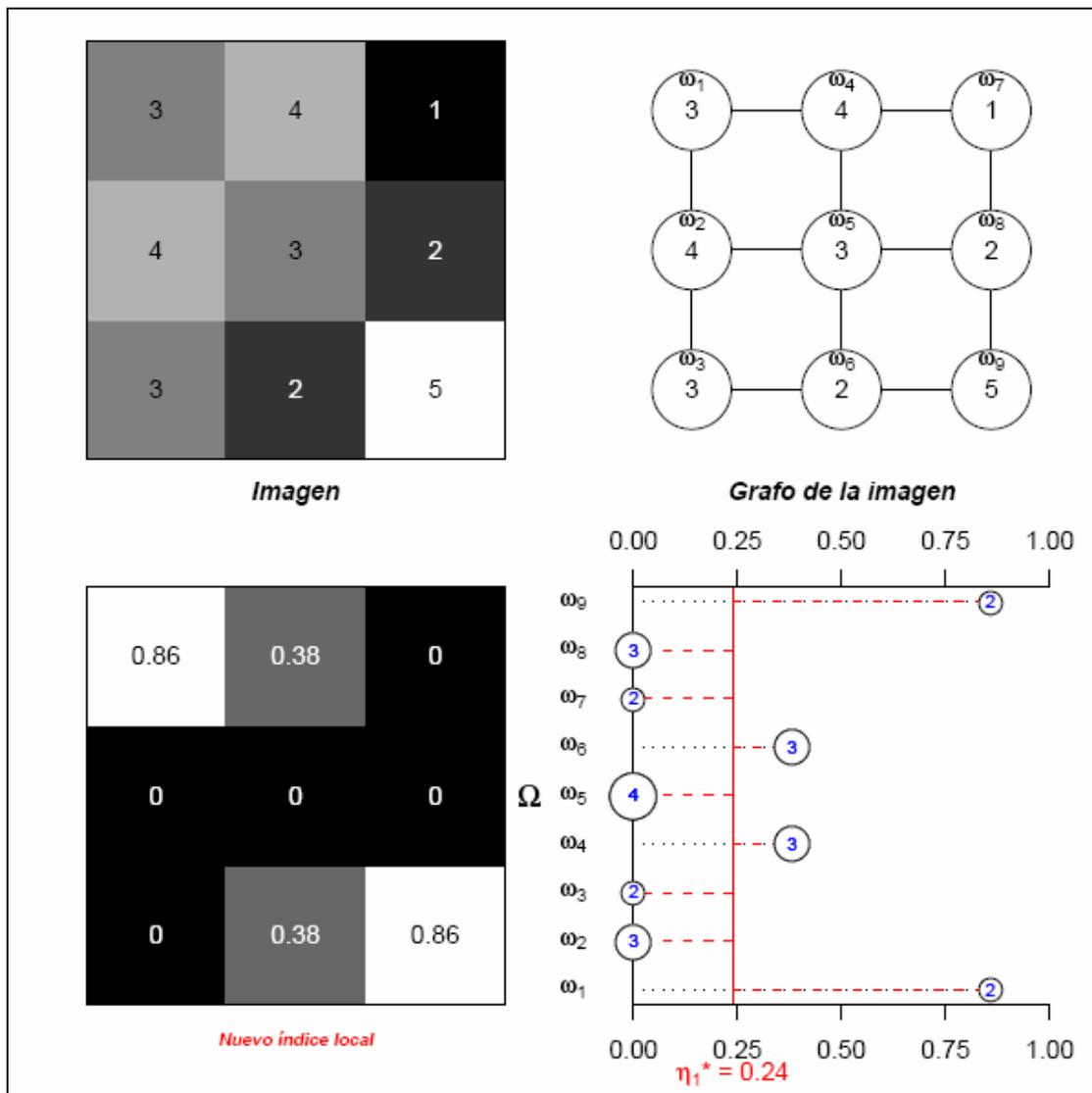


Figura 4-12 Gráfico de los índices locales η_{li}^*

El valor interpretativo de este índice no es, sin embargo, alto, en cuanto que no toma en consideración los valores asignados a los objetos mismos, sino a los de sus vecinos.

Su suma ponderada por los pesos espaciales A_i es el índice propuesto

$$\eta_l^* = \frac{1}{A_{Total}} \sum_{i=1}^n A_i \eta_{li}^*$$

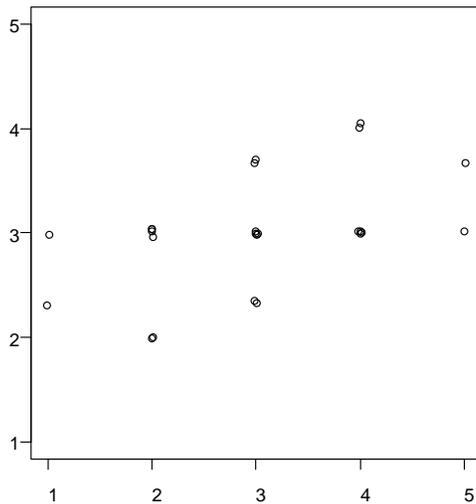
La contribución de un determinado objeto ω_i se mediría multiplicando este índice local por su propio peso espacial $A_i \eta_{li}^*$.

Por otra parte, el cociente s_i^2/s_A^2 también podría considerarse como un índice de autocorrelación parejo al índice de Geary, en cuanto que mide las diferencias resultantes de los valores medidos respecto a los de las vecindades (promediados), en lugar de hacerlo respecto a los valores medidos de la vecindad.

$$\eta_2^* = \frac{\sum_{i=1}^n \sum_{j=1}^n a_{ij} (y_i - x_j)^2}{\sum_{i=1}^n A_i (x_i - \bar{x}_A)^2} = \frac{\sum_{i=1}^n \sum_{j=1}^n a_{ij} (x_i - y_j)^2}{\sum_{i=1}^n A_i (x_i - \bar{x}_A)^2}$$

Su rango de valores es también entre 0 y 1. Si se acerca a 0 indica que hay una gran autocorrelación, y si a 1 su ausencia.

Al igual que su índice complementario, η_1^* ($\eta_1^2 + \eta_2^2 = 1$) se puede interpretar como la pendiente de la recta de regresión en un gráfico de dispersión (ver Figura 4-13), y se pueden definir índices locales.



cada valor con los valores 'alisados' de su vecindad, la covarianza cambia de signo y el índice resulta ser la pendiente de la recta de regresión de dicha variable respecto a X' .

Asociado a este índice también pueden considerarse los índices locales para cada objeto ω_i

$$\eta_{2i}^* = \frac{\frac{1}{A_i} \sum_{j=1}^n a_{ij} (y_i - x_j)^2}{s_A^2}$$

Su suma ponderada por los pesos espaciales A_i es el índice propuesto

$$\eta_2^* = \frac{1}{A_{Total}} \sum_{i=1}^n A_i \eta_{2i}^*$$

La contribución de un objeto ω_i al valor del índice es $A_i \eta_{2i}^*$.

En la Figura **4-14** se muestran los valores de los índices locales calculados en un ejemplo.

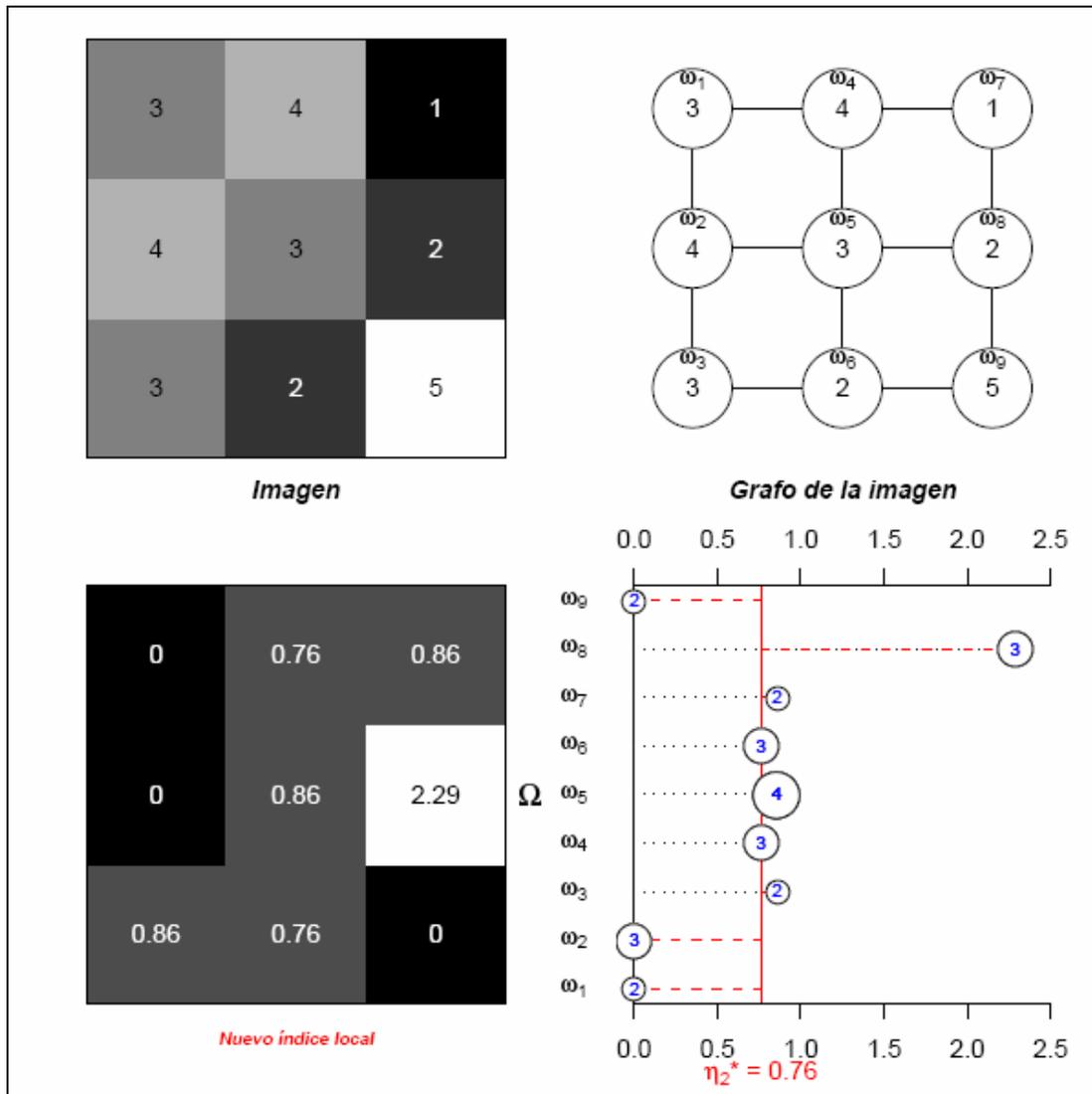


Figura 4-14 Gráfico de los índices locales η_{2i}^*

Podría definirse otro índice en relación a la doble definición de η_2^*

$$\eta_{2i}^{*'} = \frac{\frac{1}{A_i} \sum_{j=1}^n a_{ij} (x_i - y_j)^2}{s_A^2}$$

pero carece, en principio, de interés por su dificultad de ser interpretado. En general,

aunque sea $\eta_2^* = \frac{1}{A_{Total}} \sum_{i=1}^n A_i \eta_{2i}^{*'}$, es $\eta_{2i}^* \neq \eta_{2i}^{*'}$.

4.2.4.2 Índices de vecindad δ

De la otra relación entre varianzas

$$s_w^2 = s_v^2 + s_t^2$$

donde

$$s_w^2 = \frac{1}{A_{Total}} \sum_{i=1}^n \sum_{j=1}^n a_{ij} (x_i - x_j)^2, \quad s_v^2 = \frac{1}{A_{Total}} \sum_{i=1}^n A_i (x_i - y_i)^2, \quad \text{y} \quad s_t^2 = \frac{1}{A_{Total}} \sum_{i=1}^n \sum_{j=1}^n a_{ij} (y_i - x_j)^2$$

se pueden hacer consideraciones análogas a las hechas en el subapartado anterior

$$\frac{s_v^2}{s_w^2} + \frac{s_t^2}{s_w^2} = 1$$

El cociente s_v^2/s_w^2 es una proporción de lo que explican las diferencias entre los valores medios de las vecindades y los medidos, respecto a todas las diferencias.

$$\delta_1^* = \frac{\sum_{i=1}^n A_i (x_i - y_i)^2}{\sum_{i=1}^n \sum_{j=1}^n a_{ij} (x_i - x_j)^2}$$

Sus posibles valores están acotados por 0 y 1. Los valores bajos delatan una alta autocorrelación, y los altos poca autocorrelación.

Al ser s_v^2 la covarianza entre las variables $X-X'$ y $X-Y$, δ_1^* puede ser interpretado como la pendiente de la recta de regresión de $X-Y$ respecto a $X-X'$ (ver Figura 4-15).

Obviamente, también se puede entender δ_1^* como la media de los valores locales que toma en cada objeto

$$\delta_{1i}^* = \frac{(x_i - y_i)^2}{s_w^2} \text{ y}$$

$$\delta_1^* = \frac{1}{A_{Total}} \sum_{i=1}^n A_i \delta_{1i}^*$$

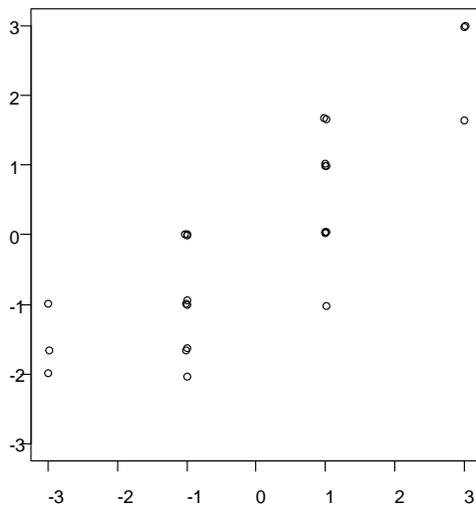
Estos índices locales no aportan ninguna información si se calculan los índices locales de Lebart, pues los numeradores en la definición son los mismos, y sólo se diferencian en los denominadores.

Por la otra parte, el cociente s_t^2/s_w^2 es una proporción de lo que explican los valores suavizados de las vecindades.

$$\delta_2^* = \frac{\sum_{i=1}^n \sum_{j=1}^n a_{ij} (y_i - x_j)^2}{\sum_{i=1}^n \sum_{j=1}^n a_{ij} (x_i - x_j)^2} = \frac{\sum_{i=1}^n \sum_{j=1}^n a_{ij} (x_i - y_j)^2}{\sum_{i=1}^n \sum_{j=1}^n a_{ij} (x_i - x_j)^2}$$

Su valores pueden resultar entre 0 a 1. Si se acerca 1 quiere decir que los valores alisados son próximos a los medidos, es decir, que hay alta autocorrelación, y si a 0, que hay baja autocorrelación.

Como s_t^2 es $\text{cov}(X', T)$ y s_w^2 la varianza de $X-X'$, δ_2^* resulta ser la pendiente de la recta de regresión de $(X-X')$ vs. $(Y-X')$ (ver Figura 4-15).



Para este índice también pueden considerarse dos índices locales para cada objeto ω_i , según se considere el ordenamiento de la descomposición del numerador

$$\delta_{2i}^{*'} = \frac{\frac{1}{A_i} \sum_{j=1}^n a_{ij} (y_i - x_j)^2}{s_w^2}$$

$$\delta_{2i}^{*''} = \frac{\frac{1}{A_i} \sum_{j=1}^n a_{ij} (x_i - y_j)^2}{s_w^2}$$

que, como puede observarse, son proporcionales a $\eta_{2i}^{*'}$ y $\eta_{2i}^{*''}$, respectivamente, y, por tanto, no aportan ninguna información novedosa.

4.2.5 Relaciones entre los diversos tipos de índices

En este trabajo se han definido diferentes índices de autocorrelación espacial basados en la noción de varianza o covarianza de determinadas variables estadísticas derivadas de una inicial. Es de esperar, por tanto, que aun cuando cada uno de ellos expresa una faceta distinta de la autocorrelación, están relacionados entre sí.

Los índices globales presentados pueden ser descompuestos como sumas ponderadas de los índices locales respectivos. De hecho, se interpretan como valores medios de los mismos. A continuación se describen las principales relaciones encontradas entre los índices globales, y entre los locales.

4.2.5.1 Relaciones entre índices globales

El índice δ_1^* (s_v^2/s_w^2) está relacionado con el índice de Lebart (s_v^2/s_A^2) por tener el mismo numerador, y, por tanto, recordando la fórmula del índice modificado de Geary ($s_w^2/2s_A^2$), se tiene

$$L^* = 2C^* \delta_1^*, \text{ o}$$

$$\delta_1^* = \frac{1}{2} \frac{L^*}{C^*}$$

De esta relación y de los límites de C^* y δ_1^* , se deducen también los límites del índice modificado de Lebart: $0 \leq L^* \leq 4$.

De la relación de igualdad $s_w^2 = s_v^2 + s_t^2$, dividiendo por s_A^2 se tiene $2C^* = L^* + \eta_2^*$. Por tanto

$$L^* = 2C^* - \eta_2^*, \text{ o}$$

$$\eta_2^* = 2C^* - L^*$$

Las relaciones $L^* = 2C^* \delta_1^*$ y $L^* = 2C^* - \eta_2^*$ pueden resultar de gran utilidad para analizar la relación entre los índices modificados de Geary y de Lebart.

Para $\eta_2^* = s_t^2/s_A^2$ y de $\delta_2^* = s_t^2/s_w^2$, al tener el mismo numerador, se establece la siguiente relación

$$\eta_2^* = 2C^* \delta_2^*$$

Resumiendo, los índices globales de autocorrelación presentados en este apartado son expresables en función de únicamente dos: C^* y L^* (ver Tabla 4-3).

Tabla 4-3 Relación entre los diferentes índices

I^*	C^*	L^*	η_1^*	η_2^*	δ_1^*	δ_2^*
$1 - C^*$	C^*	L^*	$1 - 2 \cdot C^* + L^*$	$2 \cdot C^* - L^*$	$\frac{L^*}{2 \cdot C^*}$	$\frac{2 \cdot C^* - L^*}{2 \cdot C^*}$

Los índices calculados para el ejemplo de la Figura 4-3 se muestran en la Tabla 4-4.

Tabla 4-4 Valores de los índices calculados para el ejemplo de la Figura 4-3

I^*	C^*	L^*	η_1^*	η_2^*	δ_1^*	δ_2^*
-0.29	1.29	1.81	0.24	0.76	0.70	0.30
	1.00			1.00		1.00

4.2.5.2 Relaciones entre índices locales

De la igualdad establecida en el subapartado de definiciones previas y propiedades, se tiene la igualdad análoga a $s_w^2 = s_v^2 + s_t^2$

$$\frac{1}{A_i} \sum_{j=1}^n a_{ij} (x_i - x_j)^2 = (x_i - y_i)^2 + \frac{1}{A_i} \sum_{j=1}^n a_{ij} (x_j - y_i)^2$$

Si se dividen ambas partes de la igualdad por s_A^2 , se tiene

$$2C_i^* = L_i^* + \eta_{2i}^*$$

Por otra parte, $L_i^* = 2C^* \delta_{li}^*$ y $\eta_{2i}^* = 2C^* \delta_{2i}^*$, con lo cual los índices locales C_i^* y L_i^* son suficientes para expresar η_{2i}^* , δ_{li}^* y δ_{2i}^* . Quedan al margen los índices η_{2i}^* y δ_{2i}^* , que están relacionados mediante la fórmula $\delta_{2i}^* = 2C^* \eta_{2i}^*$.

Por una analogía con la igualdad anterior, conviene señalar que, si bien la ecuación $s_A^2 = s_y^2 + s_t^2$ expresada según sus definiciones sugiere una igualdad para cada objeto ω_i

$$\frac{1}{A_{Total}} \sum_{i=1}^n A_i (x_i - \bar{x}_A)^2 = \frac{1}{A_{Total}} \sum_{i=1}^n A_i \left((y_i - \bar{x}_A)^2 + \frac{1}{A_i} \sum_{j=1}^n a_{ij} (y_i - x_j)^2 \right)$$

en general, es falso que los sumandos respectivos coincidan

$$(x_i - \bar{x}_A)^2 \neq (y_i - \bar{x}_A)^2 + \frac{1}{A_i} \sum_{j=1}^n a_{ij} (y_i - x_j)^2$$

por lo cual no es posible establecer, dividiendo por s_A^2 , una relación entre η_{li}^* y η_{2i}^* . Tampoco se puede establecer una relación funcional entre I_i^* y C_i^* .

Resumiendo se tienen, al menos, cinco índices locales no relacionados directamente: I_i^* , C_i^* , L_i^* , η_{li}^* y η_{2i}^* , y uno, η_{2i}^* , relacionado con dos de ellos.

Para el ejemplo de la Figura 4-3 se tienen los resultados que siguen (ver Tabla 4-5).

Tabla 4-5 Valores de los índices locales calculados para el ejemplo de la Figura 4-3

Ω	A_i	I_i^*	C_i^*	L_i^*	η_{1i}^*	η_{2i}^*	$\eta_{2i}^{\prime*}$
ω_1	2	0.00	0.43	0.86	0.86	0.00	0.19
ω_2	3	0.00	0.43	0.86	0.00	0.00	0.57
ω_3	2	0.00	0.43	0.00	0.00	0.86	0.19
ω_4	3	-0.57	1.57	2.38	0.38	0.76	0.57
ω_5	4	0.00	0.43	0.00	0.00	0.86	0.19
ω_6	3	-0.57	1.57	2.38	0.38	0.76	0.57
ω_7	2	0.00	2.14	3.43	0.00	0.86	2.48
ω_8	3	0.00	1.57	0.86	0.00	2.29	0.57
ω_9	2	-1.71	3.86	7.71	0.86	0.00	2.48

La matriz de correlaciones se muestra en la Tabla 4-6.

Tabla 4-6 Matriz de correlaciones entre los índices para el ejemplo de la Figura 4-3

	I_i^*	C_i^*	L_i^*	η_{1i}^*	η_{2i}^*	$\eta_{2i}^{\prime*}$
I_i^*	1.00	-0.83	-0.90	-0.69	0.32	-0.56
C_i^*		1.00	0.95	0.47	0.02	0.84
L_i^*			1.00	0.61	-0.31	0.86
η_{1i}^*				1.00	-0.50	0.25
η_{2i}^*					1.00	-0.17
$\eta_{2i}^{\prime*}$						1.00

4.2.6 Casos particulares de estructura espacial

En este apartado se consideran determinadas estructuras espaciales, caracterizadas por la matriz de adyacencias A , que resultan interesantes porque definen situaciones extremas, y sirven de referencia para la interpretación de otras estructuras.

En cada una de ellas se analiza el comportamiento de los valores de los índices de autocorrelación globales y locales detallados en los apartados anteriores.

4.2.6.1 Caso a)

La matriz de adyacencias, A , está repleta de 1s ($a_{ij} = 1$, $i, j = 1, \dots, n$, incluidos los $a_{ii} = 1$), y se designa como $A = U$. Se trata de un conjunto sin estructura espacial: $C^* = 1$ ($I^* = 0$), $L^* = 1$.

En efecto, $A_i = n$, y $A_{Total} = n^2$. El sistema de pesos espaciales tiene la propiedad de que

$$\frac{a_{ij}}{A_{Total}} = \frac{A_i}{A_{Total}} \frac{A_j}{A_{Total}}, \text{ pues } \frac{1}{n^2} = \frac{n}{n^2} \frac{n}{n^2}$$

la cual es análoga a la de independencia de sucesos en teoría de la probabilidad. Se diría que un objeto es espacialmente independiente de sus vecinos, precisamente porque es adyacente a todos, incluido a sí mismo.

Además, se tiene entonces que tanto la media como la varianza modificadas resultan ser las habituales

$$\bar{x}_A = \frac{1}{A_{Total}} \sum_{i=1}^n A_i x_i = \frac{1}{n^2} \sum_{i=1}^n n x_i = \frac{1}{n} \sum_{i=1}^n x_i = \bar{x}$$

$$s_A^2 = \frac{1}{A_{Total}} \sum_{i=1}^n A_i (x_i - \bar{x}_A)^2 = \frac{1}{n^2} \sum_{i=1}^n n (x_i - \bar{x})^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = s^2$$

El valor del índice modificado de Geary, siempre que haya cierta variabilidad ($s^2 \neq 0$), es, por tanto, $C^* = 1$

$$\begin{aligned} C^* &= \frac{1}{2} \frac{1}{A_{Total}} \sum_{i=1}^n \sum_{j=1}^n a_{ij} \left(\frac{x_i - x_j}{s_A} \right)^2 = \frac{1}{2} \frac{1}{n^2} \frac{1}{s^2} \sum_{i=1}^n \sum_{j=1}^n ((x_i - \bar{x}) - (x_j - \bar{x}))^2 = \\ &= \frac{1}{2} \frac{1}{n^2} \frac{1}{s^2} \sum_{i=1}^n \sum_{j=1}^n ((x_i - \bar{x}) - (x_j - \bar{x}))^2 = \\ &= \frac{1}{2} \frac{1}{n^2} \frac{1}{s^2} \sum_{i=1}^n \sum_{j=1}^n ((x_i - \bar{x})^2 - 2(x_i - \bar{x})(x_j - \bar{x}) + (x_j - \bar{x})^2) = \\ &= \frac{1}{2} \frac{1}{n^2} \frac{1}{s^2} \left(2 \sum_{i=1}^n n(x_i - \bar{x})^2 - 2 \sum_{i=1}^n \sum_{j=1}^n (x_i - \bar{x})(x_j - \bar{x}) \right) = \\ &= \frac{1}{s^2} \left(\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \right) = 1 \end{aligned}$$

En realidad, al estar todos los valores conectados con todos los demás, la autocorrelación espacial no tiene sentido, pues no hay tal. Se puede tratar como si fuera el caso del enfoque clásico o habitual, en el que los I s representan copertenencias a una clase (apartado 2.2.1). En el enfoque clásico no se consideran adyacencias espaciales, aunque, de hecho, desde este enfoque, resulte ser equivalente a que todos los objetos sean adyacentes espacialmente

a todos los demás. Por tanto, $I^* = 0$, lo cual denota más explícitamente la ausencia de autocorrelación espacial.

El índice modificado de Lebart, por supuesto, también toma el valor $L^* = 1$ que denota la misma situación anterior. En efecto, como

$$y_i = \frac{1}{A_i} \sum_{j=1}^n a_{ij} x_j = \frac{1}{n} \sum_{j=1}^n x_j = \bar{x},$$

se tiene que

$$L^* = \frac{\frac{1}{A_{Total}} \sum_{i=1}^n A_i (x_i - y_i)^2}{s_A^2} = \frac{\frac{1}{n^2} \sum_{i=1}^n n (x_i - \bar{x})^2}{s^2} = 1$$

Por tanto, de las relaciones encontradas, se tiene que los valores de los nuevos índices propuestos son

$$\eta_2^* = 2C^* - L^* = 1 \quad \text{y} \quad \eta_1^* = 1 - \eta_2^* = 0$$

En cuanto a los índices locales, se tiene que, como $y_i = \bar{x}$, los de Moran son nulos

$$I_i^* = \frac{(x_i - \bar{x}_A)(y_i - \bar{x}_A)}{s_A^2} = \frac{(x_i - \bar{x})(\bar{x} - \bar{x})}{s^2} = 0$$

y los de Geary son una función cuadrática de los valores típicos o estándares

$$\begin{aligned} C_i^* &= \frac{1}{2} \frac{1}{A_i} \sum_{j=1}^n a_{ij} \frac{(x_i - x_j)^2}{s_A^2} = \frac{1}{2} \frac{1}{n} \frac{1}{s^2} \sum_{j=1}^n (x_i - x_j)^2 = \frac{1}{2} \frac{1}{n} \frac{1}{s^2} \sum_{j=1}^n ((x_i - \bar{x}) - (x_j - \bar{x}))^2 = \\ &= \frac{1}{2} \left[\left(\frac{x_i - \bar{x}}{s} \right)^2 + 1 \right] = \frac{1}{2} (z_i^2 + 1) \end{aligned}$$

Por lo que se refiere a los índices locales de Lebart, es de señalar que estos son precisamente los cuadrados de los valores típicos

$$L_i^* = \frac{(x_i - y_i)^2}{s_A^2} = \frac{(x_i - \bar{x})^2}{s^2} = \left(\frac{x_i - \bar{x}}{s} \right)^2 = z_i^2$$

por lo que

$$C_i^* = \frac{1}{2}(L_i^* + 1).$$

La ausencia de autocorrelación queda también reflejada en los nuevos índices η propuestos

$$\begin{aligned} \eta_{1i}^* &= \frac{(y_i - \bar{x}_A)^2}{s_A^2} = \frac{(\bar{x} - \bar{x})^2}{s^2} = 0 \\ \eta_{2i}^* &= 2C_i^* - L_i^* = 1 \\ \eta_{2i}^{**} &= \frac{\frac{1}{A_i} \sum_{j=1}^n a_{ij} (x_i - y_j)^2}{s_A^2} = \frac{\frac{1}{n} \sum_{j=1}^n (x_i - \bar{x})^2}{s^2} = 1 \end{aligned}$$

4.2.6.2 Caso b)

El conjunto de objetos, Ω , no tiene más adyacencias que las que asocian los objetos consigo mismo. Esta situación, en el caso de que a_{ij} sea binario, 0 ó 1, se puede expresar de la siguiente manera

$$a_{ij} = \begin{cases} 1, & \text{si } i = j \\ 0, & \text{si } i \neq j \end{cases}$$

La matriz de adyacencias, A , es la matriz con la diagonal repleta de 1s y el resto de 0s, es decir, la matriz identidad: $A = D_1$.

En este caso se tiene que el peso total de cada objeto es igual a 1, es decir, $A_i = 1$, de donde se obtiene que el peso total del conjunto es igual a n , o sea, $A_{Total} = n$.

Las fórmulas de la media y la varianza definidas antes se mantienen en relación a las habituales. Así se tiene que

$$\begin{aligned} \bar{x}_A &= \frac{1}{A_{Total}} \sum_{i=1}^n A_i x_i = \frac{1}{n} \sum_{i=1}^n x_i = \bar{x} \\ s_A^2 &= \frac{1}{A_{Total}} \sum_{i=1}^n A_i (x_i - \bar{x}_A)^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = s^2 \end{aligned}$$

El índice modificado de Moran, que coincide con el coeficiente de correlación lineal de la variable consigo misma, vale 1

$$I^* = \frac{1}{A_{Total}} \sum_{i=1}^n \sum_{j=1}^n a_{ij} \left(\frac{x_i - \bar{x}_A}{s_A} \right) \left(\frac{x_j - \bar{x}_A}{s_A} \right) = \frac{1}{n} \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{s} \right)^2 = 1$$

Por tanto, $C^* = 0$, y como $y_i = x_i$, $L^* = 0$.

Este resultado indica que para un conjunto de objetos esparcidos en el espacio sin relación entre ellos excepto consigo mismo, independientemente de los valores asociados, se obtiene los valores mínimos para C^* y L^* , y, por tanto, el máximo para I^* , lo que indica que la autocorrelación es totalmente positiva.

Los valores de los nuevos índices η propuestos son

$$\eta_2^* = 2C^* - L^* = 0 \quad \text{y} \quad \eta_1^* = 1 - \eta_2^* = 1$$

Los índices δ no se pueden calcular por tener el denominador de su definición nulo. Los índices locales tampoco tienen sentido por la ausencia de vecindad.

4.2.6.3 Caso c)

La matriz de adyacencias, A , está repleta de 1 s ($a_{ij} = 1$, $i, j = 1, \dots, n$, salvo los $a_{ii} = 0$, tal y como se define habitualmente). Se puede expresar como la diferencia de las matrices de los dos casos anteriores: $A = U - D_1$, donde U es la matriz del caso **a)** anterior repleta de 1 s, y D_1 es la matriz con la diagonal repleta de 1 s y el resto de 0 s, es decir, la matriz identidad. El grafo correspondiente es completo.

Hay que señalar que éste es el caso de relación que habitualmente se considera en un análisis de datos sin estructura entre los objetos, es decir, no se consideran las adyacencias reflexivas de los objetos.

Luego, $A_i = n - 1$, y $A_{Total} = n^2 - n = n(n - 1)$.

Ello conduce a que la media y la varianza sean las habituales, como en los dos casos anteriores.

$$\bar{x}_A = \frac{1}{A_{Total}} \sum_{i=1}^n A_i x_i = \frac{1}{n(n-1)} \sum_{i=1}^n (n-1) x_i = \frac{1}{n} \sum_{i=1}^n x_i = \bar{x}$$

$$s_A^2 = \frac{1}{A_{Total}} \sum_{i=1}^n A_i (x_i - \bar{x}_A)^2 = \frac{1}{n(n-1)} \sum_{i=1}^n (n-1)(x_i - \bar{x}_A)^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = s^2$$

En cuanto al índice modificado de Moran se llega a lo siguiente

$$\begin{aligned} I^* &= \frac{1}{A_{Total}} \sum_{i=1}^n \sum_{j=1}^n a_{ij} \left(\frac{x_i - \bar{x}_A}{s_A} \right) \left(\frac{x_j - \bar{x}_A}{s_A} \right) = \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{j=1}^n (u_{ij} - i_{ij}) \left(\frac{x_i - \bar{x}}{s} \right) \left(\frac{x_j - \bar{x}}{s} \right) = \\ &= \frac{1}{n(n-1)} \left[\sum_{i=1}^n \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{s} \right) \left(\frac{x_j - \bar{x}}{s} \right) - \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{s} \right)^2 \right] = \\ &= \frac{1}{n(n-1)} \left[\sum_{i=1}^n \sum_{j=1}^n \left(\frac{x_i - \bar{x}}{s} \right) \left(\frac{x_j - \bar{x}}{s} \right) - \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{s} \right)^2 \right] = \frac{1}{n(n-1)} [0 - n] = -\frac{1}{n-1} \end{aligned}$$

El valor del índice modificado de Moran

$$I^* = \frac{-1}{n-1}$$

coincide con la esperanza del índice clásico de Moran en el caso de independencia entre los valores asignados a los objetos. Para establecer este valor se hace el supuesto de que $a_{ii}=0$. Se obtiene así el mismo valor para el índice clásico de Moran desde dos puntos de vista distintos.

Hay que tener en cuenta que, sin embargo, en el desarrollo que estamos presentando

$$\frac{a_{ij}}{A_{Total}} \neq \frac{A_i}{A_{Total}} \frac{A_j}{A_{Total}},$$

lo que indica que no son completamente independientes en ese contexto, si bien se aproxima a ello cuando n es grande,

$$\frac{1}{n(n-1)} \neq \frac{(n-1)}{n(n-1)} \frac{(n-1)}{n(n-1)}$$

Consecuentemente, el índice modificado de Geary es

$$C^* = 1 - I^* = 1 + \frac{1}{n-1} = \frac{n}{n-1}.$$

Para el cálculo del índice modificado de Lebart, se tienen en cuenta el valor medio de la vecindad

$$y_i = \frac{1}{A_i} \sum_{j=1}^n a_{ij} x_j = \frac{1}{n-1} \left(\sum_{j=1}^n x_j - x_i \right) = \frac{n}{n-1} \bar{x} - \frac{1}{n-1} x_i,$$

y su desviación respecto a x_i ,

$$x_i - y_i = x_i - \left(\frac{n}{n-1} \bar{x} - \frac{1}{n-1} x_i \right) = \frac{n}{n-1} (x_i - \bar{x}),$$

y así se tiene que

$$L^* = \frac{\frac{1}{A_{Total}} \sum_{i=1}^n A_i (x_i - y_i)^2}{s_A^2} = \frac{\frac{1}{n(n-1)} \sum_{i=1}^n (n-1) \left(\frac{n}{n-1} (x_i - \bar{x}) \right)^2}{s^2} = \left(\frac{n}{n-1} \right)^2.$$

Por su parte, se tiene que los valores de los nuevos índices propuestos son

$$\eta_2^* = 2C^* - L^* = 2 \frac{n}{n-1} - \left(\frac{n}{n-1} \right)^2 = \frac{n}{n-1} \left(2 - \frac{n}{n-1} \right) = \frac{n}{n-1} \frac{n-2}{n-1}$$

$$\text{y } \eta_1^* = 1 - \eta_2^* = \left(\frac{1}{n-1} \right)^2$$

En cuanto a los índices locales, se tiene que, como

$$y_i = \frac{n}{n-1} \bar{x} - \frac{1}{n-1} x_i, \text{ e } y_i - \bar{x} = \frac{n}{n-1} \bar{x} - \frac{1}{n-1} x_i - \bar{x} = \frac{-1}{n-1} (x_i - \bar{x}),$$

los de Moran son proporcionales a los cuadrados de los valores típicos

$$I_i^* = \frac{(x_i - \bar{x}_A)(y_i - \bar{x}_A)}{s_A^2} = \frac{-1}{n-1} \frac{(x_i - \bar{x})^2}{s^2} = \frac{-1}{n-1} z_i^2$$

Para los índices locales de Geary

$$\begin{aligned}
C_i^* &= \frac{1}{2} \frac{1}{A_i} \sum_{j=1}^n a_{ij} \frac{(x_i - x_j)^2}{s_A^2} = \frac{1}{2} \frac{1}{n-1} \frac{1}{s^2} \sum_{j=1}^n (x_i - x_j)^2 = \frac{1}{2} \frac{n}{n-1} \frac{1}{s^2} \frac{1}{n} \sum_{j=1}^n \left((x_i - \bar{x}) - (x_j - \bar{x}) \right)^2 = \\
&= \frac{1}{2} \frac{n}{n-1} \frac{1}{s^2} \frac{1}{n} \left(\sum_{j=1}^n (x_i - \bar{x})^2 - 2 \sum_{j=1}^n (x_i - \bar{x})(x_j - \bar{x}) + \sum_{j=1}^n (x_j - \bar{x})^2 \right) = \\
&= \frac{1}{2} \frac{n}{n-1} \left[\left(\frac{x_i - \bar{x}}{s} \right)^2 + 1 \right] = \frac{n}{n-1} \frac{1}{2} (z_i^2 + 1)
\end{aligned}$$

En cuanto a los índices locales de Lebart

$$L_i^* = \frac{(x_i - y_i)^2}{s_A^2} = \left(\frac{n}{n-1} \right)^2 \frac{(x_i - \bar{x})^2}{s^2} = \left(\frac{n}{n-1} \right)^2 z_i^2$$

En los nuevos índices η propuestos se tiene

$$\begin{aligned}
\eta_{ii}^* &= \frac{(y_i - \bar{x}_A)^2}{s_A^2} = \left(\frac{-1}{n-1} \right)^2 \left(\frac{x_i - \bar{x}}{s} \right)^2 = \frac{1}{(n-1)^2} z_i^2 \\
\eta_{2i}^* &= 2C_i^* - L_i^* = \frac{n}{n-1} (z_i^2 + 1) - \left(\frac{n}{n-1} \right)^2 z_i^2 = \frac{n}{n-1} - \frac{n}{n-1} z_i^2 \left(\frac{-1}{n-1} \right) = \frac{n}{n-1} \left(1 + \frac{1}{n-1} z_i^2 \right) \\
\eta_{2i}^{**} &= \frac{\frac{1}{A_i} \sum_{j=1}^n a_{ij} (x_i - y_j)^2}{s_A^2} = \frac{1}{n-1} \frac{\left(\sum_{j=1}^n (x_j - \bar{x})^2 - (x_i - \bar{x})^2 \right)}{s^2} = \\
&= \frac{\frac{n}{n-1} s^2 - \frac{1}{n-1} (x_i - \bar{x})^2}{s^2} = \frac{n}{n-1} - \frac{1}{n-1} z_i^2
\end{aligned}$$

Los valores de todos los índices locales se aproximan, lógicamente, a los valores de los índices del caso **a)** en el que todos los objetos son adyacentes a todos, incluidos ellos mismos, cuando n se hace grande.

4.2.6.4 Caso d)

En este caso se considera que la estructura de las relaciones de los objetos de Ω consiste en una partición, es decir, la matriz de adyacencias A puede verse, mediante reorganización de sus filas y columnas, como una matriz organizada diagonalmente en K bloques, cada uno de los cuales representa la conexión entre n_k objetos de determinados subconjuntos.

Si cada bloque se representa por la matriz de adyacencias A^k $k=1,2,\dots,K$, en la cual se tiene $a_{ij}^k = 0$ si ω_j no es de la misma parte, entonces

$$A = \sum_{k=1}^K A^k$$

Hay que aclarar que el hecho de organizarse en K bloques no implica que cada bloque esté necesariamente repleta de 1 s, es decir, no es necesario que el subgrafo correspondiente sea completo.

Los casos extremos $K=1$ y $K=n$ están ya analizados, pues se trata de los casos **a)** y **b)**, respectivamente. Incluso el caso **c)** puede considerarse como un caso particular extremo ($K=1$).

En general se tiene que

$$\begin{aligned}\bar{x}_A &= \frac{1}{A_{Total}} \sum_{i=1}^n A_i x_i = \frac{1}{A_{Total}} \sum_{k=1}^K \sum_{i=1}^{n_k} A_i^k x_i^k = \frac{1}{A_{Total}} \sum_{k=1}^K A_{Total}^k \bar{x}_{A^k} \\ s_A^2 &= \frac{1}{A_{Total}} \sum_{k=1}^K A_{Total}^k s_{A^k}^2 + \frac{1}{A_{Total}} \sum_{k=1}^K A_{Total}^k (\bar{x}_{A^k} - \bar{x}_A)^2\end{aligned}$$

En esta última igualdad, el primer sumando es un promedio de las varianzas dentro de cada bloque (*within*), y el segundo una varianza entre los bloques (*between*). Con estas expresiones se pueden expresar los valores globales de los índices en función de los valores que alcanzan dentro de cada bloque particular, y los valores entre los bloques.

$$\begin{aligned}I^* &= \frac{1}{A_{Total}} \frac{1}{s_A^2} \sum_{k=1}^K \sum_{i=1}^{n_k} \sum_{j=1}^{n_k} a_{ij}^k (x_i^k - \bar{x}_A)(x_j^k - \bar{x}_A) = \\ &= \frac{1}{A_{Total}} \frac{1}{s_A^2} \sum_{k=1}^K \sum_{i=1}^{n_k} \sum_{j=1}^{n_k} a_{ij}^k \left((x_i^k - \bar{x}_{A^k}) - (\bar{x}_{A^k} - \bar{x}_A) \right) \left((x_j^k - \bar{x}_{A^k}) - (\bar{x}_{A^k} - \bar{x}_A) \right) = \\ &= \frac{1}{A_{Total}} \frac{1}{s_A^2} \sum_{k=1}^K \sum_{i=1}^{n_k} \sum_{j=1}^{n_k} a_{ij}^k (x_i^k - \bar{x}_{A^k})(x_j^k - \bar{x}_{A^k}) + \frac{1}{A_{Total}} \frac{1}{s_A^2} \sum_{k=1}^K \sum_{i=1}^{n_k} \sum_{j=1}^{n_k} a_{ij}^k (\bar{x}_{A^k} - \bar{x}_A)^2 = \\ &= \sum_{k=1}^K \frac{A_{Total}^k}{A_{Total}} \frac{s_{A^k}^2}{s_A^2} I_k^* + \frac{1}{s_A^2} \sum_{k=1}^K \frac{A_{Total}^k}{A_{Total}} (\bar{x}_{A^k} - \bar{x}_A)^2\end{aligned}$$

Se observa que el valor global del coeficiente no es el valor medio del de las partes correspondientes a los bloques, sino que los valores de los índices de los bloques están afectados por unos coeficientes dependientes de las varianzas parciales $s_{A^k}^2 / s_A^2$.

Además, está el término de la varianza entre (*between*) las partes relativizada respecto a la varianza global (algo así como la razón de correlación), que es una cantidad positiva entre 0 y 1.

El índice de Geary, por su parte, se descompone así

$$C^* = \frac{1}{2} \frac{1}{s_A^2} \frac{1}{A_{Total}} \sum_{k=1}^K \sum_{i=1}^{n_k} \sum_{j=1}^{n_k} a_{ij}^k (x_i^k - x_j^k)^2 = \sum_{k=1}^K \frac{A_{Total}^k}{A_{Total}} \frac{s_{A^k}^2}{s_A^2} C_k^*$$

Aquí también es preciso a señalar que C^* no es un valor promediado por las varianzas y pesos espaciales de los índices particulares.

Al ser $C_k^* + I_k^* = 1$, se verifica que

$$\begin{aligned} C^* + I^* &= \sum_{k=1}^K \frac{A_{Total}^k}{A_{Total}} \frac{s_{A^k}^2}{s_A^2} C_k^* + \sum_{k=1}^K \frac{A_{Total}^k}{A_{Total}} \frac{s_{A^k}^2}{s_A^2} I_k^* + \frac{1}{s_A^2} \sum_{k=1}^K \frac{A_{Total}^k}{A_{Total}} (\bar{x}_{A^k} - \bar{x}_A)^2 = \\ &= \sum_{k=1}^K \frac{A_{Total}^k}{A_{Total}} \frac{s_{A^k}^2}{s_A^2} + \frac{1}{s_A^2} \sum_{k=1}^K \frac{A_{Total}^k}{A_{Total}} (\bar{x}_{A^k} - \bar{x}_A)^2 = \frac{1}{s_A^2} \sum_{k=1}^K \frac{A_{Total}^k}{A_{Total}} (s_{A^k}^2 + (\bar{x}_{A^k} - \bar{x}_A)^2) = 1 \end{aligned}$$

En cuanto al índice modificado de Lebart se tiene una relación similar

$$L^* = \frac{1}{A_{Total}} \frac{\sum_{k=1}^K \sum_{i=1}^{n_k} A_i^k (x_i^k - y_i^k)^2}{s_A^2} = \frac{1}{A_{Total}} \frac{\sum_{k=1}^K A_{Total}^k s_{A^k}^2 L_k^*}{s_A^2} = \sum_{k=1}^K \frac{A_{Total}^k}{A_{Total}} \frac{s_{A^k}^2}{s_A^2} L_k^*$$

A partir de estas fórmulas se hacen las siguientes consideraciones:

Cuando todos los valores medios de las partes conexas coinciden ($\bar{x}_{A^k} = \bar{x}_A$, $k=1, 2, \dots, K$), entonces $A_{Total} s_A^2 = \sum_{k=1}^K A_{Total}^k s_{A^k}^2$, y, por tanto, $I^* = C^*$ y L^* son los valores medios (ponderados por pesos espaciales y varianzas parciales) de los valores parciales I_k^* , C_k^* y L_k^* , respectivamente. Y viceversa.

El hecho de que las K partes, además de tener la misma media \bar{x}_{A^k} , tengan la misma varianza ($s_{A^k}^2 = s_A^2 \neq 0$), no implica necesariamente que sea $I^* = 0$ y $C^* = 1$ (ausencia de autocorrelación), ya que en ese caso serían el valor medio de los valores parciales, ponderados estos sólo por los pesos espaciales

$$I^* = \sum_{k=1}^K \frac{A_{Total}^k}{A_{Total}} I_k^* \quad C^* = \sum_{k=1}^K \frac{A_{Total}^k}{A_{Total}} C_k^* \quad L^* = \sum_{k=1}^K \frac{A_{Total}^k}{A_{Total}} L_k^*$$

Si todas las partes fueran totalmente homogéneas ($x_i^k = x_j^k, i, j = 1, 2, \dots, n_k$, $s_{A^k}^2 = 0, k = 1, 2, \dots, K$, siendo $s_A^2 \neq 0$, es decir, si hubiera al menos dos valores medios parciales distintos), entonces $I^* = 1$, $C^* = 0$ y $L^* = 0$, lo cual indica que se trata de una situación de autocorrelación positiva máxima. Esta situación es equiparable al caso **b)** de objetos no relacionados con los demás, que es un caso particular de la estructura en bloques: $K=n, n_k=1, k=1, 2, \dots, n$, con $a_{ii} \neq 0$ y las partes totalmente homogéneas.

Para establecer este resultado general hay que acudir a la definición misma de C^* , ya que las C_k^* no podrían definirse

$$C^* = \frac{1}{2} \frac{1}{A_{Total}} \sum_{k=1}^K \sum_{i=1}^{n_k} \sum_{j=1}^{n_k} a_{ij}^k \left(\frac{x_i^k - x_j^k}{s_A} \right)^2 = 0$$

Si las partes no estuvieran en absoluto autocorreladas ($I_k^* = 0, C_k^* = 1, k = 1, 2, \dots, K$) entonces

$$C^* = \sum_{k=1}^K \frac{A_{Total}^k}{A_{Total}} \frac{s_{A^k}^2}{s_A^2}$$

y el conjunto global Ω estaría autocorrelado positivamente, $0 < C^* \leq 1$, ya que

$$A_{Total} s_A^2 \geq \sum_{k=1}^K A_{Total}^k s_{A^k}^2.$$

No se podría alcanzar la máxima autocorrelación positiva ($I^* = 1, C^* = 0$) pues ello implicaría que $s_{A^k}^2 = 0, k = 1, 2, \dots, K$, lo cual es contradictorio con $C_k^* = 1, k = 1, 2, \dots, K$. En cambio, podría llegar a ser $C^* = 1$ si todos los valores medios fueran iguales $\bar{x}_{A^k} = \bar{x}_A, k = 1, 2, \dots, K$.

Un caso particular que refleja esta situación es aquella en la que en cada una de las K partes todos los objetos están relacionados con todos los demás de su parte, y es, además, el caso de una relación de adyacencia reflexiva ($a_{ii} = 1, i = 1, \dots, n$).

Entonces se tendría $A_i^k = n_k, A_{Total}^k = n_k^2, A_{Total} = \sum_{k=1}^K n_k^2$, con $C_k^* = 1, k = 1, 2, \dots, K$

(por tanto, $s_{A^k}^2 \neq 0, k = 1, 2, \dots, K$). Así

$$C^* = \frac{\frac{1}{\sum_{k=1}^K n_k^2} \sum_{k=1}^K n_k^2 s_{A^k}^2}{s_A^2}$$

Es decir, si las partes no están autocorreladas $I_k^* = 0$, $k = 1, 2, \dots, K$ en absoluto, entonces

$$I^* = \frac{\frac{1}{\sum_{k=1}^K n_k^2} \sum_{k=1}^K n_k^2 (\bar{x}_{A^k} - \bar{x}_A)^2}{s_A^2}$$

lo cual indica una autocorrelación no negativa, que sólo es nula ($I^* = 0$, $C^* = 1$) cuando los valores medios parciales coinciden.

Si la relación espacial no fuera reflexiva ($a_{ii} = 0$, $\forall i = 1, \dots, n$), las fórmulas anteriores habría que precisarlas considerando para cada parte que $A_i^k = n_k - 1$,

$$A_{Total}^k = n_k(n_k - 1) \text{ y } C_k^* = \frac{n_k}{n_k - 1}.$$

$$\text{Además } A_{Total} = \sum_{k=1}^K n_k(n_k - 1) = \sum_{k=1}^K n_k^2 - n$$

$$C^* = \frac{\frac{1}{\sum_{k=1}^K n_k(n_k - 1)} \sum_{k=1}^K n_k(n_k - 1) s_{A^k}^2}{s_A^2}$$

En cuanto a los índices locales, basta considerar las definiciones para establecer la relación entre los valores parciales y globales

$$\begin{aligned} I_i^* &= \frac{(x_i - \bar{x}_A)(y_i - \bar{x}_A)}{s_A^2} = \frac{((x_i - \bar{x}_{A^k}) + (\bar{x}_{A^k} - \bar{x}_A))((y_i - \bar{x}_{A^k}) + (\bar{x}_{A^k} - \bar{x}_A))}{s_A^2} = \\ &= \frac{(x_i - \bar{x}_{A^k})(y_i - \bar{x}_{A^k}) + (\bar{x}_{A^k} - \bar{x}_A)(y_i - \bar{x}_{A^k}) + (x_i - \bar{x}_{A^k})(\bar{x}_{A^k} - \bar{x}_A) + (\bar{x}_{A^k} - \bar{x}_A)^2}{s_A^2} = \\ &= \frac{s_{A^k}^2}{s_A^2} \frac{(x_i - \bar{x}_{A^k})(y_i - \bar{x}_{A^k})}{s_{A^k}^2} + \frac{(\bar{x}_{A^k} - \bar{x}_A)((y_i - \bar{x}_{A^k}) + (x_i - \bar{x}_{A^k}) + (\bar{x}_{A^k} - \bar{x}_A))}{s_A^2} = \\ &= \frac{s_{A^k}^2}{s_A^2} I_{ki}^* + \frac{(\bar{x}_{A^k} - \bar{x}_A)((y_i - \bar{x}_{A^k}) + (x_i - \bar{x}_A))}{s_A^2} \end{aligned}$$

lo cual indica que los valores de los índices locales parciales de Moran pueden variar mucho cuando el contexto cambia a todos los bloques.

Sin embargo, la relación es más directa cuando se trata de los valores de los índices locales de Geary

$$C_i^* = \frac{1}{2} \frac{1}{A_i} \sum_{j=1}^n a_{ij} \frac{(x_i - x_j)^2}{s_A^2} = \frac{s_{A^k}^2}{s_A^2} \frac{1}{2} \frac{1}{A_i} \sum_{j=1}^{n_k} a_{ij} \frac{(x_i - x_j)^2}{s_{A^k}^2} = \frac{s_{A^k}^2}{s_A^2} C_{ki}^*$$

Y lo mismo sucede con los de Lebart

$$L_i^* = \frac{(x_i - y_i)^2}{s_A^2} = \frac{s_{A^k}^2}{s_A^2} \frac{(x_i - y_i)^2}{s_{A^k}^2} = \frac{s_{A^k}^2}{s_A^2} L_{ki}^*$$

Para los nuevos índices η propuestos se tiene

$$\begin{aligned} \eta_{1i}^* &= \frac{(y_i - \bar{x}_A)^2}{s_A^2} = \frac{s_{A^k}^2}{s_A^2} \frac{((y_i - \bar{x}_{A^k}) + (\bar{x}_{A^k} - \bar{x}_A))^2}{s_{A^k}^2} = \frac{s_{A^k}^2}{s_A^2} \eta_{1ki}^* + \frac{(\bar{x}_{A^k} - \bar{x}_A)((y_i - \bar{x}_{A^k}) + (y_i - \bar{x}_A))}{s_A^2} \\ \eta_{2i}^* &= 2C_i^* - L_i^* = 2 \frac{s_{A^k}^2}{s_A^2} C_{ki}^* - \frac{s_{A^k}^2}{s_A^2} L_{ki}^* = \frac{s_{A^k}^2}{s_A^2} (2C_{ki}^* - L_{ki}^*) = \frac{s_{A^k}^2}{s_A^2} \eta_{2ki}^* \\ \eta_{2i}^{*'} &= \frac{\frac{1}{A_i} \sum_{j=1}^n a_{ij} (x_i - y_j)^2}{s_A^2} = \frac{s_{A^k}^2}{s_A^2} \frac{\frac{1}{A_i} \sum_{j=1}^n a_{ij} (x_i - y_j)^2}{s_{A^k}^2} = \frac{s_{A^k}^2}{s_A^2} \eta_{2ki}^{*'} \end{aligned}$$

Resumiendo, los índices locales del estilo de Geary ($C_i^*, L_i^*, \eta_{2i}^*, \eta_{2i}^{*'}$) tienen una relación directa entre los parciales y los globales, pero los del estilo del de Moran (I_i^*, η_{1i}^*), no.

4.2.7 Aplicación del índice modificado de Geary a particiones

4.2.7.1 Introducción

En este apartado se trata de aplicar los índices del apartado anterior a la estructura de partición del conjunto de objetos $\Omega = \{\omega_1, \omega_2, \dots, \omega_n\}$. Se asume que a cada objeto se le asocia cierta magnitud, x_i , $i=1, \dots, n$ y ciertas adyacencias reflejadas por la matriz A ,

cuyos elementos a_{ij} denotan la relación espacial de los objetos ω_i y ω_j (si $a_{ij}=0$, no son adyacentes).

En realidad sólo se analizan el índice modificado de Geary y sus respectivos locales, pues el de Lebart no parece muy apropiado para esta aplicación, debido a lo complejas que resultarían las expresiones que se obtienen, como se mostrará más adelante. Tampoco se consideran los índices locales de Moran, por razones análogas.

El punto de partida de este apartado podría ser el análisis de una partición cualquiera del conjunto de los enlaces o adyacencias entre los objetos de Ω , pero dado que prácticamente interesa más una partición del conjunto mismo Ω , se considerará únicamente la partición del conjunto de adyacencias inducida por una partición de Ω .

La partición inducida del conjunto de adyacencias consta de dos partes: las adyacencias entre los objetos de una misma parte de la partición de Ω (*within*, W) y las adyacencias entre objetos de partes distintas de Ω (*between*, B), es decir, entre los objetos fronterizos. Así la matriz de adyacencias, A , se descompone aditivamente en dos matrices de las mismas dimensiones (ver apartado 2.2.1)

$$A = W + B$$

Esta descomposición induce otra del índice modificado de Geary, en la cual la parte correspondiente a B , la de los objetos fronterizos, tiene un significado especialmente interesante.

4.2.7.2 Notaciones previas para una partición

Se considera una partición de Ω en K partes, es decir, K partes no vacías y disjuntas entre sí, de tamaños n_k ($k=1,2,\dots,K$) cuya unión se corresponde con Ω , con

$$n = \sum_{k=1}^K n_k$$

Cuando se define una partición de K partes sobre los objetos Ω , implícitamente se define una partición de 2 partes en el conjunto de adyacencias que definen su estructura espacial, representados por las matrices W y B . Es decir, en virtud de la partición la matriz de adyacencias, A , puede descomponerse como suma de dos matrices, una relativa a las adyacencias dentro (*within*, W) de las partes, y otra relativa a las adyacencias fuera de cada parte, o sea, entre (*between*, B) las partes: $A=W+B$ con $a_{ij} = w_{ij} + b_{ij}$, en el cual, si ω_i y ω_j son de la misma parte $b_{ij}=0$, y si de distinta $w_{ij}=0$.

Por tanto, hay que hacer notar que los pesos espaciales de cada objeto se reparten en dos clases: los correspondientes a las adyacencias internas a su parte, y los correspondientes a

las de otras partes. De forma que en los w_{ij} se toman únicamente en cuenta los pesos espaciales internos a las partes, y en los b_{ij} los pesos espaciales fronterizos (*between*) entre las partes.

Si se reorganizaran convenientemente las filas y columnas de A , de acuerdo a la partición definida, la matriz W estaría distribuida en bloques correspondientes a las partes de la partición, y la matriz B , por el contrario, se caracterizaría por tener dichos bloques repletos de 0's (ver Figura 2-3). Las notaciones que se emplearán en lo que viene se incluyen en la Tabla 4-7.

Las relaciones directas entre estas varianzas, según el teorema de descomposición de la varianza respecto a la partición, son

$$\begin{aligned}
 s_A^2 &= \frac{1}{A_{Total}} \sum_{k=1}^K A_{Total}^k s_{A^k}^2 + \frac{1}{A_{Total}} \sum_{k=1}^K A_{Total}^k (\bar{x}_{A^k} - \bar{x}_A)^2 \\
 s_W^2 &= \frac{1}{W_{Total}} \sum_{k=1}^K W_{Total}^k s_{W^k}^2 + \frac{1}{W_{Total}} \sum_{k=1}^K W_{Total}^k (\bar{x}_{W^k} - \bar{x}_W)^2 \\
 s_B^2 &= \frac{1}{B_{Total}} \sum_{k=1}^K B_{Total}^k s_{B^k}^2 + \frac{1}{B_{Total}} \sum_{k=1}^K B_{Total}^k (\bar{x}_{B^k} - \bar{x}_B)^2
 \end{aligned}$$

Tabla 4-7 Notaciones empleadas

$A_i^k = \sum_{j=1}^n a_{ij}^k = \sum_{k'=1}^K \sum_{j=1}^{n_{k'}} a_{ij}^k$	$W_i^k = \sum_{j=1}^n w_{ij}^k = \sum_{k'=1}^K \sum_{j=1}^{n_{k'}} w_{ij}^k$ $= \sum_{j=1}^{n_k} w_{ij}^k$	$B_i^k = \sum_{j=1}^n b_{ij}^k = \sum_{k'=1}^K \sum_{j=1}^{n_{k'}} b_{ij}^k$ $= \sum_{k' \neq k} \sum_{j=1}^{n_{k'}} b_{ij}^k$
	$A_i^k = W_i^k + B_i^k$	
$A_{Total}^k = \sum_{i=1}^{n_k} A_i^k$	$W_{Total}^k = \sum_{i=1}^{n_k} W_i^k$	$B_{Total}^k = \sum_{i=1}^{n_k} B_i^k$
	$A_{Total}^k = W_{Total}^k + B_{Total}^k$	
$A_{Total} = \sum_{k=1}^K A_{Total}^k$	$W_{Total} = \sum_{k=1}^K W_{Total}^k$	$B_{Total} = \sum_{k=1}^K B_{Total}^k$
	$A_{Total} = W_{Total} + B_{Total}$	
$\bar{x}_{A^k} = \frac{1}{A_{Total}^k} \sum_{i=1}^{n_k} A_i^k x_i^k$	$\bar{x}_{W^k} = \frac{1}{W_{Total}^k} \sum_{i=1}^{n_k} W_i^k x_i^k$	$\bar{x}_{B^k} = \frac{1}{B_{Total}^k} \sum_{i=1}^{n_k} B_i^k x_i^k$
	$\bar{x}_{A^k} = \frac{W_{Total}^k}{A_{Total}^k} \bar{x}_{W^k} + \frac{B_{Total}^k}{A_{Total}^k} \bar{x}_{B^k}$	
$\bar{x}_A = \frac{1}{A_{Total}} \sum_{k=1}^K A_{Total}^k \bar{x}_{A^k}$	$\bar{x}_W = \frac{1}{W_{Total}} \sum_{k=1}^K W_{Total}^k \bar{x}_{W^k}$	$\bar{x}_B = \frac{1}{B_{Total}} \sum_{k=1}^K B_{Total}^k \bar{x}_{B^k}$
	$\bar{x}_A = \frac{W_{Total}}{A_{Total}} \bar{x}_W + \frac{B_{Total}}{A_{Total}} \bar{x}_B$	
$s_{A^k}^2 = \frac{1}{A_{Total}^k} \sum_{i=1}^{n_k} A_i^k (x_i^k - \bar{x}_{A^k})^2$	$s_{W^k}^2 = \frac{1}{W_{Total}^k} \sum_{i=1}^{n_k} W_i^k (x_i^k - \bar{x}_{W^k})^2$	$s_{B^k}^2 = \frac{1}{B_{Total}^k} \sum_{i=1}^{n_k} B_i^k (x_i^k - \bar{x}_{B^k})^2$
$s_A^2 = \frac{1}{A_{Total}} \sum_{k=1}^K \sum_{i=1}^{n_k} A_i^k (x_i^k - \bar{x}_A)^2$	$s_W^2 = \frac{1}{W_{Total}} \sum_{k=1}^K \sum_{i=1}^{n_k} W_i^k (x_i^k - \bar{x}_W)^2$	$s_B^2 = \frac{1}{B_{Total}} \sum_{k=1}^K \sum_{i=1}^{n_k} B_i^k (x_i^k - \bar{x}_B)^2$

La relación entre las varianzas globales se obtiene aplicando el mismo teorema de descomposición de la varianza s_A^2 a la partición del conjunto de adyacencias en W y B . Tal y como se procede habitualmente, se tiene, por un lado, una combinación (*pooled*) o promedio de las varianzas internas (*within*),

$$\left(\frac{W_{Total}}{A_{Total}} s_W^2 + \frac{B_{Total}}{A_{Total}} s_B^2 \right)$$

y, por otro, una varianza entre las partes (*between*),

$$\left(\frac{W_{Total}}{A_{Total}} (\bar{x}_W - \bar{x}_A)^2 + \frac{B_{Total}}{A_{Total}} (\bar{x}_B - \bar{x}_A)^2 \right).$$

Es decir,

$$s_A^2 = \left(\frac{W_{Total}}{A_{Total}} s_W^2 + \frac{B_{Total}}{A_{Total}} s_B^2 \right) + \left(\frac{W_{Total}}{A_{Total}} (\bar{x}_W - \bar{x}_A)^2 + \frac{B_{Total}}{A_{Total}} (\bar{x}_B - \bar{x}_A)^2 \right)$$

Esta fórmula se puede expresar también, evitando el valor medio total, \bar{x}_A , de la siguiente manera

$$s_A^2 = \left(\frac{W_{Total}}{A_{Total}} s_W^2 + \frac{B_{Total}}{A_{Total}} s_B^2 \right) + \frac{W_{Total}}{A_{Total}} \frac{B_{Total}}{A_{Total}} (\bar{x}_W - \bar{x}_B)^2$$

Otra fórmula interesante, que pone de manifiesto las varianzas internas a las partes, y aplicando la descomposición de la varianza s_W^2 en función de la partición de objetos, es la siguiente

$$s_A^2 = \left(\sum_{k=1}^K \frac{W_{Total}^k}{A_{Total}} s_{W^k}^2 + \frac{B_{Total}}{A_{Total}} s_B^2 \right) + \sum_{k=1}^K \frac{W_{Total}^k}{A_{Total}} (\bar{x}_{W^k} - \bar{x}_W)^2 + \frac{W_{Total}}{A_{Total}} \frac{B_{Total}}{A_{Total}} (\bar{x}_W - \bar{x}_B)^2$$

4.2.7.3 Índices modificados asociados a una partición

Según la descomposición $a_{ij} = w_{ij} + b_{ij}$ el numerador del índice modificado de Geary puede así mismo descomponerse

$$\sum_{i=1}^n \sum_{j=1}^n a_{ij} (x_i - x_j)^2 = \sum_{i=1}^n \sum_{j=1}^n w_{ij} (x_i - x_j)^2 + \sum_{i=1}^n \sum_{j=1}^n b_{ij} (x_i - x_j)^2$$

Para establecer las relaciones entre el índice modificado de Geary global y los respectivos índices de las partes, y luego analizar las diversas situaciones, es necesario reescribir las fórmulas anteriores distinguiendo las clases a las que pertenecen los objetos. Para ello, atendiendo a la pertenencia de los objetos a las partes ($a_{ij}^k = w_{ij}^k + b_{ij}^k$), se tiene

$$\sum_{k=1}^K \sum_{k'=1}^K \sum_{i=1}^{n_k} \sum_{j=1}^{n_{k'}} a_{ij}^k (x_i^k - x_j^{k'})^2 = \sum_{k=1}^K \sum_{k'=1}^K \sum_{i=1}^{n_k} \sum_{j=1}^{n_{k'}} w_{ij}^k (x_i^k - x_j^{k'})^2 + \sum_{k=1}^K \sum_{k'=1}^K \sum_{i=1}^{n_k} \sum_{j=1}^{n_{k'}} b_{ij}^k (x_i^k - x_j^{k'})^2$$

que, teniendo en cuenta que $w_{ij}^k = 0$ cuando los objetos ω_i y ω_j pertenecen a distintas partes, y $b_{ij}^k = 0$ cuando pertenecen a la misma, se convierte en

$$\sum_{k=1}^K \sum_{k'=1}^K \sum_{i=1}^{n_k} \sum_{j=1}^{n_{k'}} a_{ij}^k (x_i^k - x_j^{k'})^2 = \sum_{k=1}^K \sum_{i=1}^{n_k} \sum_{j=1}^{n_k} w_{ij}^k (x_i^k - x_j^k)^2 + \sum_{k=1}^K \sum_{i=1}^{n_k} \sum_{k' \neq k} \sum_{j=1}^{n_{k'}} b_{ij}^k (x_i^k - x_j^{k'})^2$$

Dadas estas relaciones, el índice modificado de Geary se puede expresar en función de los índices de las partes

$$C^* = \sum_{k=1}^K \frac{W_{Total}^k}{A_{Total}} \frac{s_{W^k}^2}{s_A^2} C_k^* + \frac{B_{Total}}{A_{Total}} \frac{s_B^2}{s_A^2} C_B^*$$

donde

$$C_k^* = \frac{1}{2} \frac{1}{W_{Total}^k} \sum_{i=1}^{n_k} \sum_{j=1}^{n_k} w_{ij}^k \left(\frac{x_i^k - x_j^k}{s_{W^k}} \right)^2, \text{ y}$$

$$C_B^* = \frac{1}{2} \frac{1}{B_{Total}} \sum_{k=1}^K \sum_{i=1}^{n_k} \sum_{k' \neq k} \sum_{j=1}^{n_{k'}} b_{ij}^k \left(\frac{x_i^k - x_j^{k'}}{s_B} \right)^2$$

Cuando la partición definida sobre Ω tiene la propiedad de que las partes son totalmente homogéneas dentro de sí mismas (*whitin*, $x_i^k = \bar{x}_{W^k} \equiv x_{W^k}$, $s_{W^k}^2 C_k^* = 0$), entonces

$$C^* = \frac{B_{Total}}{A_{Total}} \frac{s_B^2}{s_A^2} C_B^* = \frac{\frac{B_{Total}}{A_{Total}} s_B^2}{\frac{B_{Total}}{A_{Total}} s_B^2 + \sum_{k=1}^K \frac{W_{Total}^k}{A_{Total}} (x_{W^k} - \bar{x}_W)^2 + \frac{W_{Total}}{A_{Total}} \frac{B_{Total}}{A_{Total}} (\bar{x}_W - \bar{x}_B)^2} C_B^*$$

De donde se ve que C^* es inferior a C_B^* .

Si, además, las partes son muy heterogéneas entre sí (*between*, $\sum_{k=1}^K \frac{W_{Total}^k}{A_{Total}} (\bar{x}_{W^k} - \bar{x}_W)^2 / s_A^2$ alto), C_B^* es claramente inferior a C_B^* , lo cual denota una alta autocorrelación positiva.

Por ejemplo, en el caso de una imagen real de alta resolución las vecindades de los píxeles, por lo general, son muy homogéneas lo cual implica que C_B^* se acerque a 0. La segmentación de dicha imagen en sus clases naturales agrupa los píxeles homogéneos en clases y C_B^* no tendría por qué ser bajo.

Sin tomar en cuenta los tamaños de las partes, el índice C_B^* es el que realmente caracteriza la autocorrelación de la partición definida. En efecto, atendiendo a la pertenencia de los objetos a las partes, si las partes fueran totalmente homogéneas ($x_i^k = \bar{x}_{W^k} \equiv x_{W^k}$) C_B^* se podría expresar de la siguiente manera

$$C_B^* = \frac{1}{2} \frac{1}{B_{Total}} \sum_{k=1}^K \sum_{k'=1}^K b_{kk'} \left(\frac{x_{W^k} - x_{W^{k'}}}{s_B} \right)^2$$

donde

$$b_{kk'} = \sum_{i=1}^{n_k} \sum_{j=1}^{n_{k'}} b_{ij} \text{ y } b_{kk} = 0$$

y al ser $s_{B^k}^2 = 0$ y $x_i^k = \bar{x}_{B^k} \equiv x_{W^k}$

$$s_B^2 = \frac{1}{B_{Total}} \sum_{k=1}^K B_{Total}^k s_{B^k}^2 + \frac{1}{B_{Total}} \sum_{k=1}^K B_{Total}^k (\bar{x}_{B^k} - \bar{x}_B)^2 = \frac{1}{B_{Total}} \sum_{k=1}^K B_{Total}^k (x_{W^k} - \bar{x}_B)^2$$

Es decir, C_B^* sería el índice de un grafo con menos nodos (más sencillo, reducido) en el que los nodos originales de una misma parte se fundirían en un único nodo, y las adyacencias entre los nuevos nodos serían la suma de las adyacencias iniciales entre los objetos que pertenecen a una y otra parte.

En la Figura 4-16 se muestra la partición de un retículo en 4 partes totalmente homogéneas, y su correspondiente grafo reducido en el caso de 4-vecindades.

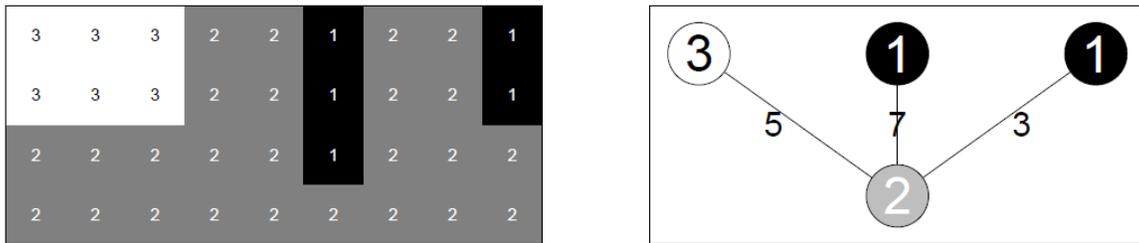


Figura 4-16 Ejemplo de grafo reducido de una imagen

Por su parte, el índice modificado de Lebart,

$$L^* = \frac{1}{A_{Total}} \frac{\sum_{i=1}^n A_i (x_i - y_i)^2}{s_A^2},$$

presenta dificultades a la hora de relacionar L^* con los L_k^* y L_B^* , pues en las vecindades de los elementos fronterizos hay que distinguir los objetos interiores a una parte y los fronterizos para realizar los cálculos correspondientes. En general, el numerador de L^* puede desglosarse en dos partes

$$\sum_{k=1}^K \sum_{i=1}^{n_k} A_i^k (x_i^k - y_i^k)^2 = \sum_{k=1}^K \sum_{i=1}^{n_k} W_i^k (x_i^k - y_i^k)^2 + \sum_{k=1}^K \sum_{i=1}^{n_k} B_i^k (x_i^k - y_i^k)^2$$

Los objetos interiores de las partes no presentan problemas porque es $B_i^k = 0$, pero en los fronterizos y_i^k , la media de la respectiva vecindad debe descomponerse en dos, la correspondiente a su parte y la de la frontera, y luego tratar se separarlos en el desarrollo del binomio cuadrado para llegar a los L_k^* y L_B^* , lo cual es muy complicado.

En lo adelante la aplicación de los índices de autocorrelación del apartado anterior se limita al índice modificado de Geary. No se descarta el uso de los índices locales en otras problemáticas.

4.2.7.4 Índice modificado de Geary asociado a una partición en un espacio estructurado.

A continuación se va a proceder a una disquisición sobre los posibles valores de C^* y de C_B^* asociado a determinada partición, en relación a distintas estructuras reflejadas por la matriz de adyacencias, A .

Si la matriz de adyacencias tiene bloques (más de uno), lo que refleja que en la estructura espacial de Ω existen varios componentes conexos, el hecho de que determinada partición considere en una misma parte a partes de distintas componentes, no tiene mucho interés desde el punto de vista operativo. Es más procedente considerar la nueva partición que se obtiene de las intersecciones no vacías por pares de las partes de la partición y las componentes conexas. Este procedimiento llevaría a considerar una estructura de particiones de componentes conexos.

Las disquisiciones siguientes se hacen distinguiendo primero, por conveniencia, diversos casos en que $K=1$, es decir, aquellos en los que todos los objetos de Ω son conexos entre sí, y luego diversos casos en los que hay K ($K>1$) componentes conexos.

4.2.7.4.1 Caso a)

Ω es un conjunto sin estructura espacial. La matriz de adyacencias, A , es $a_{ij}=1$, $i, j=1, \dots, n$ (incluidos los $a_{ii}=1$): $A=U$. Se tiene, según el apartado anterior, que $C^*=1$.

Si la partición definida sobre Ω es la menos fina ($K=1$, parte única, Ω), sus características principales son

$$\begin{array}{lllll}
 K=1 & n_1=n & a_{ij}^1 = a_{ij} & w_{ij}^1 = w_{ij} & b_{ij}^1 = b_{ij} \\
 & & a_{ij} = w_{ij} + b_{ij} & w_{ij} = 1 & b_{ij} = 0 \\
 & & A_i^1 = n & W_i^1 = n & B_i^1 = 0 \\
 & & A_{Total}^1 = n^2 & W_{Total}^1 = n^2 & B_{Total}^1 = 0
 \end{array}$$

Por tanto, $C^* = C_1^* \equiv 1$ y C_B^* no tiene sentido.

Si se define, en cambio, la partición menos fina (cada objeto ω_i constituye una parte, $\{\omega_i\}$, de Ω), se tiene que

$$\begin{array}{lll}
K = n & n_k = 1 & a_{ij}^k = a_{ij} & w_{ij}^k = w_{ij} & b_{ij}^k = b_{ij} \\
& & a_{ij} = w_{ij} + b_{ij} & w_{ii} = 1, w_{ij} = 0 \quad j \neq i & b_{ii} = 0, b_{ij} = 1 \quad j \neq i \\
& & A_i = W_i + B_i & W_i = 1 & B_i = n - 1 \\
& & A_{Total} = W_{Total} + B_{Total} & W_{Total} = n & B_{Total} = n(n - 1)
\end{array}$$

Se tiene además, que respecto al interior de las partes (*within*)

$$\begin{aligned}
x_i^k &= x_i & \bar{x}_{W^k} &= x_i \\
\bar{x}_W &= \frac{1}{n} \sum_{i=1}^n x_i = \bar{x} \\
s_{W^k}^2 &= 0
\end{aligned}$$

$$s_W^2 = \frac{1}{W_{Total}} \sum_{k=1}^K \sum_{i=1}^{n_k} W_i^k (x_i^k - \bar{x}_{W^k})^2 = \frac{1}{W_{Total}} \sum_{k=1}^K W_i (x_i - \bar{x}_W)^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = s^2$$

y respecto a las fronteras entre las partes (*between*)

$$\begin{aligned}
\bar{x}_{B^k} &= \frac{1}{B_{Total}^k} \sum_{i=1}^{n_k} B_i^k x_i^k = \frac{1}{B_i} B_i x_i = x_i \\
\bar{x}_B &= \frac{1}{B_{Total}} \sum_{k=1}^K B_{Total}^k \bar{x}_{B^k} = \frac{1}{B_{Total}} \sum_{i=1}^n B_i x_i = \frac{1}{n(n-1)} \sum_{i=1}^n (n-1)x_i = \bar{x} \\
s_{B^k}^2 &= \frac{1}{B_{Total}^k} \sum_{i=1}^{n_k} B_i^k (x_i^k - \bar{x}_{B^k})^2 = (x_i - x_i)^2 = 0
\end{aligned}$$

$$s_B^2 = \frac{1}{B_{Total}} \sum_{k=1}^K \sum_{i=1}^{n_k} B_i^k (x_i^k - \bar{x}_B)^2 = \frac{1}{B_{Total}} \sum_{i=1}^n B_i (x_i - \bar{x}_B)^2 = \frac{1}{n(n-1)} \sum_{i=1}^n (n-1)(x_i - \bar{x})^2 = s^2$$

que se corroboran en las fórmulas de descomposición de las varianzas

$$\begin{aligned}
s_W^2 &= \frac{1}{W_{Total}} \sum_{k=1}^K W_{Total}^k s_{W^k}^2 + \frac{1}{W_{Total}} \sum_{k=1}^K W_{Total}^k (\bar{x}_{W^k} - \bar{x}_W)^2 = 0 + \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = s^2 \\
s_B^2 &= \frac{1}{B_{Total}} \sum_{k=1}^K B_{Total}^k s_{B^k}^2 + \frac{1}{B_{Total}} \sum_{k=1}^K B_{Total}^k (\bar{x}_{B^k} - \bar{x}_B)^2 = 0 + \frac{1}{n(n-1)} \sum_{i=1}^n (n-1)(x_i - \bar{x})^2 = s^2
\end{aligned}$$

Por tanto,

$$\begin{aligned}
s_A^2 &= \left(\sum_{k=1}^K \frac{W_{Total}^k}{A_{Total}} s_{W^k}^2 + \frac{B_{Total}}{A_{Total}} s_B^2 \right) + \sum_{k=1}^K \frac{W_{Total}^k}{A_{Total}} (\bar{x}_{W^k} - \bar{x}_W)^2 + \frac{W_{Total}}{A_{Total}} \frac{B_{Total}}{A_{Total}} (\bar{x}_W - \bar{x}_B)^2 = \\
&= \left(\sum_{i=1}^n \frac{1}{n^2} 0 + \frac{n(n-1)}{n^2} s^2 \right) + \sum_{i=1}^n \frac{1}{n^2} (x_i - \bar{x})^2 + \frac{n}{n^2} \frac{n(n-1)}{n^2} (\bar{x} - \bar{x})^2 = s^2
\end{aligned}$$

Así el valor del índice modificado de Geary asociado a la partición es

$$\begin{aligned}
C_B^* &= \frac{1}{2} \frac{1}{B_{Total}} \sum_{k=1}^K \sum_{i=1}^{n_k} \sum_{\substack{k' \neq k \\ j=1}}^{n_{k'}} b_{ij}^k \left(\frac{x_i^k - x_j^{k'}}{s_B} \right)^2 = \frac{1}{s^2} \frac{1}{2} \frac{1}{n(n-1)} \sum_{\substack{i=1 \\ i \neq j}}^n \sum_{j=1}^n (x_i - x_j)^2 = \\
&= \frac{1}{s^2} \frac{1}{2} \frac{n^2}{n(n-1)} \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n (x_i - x_j)^2 = \frac{1}{s^2} \frac{1}{2} \frac{n^2}{n(n-1)} 2s^2 = \frac{n}{n-1}
\end{aligned}$$

que corrobora la igualdad

$$C^* = \sum_{k=1}^K \frac{W_{Total}^k}{A_{Total}} \frac{s_{W^k}^2}{s_A^2} C_k^* + \frac{B_{Total}}{A_{Total}} \frac{s_B^2}{s_A^2} C_B^*$$

en el que $s_{W^k}^2 C_k^* = 0$. En efecto,

$$C^* = 0 + \frac{n(n-1)}{n^2} \frac{s^2}{s^2} C_B^* = \frac{n-1}{n} \frac{n}{n-1} = 1$$

El valor del índice C_B^* asociado a la partición coincide con el valor de C^* en el caso de que la matriz de adyacencias, A , representa el grafo completo de Ω ($a_{ij} = 1$ si $i \neq j$ y $a_{ii} = 0$) que es la estructura espacial habitual del análisis de datos en la que los objetos no están relacionados entre sí.

En el caso de una partición cualquiera K , $1 < K \leq n$, la matriz A se descompone como la suma de dos matrices (ver apartado 2.2.1): $A = W + B$, donde W es una matriz diagonal de bloques (ver apartado 4.2.6.4), y representa la relación binaria de copertenencia de objetos. Se tiene en este caso que $C^* = 1$ y que $C_k^* = 1$, por lo que

$$1 = \sum_{k=1}^K \frac{W_{Total}^k}{A_{Total}} \frac{s_{W^k}^2}{s_A^2} 1 + \frac{B_{Total}}{A_{Total}} \frac{s_B^2}{s_A^2} C_B^*$$

de donde se obtiene

$$C_B^* = \frac{A_{Total} s_A^2 - \sum_{k=1}^K W_{Total}^k s_{W^k}^2}{B_{Total} s_B^2}$$

Se observa que en el valor del índice C_B^* toman parte las varianzas internas, fronterizas y totales, y se corrobora que $C_B^* = n/(n-1)$ en el caso $K = n$.

Por otra parte, si las clases fueran totalmente homogéneas ($s_{W^k}^2 = 0$), $C_B^* = A_{Total} s_A^2 / B_{Total} s_B^2$, en el que el numerador es constante.

4.2.7.4.2 Caso b)

Ω está completamente estructurado. Ω no tiene más adyacencias que las que asocian los objetos consigo mismo

$$a_{ij} = \begin{cases} 1, & \text{si } i = j \\ 0, & \text{si } i \neq j \end{cases}$$

La intersección de cualquier partición sobre Ω es la partición más fina de partes singulares, que no están conectadas. Se sabe que $C^* = 0$, mientras que C_B^* no tiene sentido.

4.2.7.4.3 Caso c)

La estructura espacial de Ω está representada por un grafo completo cuya matriz de adyacencias, A , es tal que ($a_{ij} = 1$ si $i \neq j$ y $a_{ii} = 0$).

Se sabe, por el apartado anterior, que $C^* = n/n-1$, por lo que Ω apenas tiene estructura espacial (tanto menos en cuanto n se hace más grande). Desde el punto de vista de la autocorrelación espacial, los problemas clásicos del análisis de datos tienen esta estructura (ver apartado 4.6.2.3).

En el caso de que la partición de Ω sea la menos fina ($K=1$), todos los objetos son interiores a la parte única, y, por tanto, C_B^* no tiene sentido. Y en el caso de la partición más fina ($K=n$), todos los objetos y todas las relaciones de adyacencia son fronterizos, por lo que $C_B^* = n/(n-1)$, es decir, C^* mismo.

Esta coincidencia es interesante, pues indica de alguna manera que las relaciones de los objetos consigo mismo no aportan ninguna información relevante. En el caso *a*) $C^* < C_B^*$, lo que indica que hay más autocorrelación positiva que la que realmente hay. Otra cuestión distinta es que la consideración de la relación reflexiva resulte más cómoda para el desarrollo de las fórmulas, pues en ese caso la fórmula resulta análoga al caso de independencia entre los objetos.

4.3 Conclusiones

En el presente capítulo se ha realizado un estudio basado en las nociones de autocorrelación espacial de Moran y Geary. A partir de considerar como unidades estadísticas las relaciones espaciales entre los objetos, en lugar de considerar los objetos en sí, se han modificado las definiciones de media y varianza.

Esto ha permitido establecer una relación de complementariedad entre los índices modificados de Moran y Geary. Se ha propuesto un índice de autocorrelación espacial a partir del concepto de variabilidad local de Lebart.

Por otro parte, las fórmulas de descomposición de la varianza han posibilitado la definición de nuevos índices de autocorrelación espacial cuyas propiedades han sido analizadas, así como sus relaciones con los mencionados anteriormente.

La metodología LISA (*Local Indicators of Spatial Spatial Association*) ha permitido asociar índices locales a cada uno de los índices globales de autocorrelación espacial. Ellos posibilitan profundizar en el conocimiento de la estructura espacial de los datos.

La aplicación del índice de Geary al estudio de las particiones de un conjunto de objetos con relaciones espaciales ha determinado la definición de un índice de autocorrelación espacial asociado a particiones, C_B^* , que toma en cuenta únicamente las fronteras entre las clases de una partición. En posteriores capítulos este índice denominado simplemente C_B .

Se han analizado casos particulares de estructura espacial y el comportamiento en estos casos de los índices mencionados.

El caso habitual de datos no estructurados ha resultado ser un caso particular en este contexto.

PARTE III

5 Clustering con restricciones

5.1 *Introducción*

Los algoritmos de clustering que han sido analizados anteriormente tratan los objetos de manera independiente ignorando cualquier tipo de relación que pudiera existir entre ellos. El criterio principal que se plantea es el de la compacidad de los clusters. Sin embargo, en el caso de la segmentación de imágenes por lo general los píxeles vecinos poseen características similares y es de esperar, por tanto, que se encuentren en el mismo cluster. Esta idea está en correspondencia con la primera ley de la geografía de Tobler que plantea que todo está relacionado, pero las cosas cercanas están más relacionadas que las cosas lejanas [140]. En el caso de las imágenes médicas esto se traduce en el hecho de que generalmente un píxel dado pertenece a la misma clase de tejido que sus vecinos; o lo que es lo mismo, es muy poco probable que una estructura anatómica consista en un único píxel de la imagen. La información espacial asociada a los píxeles vecinos puede ser muy útil para la segmentación.

Los algoritmos de clustering con restricciones imponen ciertas condiciones adicionales al conjunto de posibles soluciones. Dentro de los tipos de restricciones más comunes se encuentra la contigüidad, la cual requiere que los miembros de un cluster sean no sólo semejantes entre sí, sino que haya conexión desde el punto de vista espacial entre sus miembros.

En este capítulo se examinan las principales alternativas para la incorporación de la información espacial en los algoritmos de clustering. Una vez introducida la notación necesaria se presenta una nueva definición de desemejanza entre clusters que toma en cuenta la estructura de los datos.

Se presenta además un estudio del algoritmo de agrupamiento jerárquico aglomerativo con restricciones de contigüidad utilizando la nueva desemejanza desde el punto de vista de la ocurrencia de inversiones. Se incluyen varias estrategias de implementación unido a un análisis de la complejidad computacional.

5.2 *El agrupamiento con restricciones*

Entre los tipos más comunes de restricciones están aquellas relacionadas con la cardinalidad y la composición de los clusters, con la topología de la jerarquía en el caso del agrupamiento jerárquico, y con la contigüidad o conectividad de los objetos que forman un cluster [141, 142].

Este tipo de problemas es relativamente antiguo, particularmente el conocido como problema de la regionalización, es decir, la formación de clusters de zonas geográficas semejantes según ciertas características, socioeconómicas o de otro tipo, pero con la condición de que sean además geográficamente conexas o contiguas [143-146].

Un enfoque para resolver este problema consiste en realizar el agrupamiento tomando como base únicamente las características o rasgos de los objetos sin tener en cuenta la información espacial y confrontar esta información al final del proceso. La otra posibilidad es utilizar la información espacial durante el proceso mismo de agrupamiento.

Estas ideas son utilizadas en el método CONAN (*CONtextual ANalysis*) desarrollado como parte de un estudio para clasificar patrones de uso de suelos a partir de imágenes satelitales [147]. Este método presupone un proceso de clasificación en dos etapas. En la primera etapa los datos de las imágenes se clasifican en base a la información espectral y se obtienen los componentes de la imagen (por ejemplo, césped, pavimento, techo). Ahora bien, si el objetivo es identificar zonas comerciales y residenciales esto no bastaría. Precisamente este es el argumento para la utilización de la información espacial en este caso. Como en el caso de las zonas comerciales la distribución de los componentes tiene sus características propias (grandes edificios y zonas de parqueo, es decir, la mayoría de la zona sería pavimento y techo, y poco césped o área verde), a diferencia de las zonas residenciales, con una mayor presencia de áreas verdes, esto conduce a analizar la vecindad de cada punto y clasificarlo según el primer criterio para ver las proporciones de los otros componentes en sus alrededores. Esta es en esencia la segunda etapa donde se utiliza una medida contextual para cada píxel obtenida a partir de la mezcla de componentes que se encuentran en la vecindad la cual se define mediante una ventana alrededor de cada píxel. Precisamente, el tamaño de la ventana constituye un parámetro de este método el cual puede influir de manera significativa en los resultados que se obtienen.

Una vía más directa de utilizar la información espacial consiste en considerar las coordenadas espaciales como variables activas en el proceso de agrupamiento [148, 149].

En la Figura 5-1 se muestra un ejemplo de esta situación. Los valores mostrados se obtuvieron de la siguiente manera: los primeros 60 valores fueron derivados de una distribución normal con media igual a 15 y una varianza de 100, mientras los otros 40 puntos se generaron a partir de una distribución normal con media igual a 25 y varianza igual 4. En la Figura 5-1 (a) se han representado los datos utilizando el valor generado como abscisa y el número de orden como ordenada. La línea azul indica la división real de los puntos, es decir, por debajo de ella están los 60 primeros y por encima los otros 40 puntos. La proyección de los datos en el eje x evidencia un alto grado de solapamiento.

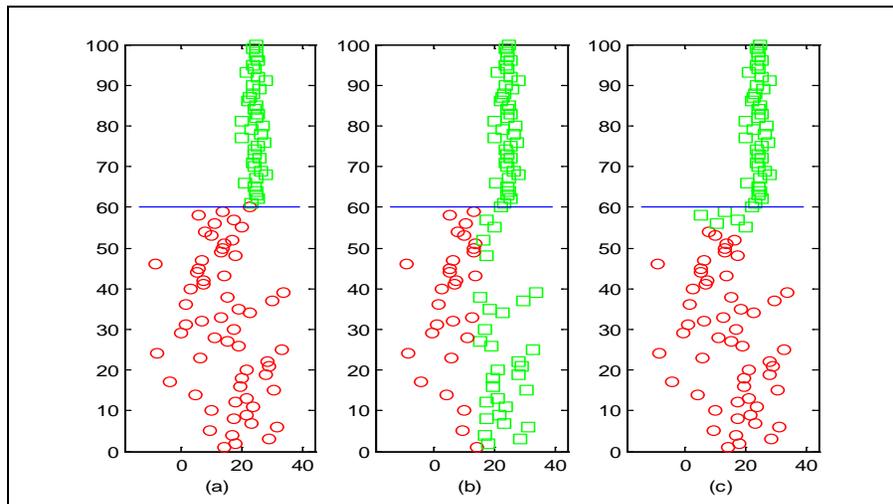


Figura 5-1 Ejemplo de inclusión de la información espacial. Los datos ordenados (a), partición sin considerar el orden (b), partición usando el número de orden (c).

En la Figura 5-1 (b) se muestra el resultado de utilizar un agrupamiento con el algoritmo *k-means* utilizando solamente el valor generado asociado a cada punto. Como se puede apreciar el cluster de la derecha (rectángulos verdes) contiene aquellos valores que se agrupan aproximadamente alrededor del valor 25 (la media de los valores generados). En la Figura 5-1 (c) se aprecia el resultado cuando se incluye el número de orden de generación como variable activa. El sentido sería expresar el hecho de que los valores cercanos en el orden de generación deben estar en el mismo cluster. Los resultados en este caso se corresponden más con el esquema de generación utilizado. Se debe destacar que este ejemplo de serie temporal, así como el uso de la información espacial (coordenadas espaciales) son casos particulares de información con restricciones de conectividad.

La inclusión de las coordenadas espaciales unidas a la información sobre el resto de las características o rasgos de los objetos plantea la problemática de la influencia de unas y otras sobre todo al utilizar alguna función de semejanza. Una forma de controlar esta influencia consiste en la asignación de pesos a cada componente [150]. Si la ponderación del componente asociado a la cercanía espacial es alta entonces los clusters resultantes estarán formados por objetos cercanos espacialmente entre sí. La principal objeción a estos métodos consiste en la especificación subjetiva de los factores de ponderación.

Otra idea de utilización de la información espacial está asociada al hecho de incluir en la representación de los datos el contexto en que se encuentra cada objeto. Usualmente el contexto se define tomando como base cierta relación de vecindad. En el caso de la segmentación de imágenes cada píxel se encuentra en una posición del retículo regular mediante el cual se puede describir la imagen [151].

Las relaciones de vecindad se pueden entonces establecer de manera más sencilla, por ejemplo, utilizando la 4-vecindad o la 8-vecindad. En ocasiones la noción de vecindad o contigüidad se define de manera más general que la simple adyacencia, por ejemplo,

considerando como contiguos aquellos objetos que no se encuentran más lejos que un umbral definido a priori [152].

Una modificación dentro del agrupamiento jerárquico aglomerativo del enlace simple con la introducción de restricciones de contigüidad lo propone Monestiez: dos objetos se unen si son contiguos de manera que los clusters que se forman son conexos [146].

Un estudio de los métodos clásicos de agrupamiento jerárquico bajo restricciones de contigüidad se presenta en [153]. Dos clusters se consideran conexos si existe al menos un par de objetos, uno de cada cluster, que sean conexos desde el punto de vista de la relación inicial de adyacencia. Se estudian dos maneras de propagar las restricciones a nivel de clusters y se analizan los casos de los métodos de enlace simple y de enlace completo bajo dos tipos de restricciones de contigüidad. Al analizar el método de enlace simple se asume un tipo de restricción fuerte: se exige que los clusters sean conexos y que el valor de desemejanza mínima entre ellos se alcance para un par de objetos conexos. Bajo estas condiciones el método de enlace simple es libre de inversiones, es decir, la sucesión de valores de desemejanza entre los clusters mezclados es monótona creciente. Esta es precisamente una de las variantes que plantea Monestiez en [146].

Al suavizar la restricción, es decir, al exigir únicamente que los clusters sean conexos se demuestra que, de los métodos clásicos, solamente el método de enlace completo es el único que es libre de inversiones. Este método ha sido aplicado recientemente al análisis de guiones cinematográficos y televisivos [154].

Además de los métodos clásicos, otros enfoques han sido ensayados para el agrupamiento bajo restricciones de contigüidad. En [155] se analiza el criterio de la separación (*split*). La separación del borde (*border split*) de un cluster se define como la menor desemejanza de un miembro del cluster y de otro elemento cualquiera externo perteneciente a una región contigua. Por su parte la separación global de los bordes de una partición se define como la menor de todas las separaciones de sus clusters.

Si no se impone la restricción de contigüidad, entonces el criterio se conoce simplemente como separación (de un cluster o de una partición). Según este punto de vista, el método de enlace simple (sin restricciones) maximiza la separación de todas las particiones a las que conduce. Por su parte, el algoritmo de Monestiez con restricciones contigüidad [146], mencionado anteriormente, maximiza la separación de todas las particiones que se van obteniendo en la jerarquía.

La utilización del criterio de separación de los bordes puede resultar de interés en diversas aplicaciones tales como la detección de bordes en imágenes, o la definición de contornos en aplicaciones de cartografía.

En general se debe tener en cuenta la diferencia entre el criterio de la separación y el de la separación del borde. En el primer caso no se impone la restricción de contigüidad, mientras en el segundo, sí. Esta diferencia pudiera conducir a una situación en la que dos

clusters posean miembros muy diferentes a través de sus fronteras y, sin embargo, los miembros dentro de cada uno de ellos sean similares.

En [155] se analiza un híbrido, es decir, se toma como base el criterio de la máxima separación, pero se analizan solamente aquellas particiones cuyos miembros se corresponden con componentes conexos del grafo de contigüidad inicial. Eso significa que cada cluster debe ser conexo según la relación de contigüidad impuesta a los datos. Sin embargo, a la hora de analizar la separación de un cluster (y de la partición) no se tiene en cuenta la restricción de contigüidad.

Más adelante en este trabajo se analiza esta problemática, partiendo de una definición diferente de la separación del borde.

Otra manera intuitiva de incluir el contexto consiste en expandir la representación de los datos de cada objeto con la información de sus vecinos. Este enfoque, aunque es bastante general, presenta algunos inconvenientes como es, por ejemplo, el alto costo de memoria y, asociado a esto, la necesidad de limitar el tamaño de la vecindad de cada objeto.

También se ha ensayado otra posibilidad de incluir la información de vecindad. En este sentido, en [156] se introduce una modificación de la función objetivo que se define mediante una combinación lineal de la compacidad (asociada a la varianza) y la contigüidad de los clusters. La contigüidad de un píxel se define allí como la fracción de sus vecinos que se encuentran en la misma clase. El peso de cada componente se define a través de un parámetro. De manera similar se procede en [157] donde se introduce un método que incorpora restricciones de contigüidad. Este enfoque penaliza cada posible solución en dependencia de su contigüidad. Aquí también se utiliza un parámetro que controla la importancia de cada componente de la función objetivo.

La definición de relaciones de vecindad es equivalente a la construcción de un grafo, donde cada objeto representa un vértice o nodo y la relación de vecindad se expresa mediante una arista entre los objetos relacionados [158]. El agrupamiento con restricciones espaciales se reduce en este caso a: (1) la construcción del grafo de vecindades, y (2) la ejecución de un algoritmo que respete las restricciones impuestas [157].

La construcción del grafo de vecindades no es una cuestión tan trivial. Por ejemplo, en el caso de las imágenes cabría hacer la pregunta sobre si debe considerarse que cada píxel posee 4 vecinos, 8 vecinos, u otra cantidad. En el caso de patrones de puntos en el plano es posible utilizar algunas estructuras disponibles dentro del campo de la geometría computacional. Así por el ejemplo, el diagrama de Voronoi asocia a cada punto aquella región que se encuentra más cercana a él que a cualquier otro punto; su dual, la triangulación de Delaunay une con un arco aquellos puntos cuyas regiones poseen una frontera común [159, 160]. Por su parte según grafo de Gabriel se unen con un arco dos puntos sí y sólo sí el resto de los puntos se encuentran fuera del círculo cuyo diámetro coincide con la longitud de dicho arco [161]. El grafo de vecindades relativas se construye especificando un arco entre dos puntos sí y sólo sí la distancia entre ellos no es mayor que

la distancia de cualquiera de ellos a cualquier otro punto [162]. Por último, es posible utilizar los árboles de expansión mínima que representan un árbol de expansión en el cual la suma de las longitudes o pesos de los arcos es mínima. Estas estructuras brindan una amplia gama de posibilidades para la construcción grafos de contigüidad [141].

Desde el punto de vista formal las restricciones pueden definirse de diversas maneras, tales como, relaciones simétricas y no simétricas [158, 163]. Estas restricciones pueden ser impuestas entre objetos o instancias debido, por ejemplo, a la disponibilidad de algún conocimiento previo mediante el cual se conoce de antemano la clase a la que pertenecen algunos objetos [164, 165]. La contigüidad geográfica es un ejemplo de relación simétrica. También son posibles restricciones a nivel de las clases.

Entre los tipos de restricciones más comúnmente utilizados se encuentran las denominadas restricciones *must-link*, las cuales involucran un par de objetos y exigen que en un agrupamiento factible ambos objetos deban estar en el mismo cluster. De manera similar se utilizan las restricciones *cannot-link* con el sentido opuesto, es decir, el par de objetos involucrados no deben estar en el mismo cluster. También es posible exigir restricciones a nivel de clusters. Por ejemplo, la restricción de mínima separación exige que la distancia entre dos objetos cualquiera que se encuentren en clusters distintos sea como mínimo igual a un valor dado lo cual influye en la variación inter-cluster. Por otra parte, la restricción de compacidad plantea que, para un cluster cualquiera que contenga más de dos elementos, se debe cumplir que cualquier objeto perteneciente a dicho cluster debe tener otro objeto del mismo cluster a una distancia que no supere el valor dado que condiciona, por su parte, la variación intra-cluster [165].

La utilización de este tipo de restricciones brinda la posibilidad de incluir conocimiento adicional del dominio. Por ejemplo, en el caso de la segmentación de imágenes médicas el conocimiento acerca de la presencia de determinados tipos de tejidos puede ser incluido y tenido en cuenta por los algoritmos que se utilicen.

La introducción de restricciones cuenta entre sus beneficios con la posibilidad de encontrar soluciones con ciertas propiedades deseadas que las funciones de desemejanzas no pueden encontrar por sí solas, y en ocasiones es posible mejorar los tiempos de ejecución de los algoritmos. Sin embargo, estos beneficios vienen con un costo adicional y es la posibilidad de no encontrar soluciones factibles que cumplan con todas las restricciones [165]. Por ejemplo, supongamos que se tienen 3 objetos a , b y c , se desean formar dos clusters, y se tienen las siguientes restricciones: *cannot-link*(a , b), *cannot-link*(b , c) y *cannot-link*(a , c) (ver Figura 5-2).

Evidentemente no existe un agrupamiento factible de dos clusters que respete todas las restricciones. Lo mismo ocurriría si estas restricciones fueran de tipo *must-link*, sólo que en este caso se podría formar un cluster, aunque estaríamos en presencia de un callejón sin salida (*dead-end*) o de un agrupamiento irreducible [166].

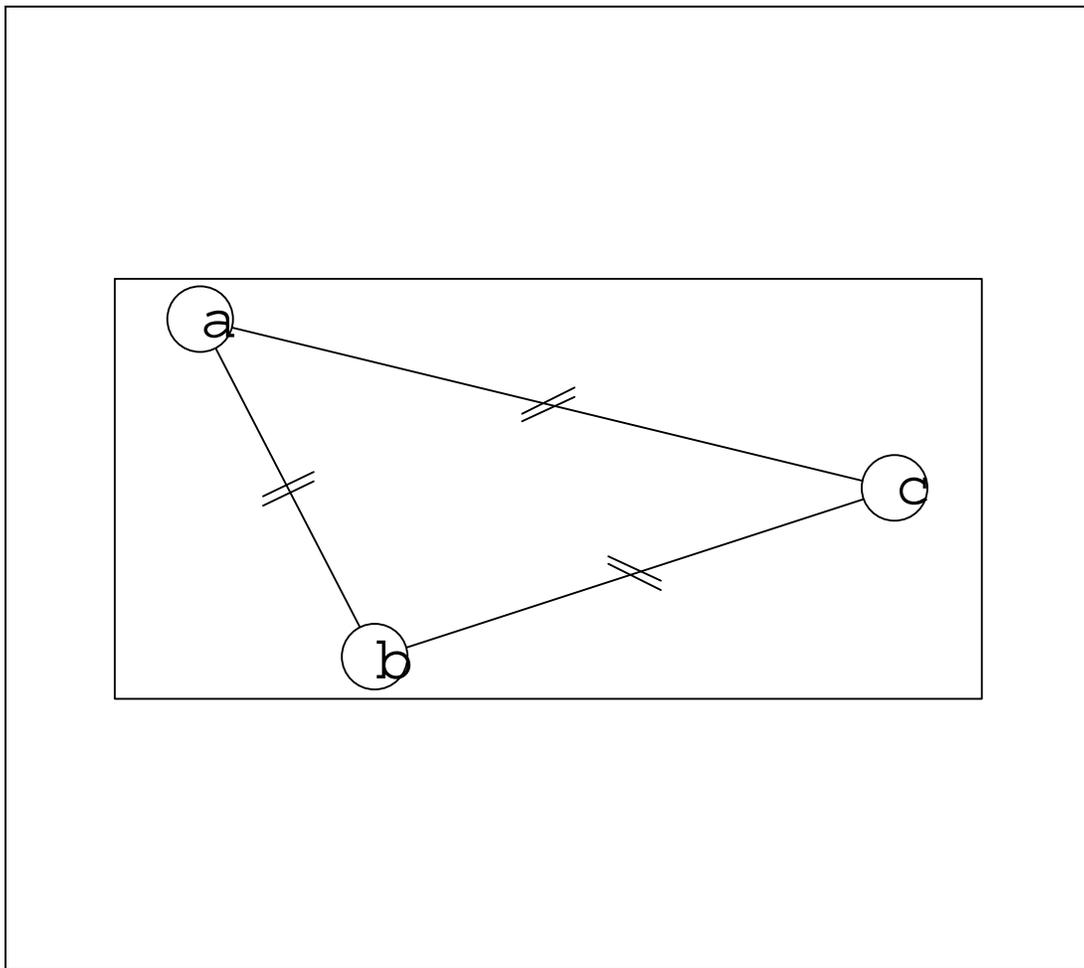


Figura 5-2 Ejemplo con restricciones *cannot-link*

El clustering con restricciones a pares, o con la inclusión de conocimiento del dominio en forma de relaciones entre objetos o instancias se denomina también agrupamiento semisupervisado y es un tema actual de investigación [164, 167, 168].

Para más obtener más detalles sobre los métodos de agrupamiento con la utilización de restricciones se recomienda consultar los artículos [141, 152, 158, 169-171].

Las interacciones espaciales pueden ser interpretadas a luz del concepto de autocorrelación, entendida ésta como la correlación entre los valores de la misma variable. La autocorrelación es una característica común a muchos conjuntos de datos. Por ejemplo, los valores de intensidad de los píxeles de una imagen digital están relacionados, en general, en función de su proximidad en la imagen misma.

En el capítulo anterior del presente trabajo se presenta un análisis de las interacciones espaciales desde el punto de vista de la autocorrelación espacial. Allí se introducen modificaciones de índices clásicos de autocorrelación espacial con el fin de equipararlos.

También se presentan sus propiedades fundamentales, se proponen otros índices, y se analizan algunos casos particulares de datos estructurados espacialmente.

5.3 Planteamiento del problema

Los algoritmos de agrupamiento con restricciones que se han comentado antes poseen una gran variedad de enfoques. En lo que sigue el análisis se centra en los casos en que se consideran restricciones de contigüidad sobre los datos y se propone un método que integra y preserva la estructura del conjunto de datos.

Consideremos un conjunto de objetos $\Omega = \{\omega_1, \omega_2, \dots, \omega_n\}$ a los cuales se les asocian ciertas magnitudes x_i , $i = 1, \dots, n$. Un objeto será identificado de manera indistinta como ω_i , o simplemente mediante el subíndice i .

Sea $A = [a_{ij}]$ una matriz simétrica de contigüidad, donde $a_{ij} = 1$ si y sólo si los objetos i y j son contiguos o adyacentes, y $a_{ij} = 0$ en caso contrario. Se considera que $\forall i, a_{ii} = 0$. Esta matriz puede ser vista como la matriz de adyacencias de un grafo no dirigido de contigüidad $G = \langle \Omega, E \rangle$ en el cual cada vértice representa un objeto, y una arista entre dos vértices indica que estos son contiguos. Luego, el conjunto de vértices coincide con el conjunto Ω . Por su parte el conjunto de aristas E está compuesto por los pares $\{\omega_i, \omega_j\}$ tales que $a_{ij} = 1$.

A continuación se incluyen algunos conceptos sobre grafos no dirigidos que serán utilizados más adelante.

Si $S \subseteq \Omega$, denotaremos mediante $G(S)$ el subgrafo de G inducido por S , es decir, el grafo cuyo conjunto de vértices es S , y cuyo conjunto de aristas es $\{\{\omega_i, \omega_j\} \in E / \omega_i, \omega_j \in S\}$.

Un camino en G es una secuencia de vértices $\{\omega_{i_1}, \omega_{i_2}, \dots, \omega_{i_p}\}$ tales que $\{\omega_{i_l}, \omega_{i_{l+1}}\} \in E$ para $l = 1, 2, \dots, p-1$, donde ω_{i_1} es el vértice inicial, y ω_{i_p} es el vértice final, del camino que los une. Si $\omega_{i_1} = \omega_{i_p}$ el camino se convierte en un ciclo. Un camino se considera elemental o simple si todos sus vértices son distintos.

Un grafo G es completo si todos los pares de vértices son adyacentes.

Un grafo G es conexo si existe un camino entre cualquier par de vértices.

En lo que sigue se considera que el grafo de contigüidad G es conexo.

Un componente conexo G' de G es un subgrafo maximal conexo. Sea $G' = \{V', E'\}$ el subgrafo asociado al componente conexo. El hecho de que G' sea un subgrafo maximal conexo significa que no existe otro subgrafo $G'' = (V'', E'')$ de G , que también sea conexo, y para el que se cumpla que $V' \subset V''$ ó $E' \subset E''$.

En el caso de las imágenes digitales una relación de contigüidad o adyacencia razonable puede ser especificada mediante los vecinos más próximos en el plano de la imagen, por ejemplo, los 4 u 8 píxeles vecinos más próximos. Generalizando esta idea también es posible definir como vecinos aquellos píxeles que se encuentren a una distancia (euclidiana), según el arreglo de los píxeles en la imagen, no mayor de un valor r dado respecto al píxel que se analiza. Así la 4 – vecindad se correspondería con $r = 1$; la 8 – vecindad con $r = \sqrt{2}$ y así sucesivamente para $r \geq 2$.

En cualquier caso la contigüidad se interpreta aquí como adyacencia espacial según la relación que se utilice y es una manera de representar la estructura de los datos.

Supongamos que se han clasificado los datos del conjunto Ω en K grupos, clusters, o clases disjuntas, es decir, se ha obtenido una partición $P = \{P_1, P_2, \dots, P_K\}$. Una clase P_k puede ser singular, es decir, estar formada por un único objeto.

Se dice que la partición P es conexa si cada uno de sus grupos o clases es conexo, es decir, si el subgrafo inducido $G(P_k)$ es conexo, $\forall P_k \in P$.

Denotemos mediante $\Phi(\Omega)$ el conjunto de particiones conexas de Ω según el grafo de contigüidad G . El conjunto particiones conexas de Ω de cardinalidad K se denotará por $\Phi_K(\Omega)$ (ver apartado 4.2.7.4).

Siguiendo la idea de [172] se puede considerar que, en el agrupamiento con restricciones, las particiones analizadas deben pertenecer a un conjunto de soluciones factibles, $\Phi(\Omega)$ en nuestro caso. El conjunto $\Phi(\Omega)$ representa el conjunto de aquellas particiones que satisfacen los requerimientos impuestos de antemano.

Si los requerimientos están asociados a la información de los vecinos se estaría en presencia de lo que se conoce como agrupamiento con restricciones de contigüidad (*contiguüity-constrained clustering*) [141, 153, 169], o también como agrupamiento condicional (*condicional clustering*) [173].

En nuestro caso se exige que para una partición $P \in \Phi(\Omega)$, el grafo inducido por cada uno de sus miembros, $G(P_k)$, sea un subgrafo conexo del grafo de contigüidad G asociado a la matriz A .

Esta condición se puede satisfacer en los algoritmos jerárquicos aglomerativos al restringir cuáles clusters se pueden mezclar, tal y como se comenta por ejemplo en [153], y es el enfoque que se asume en el presente trabajo. También es posible garantizar el cumplimiento de esta condición utilizando un algoritmo iterativo de relocalización tal y como el que se describe en [172].

El agrupamiento con restricciones de contigüidad requiere que el grafo de contigüidad G asociado a la matriz A sea conexo. Si el grafo de contigüidad no fuera conexo, la cantidad de componentes conexos definiría la cardinalidad mínima de cualquier partición válida.

En ocasiones es posible tratar los casos de grafos no conexos. Por ejemplo, si la contigüidad se define como en [152] en función de un umbral máximo, a partir del cual los objetos dejan de considerarse contiguos, es posible entonces aumentar este umbral hasta lograr un grafo conexo.

El otro supuesto que se maneja consiste en exigir que los clusters sean conexos, es decir, que no estén compuestos de dos o más regiones no contiguas. Este es un supuesto razonable al segmentar los objetos presentes en una imagen. Sin embargo, en dependencia del objetivo de la aplicación, este supuesto puede conducir a situaciones en las que un mismo tipo de tejido o de mineral sea contabilizado en clusters diferentes. Esto situación exigirá una etapa de post-procesamiento para poder identificar de manera adecuada estos casos.

A continuación se presentan varias definiciones que serán utilizadas a lo largo del trabajo.

5.4 Notación y definiciones

Como se mencionó antes la matriz simétrica $A = [a_{ij}]$ define una relación de contigüidad entre los objetos del conjunto $\Omega = \{\omega_1, \omega_2, \dots, \omega_n\}$.

Adicionalmente se considera que el grafo de contigüidad G asociado a esta matriz es contiguo. Además, una partición $P = \{P_1, P_2, \dots, P_k\}$ del conjunto Ω se considera válida si se cumple que ella representa un subgrafo válido de G , y además cada uno de sus componentes es conexo en el sentido de la relación de contigüidad definida entre los objetos.

El objetivo principal es construir una jerarquía de particiones mediante aglomeraciones sucesivas con la restricción de que solamente puedan ser mezclados clusters contiguos o adyacentes.

La relación de contigüidad entre objetos se propaga a los miembros de una partición de la siguiente manera.

Definición 1: Si $P_k, P_l \in P$, entonces los grupos o clases P_k y P_l son contiguos o adyacentes sí y sólo sí existe $(i, j) \in P_k \times P_l$ tales que $a_{ij} = 1$.

La definición anterior significa que dos grupos o clases P_k y P_l de P son contiguos o adyacentes si al menos un miembro de P_k es contiguo o adyacente al menos a un miembro de P_l , según la relación de contigüidad definida por A .

Con vistas a caracterizar la contigüidad entre dos clusters se define el grado de contigüidad entre clusters, así como sus propiedades.

Definición 2: El grado de contigüidad o adyacencia relativa entre dos grupos o clases $P_k, P_l \in P$ es la cantidad de pares que se pueden formar entre objetos adyacentes $i \in P_k$ y $j \in P_l$, y se denotará por n_{kl} .

$$n_{kl} = \left| \left\{ (i, j) \in P_k \times P_l / a_{ij} = 1, k \neq l \right\} \right|$$

Se debe notar que $n_{kl} = n_{lk}$. De manera alternativa se puede escribir que

$$n_{kl} = \sum_{\substack{i \in P_k \\ j \in P_l \\ k \neq l}} a_{ij}$$

Propiedad 1: El grado de contigüidad relativa entre grupos o clases cumple las siguientes propiedades.

- (a) $n_{kl} \neq 0 \leftrightarrow P_k$ y P_l son adyacentes
- (b) $n_{kl} = 0 \leftrightarrow P_k$ y P_l no son adyacentes

Siguiendo la definición se puede plantear que n_{kk} representa el grado de contigüidad interno del cluster P_k .

Supongamos que entre los objetos del conjunto Ω está definida una función de desemejanza $d(\omega_i, \omega_j)$, la cual toma en cuenta las mediciones asociadas a cada objeto. En el apéndice A se asume que la desemejanza entre los objetos está dada por distancia euclidiana cuadrática.

De manera análoga a lo planteado en el apartado 2.2.1 del capítulo 2, la descomposición de la matriz de conexiones conduce a una descomposición de la suma de las desemejanzas entre los objetos relacionados para una partición dada.

Es decir, para una partición $P = \{P_1, P_2, \dots, P_K\}$ se tiene que

$$\sum_{\omega_i, \omega_j \in \Omega} a_{ij} d(\omega_i, \omega_j) = \sum_{k=1}^K \sum_{\omega_i, \omega_j \in P_k} a_{ij} d(\omega_i, \omega_j) + \sum_{k=1}^K \sum_{\substack{l=1 \\ l \neq k}}^K \sum_{\omega_i \in P_k} \sum_{\omega_j \in P_l} a_{ij} d(\omega_i, \omega_j)$$

Esta descomposición incluye un componente de homogeneidad asociado a las desemejanzas internas (entre objetos relacionados de cada clase) y un componente de heterogeneidad asociado a las desemejanzas cruzadas (entre objetos relacionados de diferentes clases).

El componente asociado a la heterogeneidad puede ser considerado también como una medida de la separación entre clases. Esta separación toma en cuenta la estructura de los datos que se especifica a través de la matriz de contigüidad.

Si se toma como criterio para la formación de una partición de tamaño K dado la máxima separación entre clases, obviamente se estaría buscando también una partición de máxima homogeneidad. Es decir, maximizar el componente de desemejanzas entre clases es equivalente a minimizar el componente de desemejanzas internas.

Si se escoge como criterio de calidad de una partición el componente de homogeneidad se puede plantear que el objetivo sería encontrar una partición factible de cardinalidad K , $P^* \in \Phi_K(\Omega)$, tal que

$$J(P) = \sum_{k=1}^K J'(P_k)$$

donde

$$J'(P_k) = \sum_{\omega_i, \omega_j \in P_k} a_{ij} d(\omega_i, \omega_j)$$

Debido a su complejidad computacional, para resolver este tipo de problemas se recurre a métodos aproximados tales como métodos de optimización local y métodos aglomerativos (divisivos y aglomerativos) [174], tales como los que se tratan en el capítulo 2.

Si analizamos el caso de dos clases adyacentes P_k y P_l , tales que $P_k \cap P_l = \emptyset$ podemos reescribir la descomposición anterior de la siguiente manera:

$$\sum_{\omega_i, \omega_j \in P_k \cup P_l} a_{ij} d(\omega_i, \omega_j) = \sum_{\omega_i, \omega_j \in P_k} a_{ij} d(\omega_i, \omega_j) + \sum_{\omega_i, \omega_j \in P_l} a_{ij} d(\omega_i, \omega_j) + \sum_{\omega_i \in P_k} \sum_{\omega_j \in P_l} a_{ij} d(\omega_i, \omega_j)$$

O de forma equivalente

$$J(P_k \cup P_l) = J(P_k) + J(P_l) + \sum_{\omega_i \in P_k} \sum_{\omega_j \in P_l} a_{ij} d(\omega_i, \omega_j)$$

Esta última igualdad sugiere el siguiente esquema de aproximación. Supongamos que a partir de una partición P dada se obtiene otra partición P' mediante la fusión de los clusters P_k y P_l , es decir,

$$P' = (P \setminus \{P_k, P_l\}) \cup \{P_k \cup P_l\}$$

Entonces se puede plantear que

$$J(P') = J(P) - J(P_k) - J(P_l) + J(P_k \cup P_l)$$

Sustituyendo $J(P_k \cup P_l)$ según la expresión derivada antes se obtiene

$$J(P') = J(P) + \sum_{\omega_i \in P_k} \sum_{\omega_j \in P_l} a_{ij} d(\omega_i, \omega_j)$$

Es decir,

$$J(P') - J(P) = \sum_{\omega_i \in P_k} \sum_{\omega_j \in P_l} a_{ij} d(\omega_i, \omega_j)$$

Se puede entonces plantear la siguiente heurística: si se comienza con la partición más fina $P_0 = \{\{\omega_i\} / \omega_i \in \Omega, i = 1, \dots, n\}$ y en cada paso se mezclan aquellos clusters (contiguos) con menor separación en el sentido de la heterogeneidad se van obteniendo los menores incrementos de la función de costo y probablemente se obtengan solución (casi) óptimas. Esto coincide, en esencia, con el algoritmo clásico de agrupamiento jerárquico aglomerativo.

En el presente trabajo la separación entre clusters se define de manera un tanto diferente.

Definición 3: La desemejanza entre dos grupos $P_k, P_l \in P$ que se utilizará se basa en el concepto de heterogeneidad y se define de la siguiente manera:

$$d_{AH}(P_k, P_l) = \frac{1}{n_{kl}} \sum_{\substack{\omega_i \in P_k \\ \omega_j \in P_l}} a_{ij} d(\omega_i, \omega_j) \quad (1)$$

El factor $1/n_{kl}$ se incluye para evitar que se favorezca a aquellos pares de clusters que posean menor cantidad de enlaces. Por ejemplo, supongamos que entre los clusters P_1 y P_2 existe un único enlace con valor de desemejanza igual a 10, y entre los clusters P_3 y P_4 existen 5 enlaces cada uno con una desemejanza asociada de valor 4. En caso de no promediarse la desemejanza entre ellos sería igual a 20 y se favorecería la unión de P_1 y P_2 (menor desemejanza). Sin embargo, según la definición la nueva desemejanza es igual a 4 y en este caso la selección del par de clusters a fusionar recaería sobre P_3 y P_4 .

Se debe notar que la desemejanza entre dos grupos $P_k, P_l \in P$ está definida siempre y cuando $n_{kl} \neq 0$, es decir, cuando los grupos son contiguos.

No es difícil comprobar que en el caso límite en que todos los objetos se consideren adyacentes entre sí, la expresión (1) coincide con la que se utiliza en el método de enlace promedio.

Analicemos el caso en que se unen dos grupos contiguos $P_k, P_l \in P$ para formar uno nuevo, $P_k \cup P_l$.

Propiedad 2: El grado de contigüidad o adyacencia relativa entre $P_k \cup P_l \in P$ y otro grupo cualquiera $P_m \in P$, denotado mediante $n_{k \cup l, m}$, se calcula de la siguiente manera

$$n_{k \cup l, m} = n_{km} + n_{lm} \quad (2)$$

En efecto, existen cuatro posibles casos:

1. El grupo P_m no es adyacente ni a P_k ni a P_l . En este caso $P_k \cup P_l$ tampoco es adyacente a P_m (cf. **Def. 1**), luego, $n_{km} = 0$, $n_{lm} = 0$, y $n_{k \cup l, m} = 0$ (cf. **Prop. 1**).
2. El grupo P_m es adyacente a P_k pero no a P_l . En este caso $n_{lm} = 0$, y $n_{k \cup l, m} = n_{kl}$.
3. El grupo P_m es adyacente a P_l pero no a P_k . Este caso es equivalente al caso (2).
4. El grupo P_m es adyacente tanto a P_k como a P_l . Si se considera que $P_k \cap P_l = \emptyset$, por ser miembros de la partición P , entonces no existen enlaces repetidos, y por tanto $n_{k \cup l, m} = n_{km} + n_{lm}$.

Respecto a la desemejanza entre la clase que se forma por la unión de $P_k, P_l \in P$, es decir $P_k \cup P_l$, y otro grupo adyacente cualquiera $P_m \in P$, se tiene que (cf. **Def. 3**)

$$d_{AH}(P_k \cup P_l, P_m) = \frac{1}{n_{k \cup l, m}} \sum_{\substack{\omega_i \in P_k \cup P_l \\ \omega_j \in P_m}} a_{ij} d(\omega_i, \omega_j)$$

Es válida la siguiente propiedad.

Propiedad 3: Se cumple que

$$d_{AH}(P_k \cup P_l, P_m) = \frac{n_{km}}{n_{km} + n_{lm}} d_{AH}(P_k, P_m) + \frac{n_{lm}}{n_{km} + n_{lm}} d_{AH}(P_l, P_m) \quad (3)$$

Efectivamente, según la **Prop. 2** del grado de contigüidad relativa se tiene que

$$d_{AH}(P_k \cup P_l, P_m) = \frac{1}{n_{km} + n_{lm}} \sum_{\substack{\omega_i \in P_k \cup P_l \\ \omega_j \in P_m}} a_{ij} d(\omega_i, \omega_j)$$

Por otra parte, teniendo en cuenta que $P_k \cap P_l = \emptyset$, es posible descomponer la sumatoria asociada a los enlaces entre $P_k \cup P_l$ y P_m de la siguiente manera

$$d_{AH}(P_k \cup P_l, P_m) = \frac{1}{n_{km} + n_{lm}} \left[\sum_{\substack{\omega_i \in P_k \\ \omega_j \in P_m}} a_{ij} d(\omega_i, \omega_j) + \sum_{\substack{\omega_i \in P_l \\ \omega_j \in P_m}} a_{ij} d(\omega_i, \omega_j) \right]$$

Atendiendo a la **Def. 3** es posible reescribir esta última expresión como la suma de dos desemejanzas, es decir,

$$d_{AH}(P_k \cup P_l, P_m) = \frac{1}{n_{km} + n_{lm}} \left[n_{km} d_{AH}(P_k, P_m) + n_{lm} d_{AH}(P_l, P_m) \right]$$

De donde se obtiene inmediatamente

$$d_{AH}(P_k \cup P_l, P_m) = \frac{n_{km}}{n_{km} + n_{lm}} d_{AH}(P_k, P_m) + \frac{n_{lm}}{n_{km} + n_{lm}} d_{AH}(P_l, P_m)$$

Esta expresión permite calcular las desemejanzas entre una clase recién formada y cada una de las restantes.

Se debe notar que si P_m es contiguo sólo a uno de los dos clusters, por ejemplo P_k , entonces se tiene que:

$$d_{AH}(P_k \cup P_l, P_m) = d_{AH}(P_k, P_m) \quad (4)$$

Por último, si se tiene en cuenta que $d_{AH}(P_k \cup P_l, P_m)$ se obtiene como combinación convexa de $d_{AH}(P_k, P_m)$ y de $d_{AH}(P_l, P_m)$ es posible entonces escribir

$$\min\{d_{AH}(P_k, P_m), d_{AH}(P_l, P_m)\} \leq d_{AH}(P_k \cup P_l, P_m) \leq \max\{d_{AH}(P_k, P_m), d_{AH}(P_l, P_m)\} \quad (5)$$

5.5 El problema de las inversiones

Como se planteó antes, el proceso aglomerativo se restringe solamente a clusters contiguos o adyacentes. Inicialmente todos los clusters son singulares, es decir, están formados por un único objeto. En cada paso se fusionan los dos clusters adyacentes P_k y P_l para los que se obtiene la menor desemejanza según se definió en (1).

Los casos de empates deben ser analizados con precaución pues pueden afectar los resultados [9, 15]. En la práctica existe la posibilidad que puedan ocurrir empates en las desemejanzas, por lo que es necesario resolverlos mediante algún procedimiento (por ejemplo, usando otro criterio diferente al que generó el empate) con vistas a reducir la ambigüedad que pueda surgir.

Cuando las desemejanzas toman valores reales es raro que ocurran empates. Para el caso de datos de tipo booleano, por ejemplo los utilizados en quimioinformática, esta cuestión cobra mayor importancia. Sin embargo, cuando se realizan transformaciones previas de los datos originales (por ejemplo, inmersiones euclidianas mediante un análisis de correspondencias) es muy poco probable que se generen empates¹.

En lo adelante, a menos que se diga lo contrario, se asume que no existen empates en los valores de desemejanzas, o que han sido resueltos de manera que no existan ambigüedades.

Una vez realizada la fusión se utiliza la expresión (3) para calcular la desemejanza entre el cluster recién formado y el resto de los clusters contiguos de la partición. De esta manera se va construyendo la jerarquía de particiones. Este método será denominado método de enlace tipo promedio con restricciones de contigüidad.

¹ Correspondencia personal con F. Murtagh.

En este contexto el problema de las inversiones, es decir el crecimiento/decrecimiento no monótono de las desemejanzas, es propenso a surgir [54, 58]. Se dice que una inversión ocurre cuando se cumple que

$$d_{AH}(P_k \cup P_l, P_m) < d_{AH}(P_k, P_l) \quad (6)$$

En la expresión anterior P_k y P_l representan los clusters adyacentes que se fusionan y P_m es un cluster adyacente a $P_k \cup P_l$.

La presencia de inversiones en una jerarquía es un fenómeno no deseable pues dificulta la interpretación de las particiones y la definición de desemejanzas entre clusters. Incluso las inversiones pueden conducir a que se afecte la unicidad de las soluciones al aplicar diferentes implementaciones algorítmicas. En el caso que nos ocupa es posible garantizar la ausencia de inversiones.

Proposición 1: La jerarquía de clusters que se obtiene con el método de enlace tipo promedio con restricciones de contigüidad es libre de inversiones.

Demostración:

La demostración de esta proposición sigue el esquema de Murtagh [175] para el caso del método de enlace completo con restricciones.

Sean P_k, P_l y P_m tres clusters (posiblemente singulares). Sin pérdida de generalidad vamos a considerar que las fusiones de los clusters ocurren de la siguiente manera:

- Supuesto 1: Primeramente se fusionan los clusters P_k y P_l .
- Supuesto 2: A continuación se fusiona P_m con $P_k \cup P_l$ de manera directa, es decir, sin ningún cluster intermedio. Esto implica que P_m es adyacente a P_k y P_l , o al menos a uno de ellos.
- Supuesto 3: También sin pérdida de generalidad consideraremos que en el caso de que ambas desemejanzas estén definidas se cumple que $d_{AH}(P_k, P_m) \leq d_{AH}(P_l, P_m)$.

Consideremos tres casos.

Caso 1: $d_{AH}(P_k, P_m) \leq d_{AH}(P_l, P_m) \leq d_{AH}(P_k, P_l)$

Si el cluster P_m es adyacente a P_k y P_l , o sólo a uno de ellos, la premisa del caso indica que él debía haber sido fusionado antes que P_k y P_l y esto contradice el

Supuesto 1. Por otra parte, si P_m no es adyacente ni a P_k ni a P_l , entonces él no se puede unir directamente con $P_k \cup P_l$, excepto a través de un tercer cluster, lo que contradice el Supuesto 2.

$$\text{Caso 2: } d_{AH}(P_k, P_l) \leq d_{AH}(P_k, P_m) \leq d_{AH}(P_l, P_m)$$

A partir de (5) se tiene que

$$d_{AH}(P_k \cup P_l, P_m) \geq \min\{d_{AH}(P_k, P_m), d_{AH}(P_l, P_m)\}$$

Si el cluster P_m es adyacente a P_k y P_l entonces el Supuesto 3 conduce a que

$$d_{AH}(P_k \cup P_l, P_m) \geq d_{AH}(P_k, P_m)$$

Por otra parte si el cluster P_m fuera adyacente únicamente a P_l entonces de acuerdo a (4) se tendría que

$$d_{AH}(P_k \cup P_l, P_m) = d_{AH}(P_l, P_m)$$

En cualquier caso, según la premisa del caso se tiene que

$$d_{AH}(P_k \cup P_l, P_m) \geq d_{AH}(P_k, P_l)$$

$$\text{Caso 3: } d_{AH}(P_k, P_m) \leq d_{AH}(P_k, P_l) \leq d_{AH}(P_l, P_m)$$

Este caso es posible únicamente si P_k y P_m no son adyacentes, pues en caso contrario deben ser fusionados antes de P_k y P_l lo que llevaría a una contradicción con el Supuesto 1. Por su parte el Supuesto 2 conduce a que P_m y P_l son adyacentes.

De acuerdo a (4) se tiene entonces que

$$d_{AH}(P_k \cup P_l, P_m) = d_{AH}(P_l, P_m)$$

Finalmente, tomando en cuenta la premisa del caso se tiene que

$$d_{AH}(P_k \cup P_l, P_m) \geq d_{AH}(P_k, P_l)$$

En resumen, el caso 1 no puede ocurrir sin contradecir los supuestos. Por su parte en los casos 2 y 3 se tiene que $d_{AH}(P_k \cup P_l, P_m) \geq d_{AH}(P_k, P_l)$ por lo que, según la expresión (6), no conducen a inversiones. \square

5.6 Estrategias de implementación basadas en los vecinos más próximos para el caso de restricciones de contigüidad

Como se apuntaba anteriormente la inclusión de restricciones de contigüidad impone la necesidad de reexaminar el problema de las inversiones. De la misma manera se debe analizar la validez de las diferentes estrategias de implementación, en particular la relacionada con la utilización de cadenas de vecinos más próximos.

Solamente dos de los métodos tradicionales de agrupamiento jerárquico aglomerativo pueden ser adaptados de manera que en cada paso se fusionen clusters contiguos y a la vez se mantengan libres de inversiones. El primero es una variante del método de enlace simple, y el segundo es el método de enlace completo [158, 169, 175].

La variante del método de enlace simple con restricciones de contigüidad presupone que en cada paso se fusione aquel par de clusters que posea la menor desemejanza de enlace, siempre y cuando ésta se alcance para un par de objetos contiguos (uno de cada cluster).

En el caso del método del enlace completo con restricciones de contigüidad solamente se exige que los clusters que se fusionan sean contiguos, es decir, que exista un enlace de contigüidad entre al menos un miembro de cada cluster. A diferencia de la variante del método de enlace simple esa restricción no se impone a la hora de calcular las desemejanzas, es decir, la mayor desemejanza se puede alcanzar para un par de objetos no relacionados directamente.

Si se impone la última variante de restricción de contigüidad entonces la jerarquía de enlace completo es la única entre los métodos tradicionales, que siendo construida con la estrategia de vecinos más próximos mutuos o recíprocos, es libre de inversiones [153].

El método de enlace completo con restricciones de contigüidad ha sido utilizado en varias aplicaciones. En [176] se emplea como vía para el análisis de la estructura de la secuencia de escenas que forman un guión cinematográfico o televisivo. La restricción de contigüidad en este caso es temporal, específicamente, se consideran como contiguas aquellas escenas que preceden o suceden a una dada en el guión.

La posibilidad del uso de la estrategia de vecinos más próximos mutuos o recíprocos, en el caso del método de tipo enlace promedio con restricciones de contigüidad propuesto, queda expresada de la siguiente manera.

Proposición 2: El método de enlace tipo promedio con restricciones de contigüidad al ser utilizado con aglomeraciones de vecinos más próximos mutuos o recíprocos no conduce a inversiones.

Demostración:

Consideremos tres clusters P_k , P_l y P_m (que pueden ser singulares) pertenecientes a una cadena de vecinos más próximos de la siguiente manera

$$P_l = NN(P_k); P_m = NN(P_l); P_l = NN(P_m)$$

Como se puede notar P_l y P_m son vecinos más próximos mutuos o recíprocos.

Para obtener esta secuencia bajo la condición de contigüidad se debe cumplir que tanto P_k y P_l , como P_l y P_m sean contiguos, y además $d_{AH}(P_k, P_l) > d_{AH}(P_l, P_m)$.

¿Qué influencia sobre la jerarquía tiene el hecho de la contigüidad de P_k y P_m ?

Si P_k y P_m no son contiguos, entonces la fusión de P_l y P_m no conduce a inversiones. Esto se debe a que, según (4), $d_{AH}(P_l \cup P_m, P_k) = d_{AH}(P_k, P_l)$. Finalmente, según la premisa, se tiene que $d_{AH}(P_l \cup P_m, P_k) > d_{AH}(P_l, P_m)$.

La única posibilidad de inversión pudiera ocurrir cuando P_k y P_m son contiguos. Supongamos que la inversión ocurre, es decir,

$$d_{AH}(P_l \cup P_m, P_k) < d_{AH}(P_l, P_m)$$

Según (5) se tiene que

$$d_{AH}(P_l \cup P_m, P_k) \geq \min(d_{AH}(P_k, P_l), d_{AH}(P_k, P_m))$$

Aquí se tienen en principio dos posibilidades:

- a) $d_{AH}(P_k, P_l) = \min(d_{AH}(P_k, P_l), d_{AH}(P_k, P_m))$. Atendiendo a las premisas se puede escribir entonces

$$d_{AH}(P_l \cup P_m, P_k) \geq d_{AH}(P_k, P_l) > d_{AH}(P_l, P_m)$$

Es decir, no existe inversión.

- b) $d_{AH}(P_k, P_m) = \min(d_{AH}(P_k, P_l), d_{AH}(P_k, P_m))$. Este caso no sería posible pues sería equivalente a decir que $P_m = NN(P_k)$, es decir, que P_m , y no P_l , es el vecino más próximo de P_k , lo que contradice la premisa de que $P_l = NN(P_k)$.

Es decir, bajo las condiciones de una cadena completa una fusión del par de vecinos más próximos mutuos según el método de enlace tipo promedio con restricciones de contigüidad no conduce a inversiones. \square

Si asumimos la estrategia de construcción del dendrograma descrita en apartado 2.2, es decir, en cada etapa se fusionan los dos clusters adyacentes P_k y P_l para los que se obtiene la menor desemejanza, según se definió en (1), se puede comprobar que se cumple lo siguiente.

Proposición 3: En cada etapa del método de enlace tipo promedio con restricciones de contigüidad, al ser utilizado con aglomeraciones de pares de clusters contiguos de desemejanza mínima, se fusionan pares de clusters que son vecinos más próximos mutuos o recíprocos.

Demostración:

Sean P_k y P_l el par de clusters adyacentes que se fusionan en una etapa dada. Esto significa que ellos realizan la desemejanza mínima. Supongamos que P_l no es un vecino más próximo de P_k . Sea P_m el cluster contiguo vecino más próximo de P_k . Eso significa que $d_{AH}(P_k, P_m) < d_{AH}(P_l, P_m)$ lo que contradice el supuesto. Luego P_l es un vecino más próximo de P_k .

De la misma manera se demuestra que P_k es un vecino más próximo de P_l , es decir, P_k y P_l son vecinos más próximos mutuos o recíprocos. \square

Siguiendo las ideas expuestas en [59] es posible demostrar que, para el caso del método de enlace tipo promedio con restricciones de contigüidad, se cumple además que los pares de vecinos más próximos se mantienen como tales durante la construcción del dendrograma hasta el momento en que son fusionados.

Proposición 4: El método de enlace tipo promedio con restricciones de contigüidad, al ser utilizado con aglomeraciones de pares de clusters contiguos de desemejanza mínima, garantiza que los pares de vecinos más próximos mutuos o recíprocos se mantienen como tales hasta que son eventualmente fusionados en una etapa.

Demostración:

Supongamos que en una etapa dada se tienen los clusters (posiblemente singulares) $P_{i_1}, P_{i_2}, P_{i_3}, \dots, P_{i_{r-1}}$ y P_{i_r} . Sin pérdida de generalidad consideremos que P_{i_1} y P_{i_2} , así como $P_{i_{r-1}}$ y P_{i_r} forman pares de vecinos más próximos mutuos y que en la etapa dada se fusionan $P_{i_{r-1}}$ y P_{i_r} . Se asume que no existen empates en las desemejanzas (ver apartado 4.5).

Según (5) para cualquier P_{i_r} se cumple que

$$d_{AH}(P_{i_{r-1}} \cup P_{i_r}, P_{i_r}) \geq \min \{d_{AH}(P_{i_{r-1}}, P_{i_r}), d_{AH}(P_{i_r}, P_{i_r})\}.$$

Esto último significa que el cluster recién formado, $P_{i_{r-1}} \cup P_{i_r}$, no estará más cercano a ningún otro cluster existente P_{i_r} de los que estaban sus miembros $P_{i_{r-1}}$ y P_{i_r} .

En particular, según los supuestos, se tiene que ni $P_{i_{r-1}}$, ni P_{i_r} están más cerca de P_{i_1} de lo que está P_{i_2} , es decir,

$$\begin{aligned} d_{AH}(P_{i_{r-1}}, P_{i_1}) &> d_{AH}(P_{i_1}, P_{i_2}), \text{ y} \\ d_{AH}(P_{i_r}, P_{i_1}) &> d_{AH}(P_{i_1}, P_{i_2}) \end{aligned}$$

Es posible entonces escribir que

$$d_{AH}(P_{i_{r-1}} \cup P_{i_r}, P_{i_1}) \geq \min \{d_{AH}(P_{i_{r-1}}, P_{i_1}), d_{AH}(P_{i_r}, P_{i_1})\} > d_{AH}(P_{i_1}, P_{i_2})$$

Es decir, $P_{i_{r-1}} \cup P_{i_r}$ no será vecino más próximo mutuo de P_{i_1} . Lo mismo se puede decir de P_{i_2} . Esto demuestra que el par de vecinos más próximos mutuos P_{i_1} y P_{i_2} no puede ser disuelto por el cluster que se obtenga de una agregación. Luego, el par de vecinos más próximos mutuos P_{i_1} y P_{i_2} será preservado de etapa en etapa hasta que eventualmente sea fusionado.

Lo anterior significa que en la próxima etapa se puede fusionar un par de vecinos más próximos mutuos, por ejemplo P_{i_1} y P_{i_2} , o también el cluster recién formado $P_{i_{r-1}} \cup P_{i_r}$ con un cluster cualquiera que no forme parte de un par de vecinos más próximos mutuos. \square

Las Proposiciones 3 y 4 permiten sugerir una estrategia de agregaciones múltiples mediante la cual en cada etapa se fusionan todos los pares de vecinos más próximos mutuos.

5.7 *Análisis de la complejidad computacional. El caso de las imágenes digitales*

Como se mencionó anteriormente la estrategia basada en cadenas de vecinos más próximos para el caso sin restricciones conduce a una complejidad temporal $O(n^2)$.

A continuación se analiza un algoritmo basado en la estrategia de las cadenas de vecinos más próximos con restricciones de contigüidad para la segmentación de imágenes digitales [153].

1. Inicialmente se construye un arreglo de las mismas dimensiones de la imagen y se inicializa de manera que cada píxel reciba una etiqueta única. A continuación se comienza la construcción de una cadena de vecinos más próximos, en el sentido de la desemejanza utilizada, teniendo en cuenta únicamente los vecinos espaciales de cada píxel.

Este proceso incluye:

- a. La selección del píxel.
 - b. La determinación de sus vecinos espaciales.
 - c. La exclusión de aquellos vecinos espaciales que poseen la misma etiqueta.
 - d. Calcular las desemejanzas entre la región a la que pertenece el píxel analizado y todas las regiones a las que pertenecen los píxeles vecinos espaciales de etiquetas diferentes.
2. Una vez que se encuentre un par de vecinos más próximos mutuos se procede a su fusión, y se extraen de la cadena. En este momento se debe incluir la información referida a la fusión en la estructura de datos asociada a la jerarquía que se construye. Adicionalmente se deben actualizar las etiquetas de manera apropiada.

La complejidad temporal de este algoritmo depende de la cantidad de expansiones y contracciones de la cadena de vecinos más próximos. En cada expansión de la cadena se debe encontrar un vecino más próximo.

La búsqueda del vecino más próximo de una región dada en una etapa cualquiera de la construcción de la jerarquía depende de cuántos vecinos espaciales puede tener una región. Aún en el caso de restricciones espaciales la cantidad de vecinos espaciales en el peor caso es $O(n)$ [153].

Esto conduce a que también en el caso del uso de restricciones la complejidad temporal es $O(n^2)$ en el peor caso.

Sin embargo, no siempre estaremos en presencia del peor caso. En particular resulta de interés también analizar el caso promedio.

5.8 Comportamiento del algoritmo jerárquico aglomerativo con restricciones de contigüidad

Como se comentó antes, la complejidad computacional está asociada directamente a la cantidad de vecinos espaciales que puede tener una región en una etapa cualquiera de la construcción de la jerarquía.

A continuación se analiza el comportamiento de la cantidad de desemejanzas y comparaciones que se requieren a la hora de procesar los vecinos espaciales de una región cualquiera. Una problemática similar para el caso particular de vecinos espaciales dados por la relación de 8 vecindad ha sido tratada en [177].

Consideremos una imagen que contiene n píxeles. No situaremos en el caso de que la relación de vecindad espacial de un píxel esta dada por el conjunto de píxeles que se encuentran a una distancia euclidiana no mayor que un valor r dado.

Denotemos por $A^{(0)}$ la matriz de conexión asociada a esta relación en la etapa inicial del algoritmo, es decir, en el momento en que cada píxel representa una clase singular. Se tiene entonces que el elemento $a_{ij}^{(0)}$ de la matriz $A^{(0)}$ es igual a 1 si los píxeles i y j son vecinos según la relación dada.

Si se considera que en cada paso del algoritmo se mezclan solamente dos clases, se requieren entonces $n-1$ etapas para la construcción del dendrograma.

En cada etapa se tiene una partición $P^t = \{P_1, P_2, \dots, P_{K_t}\}$ del conjunto Ω .

Denotemos por $A^{(t)}$, $0 \leq t \leq n-1$, la matriz de contigüidad que se obtiene en la etapa t de la construcción del dendrograma.

La matriz $A^{(0)}$ se puede considerar como la matriz de adyacencias de un grafo no dirigido en el cual cada píxel representa un vértice. Por su parte $A^{(t)}$, $0 \leq t \leq n-1$, se obtiene mediante fusiones de clases conexas (vértices) a partir de $A^{(t-1)}$, $t \geq 1$. Es decir, un vértice del grafo $A^{(t)}$ está formado por un agregado de vértices de $A^{(t-1)}$. En lo adelante utilizaremos $A^{(t)}$ para denotar también el grafo asociado.

Cada arista del grafo $A^{(t)}$ se considera ponderada con la desemejanza entre las clases representadas por los correspondientes vértices.

De esta manera el proceso de construcción de la jerarquía de clasificación está asociado a la construcción de $n-1$ grafos $A^{(t)}$, $0 \leq t \leq n-1$. Cada grafo $A^{(t)}$ contiene $n-t$ vértices.

Para cada vértice se corresponde con una clase o cluster. Denotaremos mediante $V^t(P_k)$ la vecindad de contigüidad de un cluster cualquiera P_k según $A^{(t)}$, y por $\#(V^t(P_k))$ la cardinalidad de esta vecindad, es decir, la cantidad de vecinos espaciales.

El valor $\#(V^t(P_k))$ lo denominaremos también grado de contigüidad del vértice (clase) P_k . Notar la diferencia con el grado de contigüidad entre clases de la definición. 2. Aquí se refiere a la cantidad de vecinos en términos de clusters, mientras que la noción que se maneja es la cantidad de conexiones entre miembros de clusters, es decir, cantidad de enlaces.

A la hora de fusionar dos clases contiguas P_k y P_l en una etapa t dada de la construcción del árbol o dendrograma la cantidad de actualizaciones y comparaciones depende precisamente de $\#(V^{t+1}(P_k \cup P_l))$.

Respecto a la vecindad de la fusión de dos clases contiguas P_k y P_l y su cardinalidad se cumplen las siguientes propiedades.

Propiedad 4: Al fusionar dos clases contiguas P_k y P_l en una etapa t de la construcción del árbol o dendrograma se cumple que

$$V^{t+1}(P_k \cup P_l) = V^t(P_k) \cup V^t(P_l) - \{P_k, P_l\}$$

Efectivamente, después de la fusión P_k y P_l dejan de ser vértices de $A^{(t+1)}$, en su lugar aparece el agregado $P_k \cup P_l$, y con ellos desaparece la arista que los une. Además todo vértice P_m que sea miembro de $V^t(P_k)$ y sea diferente de P_l será también miembro de $V^{t+1}(P_k \cup P_l)$. De la misma manera todo vértice P_r que sea miembro de $V^t(P_l)$ y sea diferente de P_k será también miembro de $V^{t+1}(P_k \cup P_l)$.

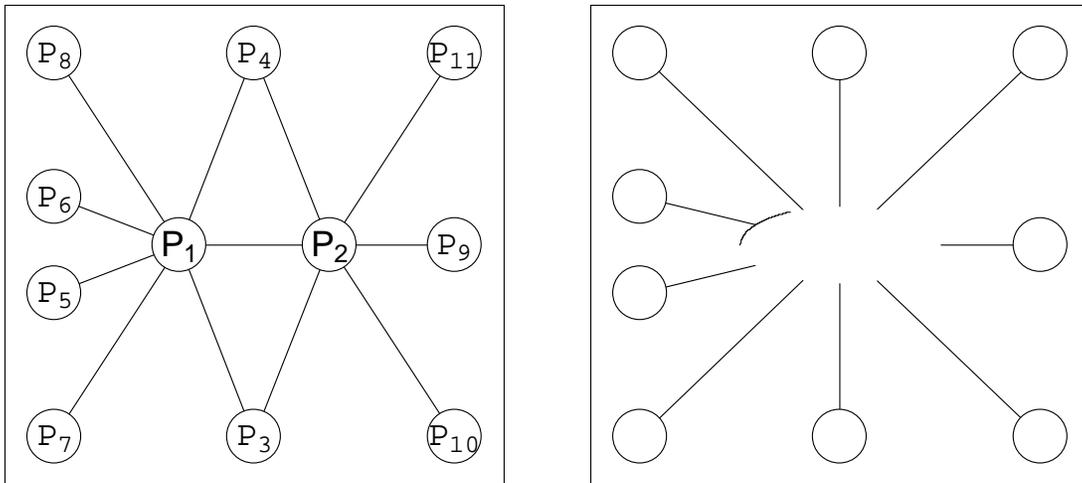
Propiedad 5: El grado de contigüidad de $P_k \cup P_l$ está dado por la siguiente expresión

$$\#(V^{t+1}(P_k \cup P_l)) = \#(V^t(P_k)) + \#(V^t(P_l)) - \#(V^t(P_k) \cap V^t(P_l)) - 2$$

donde $\#(V^t(P_k) \cap V^t(P_l))$ representa la cardinalidad de $V^t(P_k) \cap V^t(P_l)$, es decir, la cantidad de vecinos comunes tanto a P_k como a P_l .

En efecto, todos los miembros de $V^{t+1}(P_k \cup P_l)$ se cuentan en la expresión anterior. Los vértices vecinos que pertenecen a $V^t(P_k) \cap V^t(P_l)$ se cuentan dos veces, una como vecinos de P_k y otra como vecinos de P_l , de ahí que se debe restar esa cantidad en la expresión de la derecha. Por último, los vértices que se fusionan aparecen uno como vecino del otro; pero no serán miembros de $V^{t+1}(P_k \cup P_l)$ y no deben ser tomados en cuenta en el cálculo.

En el ejemplo de la Figura 5-3 se presenta una partición de 11 clusters con los respectivos enlaces entre clusters vecinos. Supongamos que en la etapa que se analiza se fusionan los clusters P_1 y P_2 .



Una vez ocurrida la fusión se obtiene el nuevo cluster $P_1 \cup P_2$ cuya vecindad es

$$V^t(P_1 \cup P_2) = \{P_3, P_4, P_5, P_6, P_7, P_8, P_9, P_{10}, P_{11}\}$$

Finalmente, según la propiedad 5 se tiene

$$\begin{aligned} \#(V^{t+1}(P_1 \cup P_2)) &= \#(V^t(P_1)) + \#(V^t(P_2)) - \#(V^t(P_1) \cap V^t(P_2)) - 2 = \\ &= 7 + 6 - 2 - 2 = 9 \end{aligned}$$

El algoritmo jerárquico ascendente con restricciones de contigüidad presupone una etapa de inicialización y luego la serie de iteraciones para formar el dendrograma.

En la etapa de inicialización cada píxel se considera una clase. Atendiendo a la relación de vecindad definida anteriormente, en el grafo $A^{(0)}$ cada píxel puede tener una cantidad máxima de vecinos, que denotaremos por $\#_{\max}^0$, y que depende del radio r de la vecindad pero no de n , la cantidad de píxeles. En esta etapa se realizan las siguientes operaciones.

- a. Obtención de la vecindad espacial de cada píxel.
- b. Cálculo de las desemejanzas entre cada píxel y sus vecinos espaciales.
- c. Determinación de los vecinos más próximos (en el sentido de la desemejanza utilizada) de cada píxel dentro de su vecindad espacial, así como el valor de la menor desemejanza.
- d. Fusionar el par de vecinos más próximos que posean el menor valor de desemejanza entre ellos.

Se debe notar que en el paso (d), en virtud de la **Proposición 3**, se fusiona un par de vecinos más próximos recíprocos.

En esta etapa los pasos (a)-(c) poseen una complejidad computacional $O(\#_{\max}^0)$. Ahora, como el valor $\#_{\max}^0$ es constante para un r fijo, y no depende de n , la complejidad de la etapa en su conjunto se mantiene $O(n)$.

En lo adelante se deben ir obteniendo los grafos $A^{(t+1)}$ a partir de $A^{(t)}$, $0 \leq t \leq n-2$, mediante fusiones sucesivas de dos clases vecinas más próximas (recíprocas) v y v' de desemejanza mínima. La fusión presupone una serie de actualizaciones.

1. Construcción de la vecindad $V^{t+1}(P_k \cup P_l)$.
2. Para cada clase P_m vecina de P_k sustituir en $V^t(P_m)$, que ahora se convierte en $V^{t+1}(P_m)$, la clase P_k por $P_k \cup P_l$.

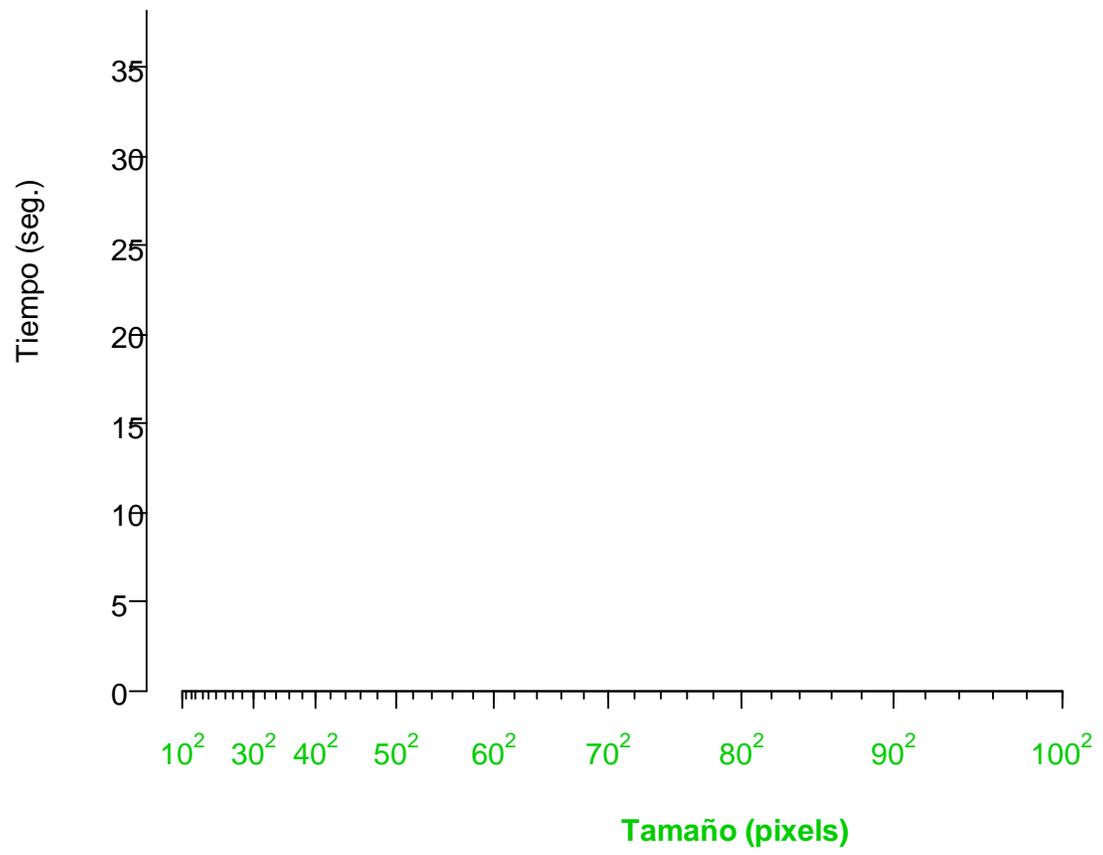
3. Para cada clase P_m vecina de P_l sustituir en $V^t(P_m)$, que ahora se convierte en $V^{t+1}(P_m)$, la clase P_l por $P_k \cup P_l$.
4. Suprimir $V^t(P_k)$ y $V^t(P_l)$.
5. Para todos los miembros de $V^{t+1}(P_k \cup P_l)$ actualizar las desemejanzas con $P_k \cup P_l$.
6. Encontrar en $V^{t+1}(P_k \cup P_l)$ los vecinos más próximos de $P_k \cup P_l$.
7. Para cada clase s vecina analizar si $P_k \cup P_l$ se encuentra entre sus vecinos más próximos en cuyo caso $P_k \cup P_l$ y s son vecinos más próximos recíprocos.

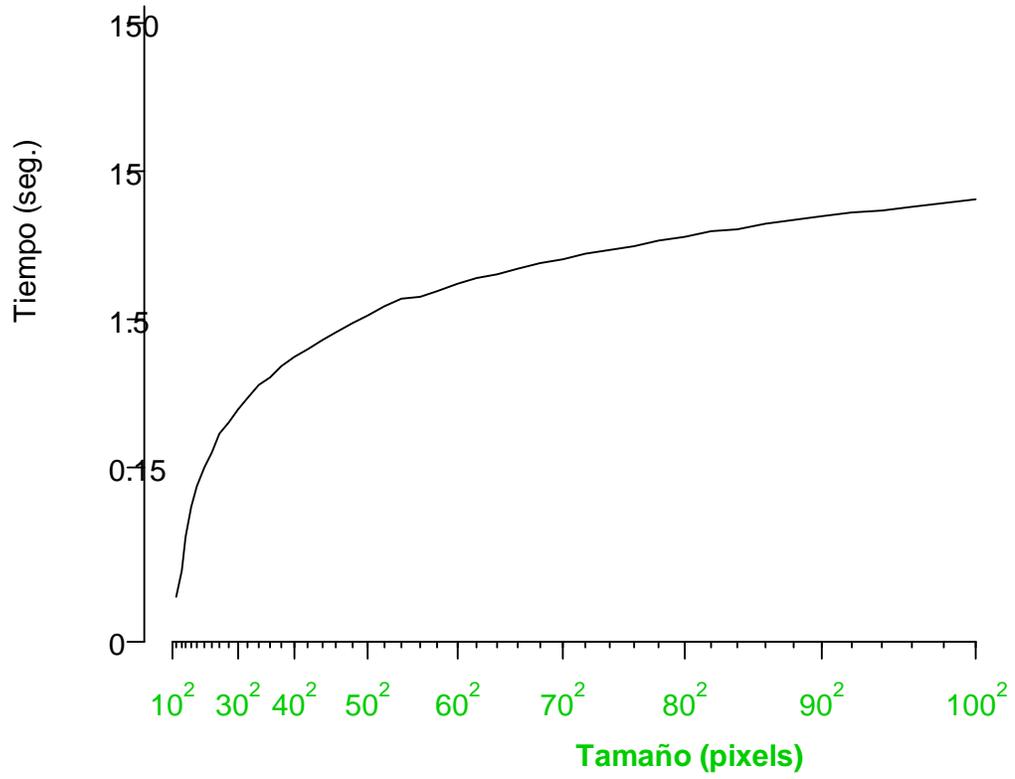
Aquí se tiene que la complejidad computacional está determinada por la manera en que evoluciona el tamaño de la vecindad $V^{t+1}(P_k \cup P_l)$, es decir, el valor $\#^{t+1}(P_k \cup P_l)$ a lo largo de las etapas de construcción del dendrograma.

El tamaño de la vecindad puede aumentar producto de las fusiones en las primeras etapas (y posiblemente a tomar el valor n), para después ir disminuyendo hasta cero cuando se hayan fundido todas las clases en una, lo que llevaría en el peor caso a una complejidad temporal $O(n^2)$. Aunque este análisis es correcto quizás pueda ser refinado mediante un análisis amortizado [178].

A manera de ilustración se incluye a continuación un gráfico que muestra el comportamiento de los tiempos de ejecución de la implementación realizada del algoritmo para diferentes tamaños de imágenes y diferentes valores de radios de vecindad (ver Figura. 5-4).

En el gráfico se incluyen las rectas de regresión lineal. Las líneas de regresión lineal del tiempo en función del tamaño de la imagen de manera describen de manera aproximada el comportamiento. Sin embargo, se debe notar que tanto al inicio de los datos como al final las líneas de regresión dejan los puntos por encima.





La imagen original está compuesta de dos patrones similares de píxeles ubicados en un fondo negro (ver Figura 5-6a). En las Figuras 5-6b – 5-6e se presentan los resultados que se obtienen al aplicar el método de enlace promedio clásico. La partición de dos clusters define los dos patrones similares como una clase y el fondo como otra. Por su parte la partición de cuatro clusters recupera la imagen original. En la partición de tres clusters el fondo sigue formando una clase, pero los patrones quedan subdivididos.

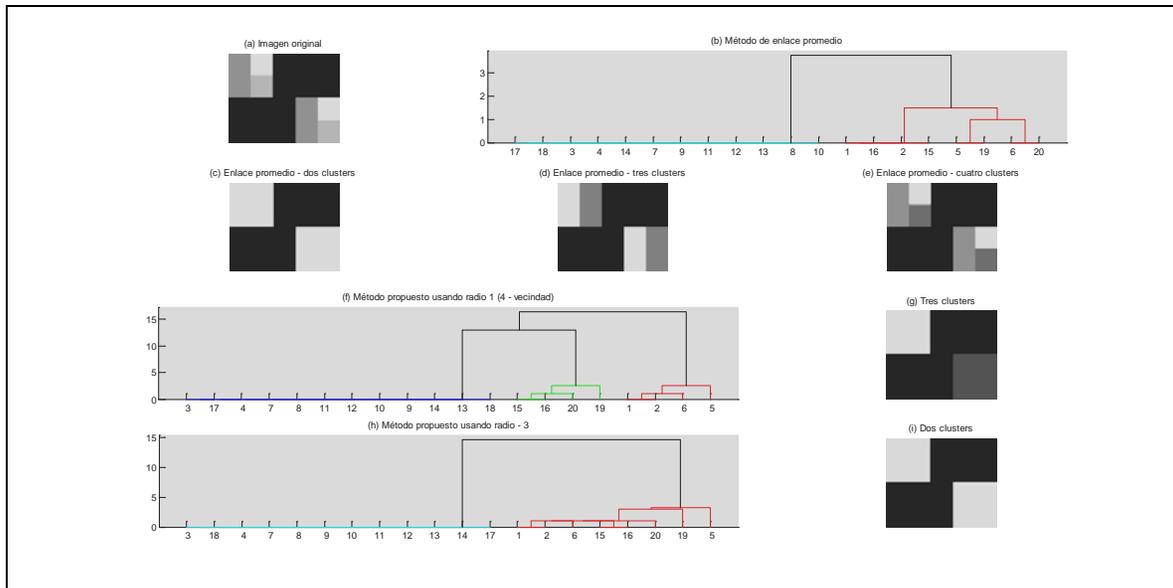


Figura 5-6 Ejemplo artificial: (a) Imagen original con dos patrones similares; (b) dendrograma obtenido con el método de enlace promedio; (c)-(d) particiones de dos y tres clusters obtenidas con el método de enlace promedio; (e)-(f) una partición de tres clusters obtenida al utilizar la 4 – vecindad; (g)-(h) una partición de tres clusters obtenida al utilizar una vecindad de radio 3.

Las Figuras 5-6f – 5-6g muestran los resultados obtenidos al aplicar el método propuesto utilizando dos tipos de vecindad: radio de un píxel (4 – vecindad), y una vecindad con radio de 3 píxeles.

Al utilizar la 4 – vecindad, se revela claramente la partición de tres clusters que se obtiene al realizar un corte del dendrograma a un nivel 10 (ver Figura 5-6g). Esta partición contiene un cluster para el fondo, y dos clusters no contiguos, una para cada uno de los patrones.

Los resultados de las Figuras 5-6h y 5-6i se obtuvieron al utilizar un radio de vecindad igual a 3 píxeles. En este caso los dos patrones similares se revelan como un único cluster. El incremento del radio de la vecindad permite establecer un enlace entre los píxeles contenidos en estos patrones. Este resultado ilustra el efecto que puede tener el radio de la vecindad utilizada en los resultados de la segmentación.

La ausencia de un enlace (una arista en el grafo de contigüidad) entre dos píxeles es una restricción dura, en el sentido de que ellos no se pueden fusionar a menos que se fusionen

todos los píxeles intermedios entre ellos. Así, por ejemplo, en la Figura **5-6g** los dos clusters asociados a los patrones semejantes, que fueron obtenidos utilizando la 4 – vecindad, no pueden ser fusionados sin incluir el fondo. Es por esa razón que se requiere una mayor cantidad de clusters para que la solución sea comparable con las obtenidas en las Figuras **5-6c** y **5-6i**.

Finalmente, se debe señalar que los tiempos de procesamiento al utilizar la 4 – vecindad (0.000264 segundos) y al utilizar una vecindad de radio 3 (0.000312 segundos) son muy inferiores al tiempo requerido por el método de enlace promedio (1.9608 segundos) para la misma tarea sobre la imagen del ejemplo. Se debe acotar aquí que se utilizó la implementación del enlace promedio de Matlab, y el ejemplo fue ejecutado en una máquina con procesador Pentium-D 820 a 2.8 GHz y 2 GB de memoria.

5.10 Conclusiones

En este capítulo se ha presentado el estudio teórico de un método de agrupamiento jerárquico aglomerativo con restricciones de contigüidad. Las restricciones de contigüidad se modelan mediante una matriz o grafo de contigüidad.

Se introduce una función de desemejanza que toma en cuenta las relaciones de contigüidad impuestas sobre los datos y se analizan sus propiedades. Se demuestra que la jerarquía de clusters que se obtiene con el método propuesto es libre de inversiones.

Se proponen varias alternativas de implementación y se realiza un análisis de la complejidad computacional. El análisis de los tiempos de ejecución al procesar imágenes permite plantear, de manera empírica, que el comportamiento del método propuesto es sub-cuadrático.

Mediante un ejemplo en el campo de la segmentación de imágenes se ilustran las principales características del método propuesto.

6 Experimentación

Con el objetivo de analizar el comportamiento del algoritmo de agrupamiento aglomerativo jerárquico de enlace tipo promedio con restricciones de contigüidad se ha seleccionado como área de aplicación la segmentación de imágenes médicas, específicamente las imágenes del cerebro. Los objetos de interés a clasificar dentro de una imagen lo constituyen los píxeles. Por su parte, los clusters o clases que se obtienen deben corresponderse con regiones de interés dentro de la imagen que pueden ser, por ejemplo, determinados órganos, o tipos de tejidos.

Un elemento a favor de este tipo de aplicación está dado por el hecho de que las relaciones de vecindad pueden ser especificadas directamente teniendo en cuenta la disposición espacial de los píxeles dentro de la imagen. Si consideramos la distancia euclidiana en el plano una imagen bidimensional, una definición razonable de contigüidad lo constituye el conjunto de los cuatro vecinos más próximos (4-vecindad) situados al norte, sur este y oeste, cuya distancia respecto a un píxel dado es la unidad. De igual manera se pudieran considerar los ocho vecinos más próximos (8-adyacencia).

Esta definición se puede generalizar a una vecindad circular alrededor de un píxel dado la que está compuesta por el conjunto de píxeles que se encuentran a una distancia menor o igual que un radio r dado. Con esta definición la vecindad de cada píxel, si se utiliza la distancia euclidiana, se corresponde con una región circular de radio r . En caso de utilizar la distancia de Chebyshev (ver Apartado 2.1.2 del Capítulo 2), la vecindad se corresponde con una región cuadrada de lado $2*r+1$ con centro en el píxel dado. Este radio o tamaño de la vecindad se mantiene constante en toda la imagen (excepto en los bordes).

A continuación se describen los principales retos asociados a la segmentación de imágenes de resonancia magnética. También se exponen las alternativas usadas para la obtención y evaluación de los resultados de la segmentación.

Los experimentos fueron ejecutados en una máquina con procesador Pentium-D 820 a 2.8 GHz y 2 GB de memoria.

6.1 *Las imágenes de resonancia magnética*

La modalidad de resonancia magnética nuclear es una tecnología no invasiva e inocua de adquisición de imágenes. Félix Bloch y Edward Purcell descubren el fenómeno de la resonancia magnética en el año 1946 de manera independiente, y por ese descubrimiento les fue concedido el premio Nóbel en el año 1952.

El fenómeno de resonancia magnética se basa en una propiedad que poseen las partículas elementales, en particular el hidrógeno, denominada espín, la cual provoca que se produzca una señal de resonancia magnética nuclear. En esencia este fenómeno consiste en que, en presencia de un campo magnético intenso, los núcleos atómicos pueden absorber energía de manera selectiva en forma de radiación electromagnética de radio – frecuencia (fenómeno de resonancia), la cual devuelven al retornar al estado de equilibrio (fenómeno de relajación), generando una señal eléctrica en una antena receptora, que una vez analizada y procesada proporciona las imágenes de resonancia magnética.

La descripción cuantitativa de la señal que se obtiene, conocida como señal de inducción libre, contiene información sobre tres parámetros importantes cada uno de los cuales refleja distintas propiedades de la materia: la densidad protónica (DP), y las constantes de tiempo de relajación T1 y T2. Así, por ejemplo, en un campo magnético de 1.5T (Tesla – unidad de inducción magnética), la sustancia gris posee un tiempo T1 \approx 900 ms, para la sustancia blanca T1 \approx 700 ms, mientras que para el líquido cefalorraquídeo T1 \approx 4000 ms. Por su parte, las constantes de tiempo T2 son aproximadamente de 70 ms, 90 ms y 400 ms para la sustancia blanca, gris y líquido cefalorraquídeo, respectivamente [179].

Teniendo en cuenta los elementos anteriores, y mediante un ajuste apropiado de los parámetros de los tomógrafos de resonancia magnética, es posible obtener imágenes donde los diferentes tipos de tejidos aparecen contrastados. Así, las imágenes más comunes son las imágenes de ponderación T1, T2 y de densidad protónica.

6.2 El problema de la segmentación de imágenes de resonancia magnética

Como se ha dicho antes, la segmentación de imágenes puede ser planteada como el problema de obtener una partición en regiones homogéneas en cierto sentido. Cada región puede consistir en un único componente conexo, o quizás en un grupo de componentes conexos, con propiedades similares.

La segmentación de imágenes de resonancia magnética, específicamente, consiste en la clasificación de los vóxeles en regiones que se correspondan con los diferentes tipos de tejidos. Se asume en este caso que cada clase de tejido posee un valor específico de la magnitud medida. Por ejemplo, esta magnitud puede estar asociada a la intensidad de la señal, correspondiente con el tipo de tejido, que es medida en las imágenes de ponderación T1, T2 o densidad protónica. En este caso ideal la segmentación de la imagen consistiría en un conjunto de regiones con valores constantes de la magnitud medida según fuera el tipo de tejido.

Aunque este enfoque resulta muy simplificado, toda vez que aún dentro de un mismo tipo de tejido existen variaciones asociadas a su estructura biológica, aquí se va a asumir que las variaciones de origen biológico dentro de cada tipo de tejido no son demasiado influyentes

en proceso de segmentación y no afectan de manera sensible la capacidad de discriminar los diferentes tipos de tejidos.

Otro problema que surge en la práctica está asociado a la limitada resolución espacial de los tomógrafos lo que genera un emborronamiento o desenfoque a lo largo de los bordes entre las regiones, así como en los cortes extremos de los volúmenes. Este fenómeno, conocido como efecto de volumen parcial, se origina por el hecho de que un vóxel dado puede contener una mezcla de tejidos por lo que intensidad de la señal que se mide estaría afectada por cantidad y las características individuales de cada tipo de tejido presente y la intensidad final se obtiene como una media ponderada. Por ejemplo, la corteza cerebral es una capa irregular con un espesor casi constante. Algunas mediciones experimentales han estimado que su espesor se encuentra en 1.5 y 5.5 mm [180]. Los surcos de la corteza, que normalmente están llenos de líquido cefalorraquídeo, poseen dimensiones comparables con la resolución espacial de los sistemas de adquisición y esto provoca un efecto de volumen parcial que reduce el contraste de intensidad y dificulta la obtención de resultados exactos.

Aunque el artefacto del volumen parcial es un factor que debe ser tenido en cuenta al valorar la segmentación, este efecto está confinado solamente a los bordes de los tejidos o a los cortes extremos a diferencia de otros fenómenos que son más globales.

En este sentido el artefacto más significativo desde el punto de vista de la segmentación es el que se conoce como heterogeneidades de intensidad. Este fenómeno está asociado a la no homogeneidad del campo magnético utilizado, aunque también puede ser provocado por interacciones del sistema de adquisición con el sujeto. La manifestación de este artefacto consiste en las variaciones de la intensidad de la señal para un mismo tipo de tejido en diferentes zonas de la imagen. Este efecto combinado puede originar variaciones de intensidad del orden del 10-20% de las intensidades para un campo magnético de 1.5T.

Aunque estas variaciones por lo general no afectan la precisión de los diagnósticos, sí pueden ser relevantes para los métodos de segmentación que se basan en las distribuciones de intensidad de los vóxeles de una imagen. Cuando estas variaciones son significativas comparadas con el contraste de la imagen pueden muy bien afectar la interpretación de una imagen de resonancia magnética.

Por último, todas las imágenes de resonancia magnética están afectadas por ruidos aleatorios. Este ruido proviene de diferentes tipos corrientes que se generan en el sistema de adquisición. Cuando el nivel ruido es significativo en una imagen de resonancia magnética esta puede resultar oscurecida y de apariencia granulada, y los tejidos con contrastes similares pudieran no ser delineados de una manera efectiva provocando errores en la segmentación.

6.3 Fuentes para la evaluación de los métodos de segmentación

Con vistas a la evaluación de un algoritmo de segmentación se requiere de algún criterio objetivo. Entre los criterios más utilizados se encuentran la eficiencia y la exactitud. La eficiencia se refiere a la viabilidad práctica del proceso y está asociada al tiempo de ejecución que puede incluir, además del tiempo computacional, el tiempo requerido por un operador humano en completar el proceso de segmentación.

La exactitud, por su parte, tiene que ver con el grado de concordancia de los resultados de la segmentación con la verdad, o con cierto sustituto de ésta. En este sentido se utiliza generalmente cierta segmentación de referencia (*gold standard*).

Las fuentes para segmentaciones de referencia son variadas e incluyen modelos físicos, cadáveres y pacientes vivos, y modelos computacionales. Cada uno de ellos tiene sus puntos fuertes y débiles.

Los modelos físicos presentan la dificultad asociada a la ausencia de una variabilidad anatómica realista, aunque sí poseen características reales del proceso de adquisición. Los cadáveres pueden brindar modelos muy reales, aunque se debe estimar la segmentación por medio de una autopsia [181, 182]. Esta posibilidad no siempre está disponible.

Las imágenes de pacientes vivos, además de ser las más abundantes, son las más reales en todos los sentidos. Sin embargo no existe una vía práctica para determinar la segmentación real de los tejidos. En su lugar se acude a ciertos sustitutos. En particular es común utilizar una combinación de segmentaciones manuales realizadas por expertos con el fin de obtener una segmentación de referencia [183, 184].

Con los modelos computacionales es posible conocer de antemano la segmentación exacta, y aunque las imágenes que se obtienen tienen a ser menos reales en el sentido de la variabilidad anatómica y de las características del sistema de adquisición, constituyen un medio útil para la evaluación por la posibilidad de controlar algunos parámetros del proceso de generación de las imágenes.

6.4 Índices de calidad de las segmentaciones

Una vez que se aplique el método de clasificación automática, descrito en el capítulo anterior, se obtiene una jerarquía o dendrograma. Mediante el corte del dendrograma a un nivel dado se obtiene una partición que coincide con una posible segmentación de la imagen.

En este sentido los métodos de clustering representan una herramienta para el análisis exploratorio de los datos. Sin embargo, el usuario debe realizar la validación de los resultados obtenidos, es decir, debe determinar la cantidad “correcta” de grupos, así como

su interpretación. La noción de validación de los clusters se refiere a los conceptos y métodos utilizados para la valoración cuantitativa y objetiva de los resultados de un algoritmo de clustering [6, 15]. Esta problemática se considera un problema abierto de gran significación pues de la decisión que se tome dependen los resultados que se obtienen.

Teniendo en cuenta la aplicación descrita nos centraremos aquí en el caso de la validación de particiones. Se pueden plantear dos escenarios. En el primero de ellos se asume que se dispone de información adicional, como puede ser una partición (segmentación) de referencia, y es posible entonces valorar hasta qué punto los resultados de la clasificación brindan información válida acerca de los datos y en qué medida la estructura recuperada refleja el carácter de éstos. En este caso es posible realizar la comparación de las posibles particiones que se derivan de la jerarquía obtenida con la partición de referencia y escoger aquélla que mejor se ajusta, valorando la calidad de ese ajuste.

El segundo escenario es aquél en el que no se dispone de información adicional. Aquí es necesario disponer de alguna medida de la calidad de una partición dada y la cantidad correcta de clusters se determinaría a partir de la “mejor” partición según la medida utilizada.

Se han propuesto diversos métodos para la determinación de la cantidad de clusters [185]. Una reseña de estos métodos la brinda Gordon en [186, 187], donde clasifica los métodos en globales y locales. Los primeros evalúan cierta medida sobre todo el conjunto de datos y la optimizan respecto a la cantidad de clusters. Los segundos consideran pares individuales de clusters y analizan si deben ser mezclados o no.

Por su parte Jain y Dubes clasifican los índices de validación de particiones en índices internos, externos y relativos [15]. Los índices externos permiten evaluar en qué medida dos particiones de un conjunto de objetos concuerdan. Una de las particiones proviene del algoritmo utilizado, por ejemplo, al cortar el dendrograma generado por un algoritmo de jerárquico, mientras la segunda es asignada a priori. Este tipo de índices puede ser útil para el primer escenario descrito con anterioridad.

Los índices internos evalúan la correspondencia entre la partición obtenida por un algoritmo y los datos y pueden tener en cuenta cuan probable es la ocurrencia de un valor dado del índice. Este último objetivo se puede lograr, por ejemplo, comparando la distribución del índice obtenida con la esperada (ver, por ejemplo, [188]).

Por último, los índices relativos responden a la pregunta: ¿cuál de los agrupamientos obtenidos se ajusta mejor a los datos? En principio cualquiera de los índices internos puede servir para este objetivo. Un índice relativo se diferencia por la manera en que se aplica. Por ejemplo, para una secuencia de agrupamientos, obtenidos a partir de una jerarquía, se obtiene una secuencia de valores, una para cada agrupamiento. En el gráfico resultante se busca algún aspecto significativo como puede ser el valor máximo, mínimo, o algún salto abrupto [15, 185, 186].

La mayoría de los índices que se utilizan para comparar particiones pueden ser descritos utilizando tablas de contingencia. Si se tienen dos particiones P y P' , no necesariamente de la misma cardinalidad, es posible analizar cada par de instancias de objetos respecto a su clasificación en la partición P y en la partición P' . Aquí son posibles cuatros casos:

1. Los dos objetos pertenecen a la misma clase en ambas particiones.
2. Los dos objetos están juntos en P , pero separados en P' .
3. Los dos objetos están juntos en P' , pero separados en P .
4. Los dos objetos pertenecen a clases diferentes en ambas particiones.

Denotando por a , b , c y d la cantidad de pares que cumplen las condiciones 1), 2), 3) y 4), respectivamente, se puede construir la siguiente tabla.

		P'	
		Agrupados	Desagrupados
P	Agrupados	a	b
	Desagrupados	c	d

Entre los índices más usados que se definen a partir de esta tabla se encuentra el índice de Rand (y su versión ajustada) [189, 190], y el índice de Jaccard [15]. El índice de Rand se obtiene mediante la expresión $(a+d)/(a+b+c+d)$, mientras el índice de Jaccard se calcula según $a/(a+b+c)$.

La versión normalizada o ajustada del índice de Rand hace que el índice tome valores entre 0 y 1 [190]. Cuando ambas particiones son idénticas el índice toma el valor 1. Por su parte, valores del índice cercanos a 0 apuntan a una baja correspondencia entre las particiones. Lo mismo es válido para el índice de Jaccard.

Para el caso de los índices internos hemos seleccionado el índice de Calinski y Harabasz [191, 192] y el índice de Hartigan [8, 192] que están relacionado por el hecho de utilizar las matrices de dispersión.

Según el estudio realizado por Milligan y Cooper [185] (ver también [188]) de los 30 índices comparados el de mejor desempeño resultó ser el de Calinski y Harabasz.

Si se tiene una partición con k clusters el índice de Calinski y Harabasz (CH) se calcula de la siguiente manera [191]

$$CH(k) = \frac{tr(S_B)/(k-1)}{tr(S_W)/(n-k)}$$

Aquí S_W y S_B representan las matrices de dispersión interna y externa, respectivamente, tal y como fueron definidas en el apartado 2.2.1, es decir

$$S_T = S_W + S_B$$

donde

$$S_T = \sum_{i=1}^n (x_i - \bar{x})^T (x_i - \bar{x})$$

$$S_W = \sum_{k=1}^K \sum_{i \in P_k} (x_i - \bar{x}_k)^T (x_i - \bar{x}_k)$$

$$S_B = \sum_{k=1}^K n_k (\bar{x}_k - \bar{x})^T (\bar{x}_k - \bar{x})$$

Para una secuencia de particiones (con diferentes cantidades de clusters) se puede obtener una secuencia de valores del índice. La cantidad óptima de clusters se define como aquella que maximiza las segundas diferencias de $CH(k)$ [193]. De esta manera se puede determinar el corte óptimo del dendrograma (en el sentido del índice CH).

El valor mínimo de las segundas diferencias se obtiene a partir de la siguiente expresión

$$\min_k \left\{ (CH(k+1) - CH(k)) - (CH(k) - CH(k-1)) \right\} = \\ \min_k \left\{ CH(k+1) + CH(k-1) - 2CH(k) \right\}$$

y mide la ocurrencia de “codos” (*elbows*) en la secuencia de valores.

Por su parte el índice de Hartigan se calcula de la siguiente manera

$$H(k) = \log \left(\frac{tr(S_B)}{tr(S_W)} \right)$$

Este índice resulta parecido al de Calinski-Harabsz excepto por el efecto de la ponderación. Para una secuencia de valores del índice, asociados a diferentes particiones, la cantidad óptima de clusters se define como aquella que maximiza las segundas diferencias de $H(k)$ [193].

La razón $tr(S_B)/tr(S_W)$ se puede ver de la siguiente manera

$$\frac{tr(S_B)}{tr(S_W)} = \frac{tr(S_B)/tr(S_T)}{1 - tr(S_B)/tr(S_T)}$$

Esto apunta al hecho de que valores grandes de $tr(S_B)/tr(S_W)$ se corresponden con clusters compactos, o de manera equivalente, $tr(S_B)/tr(S_T)$ tiende a 1. Esta relación en forma de porcentajes hace más fácil interpretación.

En este capítulo se propone utilizar también en calidad de índice relativo el índice C_B definido en el capítulo 4. Su sentido intuitivo es el siguiente. Debido a la relación que se establece entre los píxeles fronterizos de los clusters, su valor caracteriza la autocorrelación espacial de las fronteras. El índice tiene en cuenta además las relaciones de contigüidad espacial impuestas a los datos.

Este índice toma valores en el rango de 0 a 2. Un valor cercano a 2 se interpreta como una frontera con autocorrelación espacial negativa, es decir, fronteras entre clusters bien definidas. Por otra parte, un valor cercano a 0 indicaría fronteras con autocorrelación espacial positiva, es decir, píxeles semejantes a ambos lados de la frontera y por tanto ausencia de una clara separación entre clusters.

Los valores de este índice tienden a decrecer (aunque no de manera monótona) a medida que aumenta la cantidad de clusters. Se propone entonces utilizar el índice C_B como un índice relativo en el sentido de [15].

Más específicamente, para una secuencia de particiones obtenidas a partir de cortes en el dendrograma la cantidad óptima de clusters se define como aquella que minimiza las segundas diferencias de los valores del índice. Esto coincide con la ocurrencia de “codos” significativos en la secuencia. Más aún, se propone también considerar como particiones de interés aquellas que se correspondan con mínimos locales de las segundas diferencias de los valores de índice C_B . Estas ideas son analizadas en los experimentos realizados.

6.5 Sobre las bases de datos de imágenes

Para realizar la experimentación se seleccionaron imágenes de dos fuentes: la base de imágenes simuladas del cerebro *BrainWeb* [194, 195], disponible en <http://www.bic.mni.mcgill.ca/brainweb/>, y la base de datos IBSR del Centro para el Análisis Morfométrico del Hospital General de Massachussets (*Center for Morphometric Analysis at Massachussets General Hospital*), que puede ser obtenida en <http://www.cma.mgh.harvard.edu/ibsr/>

6.5.1 Base de datos de imágenes simuladas (BrainWeb)

Para la obtención de resultados bajo condiciones controladas se utilizó el simulador *BrainWeb* [196], accesible a través de una interfaz que permite seleccionar los parámetros de adquisición de las imágenes. Se han seleccionado imágenes de ponderación T1 de sujetos normales simuladas con un 3% de ruido, y valores de heterogeneidades de intensidad del 20% y una resolución de 1 mm³, es decir, 1 mm de espesor de corte y un 1 mm de resolución en ambos planos X e Y. Estos valores son los más comúnmente utilizados en la literatura [197-200]. En particular el valor de 3% de ruido es considerado un valor típico [200]. Por su parte, el valor de 20% de heterogeneidad de intensidad se considera el valor mayormente encontrado en imágenes de ponderación T1 [199]. Ambos representan los valores por defecto de *BrainWeb* y serán los utilizados, pues facilitan la comparación de los resultados.

Unido a las imágenes también se brinda un modelo anatómico digital del cerebro obtenido a partir de 27 estudios de alta resolución, convenientemente procesados y clasificados, realizados a un mismo individuo. Este modelo promediado muestra detalles de las estructuras principales del cerebro que normalmente son enmascaradas por el ruido en las imágenes [201].

Este modelo consiste en una partición de todo el volumen de la cabeza en 10 grupos de tejidos (líquido cefalorraquídeo, sustancia gris, sustancia blanca, grasa, músculo/piel, piel, cráneo, sustancia neuroglial (interfaz entre el líquido cefalorraquídeo y la sustancia blanca, con una intensidad que se encuentra entre la de la sustancia gris y la sustancia blanca), tejido conectivo, y fondo. El volumen intracraneal está compuesto por el líquido cefalorraquídeo, la sustancia gris, la sustancia blanca y la sustancia neuroglial.

Entre las aplicaciones de este modelo se encuentra la generación de imágenes de resonancia magnética simuladas. En el proceso de generación de imágenes simuladas se toma como base el modelo anatómico para la obtención de las intensidades de cada tipo de tejido. Posteriormente el simulador toma en cuenta el efecto de diferentes parámetros que caracterizan el proceso de adquisición de las imágenes incluyendo artefactos tales como el volumen parcial, el ruido y las heterogeneidades de intensidad [201].

De esta manera es posible evaluar un determinado algoritmo de clasificación respecto a los parámetros que brinda el simulador. También es posible realizar la comparación de diferentes algoritmos sobre el mismo conjunto de imágenes simuladas, evaluando sus resultados con respecto al modelo anatómico de referencia.

6.5.2 Base de datos de imágenes reales (IBSR)

Por su parte la base de datos IBSR (*Internet Brain Segmentation Repository*) del Centro para el Análisis Morfométrico (CAM) del Hospital General de Massachussets (*Center for*

Morphometric Analysis at Massachussets General Hospital) brinda imágenes de prueba con las correspondientes segmentaciones manuales.

Se ha seleccionado un conjunto de datos que incluye imágenes volumétricas de 20 individuos normales. Las segmentaciones manuales fueron realizadas por investigadores entrenados de CAM con la ayuda de un método semiautomático de trazado de contornos a partir de los valores de intensidad. Aunque las segmentaciones no pueden ser consideradas como la “verdad absoluta” (*ground truth*), sí brindan una vía razonable para la comparación de los métodos de segmentación. Los 20 estudios fueron seleccionados por IBSR ya que cubren un amplio espectro de calidades de imágenes.

Las imágenes de esta base de datos permiten analizar el comportamiento de uno o varios métodos de clasificación sobre las mismas imágenes reales, comparando los resultados con la segmentación realizada por expertos. IBSR también brinda los resultados de la segmentación de los 20 estudios usando 6 métodos automatizados: *adaptive MAP*, *biased MAP*, *fuzzy c-means*, *Maximum A Posteriori (MAP)*, *tree-structure k-means*, así como los resultados promediados de 2 expertos sobre las imágenes de 4 sujetos del conjunto [202]. Los resultados se reportan respecto a la detección de la sustancia gris y la sustancia blanca. Todos los métodos automatizados utilizaron las imágenes que contienen únicamente el tejido cerebral, es decir, después de haberse eliminado el cráneo, la piel y grasa, externas al cerebro.

6.5.3 Comentarios generales

Para ambas fuentes de imágenes se utilizó el formato *Analyze 7.5*, desarrollado por *The Biomedical Imaging Resource at the Mayo Foundation*, que puede ser procesado por diferentes sistemas. El almacenamiento de las imágenes según este formato utiliza dos archivos: un archivo de encabezamiento con extensión *.hdr*, y un archivo con los píxeles de la imagen cuya extensión es *.img*. En el archivo de encabezamiento se almacena la historia y las dimensiones de los datos. Aquí se incluye información acerca de las dimensiones de la imagen y de sus píxeles, así como el tipo de datos utilizado para almacenar los valores de la imagen, entre otros. El archivo de la imagen es sencillo y contiene los valores de los píxeles según el tipo de datos especificado. Más detalles pueden ser consultados en <http://www.mayo.edu/bir/PDF/ANALYZE75.pdf>.

Desde el punto de vista de sus características, existen analogías y diferencias entre las dos fuentes de imágenes:

- En ambos casos se tienen imágenes del cerebro.
- La resolución espacial de las imágenes de *BrainWeb* es de 1 mm^3 , mientras que en el conjunto de datos IBSR 3 mm^3 .

- Para ambos casos se dispone de una partición de referencia con la cual comparar los resultados. En un caso la referencia es un modelo anatómico (*BrainWeb*), y en el otro una segmentación realizada por expertos.
- En las imágenes de *BrainWeb* se puede controlar el nivel de ruido y de heterogeneidades de intensidad que caracterizan el proceso de adquisición lo que permitiría analizar el comportamiento del método propuesto bajo diferentes condiciones.
- Para el conjunto de datos IBSR los niveles de ruido y de heterogeneidades de intensidad no son dados. En este caso el análisis se enfocaría a la correspondencia de los resultados que brinda el método propuesto con el que se tiene de los expertos.

Por último, se debe aclarar que en todos los experimentos no se realizó ningún preprocesamiento de las imágenes utilizadas. Por otra parte, para medir la desemejanza entre píxeles se utilizaron las diferencias cuadráticas de los niveles de intensidad, es decir, la distancia euclidiana en el espacio de rasgos.

Tanto el preprocesamiento de las imágenes, como la utilización de otros rasgos posiblemente derivados de los valores de intensidad, pueden tener su influencia en el proceso de segmentación. Sin embargo, estas problemáticas no están incluidas en los objetivos de los análisis que siguen.

6.6 Experimentos con imágenes simuladas de *BrainWeb*. Resultados

A continuación se describe el primer experimento realizado en el campo de la segmentación de imágenes médicas y se discuten los resultados.

6.6.1 Objetivo

En este experimento se utilizaron las imágenes de ponderación T1 simuladas de sujetos normales, obtenidas en *BrainWeb* tal y como se describió antes, con el objetivo particular de evaluar el efecto del tamaño del radio de la vecindad en la calidad de la segmentación que se obtiene con la aplicación del método propuesto.

6.6.2 Requisitos para abordar el objetivo

De las imágenes disponibles se escogió el corte central del volumen simulado (imágenes T1, vóxeles de 1 mm^3 , 3% de ruido y un valor del 20% de heterogeneidades de intensidad). La razón que motivó esta selección está dada por el hecho que en este corte aparecen los tejidos fundamentales de interés en nuestro caso, es decir, la sustancia blanca, la sustancia gris y el líquido cefalorraquídeo.

La razón en centrar la atención en sólo tres tejidos está dado por el hecho de ser estos los que predominan en la composición del cerebro. En ciertas aplicaciones médicas es de interés conocer los volúmenes de estos tejidos con vistas a estimar los daños provocados por una enfermedad. Este es el caso, por ejemplo, de la estimación de la atrofia cerebral. Este fenómeno aparece en varias enfermedades del sistema nervioso central tales como la esclerosis múltiple, la enfermedad de Alzheimer, demencia, entre otras, y provoca pérdida de masa encefálica (sustancia blanca y gris) y aumento de líquido cefalorraquídeo, alterando de esta manera la proporción de líquido cefalorraquídeo respecto al volumen intracraneal [203-206].

En la Figura 6-1 se muestra el corte seleccionado, (a), así como el modelo anatómico que sirve de base en el proceso de generación de las imágenes (b), y que es utilizado como segmentación de referencia para evaluar los resultados del método aplicado.

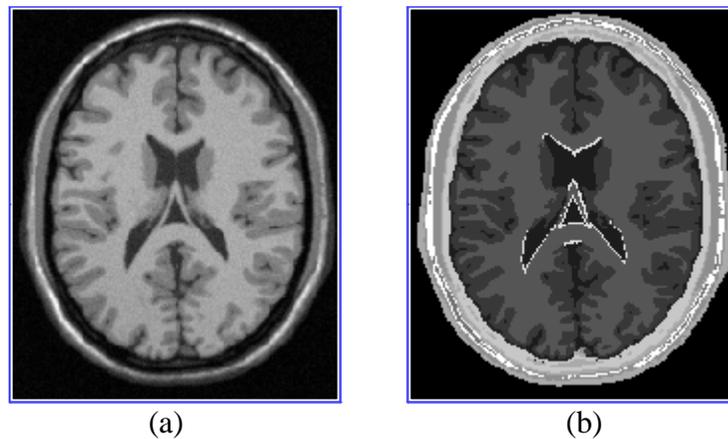


Figura 6-1 Corte central simulado obtenido de BrainWeb con 3% de ruido y 20% de heterogeneidades de intensidad (a), y el modelo anatómico correspondiente (b).

6.6.3 Elección de la región de interés

Dentro del corte seleccionado se trabajó en una región de 75x70 píxeles (ver Figura. 5-2). La razón de haber seleccionado esta región se basa en el hecho de que en un ambiente clínico normalmente el especialista decide la región que analiza. Además de razones computacionales, esta selección evita el hecho de involucrarse con el problema de la identificación y separación de los tejidos externos al cerebro, tales como el cráneo, la piel, grasa, etc. Por otra parte, en esta región aparecen los tejidos principales del cerebro.

En la Figura. 5-3 se muestra la región de interés ampliada tanto en el corte central seleccionado (a), como en el modelo anatómico (b).

En el modelo anatómico las tonalidades se corresponden con los tejidos de interés: negro – líquido cefalorraquídeo, gris oscuro – sustancia gris, gris claro – sustancia blanca, y blanco intenso – sustancia neuroglial.

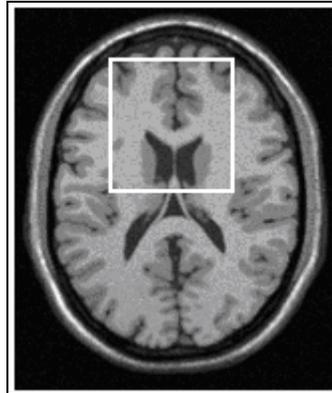


Figura 6-2 Región de interés de dimensión 75x70 píxeles en el corte seleccionado.

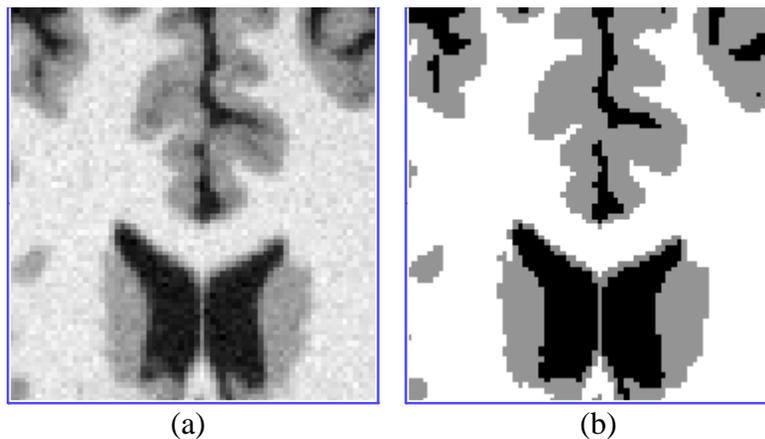


Figura 6-3 Región de interés ampliada en el corte seleccionado (a), y en el modelo anatómico (b).

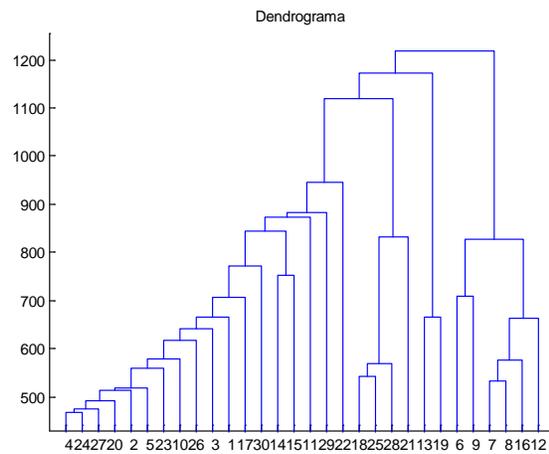
6.6.4 Selección del radio de la vecindad

El método propuesto toma en cuenta la influencia de los objetos vecinos a la hora de realizar la clasificación. Esta influencia se especifica mediante una relación de vecindad entre los objetos analizados.

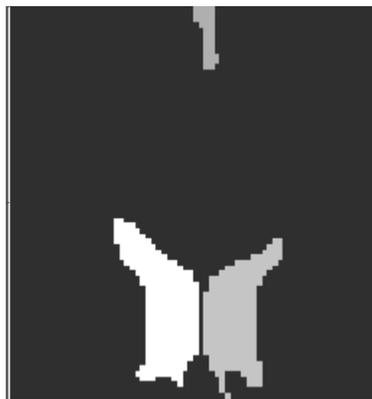
En el problema que nos ocupa, la segmentación de una imagen médica, el objetivo principal consiste en detectar regiones homogéneas en cuanto a los valores de intensidad que, es de esperar, se correspondan con zonas ocupadas por determinados tipos de tejidos.

Cuando una persona examina una imagen pudiéramos pensar que realiza un barrido visual, detectando estas zonas homogéneas de interés como zonas de baja variabilidad en los niveles de intensidad. El nivel de detalle que se aprecia depende en gran medida de la “escala” que se utiliza. Cuando la imagen se observa de cerca se pueden detectar pequeños detalles. A medida que la distancia del observador a la imagen es mayor los pequeños detalles dejan de ser perceptibles y comienzan a destacarse los patrones más generales. Así, por ejemplo, si nos enfocamos en pequeñas zonas de un paisaje podemos identificar un árbol con sus hojas y frutos. Por otra parte, si nos alejamos apreciaríamos quizás un bosque, un fondo de montañas, el cielo, etc.

En las Figuras 5-4 y 5-5 se muestran ejemplos de éste fenómeno para el caso de la imagen de la Figura 6-3a.



(a)



(b)



(c)

Figura 6-4 Dendrograma obtenido al segmentar la región de interés de la Fig. 3 utilizando un radio = $\sqrt{2}$ (8-vecindad) (a); partición de 4 clusters (b); y de 16 clusters (c).

En la Figura 6-4 se muestra el dendrograma obtenido al aplicar el método propuesto utilizando un radio = $\sqrt{2}$ (8-vecindad). Las Fig. 5-4b y 5-4c muestran dos particiones de 4 y 16 clusters, respectivamente, obtenidas al cortar al dendrograma. Los colores han sido utilizados únicamente con la finalidad de resaltar los diferentes clusters de cada partición.

Como se puede apreciar, el uso de una vecindad pequeña hace que se vayan detectando primero los pequeños detalles. Sin embargo, la correspondencia con el modelo no es tan elevada como lo indica, por ejemplo, el índice ajustado de Rand: 0.22 y 0.31 para la partición de 4 y de 16 clusters, respectivamente.

La Figura 6-5a muestra el dendrograma obtenido al utilizar una vecindad con radio 20 píxeles. La utilización de este valor aquí es arbitraria y sirve como elemento de ilustración. Una propuesta para la selección del radio será comentada más adelante. A diferencia del caso anterior, aquí se aprecia cómo los detalles más generales son detectados con la ayuda de una menor cantidad de clusters. Visualmente se aprecia una correspondencia entre los resultados de la Figura 6-5b y los de la Figura 6-3b, lo que es corroborado por el índice ajustado de Rand que toma el valor 0.88.

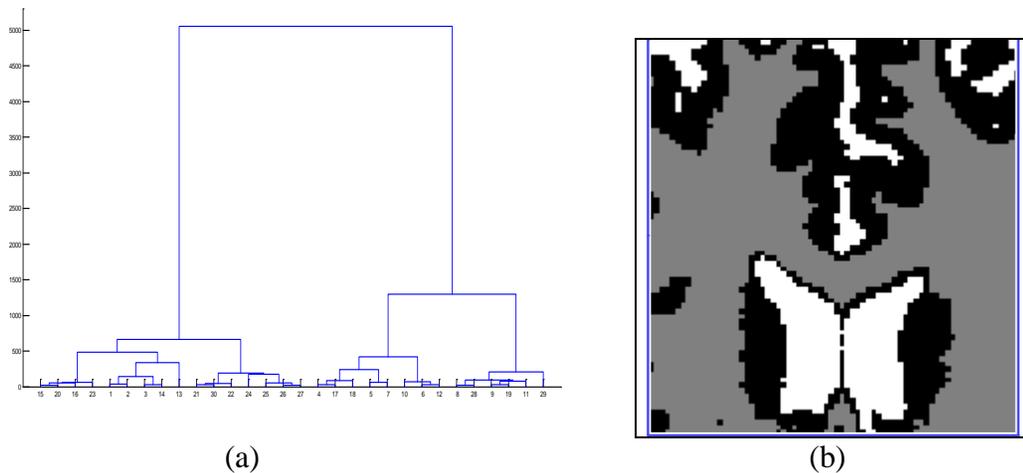


Figura 6-5 Dendrograma obtenido al segmentar la región de interés de la Fig. 3 utilizando una vecindad con radio de 20 píxeles (a); partición de 3 clusters.

Surge entonces la pregunta: ¿qué radio(s) de vecindad se debe(n) utilizar de manera que con el método propuesto se puedan resumir los patrones generales sin una cantidad excesiva de clusters? Evidentemente, la respuesta a esta pregunta es dependiente de la aplicación y resulta muy difícil emitir una recomendación universal.

Sin embargo, los razonamientos y ejemplos anteriores permiten sugerir una manera intuitiva de medir el grado de agrupamiento de los píxeles en la imagen con vistas a sugerir un radio vecindad. La idea sería utilizar una ventana de un tamaño dado y desplazarla por toda la imagen (ver Figura 6-6).

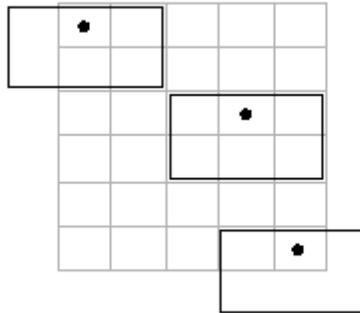


Figura 6-6 Esquema de barrido de una imagen utilizando una ventana deslizante.

En cada posición de la imagen la ventana deslizante define una vecindad del píxel central (Ver apartado 3.1). La varianza calculada en esa vecindad representa una medida de la variabilidad local de los niveles de intensidad en un píxel dado. Así en cada localización (píxel) de la imagen se obtiene un valor asociado a la variabilidad local (cf. [130]). La varianza de todos los valores así obtenidos en cada píxel es utilizada para caracterizar los cambios de la variabilidad local de las diferentes regiones de la imagen.

De esta manera a una imagen se le asocia una medida de la diversidad de la variabilidad local para un tamaño de ventana dado. Este valor se presta a la siguiente interpretación: valores altos para una ventana de un tamaño dado son un indicio de la presencia tanto de zonas homogéneas como de zonas menos homogéneas. Estas últimas pueden ser zonas donde se encuentren fronteras de regiones homogéneas.

Teniendo en cuenta la interpretación anterior se desarrolló un procedimiento para analizar el comportamiento del cambio de la varianza de los valores de variabilidad local para diferentes tamaños de la ventana deslizante que se resume de la siguiente manera.

1. Para una ventana deslizante de tamaño dado realizar un barrido de la imagen, calculando en cada localización de la imagen la variabilidad local (varianza dentro de la ventana).
2. Calcular la varianza de los valores obtenidos en el paso anterior (diversidad de la variabilidad local).
3. Repetir los pasos 1 y 2 para diferentes tamaños de la ventana deslizante.

A modo de ilustración, en la Figura 6-7 se muestra un ejemplo una ventana obtenida al utilizar un radio igual a 2 píxeles y la distancia euclidiana para medir las distancias

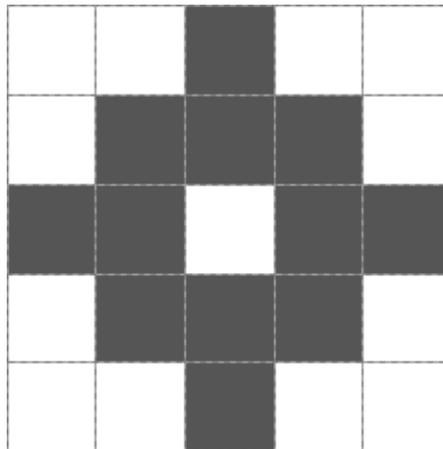


Figura 6-7 Ventana deslizante de radio 2 según la distancia euclidiana.

Empíricamente se obtuvo que el mayor pico de la curva que relaciona el radio de la vecindad con el valor de la diversidad local permite sugerir un radio de vecindad con el que el método propuesto brinda resultados aceptables.

En la Figura 6-8 se muestran los valores obtenidos para el ejemplo de la Figura 6-3a.

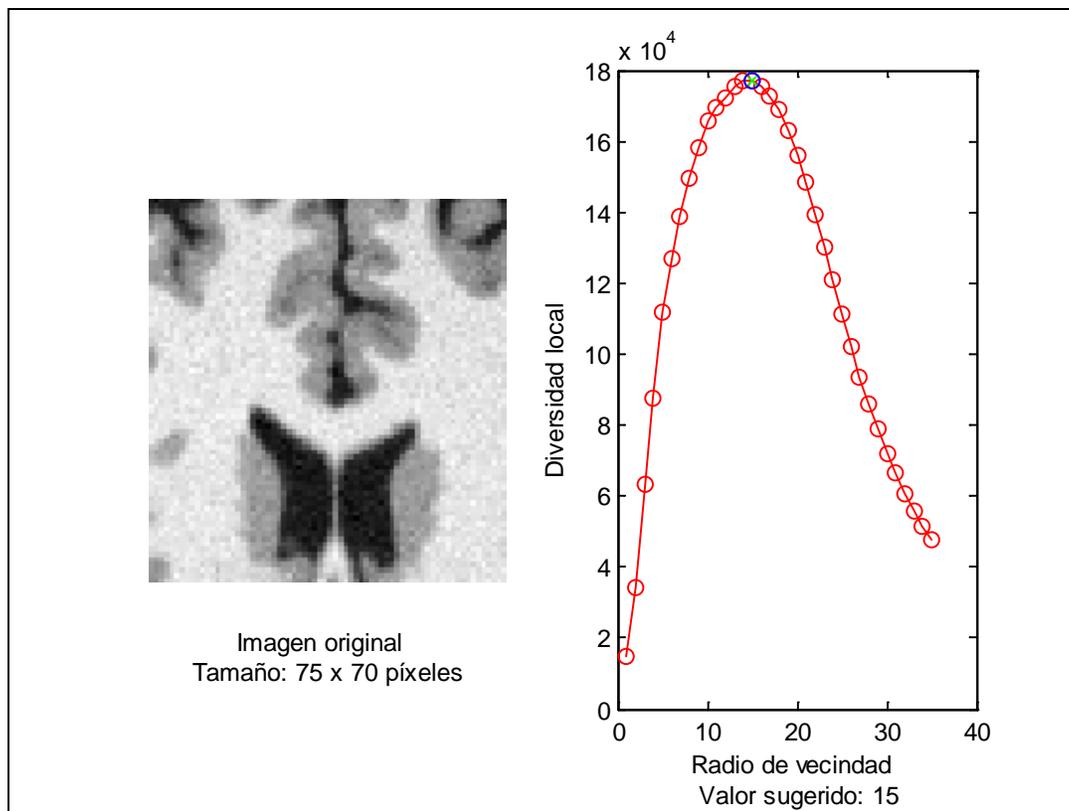


Figura 6-8 Gráfico que relaciona el radio de la vecindad deslizante y la diversidad local en el ejemplo de la Figura 6-3a.

El pico del gráfico se corresponde con un radio de 15 píxeles. A continuación se muestra el dendrograma obtenido para un radio de 15 píxeles así como la correspondiente partición seleccionada (ver Figura 6-9).

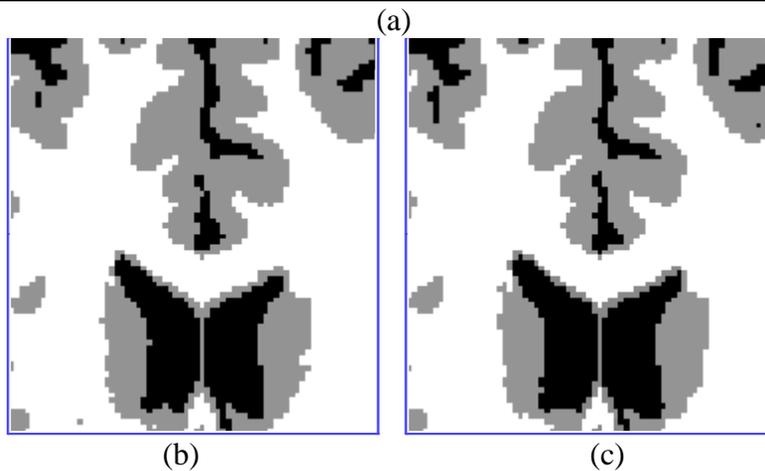
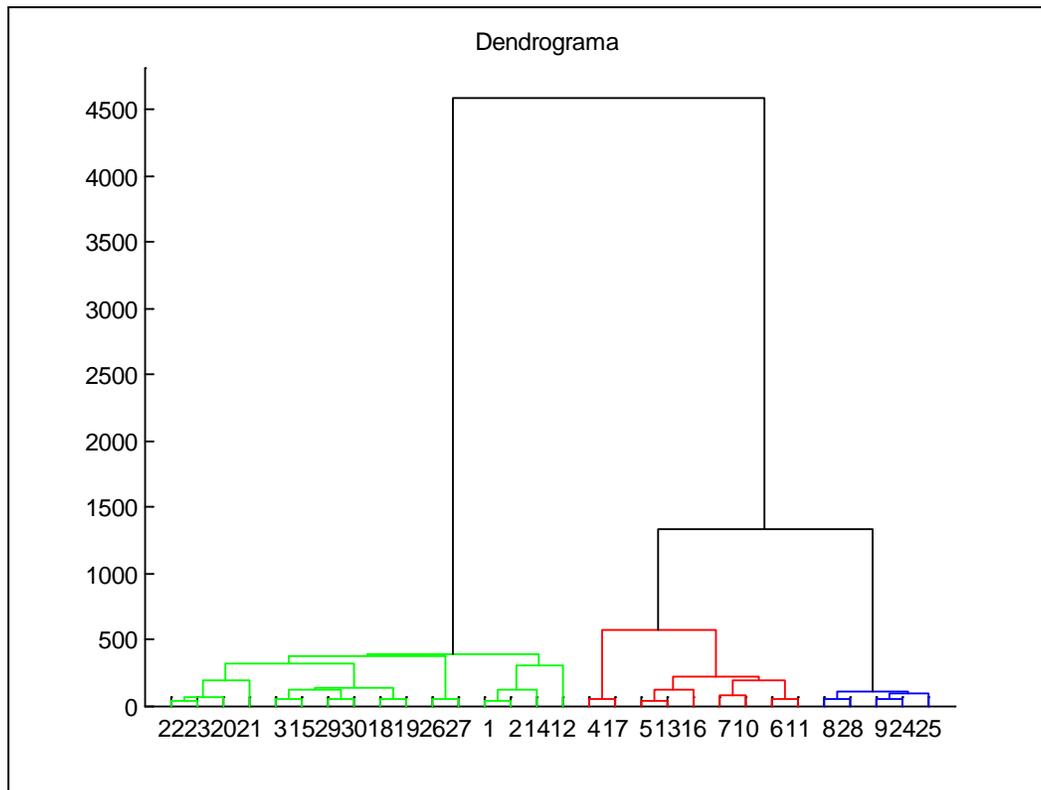
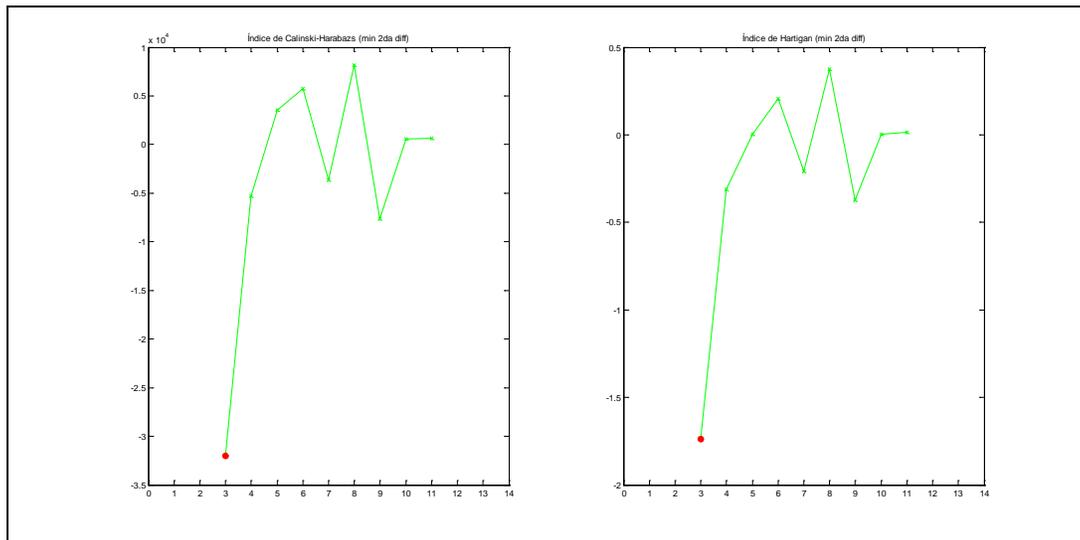


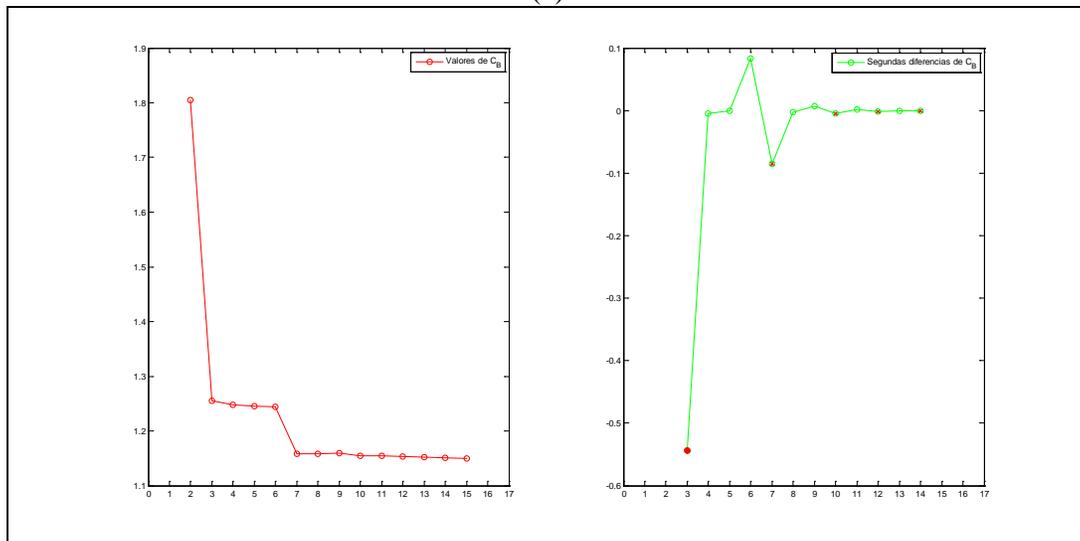
Figura 6-9 Dendrograma obtenido al segmentar la región de interés de la Fig. 3 utilizando una vecindad con radio de 15 píxeles (a); partición de 3 clusters (b); modelo anatómico (c).

Como se aprecia en las Figuras 5-9b y 5-9c la partición obtenida y el modelo anatómico son muy similares. El valor del índice de ajustado de Rand en este caso es 0.92.

Al utilizar el índice de Calinski-Harabazs y el índice de Hartigan para cortar el dendrograma la cantidad óptima sugerida de clusters en ambos es 3 (ver Figura 6-10a). Por su parte el índice C_B coincide en cuanto a la sugerencia de 3 clusters como mejor partición (ver Figura 6-10b).



(a)



(b)

Figura 6-10 Partición de 3 clusters obtenida utilizando el índice de Calinski-Harabazs y el índice de Hartigan para realizar el corte del dendrograma de la Fig. 9 (a); comportamiento del índice C_B y valor sugerido de 3 clusters (b).

Los resultados obtenidos brindan una evidencia empírica de la calidad de las segmentaciones que se obtienen mediante la utilización del procedimiento heurístico de selección de (los) radio(s) de vecindad a utilizar con el método propuesto.

Ahora bien, si el objetivo es realizar un análisis exploratorio con vistas a obtener varias particiones de los datos, se sugiere utilizar no solamente el valor asociado al pico del gráfico, si no también un conjunto de radios alrededor del valor sugerido.

En el siguiente apartado se comenta sobre la influencia del radio de la vecindad en los resultados obtenidos.

6.6.5 Efecto del radio de vecindad utilizado en los resultados

Como se apunta en el apartado anterior, el radio de vecindad sugerido por el procedimiento heurístico diseñado debe ser considerado como una alternativa ante la ausencia de información adicional.

Para mostrar de manera cuantitativa la bondad del valor (o valores) sugerido(s) en el ejemplo de la Figura 6-3a, se analizó un rango de valores de radios de vecindad alrededor del valor obtenido según el procedimiento. Concretamente se analizaron 7 valores diferentes de radios de vecindad alrededor del valor 15 sugerido (ver Figura 6-8) que van desde un radio de vecindad igual a 11 píxeles hasta un radio de vecindad igual a 20 píxeles.

Para cada radio de vecindad se aplicó el método propuesto y se seleccionó una partición a partir del dendrograma. Cada una de las 7 particiones obtenidas se comparó con el modelo anatómico utilizando el índice de Rand Ajustado. Los resultados se muestran en la Tabla 6-1.

Tabla 6-1 Particiones obtenidas con diferentes radios de vecindad

Radio	CANTIDAD DE CLUSTERS							
	2	3	4	5	6	7	8	9
11	0.35	0.90*	0.90	0.90	0.85	0.84	0.84	0.82
12	0.34	0.91*	0.89	0.85	0.85	0.85	0.85	0.83
13	0.34	0.91*	0.89	0.88	0.88	0.88	0.88	0.86
14	0.26	0.28	0.29	0.89*	0.88	0.88	0.87	0.85
15	0.34	0.92*	0.90	0.90	0.90	0.88	0.88	0.83
16	0.35	0.91*	0.90	0.90	0.85	0.85	0.85	0.83
17	0.34	0.91*	0.89	0.89	0.89	0.86	0.85	0.85
18	0.25	0.27	0.28	0.80*	0.80	0.80	0.80	0.80
19	0.34	0.82	0.83*	0.81	0.81	0.77	0.77	0.77
20	0.35	0.88*	0.87	0.87	0.86	0.86	0.82	0.82

(*) Mayor índice ajustado de Rand para el radio dado.

Como se aprecia en los resultados presentados en la Tabla 6-1, la gran mayoría de las particiones obtenidas para diferentes radios de vecindad poseen una concordancia elevada con el modelo anatómico según el índice ajustado de Rand. Se puede constatar también que con un radio de 15 píxeles se obtiene la partición con la mayor concordancia con el modelo.

Este experimento se repitió para varios cortes contenidos en el volumen bajo estudio. En el siguiente apartado se presentan y analizan los resultados obtenidos.

6.6.6 Resultados para diferentes cortes en el volumen estudiado

En este experimento se analizaron varios cortes dentro del volumen estudiado. Concretamente se seleccionaron 10 cortes por encima y 10 cortes por debajo del corte central que se había utilizado en los apartados anteriores los que conforman un volumen de $70 \times 75 \times 21 \text{ mm}^3$ (ver Figura 6-11).

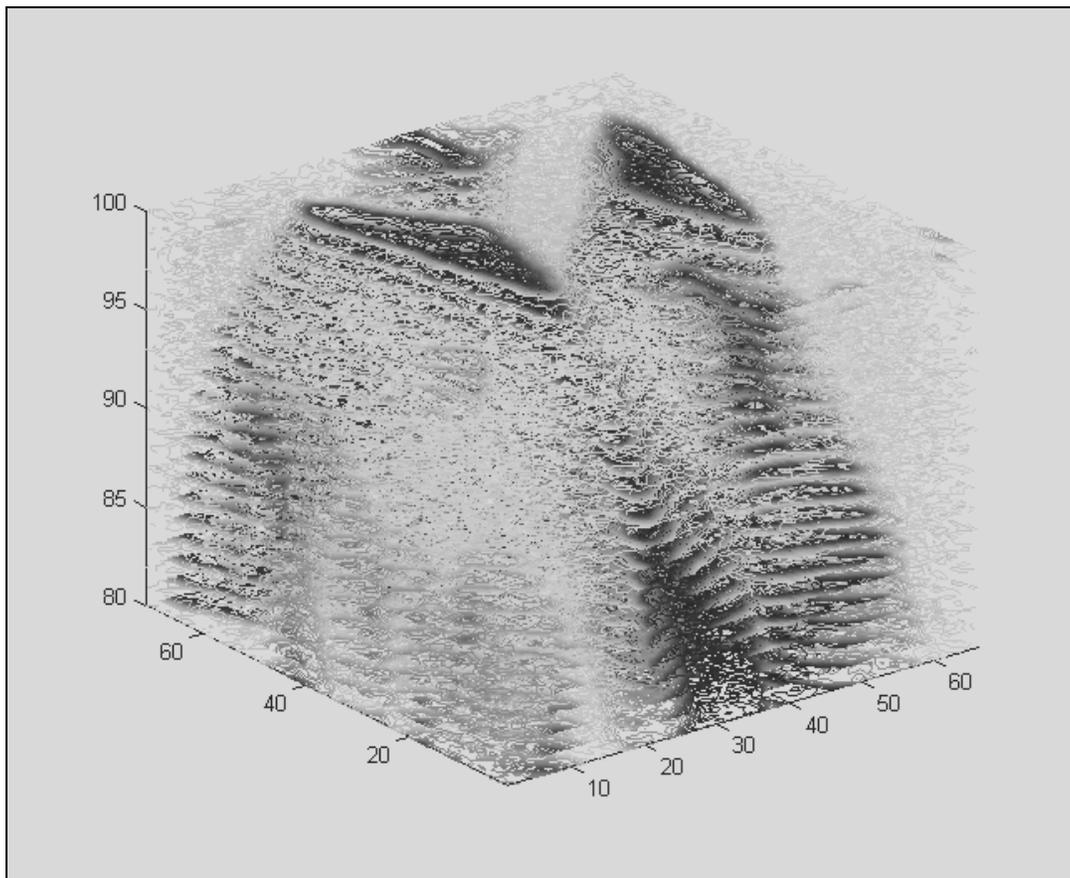


Figura 6-11 Volumen compuesto por 21 cortes en la zona central seleccionada del cerebro.

En cada corte se aplicó el algoritmo de agrupamiento aglomerativo jerárquico de enlace tipo promedio con restricciones de contigüidad. El radio de vecindad utilizado se obtuvo con la ayuda del procedimiento heurístico descrito en el apartado 6.6.4.

En la Tabla 6-2 se presenta el comportamiento de los radios de vecindad sugeridos por el procedimiento propuesto para su selección. La variación de los radios resulta bastante suave. Esto se puede explicar por el hecho de que el espesor de los cortes es de un 1mm. lo que provoca que de un corte al siguiente las variaciones de las estructuras anatómicas no sean tan bruscas.

Tabla 6-2 Comportamiento de los radios de vecindad en los 21 cortes

CORTE	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1
Radio	11	11	11	11	14	15	15	16	15	15

CORTE	0	1	2	3	4	5	6	7	8	9	10
Radio	15	15	15	15	14	13	13	12	11	10	9

Las particiones de interés asociadas a los 21 cortes se contrastaron con el modelo anatómico disponible utilizando el índice ajustado de Rand. Los resultados obtenidos se muestran en la Figura 6-12.

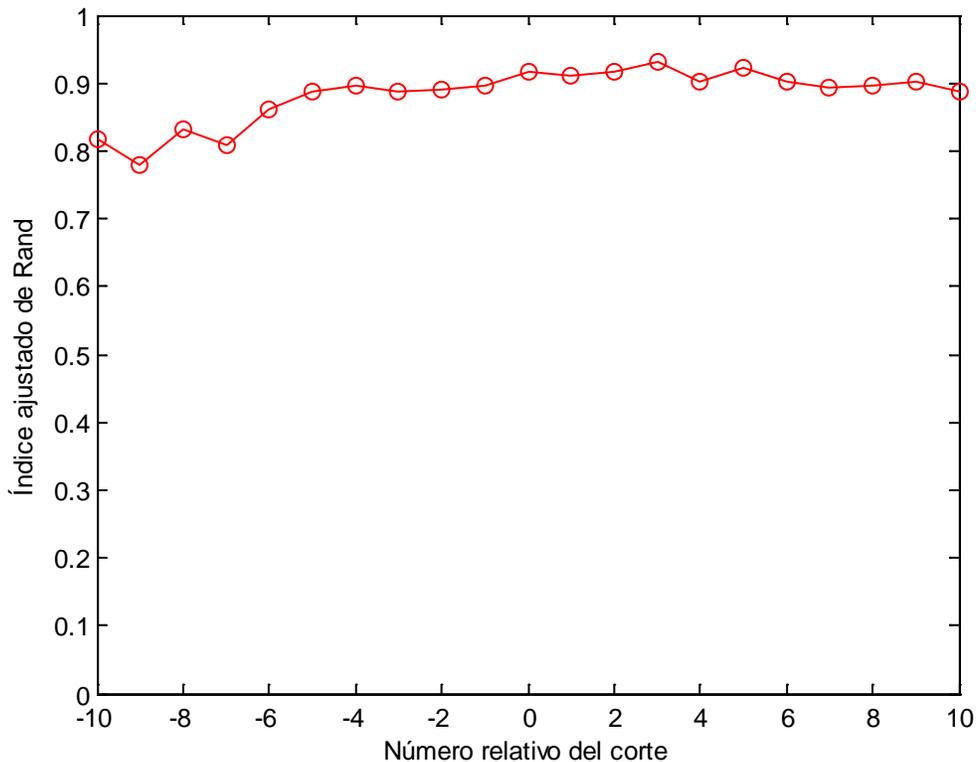


Figura 6-12 Índice ajustado de Rand para las particiones obtenidas en cada uno de los 21 cortes seleccionados.

Como se aprecia en la Figura 6-12, las particiones obtenidas en el volumen son de alta calidad en comparación con el modelo anatómico según el índice ajustado de Rand.

6.6.7 Conclusiones

En este apartado se ha estudiado el comportamiento del método propuesto en el caso de imágenes médicas simuladas.

Se propone un procedimiento heurístico para la selección del radio de vecindad a utilizar en el proceso de clasificación en este tipo de aplicación. Los resultados obtenidos corroboran, de manera empírica, la calidad de las particiones obtenidas.

El índice C_B utilizado para realizar el corte de los dendrogramas se comporta de manera satisfactoria. Los resultados que se obtienen en este caso están en concordancia con los que sugieren el índice de Calinski-Harabazs y el de Hartigan. Un elemento a favor de la calidad de las particiones resultantes lo constituyen los altos valores del índice ajustado de Rand.

El esquema de trabajo presentado permite sugerir la idea de utilizar el método propuesto como base para la construcción de un ambiente interactivo de análisis exploratorio de imágenes que permita obtener particiones con diferentes grados de detalles según las necesidades del usuario.

6.7 Experimentos con imágenes reales de IBSR. Resultados

En este apartado se presentan los experimentos realizados con imágenes de resonancia magnética reales obtenidas de la base de datos IBSR (*Internet Brain Segmentation Repository*) del Centro para el Análisis Morfométrico (CAM) del Hospital General de Massachussets [202].

Se obtuvieron 20 juegos de datos de personas normales junto con las correspondientes segmentaciones manuales. Cada uno de los juegos de datos se identifica con una combinación numérica. Así se tienen los sujetos 5_8, 4_8, 2_4, 6_10, 15_3, 16_3, 17_3, 8_4, 7_8, 110_3, 111_2, 112_2, 100_23, 202_3, 191_3, 12_3, 13_3, 1_24, 205_3 y 11_3.

Estos 20 estudios cubren un amplio espectro de calidades de imágenes, razón por la cual han sido escogidas por IBSR.

Cada estudio se compone de un conjunto imágenes de dimensión 256x256 píxeles correspondientes a cortes consecutivos sin separación de espesor igual a 3.0 mm. Hemos utilizadas los juegos de datos transformados a 8 bits que contienen solamente la región del cerebro.

Nuestro interés principal ha sido la segmentación de la sustancia gris y la sustancia blanca pues estos son los tipos de tejidos predominantes en el cerebro, junto al líquido cefalorraquídeo. Para un corte dado se selecciona una región de interés donde estén presentes los tejidos principales de interés. El procesamiento de una determinada zona de interés está dado principalmente por dos razones: (a) se requiere menor tiempo computacional; (b) en un ambiente clínico los usuarios están normalmente interesados en ciertas zonas, por ejemplo, la región donde se localiza un tumor.

6.7.1 Objetivo

El objetivo principal de estos experimentos es evaluar, en el caso de imágenes reales, la calidad de las particiones que se obtienen en comparación con el conocimiento disponible, tanto las segmentaciones manuales como los resultados obtenidos por otros métodos de segmentación.

6.7.2 Requisitos para abordar el objetivo

Para evaluar la calidad de las particiones obtenidas con el método propuesto se utilizó el índice de Jaccard descrito con anterioridad. En el caso de las imágenes, este índice puede ser visto como una medida de solapamiento.

Si denotamos por R_A una región segmentada por un método automatizado, y por R_M la correspondiente zona que ha sido segmentada manualmente, entonces el índice puede ser escrito de la siguiente forma (cf. apartado 6.4):

$$J(R_A, R_M) = \frac{|R_A \cap R_M|}{|R_A \cup R_M|} = \frac{a}{a+b+c}$$

El índice representa la fracción de píxeles que son clasificados como pertenecientes a la región que se analiza en ambas segmentaciones respecto a todos los píxeles clasificados como pertenecientes a la región en al menos una de las segmentaciones.

Se debe notar que en esta definición el universo de análisis lo constituye la unión de las regiones y, por tanto, en el esquema dado en el apartado 6.4, y que se reproduce a continuación levemente modificado,

		R_M	
		Agrupados	Desagrupados
R_A	Agrupados	a	b
	Desagrupados	c	d

la cantidad d no se toma en cuenta, y más aún es igual a 0 en este contexto. Allá se comparan particiones y aquí sólo dos regiones y al ser d igual a 0 el índice de Jaccard y el el Rand resultan iguales desde el punto de vista computacional.

El índice toma el valor 1 cuando ambas segmentaciones coinciden completamente, mientras el valor 0 evidencia un desacuerdo total.

6.7.3 Elección de la región de interés

Aquí se procedió de manera similar al experimento anterior. En cada uno de los cortes procesados se seleccionó una región de interés de 30x30 píxeles que contuviera los tejidos de interés. En la Figura 6-13 se muestra, a modo de ejemplo, el corte 30 del sujeto 191_3.

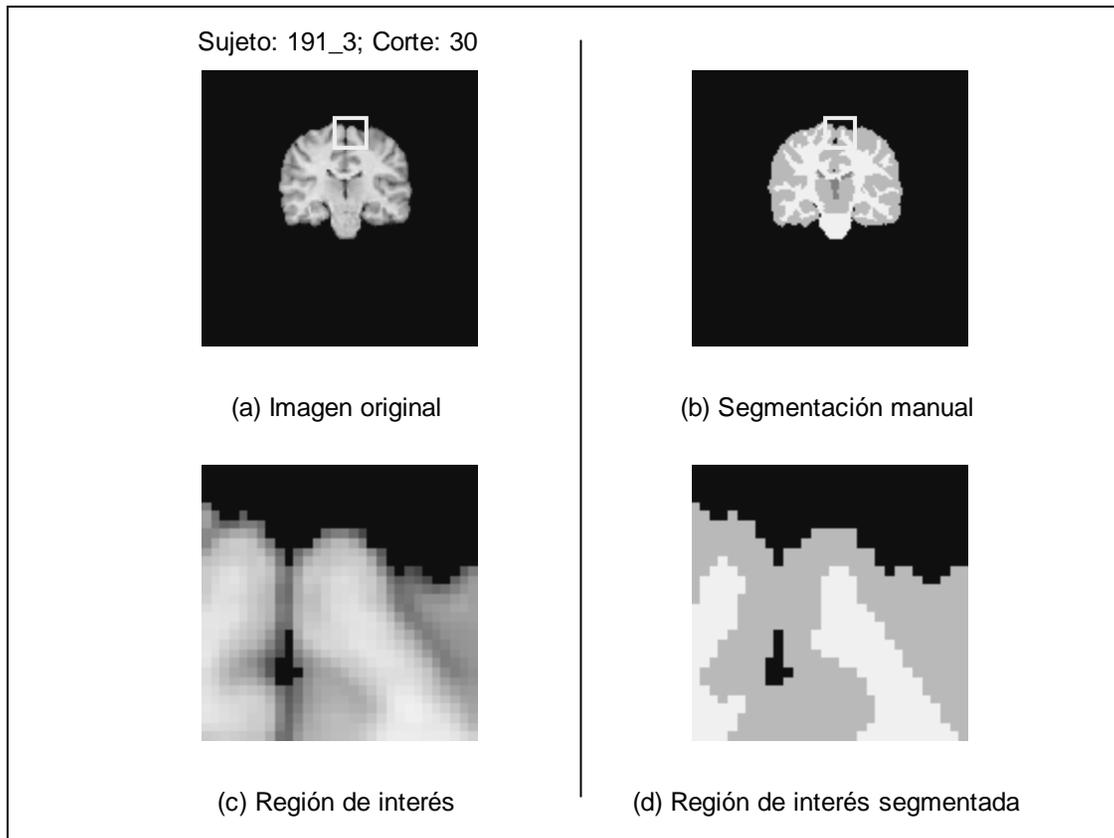


Figura 6-13 Corte de ejemplo (a); la segmentación manual correspondiente (b); la región de interés seleccionada (c); y la región de interés segmentada manualmente.

La región seleccionada, además de ser representativa de los tejidos principales, incluye fronteras entre ellos que no aparecen claramente delineadas.

6.7.4 Selección del radio de la vecindad

Como se ha evidenciado antes, el tamaño de la vecindad tiene un impacto importante en los resultados que se obtienen con el método propuesto. En general radios pequeños tienden a producir resultados peores que los que se obtienen con radios mayores.

En este caso se optó por utilizar los valores sugeridos por el procedimiento empírico introducido en el apartado 6.6.4. En la Figura 6-14 se muestra el radio de vecindad sugerido para el ejemplo de la Figura 6-13.

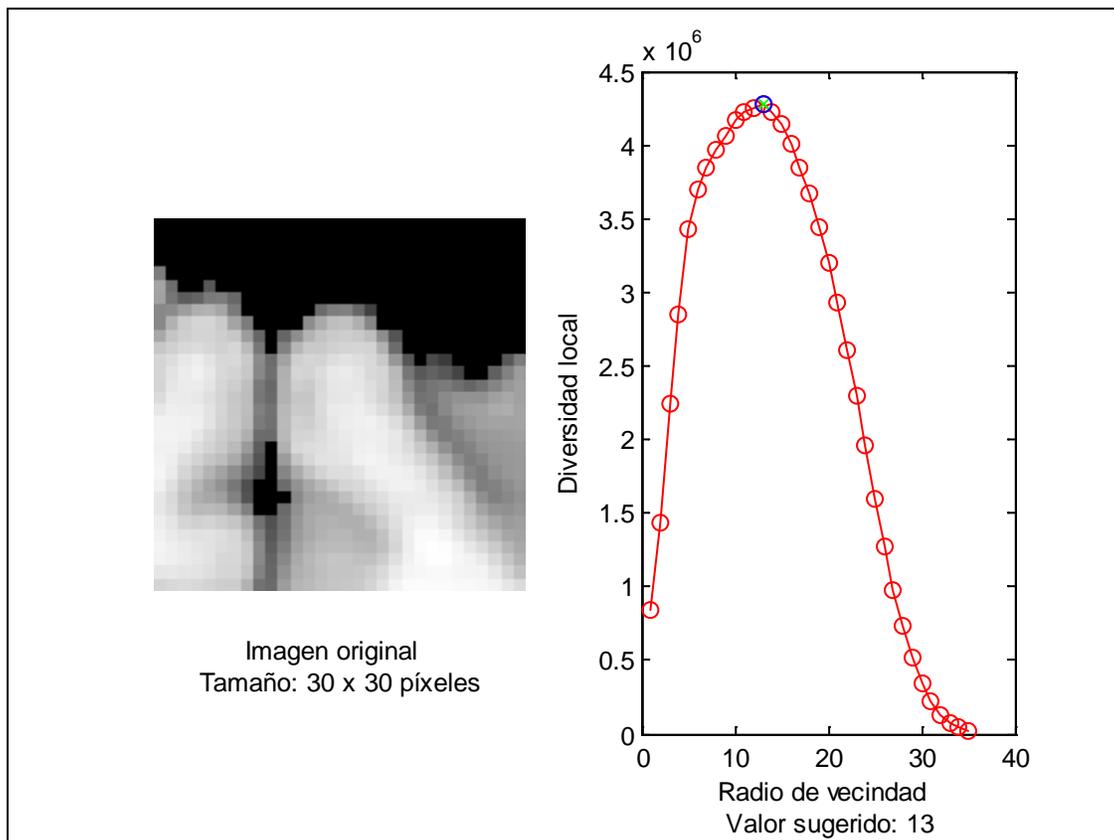


Figura 6-14 Valores de diversidad local para diferentes radios de vecindad en el ejemplo de la Figura 6-13.

Se debe enfatizar que la selección de este parámetro es un problema práctico muy dependiente de la aplicación. De ahí que la interpretación de los resultados que se obtengan debe tener presente estas cuestiones.

6.7.5 Análisis de los resultados en un ejemplo

Con vistas a ilustrar el procedimiento seguido, se presentan y analizan a continuación los resultados obtenidos con el ejemplo de la Figura 6-13.

El método propuesto fue utilizado para procesar la región de interés de la Figura 6-13c.

El dendrograma obtenido se muestra en la Figura 6-15. A partir del dendrograma es posible obtener una representación de la imagen a diferentes escalas.

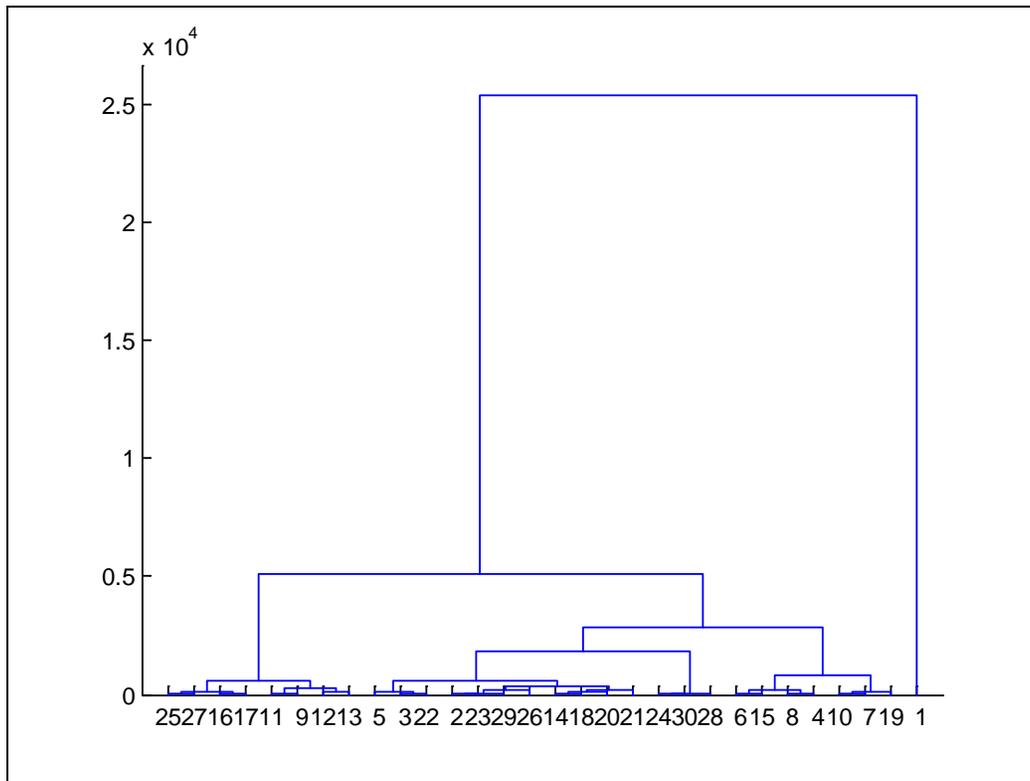


Figura 6-15 Dendrograma obtenido para el caso de ejemplo de la Figura 6-13.

El índice C_B se decanta por una partición de 3 clusters (ver Figura 6-16).

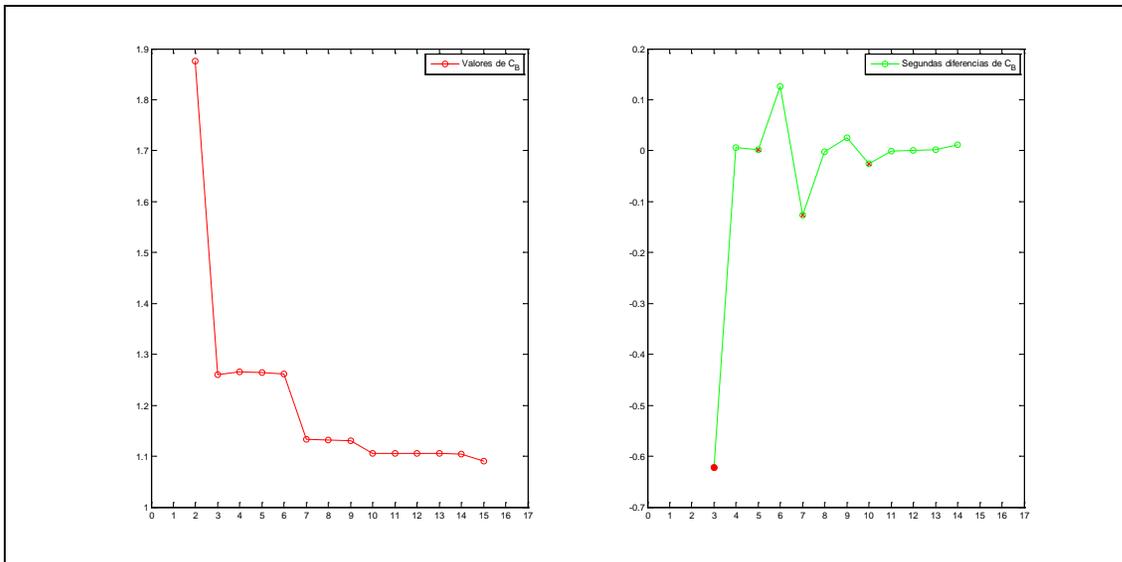


Figura 6-16 Comportamiento de índice C_B el ejemplo de la Figura 6-13.

En este caso se define con claridad una partición de 3 clusters que, por demás, guarda relación directa con la cantidad de tejidos a identificar en la región.

En la Figura 6-16 se muestra la segmentación obtenida junto a la imagen de la región de interés y la segmentación manual

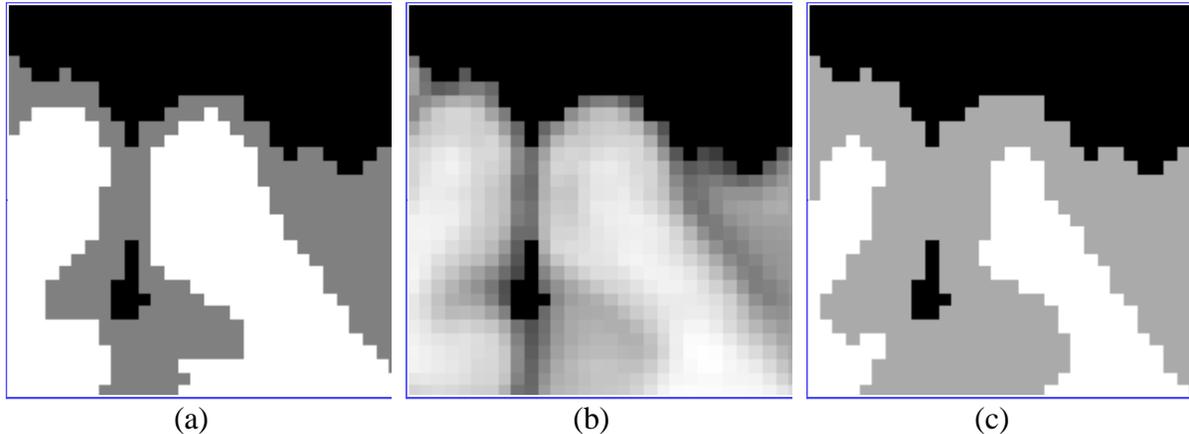


Figura 6-17 Segmentación obtenida (a); original (b); segmentación manual (c).

Los valores del índice de Jaccard para los tejidos de interés, sustancia gris, sustancia blanca, así como el fondo son 0.64269, 0.56125 y 1, respectivamente. Estos valores de solapamiento de los clusters pueden parecer relativamente bajos, excepto para el caso del fondo. Aquí se debe tener en cuenta que la partición de referencia que se toma es la segmentación manual. Como se comentaba anteriormente, la segmentación manual no debe ser asumida como la “verdad absoluta” pues su proceso de obtención pasa por la subjetividad humana.

La Figura 6-17 permite realizar una valoración comparativa y evidencia la dificultad que surge al identificar la frontera entre la sustancia gris y la sustancia blanca. Sin embargo, desde el punto de vista visual nos parece que la segmentación del algoritmo se corresponde mejor con la imagen original que la segmentación manual.

Aquí se debe tener en cuenta también la resolución con la que fueron obtenidas las imágenes. El espesor de los cortes es de 3 mm. y esto provoca que los píxeles cercanos a las fronteras entre diferentes tipos de tejidos representen en realidad una combinación de estos tejidos. Este fenómeno es conocido como efecto de volumen parcial y, como se mencionó en el apartado 1.4, es un elemento que incide en el proceso de segmentación.

Para completar la discusión de los resultados, en la Figura 6-18 se muestran las particiones de 2, 3, 4 y 5 clusters obtenidas a partir del dendrograma.

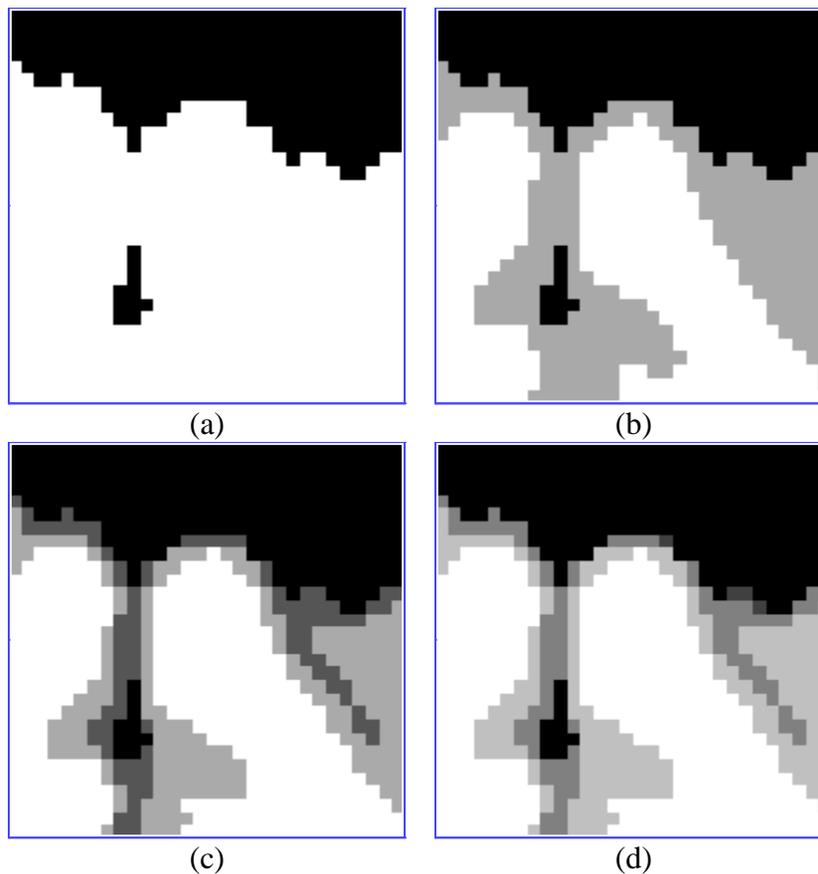


Figura 6-18 Segmentación de 2 clusters (a), de 3 clusters (b), de 4 clusters (c), y de 5 clusters (d).

La segmentación en dos clusters identifica claramente el fondo y el tejido cerebral. En el caso de la Figura **6-18c** se aprecia que la región ocupada por la sustancia gris, identificada en la Figura **6-18a**, aparece ahora subdividida. La Figura **6-18d** es muy similar a la de tres clusters, excepto por el hecho de la subdivisión, casi imperceptible, que aparece en las partes central y derecha.

El nuevo cluster que aparece en la Fig. **6-18c** parece tener una explicación razonable desde el punto de vista médico. El tejido cerebral en su conjunto está rodeado por líquido cefalorraquídeo que ocupa las cavidades o espacio subaracnoideo [207] (ver Figura **6-19**). Esto incluye las hendiduras o cavidades de la corteza conocidas como surcos. Esta fina capa de tejido es la que parece corresponderse con el líquido cefalorraquídeo que no aparece identificado en la segmentación manual (cf. Figura **6-18b**). Lo anterior está en correspondencia con lo planteado por otros acerca del hecho de que los datos de IBSR no identifican todo el líquido cefalorraquídeo [198] y resulta un hallazgo interesante que está asociado a la posibilidad de visualizar los datos a diferentes escalas o resoluciones.

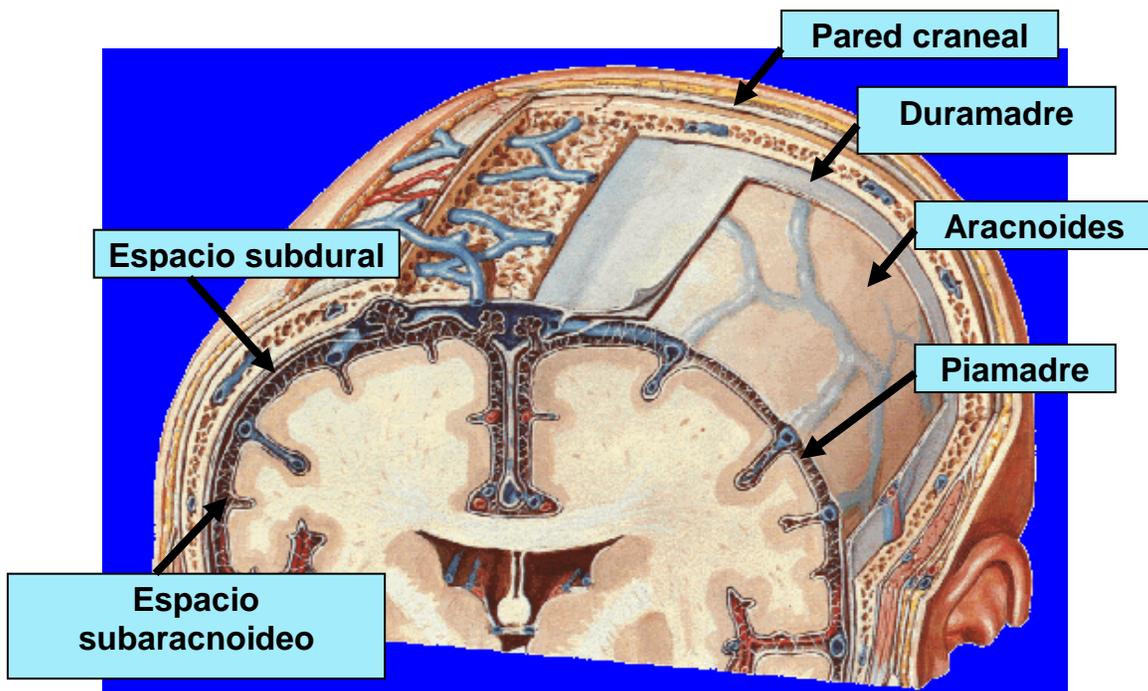


Figura 6-19 Esquema de las estructuras que cubren el encéfalo del hombre.

6.7.6 Resultados generales

Siguiendo el procedimiento descrito antes se procesaron imágenes de cada uno de los 20 conjuntos de datos de IBSR. Los valores del índice de Jaccard fueron calculados en cada caso para la sustancia gris y la sustancia blanca en todas las regiones de interés seleccionadas. Esta es una estrategia común en un ambiente clínico, donde el especialista se centra en determinadas zonas de interés.

Aparte de razones computacionales, en nuestra opinión las zonas seleccionadas son representativas de los tejidos cerebrales y pueden ser tomadas como base para la comparación. El procesamiento de los volúmenes completos resultaría más costoso aunque no imposible. Una posibilidad sería realizar análisis de las imágenes por bloques.

El índice de Jaccard promedio obtenido al utilizar el método propuesto es de 0.657 ± 0.178 para la sustancia gris, y 0.712 ± 0.120 para la sustancia blanca. Los resultados obtenidos para cada uno de los sujetos en las regiones de seleccionadas se muestran en la Tabla 6-3. Estos valores pueden ser considerados como una primera aproximación a la comparación con los reportados por los seis métodos de segmentación y los resultados de los dos expertos sobre cuatro cerebros de sujetos normales.

Tabla 6-3 Valores de índice de Jaccard para cada sujeto por tipo de tejido

SUJETOS	GRIS	BLANCA
5_8	0,88048	0,80471
4_8	0,85598	0,657
2_4	0,83511	0,8913
6_10	0,74556	0,83591
15_3	0,76512	0,73753
16_3	0,645	0,71166
17_3	0,31024	0,56546
8_4	0,6654	0,59839
7_8	0,864	0,87137
110_3	0,41075	0,71698
111_2	0,5436	0,8814
112_2	0,71875	0,69748
100_23	0,49206	0,69288
202_3	0,71399	0,61972
191_3	0,64269	0,56125
12_3	0,86647	0,86765
13_3	0,36909	0,63031
1_24	0,62675	0,76812
205_3	0,75267	0,66571
11_3	0,44598	0,47149
Promedio	0,657485	0,712316

En la Figura 6-20 se muestran los índices promedio para la sustancia gris y la sustancia blanca obtenidos por varios métodos de segmentación facilitados por IBSR. Ellos incluyen los siguientes: *adaptive MAP*, *biased MAP*, *fuzzy c-means*, *Maximum A Posteriori Probability (MAP)*, *Maximum-Likelihood*, *tree-structure k-means*, y segmentaciones manuales de otros 2 expertos sobre 4 cerebros de sujetos normales [194].

Los resultados pueden parecer bajos. Sin embargo, como ya se ha comentado, se debe tener en cuenta que los datos contienen imágenes de diferentes calidades. Por otra parte, el espesor de 3 mm. de los cortes tiene un impacto en el desempeño de los métodos de segmentación. Es por esa razón que estos resultados deben ser considerados válidos en este contexto únicamente.

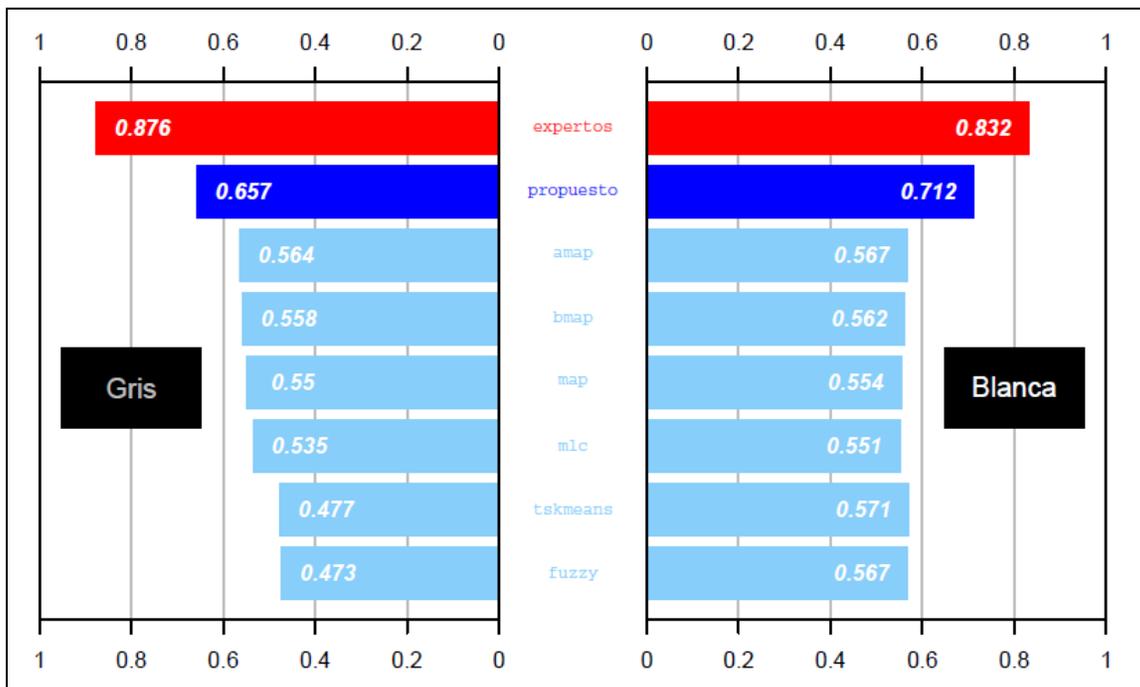


Figura 6-20 Índice promedio para la segmentación de la sustancia gris y de la sustancia blanca.

6.7.7 Conclusiones

En este apartado se ha estudiado el comportamiento del método propuesto en el caso de imágenes médicas reales.

Se aprecia que el índice C_B propuesto se comporta de manera satisfactoria y sus resultados se acercan en gran medida a los “reales” disponibles en forma de segmentaciones manuales.

Los resultados obtenidos evidencian, de manera empírica, que el comportamiento del método en el contexto de aplicación seleccionado es satisfactorio tanto desde el punto de vista de las segmentaciones manuales disponibles, como en comparación con otros métodos de segmentación. Para una comparación más completa se requieren experimentos adicionales.

6.8 Experimentos para acelerar la ejecución del algoritmo. Resultados

Según se analizó en el capítulo anterior el comportamiento del método de agrupamiento jerárquico aglomerativo de tipo enlace promedio con restricciones de contigüidad posee un comportamiento temporal cuadrático en el peor caso en función de la cantidad de píxeles a procesar.

Independientemente de que empíricamente se mostró que este comportamiento puede ser en realidad sub-cuadrático (ver apartado 5.8), a medida que crece la cantidad de píxeles el procesamiento puede llegar a ser costoso computacionalmente.

Por otra parte, el método propuesto conduce a soluciones en las cuales la separación (desemejanza) promedio entre clusters, teniendo en cuenta las restricciones de contigüidad espacial, es alta. De manera implícita la homogeneidad de los clusters tiende a ser alta.

En este apartado se explora un heurístico con el fin de disminuir el tiempo de procesamiento para obtener a la solución según el método propuesto. El heurístico está concebido para la reducción de la cantidad de objetos que se someten al algoritmo de agrupamiento jerárquico aglomerativo con restricciones, cuestión ésta que debe redundar en la disminución del tiempo de procesamiento.

6.8.1 Objetivo

El objetivo principal de estos experimentos es evaluar el efecto del heurístico sugerido en las particiones que se obtienen en ausencia de referentes para la comparación en el caso de imágenes reales de pacientes de esclerosis múltiple (EM) en cuanto al tiempo de procesamiento requerido para su obtención, así como en su calidad intrínseca.

6.8.2 Requisitos para abordar el objetivo

Para los análisis que siguen se han seleccionado imágenes de RM de pacientes de EM. La EM es una enfermedad del sistema nervioso central que afecta fundamentalmente la sustancia blanca en personas adultas jóvenes. Caracterizada por primera vez por Charcot [208], puede provocar discapacidades severas en los pacientes, e incluso causar la muerte. Aunque sus causas no están plenamente establecidas, se plantea que está condicionada por factores medioambientales y por cierta susceptibilidad genética.

Tradicionalmente se ha considerado que la EM es una enfermedad que ataca la mielina presente en el cerebro y la médula ósea. Sin embargo, se conoce que también las fibras nerviosas (axones) son objeto de daño, desapareciendo junto con las células a las cuales

están enlazadas. Desafortunadamente no existen marcadores de laboratorio que midan de manera exacta, tanto la desmielinización, como el daño axonal en las etapas tempranas de la enfermedad cuando no abundan aún muchos síntomas clínicos.

El daño acumulado producto de la pérdida de mielina, el daño axonal, y la pérdida de éstos en su conjunto se manifiesta a través de la “desaparición de tejido”, tanto en el cerebro como en la médula ósea. A esta pérdida de tejido se le denomina atrofia. Los espacios que se van generando son ocupados por el líquido cefalorraquídeo.

En la Figura 6-21 se muestran imágenes de RM de pacientes de EM obtenidas según el protocolo FLAIR (*Fluid Attenuated Inversion Recovery*) mediante el cual es posible suprimir la intensidad del líquido cefalorraquídeo.

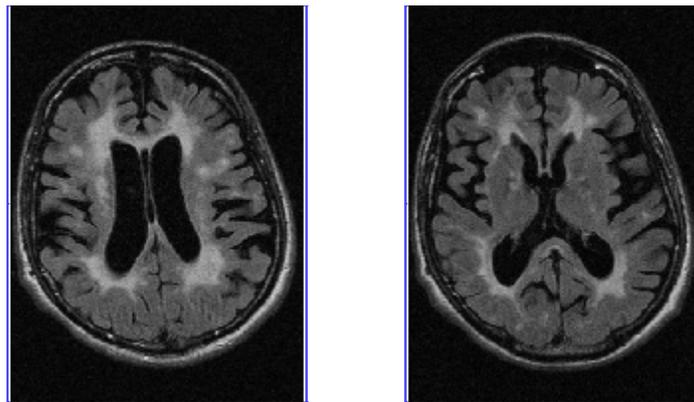


Figura 6-21 Imágenes de pacientes de EM.

En las imágenes se aprecia la disminución del volumen del tejido cerebral provocado por la enfermedad. Las zonas de alta intensidad están asociadas a las lesiones presentes en la sustancia blanca.

6.8.3 Descripción del heurístico

El algoritmo propuesto comienza el procesamiento considerando cada píxel como un cluster. Empíricamente, en el capítulo anterior (ver apartado 5.8) se mostró que la complejidad computacional puede considerarse sub-cuadrática en función de la cantidad de píxeles, aunque no es menos cierto que a medida que crece la cantidad de píxeles el procesamiento puede llegar a ser costoso.

La idea es utilizar entonces el siguiente heurístico. En la primera fase del algoritmo se mezclan todos aquellos píxeles vecinos cuya desemejanza se encuentre por debajo de un valor ε definido de antemano. De esta manera la cantidad de clusters con la que comienza el algoritmo se reduce. Se debe señalar que esta reducción se obtiene a costa de perder los

primeros niveles del dendrograma original. Sin embargo, estos primeros niveles por lo general no son tan importantes desde el punto de vista del análisis de los datos. Más aún, desde el punto de vista práctico usualmente resultan de interés aquellas particiones que contengan una cantidad pequeña de clusters. Precisamente estas particiones se obtienen en las últimas etapas del dendrograma [209, 210].

Al finalizar el proceso de reducción y con vistas a aplicar el método propuesto es necesario mantener actualizada la relación de vecindad entre los clusters resultantes. Con este fin se pueden aplicar dos estrategias: (a) partir de la relación de vecindad inicial entre clusters singulares e ir actualizando la relación de adyacencia; (b) realizar las fusiones y luego calcular las relaciones de vecindad

La primera estrategia presupone el análisis de cada par de vecinos y su fusión en el caso que la desemejanza se encuentre por debajo del umbral establecido. Luego de la fusión se realiza la actualización de las relaciones de vecindad entre el cluster recién modificado y el resto. Esta estrategia es muy similar al método propuesto, excepto en el hecho de que en cada paso no es necesario fusionar el par de clusters que posea desemejanza mínima.

Para el caso de la segunda estrategia se puede proceder de la siguiente manera. Se construye un grafo no dirigido que contenga un vértice por cada píxel y una arista entre dos vértices si y sólo si ellos son vecinos y además la desemejanza entre ambos no sobrepasa el umbral ε establecido. La etapa de reducción es equivalente entonces a la obtención de los componentes conexos del grafo así creado. La obtención de los componentes conexos se puede lograr en un tiempo $O(n+m)$, donde n es la cantidad de vértices y m la cantidad de aristas [211, 212]. Una vez obtenidos los componentes conexos se hace necesario propagar la relación de vecindad inicial. Debemos recordar aquí que para la utilización del método propuesto se requiere conocer el grado de contigüidad relativa entre clusters con vistas al cálculo de la desemejanza entre ellos (ver apartado 5.4).

6.8.4 Evaluación experimental del heurístico

Para analizar el comportamiento del heurístico de reducción de los datos se ha procedido de manera similar a los apartados anteriores. Primeramente se ha seleccionado una región de interés y luego se ha aplicado el procedimiento de selección del radio de vecindad descrito antes.

En la Figura 6-22 se presenta a modo de ilustración una imagen utilizada con la región de interés destacada.

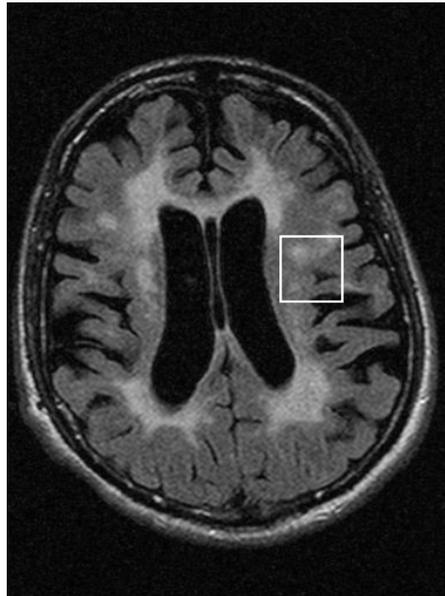


Figura 6-22 Región seleccionada en una imagen de un paciente de EM.

La Figura 6-23 muestra la región de interés ampliada, así como el radio de vecindad sugerido por el procedimiento de selección. El valor de radio escogido se utiliza para definir la relación de vecindad entre píxeles durante todo el procesamiento.

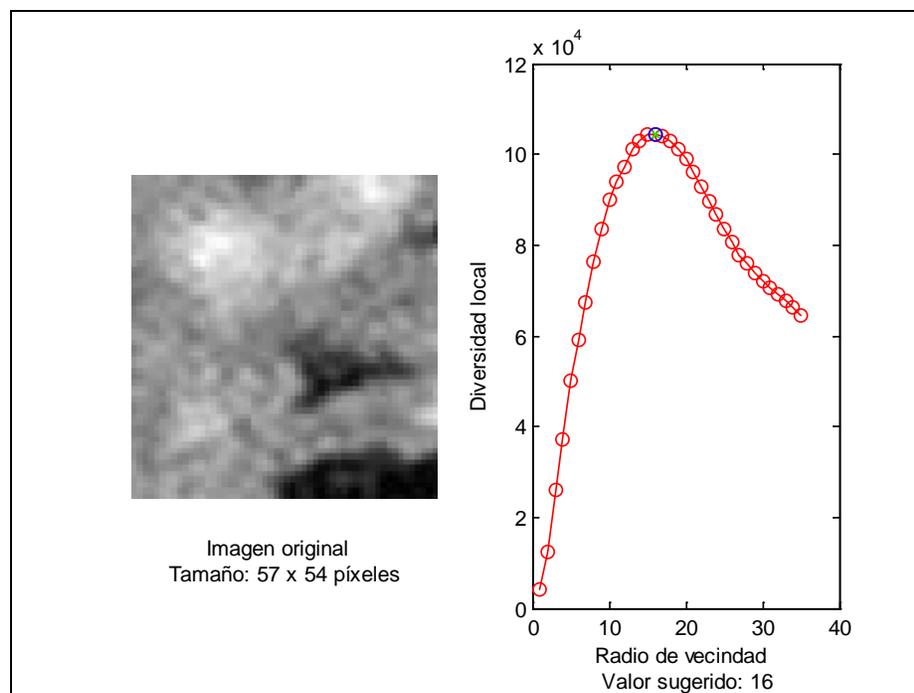


Figura 6-23 Radio de vecindad sugerido para la región de interés seleccionada.

Un aspecto práctico a resolver para la aplicación de heurístico consiste en la determinación del umbral ϵ . Es evidente que a medida que este umbral aumenta mayor será la cantidad de píxeles que se fusionan en clusters como resultado de la aplicación del heurístico limitando de esa manera la cantidad de particiones posibles a obtener al aplicar el método propuesto.

Varios procedimientos pudieran ser sugeridos para buscar un umbral. Así por ejemplo, es posible analizar la distribución de las desemejanzas entre pares de píxeles y seleccionar el umbral como la desemejanza asociada a un determinado valor. La utilización de este procedimiento permite controlar qué por ciento de enlaces entre píxeles son suprimidos. Sin embargo, la reducción efectiva en cuanto a la cantidad de clusters que se obtienen guarda estrecha relación con el radio de vecindad utilizado en el procesamiento.

En el contexto de esta aplicación se ha observado de manera empírica que con umbrales pequeños, por ejemplo diferencias de 0 y 1 unidades de intensidad de los píxeles, se obtienen resultados aceptables. La utilización de estos umbrales es equivalente a tener en cuenta en la etapa asociada al heurístico solamente aquellos píxeles que sean vecinos entre sí y cuyas diferencias de intensidades sean a lo sumo igual a 1. A continuación se discuten estos resultados.

6.8.4.1 Escenarios experimentales

La evaluación experimental del heurístico toma como base tanto las particiones obtenidas como los tiempos de ejecución.

Se han analizado los siguientes escenarios de procesamiento:

- a) Método de enlace promedio (*average-link*).
- b) Método propuesto sin heurístico.
- c) Método propuesto con heurístico y umbral igual a 0.
- d) Método propuesto con heurístico y umbral igual a 1.
- e) Método *k-means*.

Los escenarios (a) y (e), se refieren a métodos clásicos muy utilizados, que aunque funcionan sin tener en cuenta la estructura espacial de los píxeles, pueden servir útiles como referentes de comparación. El caso (a) es el equivalente del método propuesto y el (e) es un método rápido desde el punto de vista computacional. Los resultados que se reportan se refieren a las implementaciones de Matlab.

El escenario (b) ocurre bajo las mismas condiciones descritas en los apartados anteriores, es decir, se ejecuta el método propuesto partiendo de la partición formada por clusters singulares.

Los escenarios (c) y (d), por su parte, se refieren a la aplicación del heurístico con diferentes umbrales. En el caso (c) se activan enlaces entre píxeles vecinos que sean exactamente iguales en términos de intensidad. El caso (d) permite además de las condiciones anteriores que se consideren píxeles cuya diferencia de intensidades sea menor o igual que 1.

6.8.4.2 Análisis de las particiones obtenidas

6.8.4.2.1 Escenario (a)

El escenario (a) consiste en una corrida del método de enlace promedio (*average-link*) sobre los píxeles de la imagen sin considerar la estructura espacial.

Una vez obtenido el dendrograma (ver Figura 6-24) se muestran las particiones asociadas a las últimas etapas (ver Figura 6-25).

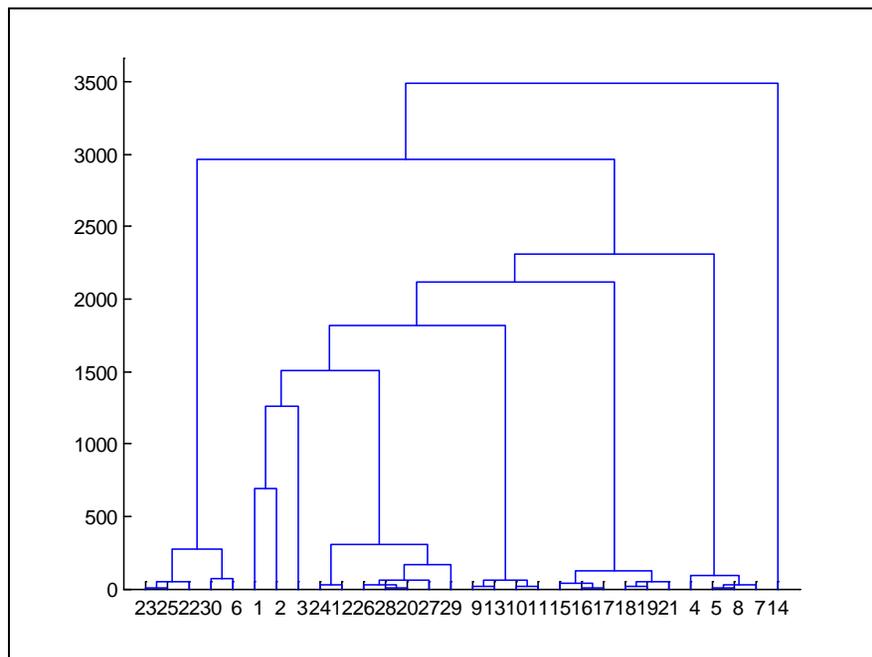


Figura 6-24 Dendrograma obtenido para el escenario (b).

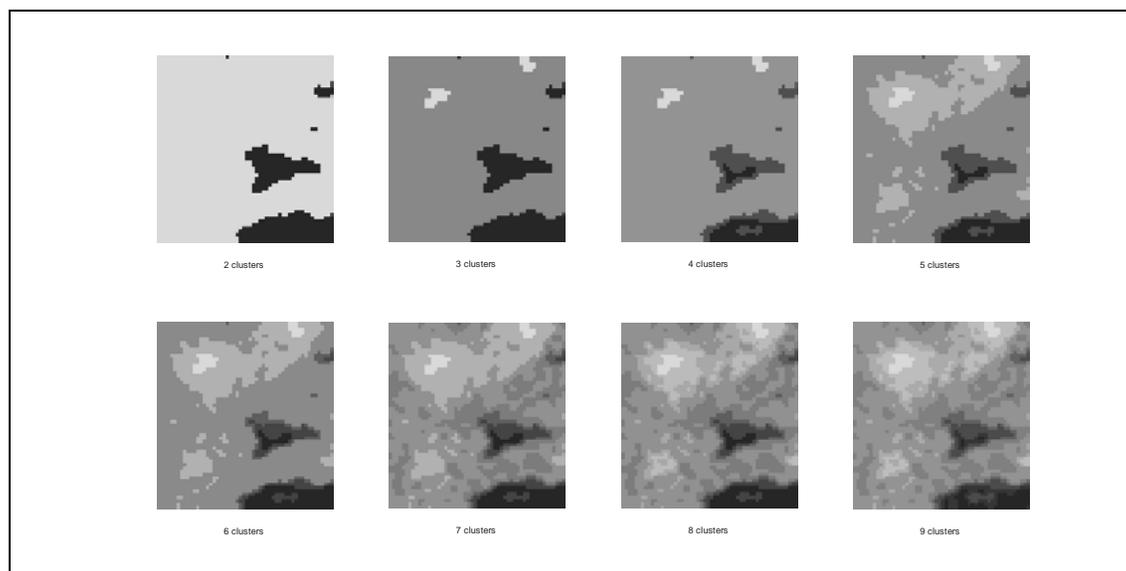


Figura 6-25 Particiones obtenidas en el escenario (a).

En este caso se puede señalar que las particiones de 2 a 6 clusters capturan de manera aceptable las principales regiones en la imagen. Ya a partir de la partición de tamaño 7 se comienza a apreciar cierto grado de sobre-segmentación.

Un elemento no favorable a señalar consiste en el elevado tiempo de procesamiento (209.233 segundos \approx 3.5 minutos, incluyendo el cálculo de las distancias) requerido para obtener el dendrograma.

6.8.4.2.2 Escenario (b)

A continuación se presentan los resultados asociados al escenario (b), es decir, la aplicación del método propuesto sin la utilización del heurístico. En este, y el resto de los escenarios que siguen se utilizó la vecindad de 16 píxeles sugerida por el procedimiento de selección (ver apartado 6.6.4).

La Figura 6-26 muestra el dendrograma obtenido. Aquí como en los demás casos nuestro interés se concentra en las particiones que se derivan de las últimas etapas de construcción del dendrograma. Estas son, usualmente, las de mayor interés práctico pues permiten resumir los datos en una pequeña cantidad de grupos.

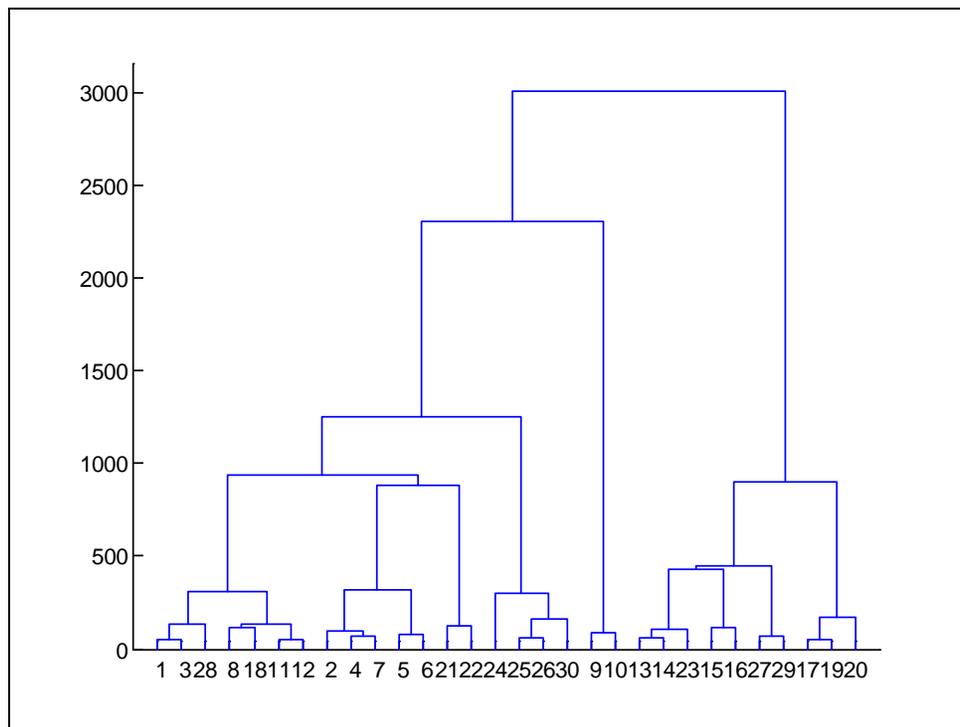


Figura 6-26 Dendrograma obtenido para el escenario (b).

El análisis del dendrograma utilizando el índice C_B apunta hacia la partición de 3 clusters, aunque también como segunda opción sugiere la partición de 5 clusters (ver Figura 6-27).

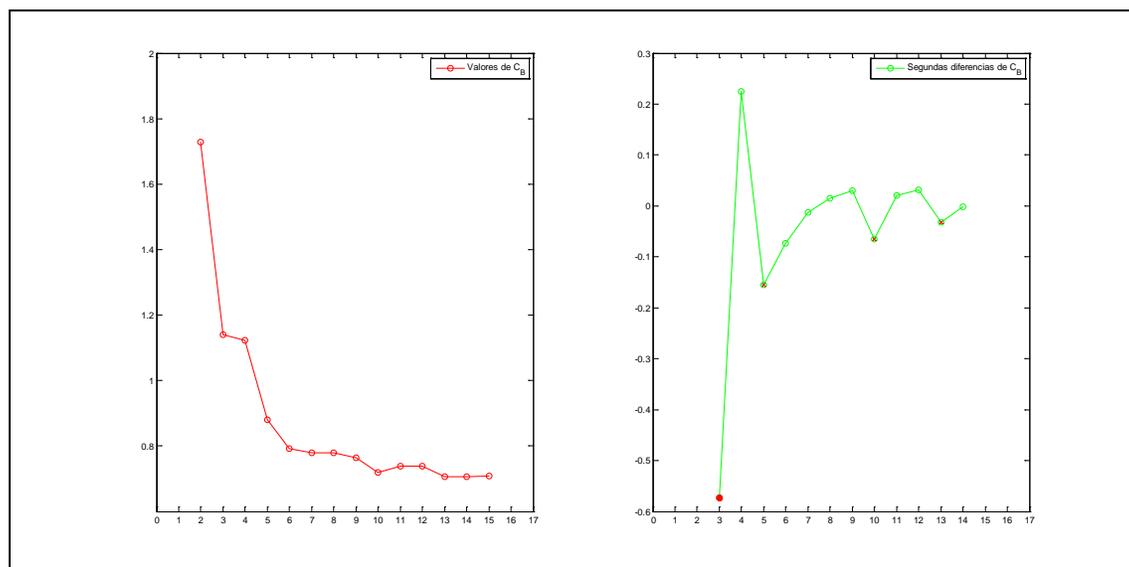


Figura 6-27 Comportamiento del índice C_B en el Escenario (b).

En la Figura 6-28 se muestran las particiones con cardinalidades que van desde 2 hasta 9 clusters obtenidas a partir del dendrograma.

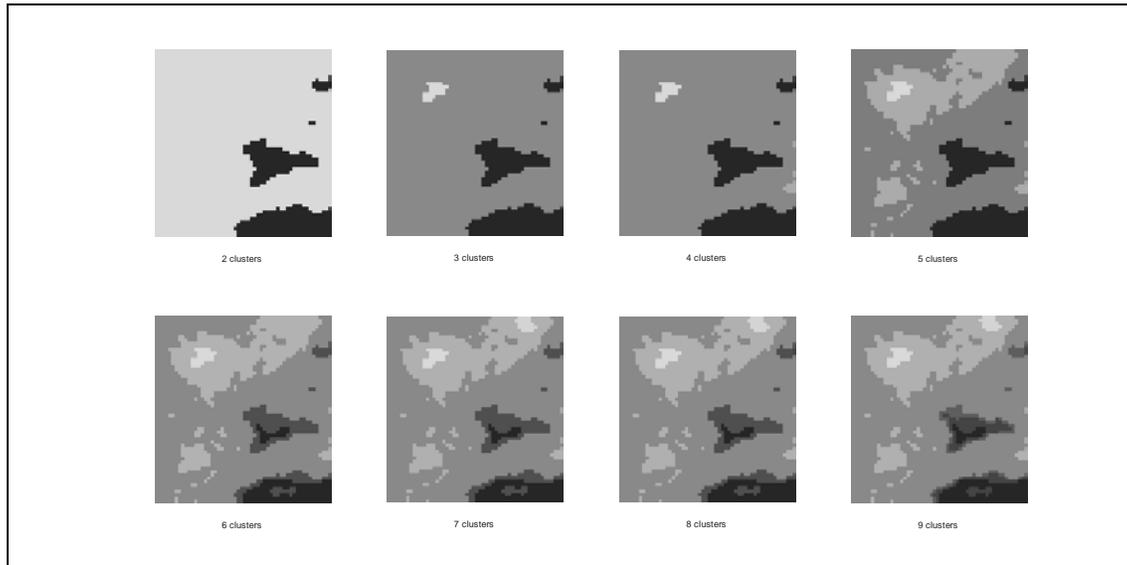


Figura 6-28 Particiones obtenidas en el escenario (b).

La partición que consiste en dos clusters separa las partes oscuras de la imagen de las claras. Ya en las particiones de 3 y 4 clusters se comienzan a destacar las zonas de más alta intensidad.

En la partición de 5 clusters se aprecian claramente las zonas de intensidad media alta. Ya a partir de la partición de 6 clusters comienzan a aparecer más detalles en los grupos oscuros.

En comparación con el Escenario (a) el tiempo de ejecución disminuye a 47.252 segundos, es decir, a menos de un minuto. Si le agregamos el tiempo requerido para el procesamiento inicial de las vecindades de los píxeles (8.259 segundos) el tiempo total sería 55.511 segundos. La implementación (en Matlab) de la obtención de los vecinos en la etapa inicial es susceptible de un refinamiento con vista a optimizar el uso de la memoria y hacerlo menos costoso en tiempo. Estos comentarios son válidos también para los escenarios (c) y (d).

En general la disminución del tiempo está en correspondencia con los resultados del apartado 5.8.

6.8.4.2.3 Escenario (c)

En el escenario (c) el método propuesto es aplicado luego de efectuar una reducción de los datos según el heurístico. En la Figura 5-29 se muestra la distribución de las distancias entre los píxeles.

La etapa de reducción en este caso utiliza un umbral igual a 0. Se observa que no es muy frecuente pero la disminución del número de aristas en el grafo correspondiente que se logra acelera de manera importante la ejecución del algoritmo propuesto.

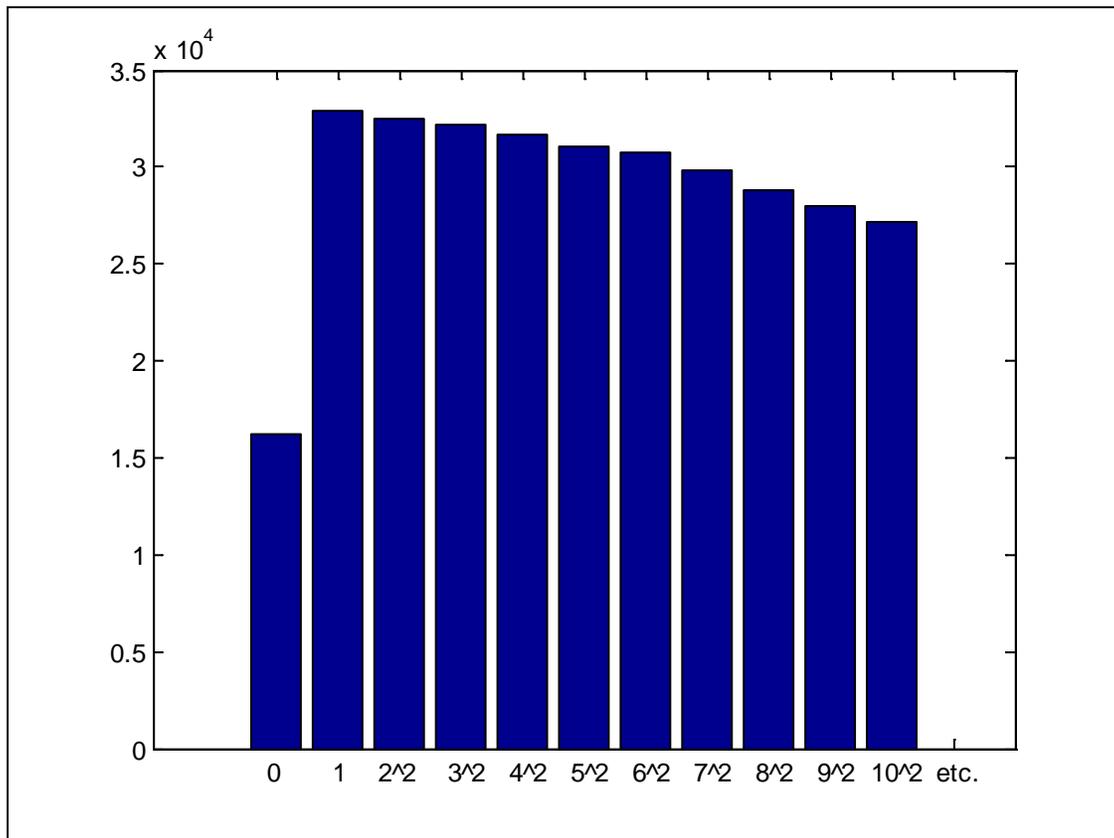


Figura 6-29 Diagrama de barras de distancias entre píxeles.

A continuación se muestran los resultados obtenidos.

- Radio de vecindad escogido: 16
- Total de aristas iniciales: 942170
- Total de aristas que cumplen la condición impuesta por el umbral: 16174 (1.72%)
- Cantidad inicial de clusters = 3078 (57*54 píxeles)
- Componentes conexos obtenidos = 280 (clusters)
- Reducción (% de clusters eliminados) = 90.90%
- Tiempo de obtención de componentes conexos = 0.086 segundos.

Como se puede apreciar la utilización de un umbral igual a 0 provoca una reducción importante de los datos del 90% aproximadamente. Los 280 clusters resultantes son los componentes conexos del grafo obtenido al imponer las condiciones de vecindad entre píxeles y de igualdad de intensidades.

Estos 280 clusters son sometidos entonces al algoritmo jerárquico propuesto, pasando por una etapa previa de ajuste de la relación de vecindad entre estos nuevos clusters tomando como base la relación de vecindad inicial entre píxeles.

Los tiempos de estas dos etapas son los siguientes:

- Preparación de los datos: 0.255 segundos
- Ejecución del algoritmo: 0.165 segundos

Si se incluye la etapa de procesamiento de las vecindades (8.259 segundos), el tiempo de obtención de los componentes conexos (0.086 segundos), así los tiempos de algoritmo (0.255+0.165 segundos), el tiempo total resultante sería igual a 8.765 segundos que es significativamente menor que el caso de los Escenarios (a) y (b).

En la Figura 6-30 se muestra el dendrograma obtenido.

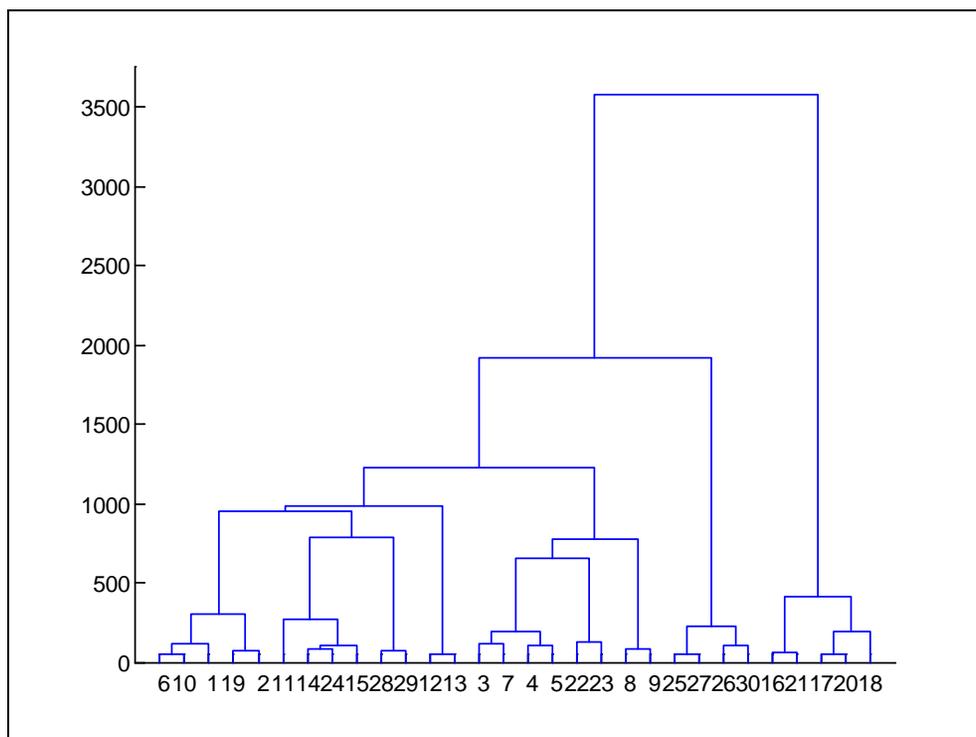


Figura 6-30 Dendrograma obtenido para el escenario (c).

La sugerencia del índice C_B (ver Figura 6-31) descansa en la partición de 4 clusters donde se destacan las zonas de baja media y alta intensidad presentes en la imagen.

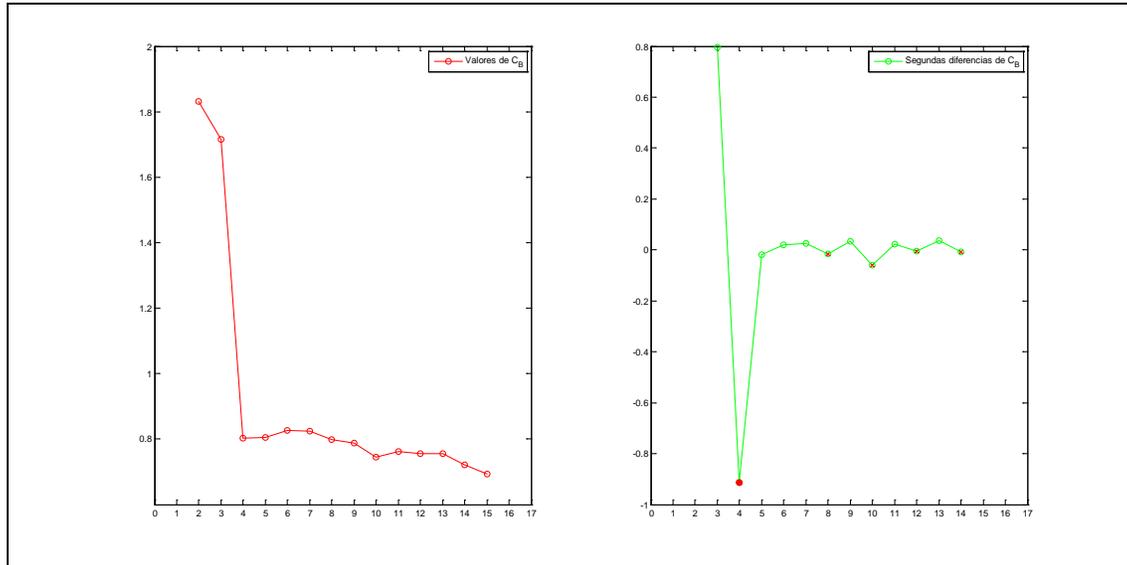


Figura 6-31 Comportamiento del índice C_B en el Escenario (c).

En la Figura 6-32 contiene las particiones asociadas a las etapas finales. Según se puede apreciar, las particiones obtenidas guardan cierta semejanza con las obtenidas en el Escenario (b).

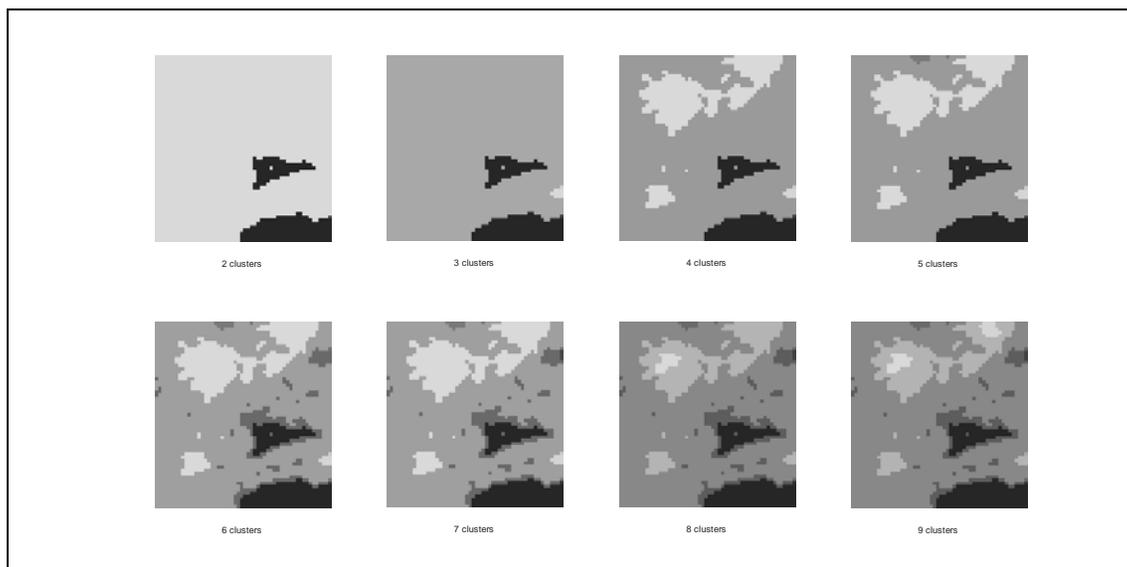


Figura 6-32 Particiones obtenidas en el escenario (c).

Las regiones intensidad media se comienzan a revelar en la partición de 4 clusters de manera similar a lo que ocurre en el Escenario (b). Las zonas de alta intensidad se vislumbran a partir de la partición de 8 clusters, a diferencia del Escenario (c) donde se aprecian ya en la partición de 3 clusters.

Esto apoya la idea de que la pérdida de información asociada a la etapa de reducción utilizando un umbral igual a 0 no influye de manera muy significativa en los resultados finales y permite obtener una disminución sustancial del tiempo de procesamiento.

6.8.4.2.4 Escenario (d)

En este escenario la única modificación consiste en llevar el umbral hasta el valor 1, es decir, se permite establecer enlaces entre aquellos píxeles que además de ser vecinos posean una diferencia de intensidad a lo sumo igual a 1. Los resultados de la ejecución se muestran a continuación.

- Radio de vecindad escogido: 16
- Total de aristas iniciales: 942170
- Total de aristas que cumplen la condición impuesta por el umbral: 49109 (5.21%)
- Cantidad inicial de clusters = 3078 (píxeles)
- Componentes conexos obtenidos = 35 (clusters)
- Reducción (% de clusters eliminados) = 98.86%
- Tiempo de obtención de componentes conexos = 0.328 segundos.

En este caso el porcentaje de aristas que cumplen la condición impuesta por el umbral ha pasado de 1.72% (Escenario (c)) a 5.21%. En cuanto al porcentaje de reducción en la cantidad de clusters éste varía del 90.90% (Escenario (c)) al 98.86%.

Por su parte los tiempos de ejecución del método propuesto junto a la etapa de preparación de los datos son los siguientes.

- Preparación de los datos: 0.237 segundos.
- Ejecución del algoritmo: 0.0004 segundos.

Como en los casos anteriores el procesamiento incluye varios componentes: procesamiento de la vecindad, obtención de los componentes conexos y las etapas del algoritmo. Esto hace que el tiempo de ejecución sea de 8.824 segundos lo que representa un aumento en el tiempo total de cerca del 1% respecto al Escenario (c). Este aumento se explica, específicamente, por la etapa de reducción. Esta etapa resulta más costosa pues el grafo resultante de aplicar el umbral contiene mayor cantidad de aristas.

Sin embargo, si se analizan los tiempos de ejecución del algoritmo propiamente dicho, es decir, los tiempos de preparación de los datos que se le suministran y de construcción de la

jerarquía, se aprecia una reducción de cerca del 50% (0.2374 segundos aquí y 0.420 segundos en el Escenario (c)).

Aquí el aumento del umbral lleva asociado, como es de esperar, un aumento en la cantidad de enlaces entre píxeles vecinos que se activan. Ello a su vez conduce a una elevada reducción del 98% de los datos iniciales: de 3078 clusters singulares (píxeles), hasta solamente 35 clusters.

Debemos hacer notar que los píxeles que se ubican como miembros de un componente conexo en la etapa de reducción nunca se separarán de éste. La Figura 6-33b muestra los 35 clusters resultantes de la etapa de reducción. Como se puede apreciar las zonas de intensidad medio alta han sido ‘fundidas’ con el gran fondo gris desde un inicio y por tanto nunca serán subdivididas. Es por esta razón que en general una excesiva reducción de los datos no es un fenómeno muy deseado.

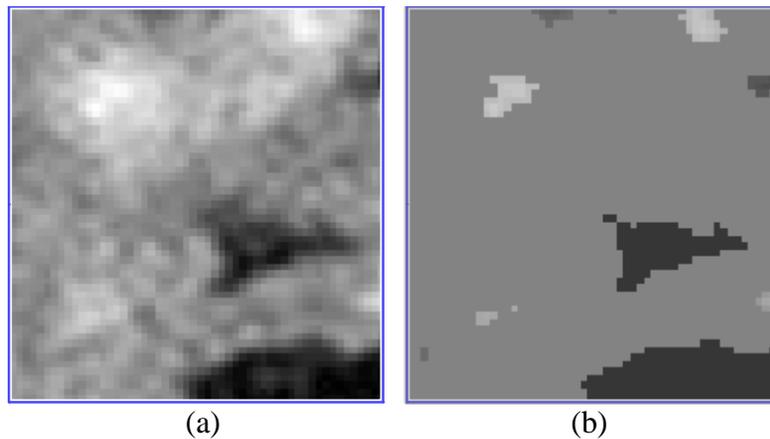


Figura 6-33 Imagen original (a); imagen después de la reducción (b).

En la Figura 6-34 se muestra el dendrograma obtenido en el caso del Escenario (d).

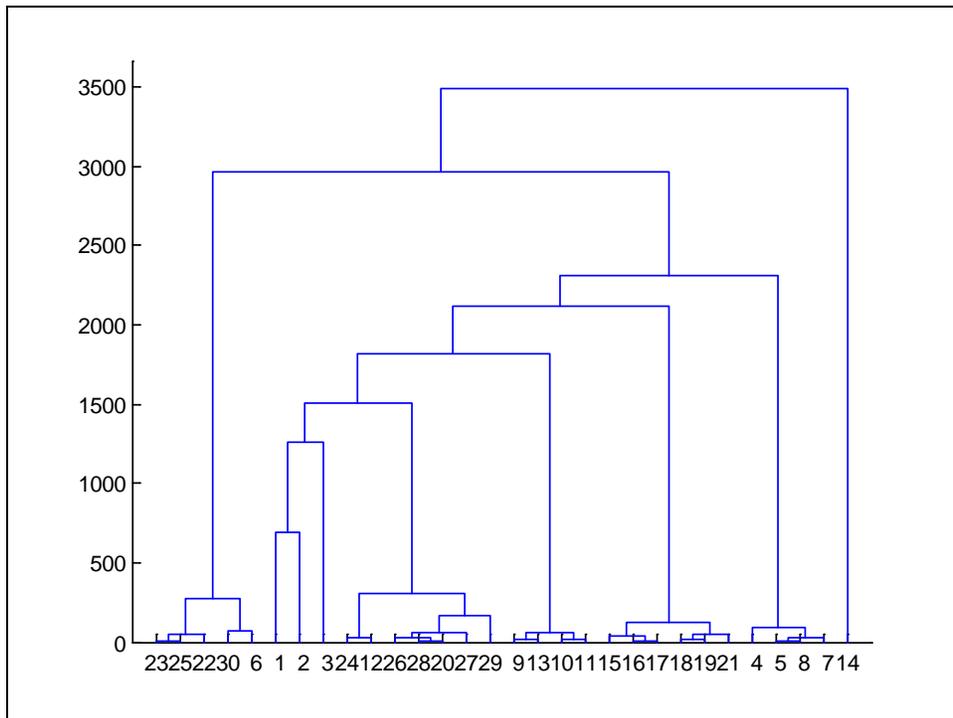


Figura 6-34 Dendrograma obtenido para el escenario (d).

Como se aprecia en la Figura 6-35 el índice C_B apunta hacia la partición de 4 clusters.

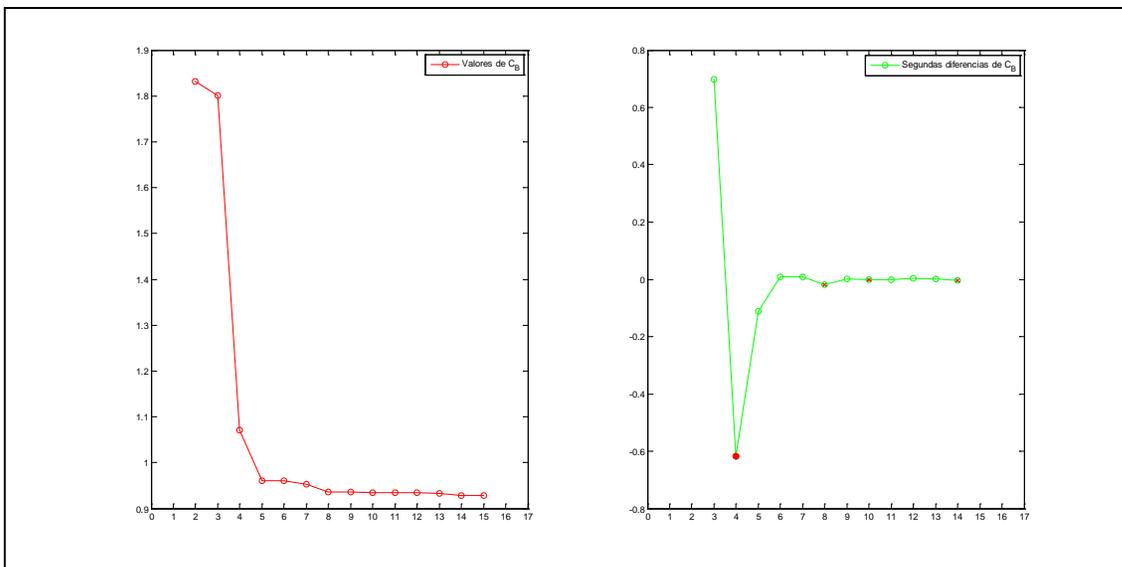


Figura 6-35 Comportamiento del índice C_B en el Escenario (d).

La Figura 6-36 muestra las particiones obtenidas en este caso. Esta vez las particiones resultan más “nítidas” y resaltan bien las zonas oscuras y las de más alta intensidad.

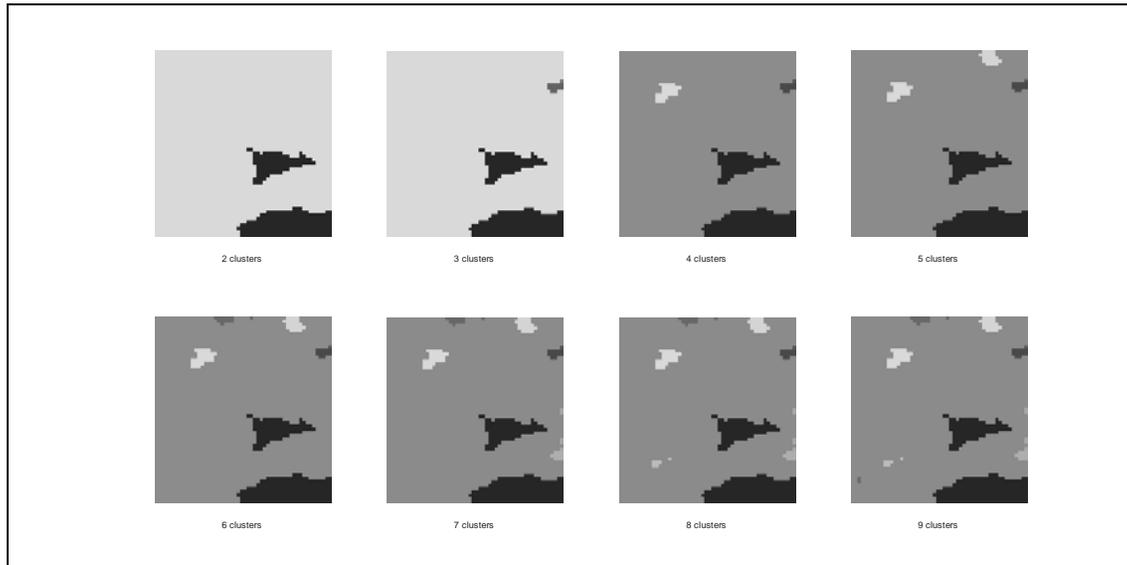


Figura 6-36 Particiones obtenidas en el Escenario (e).

Como se comentó antes, y por efecto de la etapa de reducción, las particiones subsiguientes solamente incluirán refinamientos de los clusters ya existentes teniendo como tope los 35 clusters generados en la etapa de reducción.

6.8.4.2.5 Escenario (e)

Este escenario consiste en la corrida del método *k-means* sobre los píxeles de la imagen, considerando sólo los valores de intensidad sin tener en cuenta la estructura espacial. La implementación utilizada realiza 10 corridas del algoritmo con centroides iniciales aleatorios y devuelve como resultado la partición de menor error cuadrático.

El método fue corrido para obtener las particiones con cardinalidad desde 2 hasta 9 clusters (ver Figura 6-37).

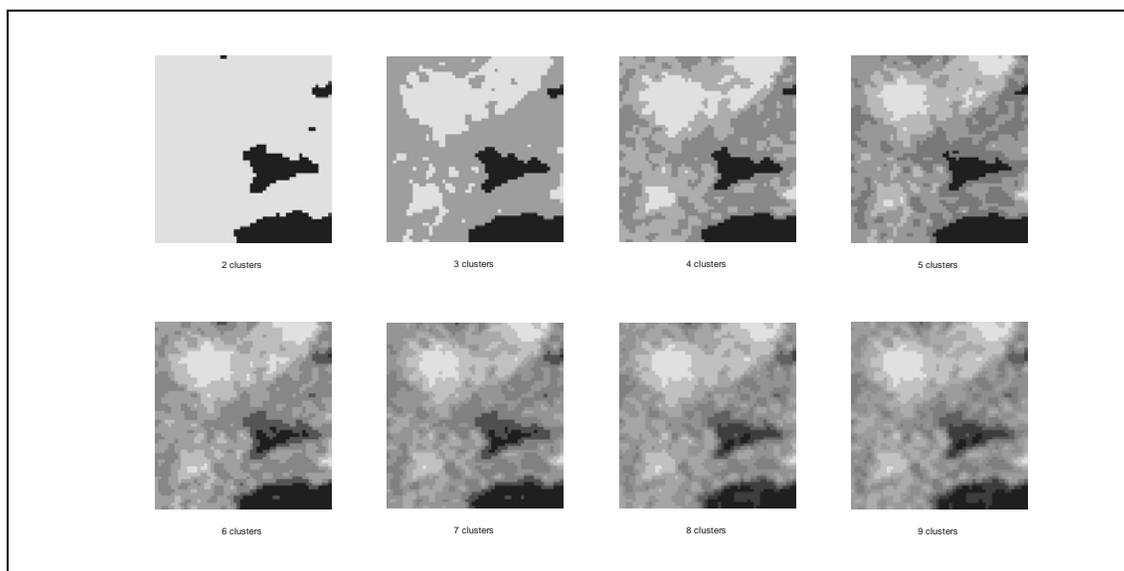


Figura 6-37 Particiones obtenidas en el escenario (a).

Los tiempos obtenidos se muestran en la Tabla 6-4.

Tabla 6-4 Tiempos del algoritmo *k-means*.

K	2	3	4	5	6	7	8	9	PROMEDIO
Tiempo (segundos)	0.021	0.022	0.015	0.020	0.057	0.064	0.028	0.074	0.0376

Las particiones más naturales, desde el punto de vista de la segmentación, obtenidas con el método *k-means* son las de 2 y 3 clusters. En la partición de cardinalidad 2 se diferencian las regiones oscuras y las más claras, mientras en la partición de cardinalidad 3 se revela además la zona de intensidad más alta.

Un aspecto positivo a resaltar son los bajos tiempos de ejecución, inferiores a todos los casos examinados anteriormente. En cuanto a la calidad de las segmentaciones se aprecia que ya partir de la partición de 4 clusters se comienza a manifestar un cierto efecto de sobre-segmentación. De hecho, la partición de 9 clusters es muy similar visualmente con la imagen original.

6.8.4.2.6 Resultados

La utilización del heurístico como paso previo a la aplicación del método propuesto, independientemente de que presupone cierta pérdida de información acerca de las primeras

etapas de la construcción del dendrograma, permite reducir de manera significativa la cantidad de datos a procesar.

Como resultado de la reducción de los datos que provoca la utilización del heurístico, se obtienen reducciones importantes del tiempo de ejecución. En la Tabla 6-5 se resumen los tiempos de ejecución asociados a cada uno de los escenarios analizados.

Tabla 6-5 Tiempos de ejecución para cada uno de los escenarios analizados.

ESCENARIO	TIEMPO DE PROCESAMIENTO (SEGUNDOS)	
<i>Average-link</i>	209.233	
Método propuesto sin heurístico	55.511	*
Método propuesto con heurístico (umbral 0)	8.765	*
Método propuesto con heurístico (umbral 1)	8.824	*
<i>k-means</i> (tiempo promedio)	0.038	

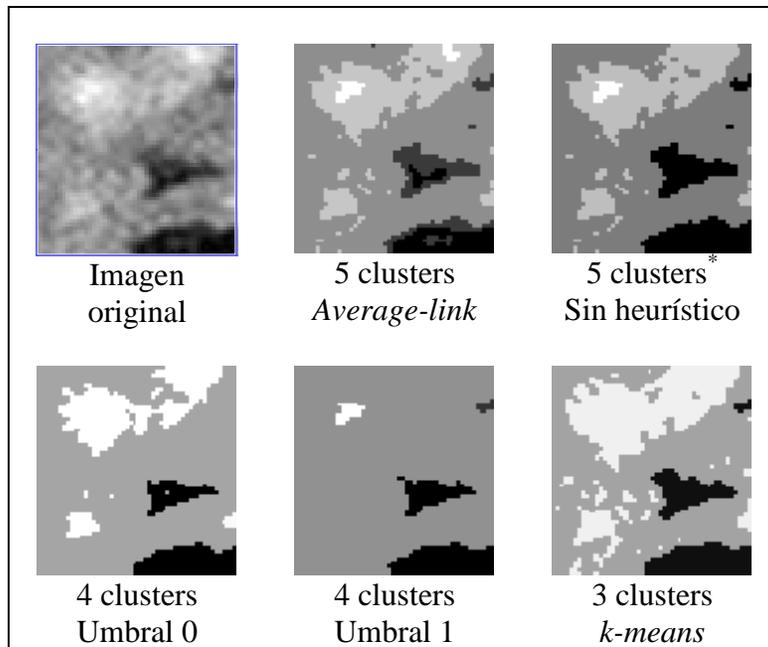
(*) Incluye los tiempos de las etapas previas al algoritmo propuesto

A partir de los dendrogramas construidos utilizando el heurístico se obtienen particiones coherentes respecto a aquellas que resultan de la aplicación del método directamente y que son además comparables con la de los métodos clásicos analizados, aún cuando estos últimos no tienen en cuenta las restricciones de contigüidad espacial impuestas a los datos.

El aumento del umbral conduce a un aumento en la cantidad de enlaces entre píxeles vecinos que se activan. Ello a su vez conduce a una excesiva reducción de los datos que, por lo general, no es un fenómeno muy deseado.

Finalmente, en la Tabla 6-6 se presentan a manera de comparación visual las particiones principales obtenidas en cada uno de los 5 escenarios analizados.

Tabla 6-6 Particiones seleccionadas en cada uno de los escenarios analizados.



(*) Segunda opción (ver apartado 6.8.4.2.3)

6.8.5 Conclusiones

En este apartado se ha estudiado el heurístico de reducción de aristas para acelerar del método propuesto en el caso de imágenes médicas reales.

El umbral mínimo utilizado ha dado resultados satisfactorios en cuanto a la reducción del tiempo de ejecución sin pérdida de calidad visual de las particiones que se obtienen.

El aumento mínimo del umbral no ha supuesto una mejora en el tiempo de ejecución ni tampoco en la calidad (visual) de la segmentación. Tomar umbrales mayores ocasiona una reducción excesiva en la cantidad de clusters, y esto puede conllevar eventualmente a la fusión de todos los píxeles en único cluster. Este efecto está asociado al radio de vecindad.

Colateralmente se ha observado la influencia de la definición del radio de vecindad en el tiempo de ejecución del algoritmo.

El algoritmo *k-means* es, con diferencia, el más rápido pero, la solución que ofrece es de baja calidad en comparación con los resultados obtenidos por el método propuesto en sus distintas variantes.

PARTE IV

7 Conclusiones y futuras investigaciones

En este trabajo se ha realizado un estudio sobre las relaciones espaciales entre los datos y la posibilidad de incluir esta información en el proceso de clasificación automática de los mismos. Se ha propuesto un nuevo método de agrupamiento jerárquico aglomerativo que tiene en cuenta las relaciones espaciales de los datos. Los capítulos 4, 5 y 6 contienen los principales análisis y resultados. En este capítulo se resumen los resultados principales logrados y se comentan varias posibilidades de trabajos futuros.

7.1 Conclusiones

En la Parte I se presenta una revisión de los principales conceptos relacionados con las imágenes digitales, en particular las imágenes médicas. Se ofrece también una panorámica de los métodos de clustering y de los métodos de segmentación de imágenes. Se destaca la idea de que la segmentación de una imagen se puede ver como la partición de los píxeles en clases homogéneas y en las semejanzas como las vías para su obtención.

La Parte II está dedicada al estudio de las interacciones espaciales entre los objetos de un conjunto. Se ha realizado un estudio basado en las nociones de autocorrelación espacial de Moran y Geary. A partir de considerar como unidades estadísticas las relaciones espaciales entre los objetos, en lugar de considerar los objetos en sí, se han modificado las definiciones de media y varianza utilizando los pesos espaciales como elementos de ponderación.

Utilizando las versiones modificadas de la media y la varianza se introducen las versiones modificadas de los índices de autocorrelación espacial de Moran y Geary. Se logra entonces establecer la complementariedad entre ambos índices modificados. Además se propone un índice de autocorrelación espacial a partir del concepto de variabilidad local de Lebart.

Las fórmulas de descomposición de la varianza que se han derivado brindan la posibilidad de definir nuevos índices de autocorrelación espacial cuyas propiedades han sido analizadas, así como sus relaciones con los índices de autocorrelación espacial mencionados anteriormente.

La metodología LISA (Local Indicators of Spatial Spatial Association) ha permitido asociar índices locales a cada uno de los índices globales de autocorrelación espacial. Ellos posibilitan profundizar en el conocimiento de la estructura espacial de los datos.

La aplicación del índice de Geary al estudio de las particiones de un conjunto de objetos con relaciones espaciales ha determinado la definición de un índice de autocorrelación espacial asociado a particiones, C_B , que toma en cuenta únicamente las fronteras entre las clases de una partición.

El sentido intuitivo del índice C_B es el siguiente. Debido a la relación que se establece entre los objetos fronterizos de los clusters, su valor caracteriza la autocorrelación espacial de las fronteras.

Se han analizado casos particulares de estructura espacial y el comportamiento en estos casos de los índices de autocorrelación espacial bajo estudio.

El caso habitual de datos no estructurados ha resultado ser un caso particular en este contexto.

La Parte III está dedicada a los métodos de clustering con restricciones.

En el capítulo 5 se presenta el estudio teórico de un método de agrupamiento jerárquico aglomerativo con restricciones de contigüidad. Las restricciones de contigüidad se modelan mediante una matriz o grafo de contigüidad.

Se introduce una función de desemejanza que toma en cuenta las relaciones de contigüidad impuestas sobre los datos y se analizan sus propiedades. Se demuestra que la jerarquía de clusters que se obtiene con el método propuesto al utilizar una estrategia aglomerativa restringida a la fusión de clusters adyacentes, es libre de inversiones.

Varias alternativas de implementación son propuestas y se realiza, además, un análisis de la complejidad computacional. El análisis de los tiempos de ejecución al procesar imágenes permite plantear, de manera empírica, que el comportamiento del método propuesto es sub-cuadrático. Mediante un ejemplo en el campo de la segmentación de imágenes se ilustran las principales características del método propuesto.

En el capítulo 6 se evalúa el método propuesto tomando como área de aplicación la segmentación de imágenes médicas, específicamente las imágenes del cerebro. Los objetos a clasificar, en este caso, son los píxeles de una imagen. Las relaciones de contigüidad tienen en cuenta la disposición espacial. En esta evaluación se han utilizado tanto imágenes simuladas como imágenes reales.

Con vistas a la selección del radio de vecindad a utilizar en el proceso de clasificación en este tipo de aplicación, se propone un procedimiento heurístico que se basa en la idea de la variabilidad local. Los resultados obtenidos corroboran, de manera empírica, la calidad de las particiones obtenidas.

El índice C_B , asumido y utilizado para realizar el corte de los dendrogramas que se obtienen tanto en el caso de imágenes simuladas como en imágenes reales, se comporta de manera satisfactoria. Los resultados que se obtienen están en concordancia con los que sugieren el índice de Calinski-Harabazs y el de Hartigan.

Un elemento a favor de la calidad de las particiones resultantes lo constituyen los altos valores del índice ajustado de Rand en el caso de las imágenes simuladas. Para las imágenes reales las segmentaciones que se obtienen se acercan en gran medida a los resultados disponibles tanto en forma de segmentaciones manuales, como en comparación con otros métodos de segmentación.

Se ha propuesto un heurístico de reducción de clases para acelerar del método propuesto en el caso de imágenes médicas reales. Como resultado de la reducción de los datos se obtienen disminuciones importantes del tiempo de ejecución sin pérdida de calidad visual de las particiones que se obtienen.

7.2 *Futuras investigaciones*

Durante la ejecución del trabajo han surgido varias cuestiones interesantes que pueden ser objetos de futuras investigaciones.

7.2.1 **Capítulo 4**

Los resultados presentados en el capítulo 4 pueden ser estudiados desde la perspectiva multivariante. También es posible generalizar la interpretación de los pesos espaciales a_{ij} al intervalo $[0,1]$.

En el apartado 2.2 se interpreta la descomposición de la varianza ($T=W+B$) como una suma de matrices. En el caso c) del apartado 4.2.6.3 se considera que el análisis de datos habitual en que los objetos no están relacionados entre sí, es el caso en que todos los objetos son adyacentes entre sí, salvo consigo mismo. Este paralelismo sugiere la posibilidad de analizar el problema de la búsqueda de la partición óptima en el análisis clásico mediante el criterio C_B .

En este trabajo se ha considerado la búsqueda de la partición óptima (en general, local) mediante un criterio parecido a C_B (su numerador promediado), y la estrategia del algoritmo jerárquico aglomerativo del enlace tipo promedio con restricciones.

Posteriormente se ha analizado el corte del dendrograma mediante C_B , que se ha mostrado interesante. Proceder directamente con algoritmos aglomerativos y el criterio (una desemejanza basada en) C_B , directamente, plantea los problemas de la posibilidad de inversiones y el de la complejidad computacional, que de ser resueltos positivamente ofrecerían a la comunidad una nueva herramienta de análisis de datos que integraría los datos clásicos con los temporales, espaciales y en general con todos aquellos que pudieran representarse mediante grafos.

7.2.2 Capítulo 5

En el apartado 5.8 se analiza el comportamiento del algoritmo jerárquico aglomerativo con restricciones contigüidad. Para una clase dada se define su vecindad $V^i(P_k)$ y la cardinalidad de ésta: $\#(V^i(P_k))$. Aquí surge un problema interesante acerca de la evolución de la cardinalidad promedio para un grafo arbitrario partiendo de que se conozca además $\#_{\max}^0$ y sometido a contracciones sucesivas.

Queda abierta la posibilidad de trabajar en el refinamiento de las implementaciones usadas tanto en el procesamiento de las vecindades, como del algoritmo de clasificación usado.

7.2.3 Capítulo 6

El trabajo experimental que se realiza en el capítulo 6 puede profundizarse y ampliarse.

Los experimentos utilizando imágenes simuladas se pueden extender a la valoración del ruido y heterogeneidades de intensidad en los resultados que se obtienen con el método propuesto.

Un análisis más completo puede ser realizado en el apartado 6.7.6 con el fin de poder comparar más detalladamente el comportamiento del método propuesto en este contexto con los resultados disponibles de otros métodos.

Un estudio adicional se puede realizar acerca del comportamiento de la diversidad local en función de los radios y del contenido de las imágenes, es decir, la distribución y tamaño de los objetos presentes en ella.

Los índices locales de autocorrelación espacial analizados en el capítulo 4 pueden ser incluidos en el proceso de segmentación de las imágenes utilizando el método propuesto con vistas a la evaluación de sus comportamientos. También es posible aplicar los índices locales en las fronteras de una segmentación con el objetivo de intentar una mejora mediante un proceso de reasignación de las pertenencias de los píxeles.

Como trabajos futuros también se recomienda la evaluación del método propuesto en otros contextos de aplicación como son los casos, por ejemplo, del análisis de guiones cinematográficos y televisivos que se comenta en el cuerpo del trabajo o de redes sociales.

Referencias

- [1] A. J. Worth, N. Makris, V. S. Caviness, and D. N. Kennedy, "Neuroanatomical Segmentation in MRI: Technological Objectives," *Journal of Pattern Recognition and Artificial Intelligence: Processing of MR Images of Human Brain*, 1997.
- [2] R. M. Haralick, K. Shanmugan, and I. Dinstein, "Texture features for image classification," *IEEE Transactions on Systems, Man, Cybernetics*, vol. 3, pp. 610-612, 1973.
- [3] R. M. Haralick and L. G. Shapiro, "Survey, image segmentation techniques," *Computer Vision, Graphics, Image Processing*, vol. 29, pp. 100-132, 1985.
- [4] T. Pavlidis, *Structural Pattern Recognition*. New York: Springer, 1997.
- [5] L. Zadeh, "Fuzzy Sets," *Information and Control*, vol. 8, pp. 338-353, 1965.
- [6] A. K. Jain, M. N. Murthy, and P. J. Flynn, "Data Clustering: A Review," *ACM Computing Surveys*, vol. 31, pp. 264-323, 1999.
- [7] E. Concepción and Y. Yurramendi, "Consideraciones sobre la segmentación de imágenes médicas mediante métodos de clasificación no supervisada," in *II Conferencia Científica de la Universidad de las Ciencias Informáticas, UCIENCIA* La Habana, Jul, 2006.
- [8] J. Hartigan, *Clustering algorithms*. New York: Wiley, 1975.
- [9] N. Jardine and R. Sibson, *Mathematical taxonomy*. London: Wiley, 1971.
- [10] J. Puzicha, T. Hofmann, and J. Buhmann, "A theory of proximity based clustering: Structure detection by optimization," *Pattern Recognition*, vol. 33, pp. 617-634, 1999.
- [11] W. Wright, "A formalization of cluster analysis," *Pattern Recognition*, vol. 5, pp. 273-282, 1973.
- [12] J. Kleinberg, "An impossibility theorem for clustering," in *Advances in Neural Information*, Cambridge, MA, 2003, pp. 446-453.
- [13] V. Hosel and S. Walcher, "Clustering Techniques: A Brief Review," 2001.
- [14] U. von Luxburg and S. Ben-David, "Towards a statistical theory for clustering," in *PASCAL Workshop on Statistics and Optimization of Clustering Workshop*, London, U.K., 2005.
- [15] A. K. Jain and R. C. Dubes, *Algorithms for Clustering Data*. Englewoods Cliffs, NJ: Prentice Hall, 1988.
- [16] K. S. Fu and J. K. Mui, "A survey on image segmentation," *Pattern Recognition*, vol. 13, pp. 3-16, 1981.
- [17] R. P. W. Duin, F. Roli, and D. De Ridder, "A note on core research issues for statistical pattern recognition," *Pattern Recognition Letters*, vol. 23, pp. 493-499, 2002.
- [18] S. Edelman, "Representation and Recognition in Vision," MIT Press 1999.
- [19] R. P. W. Duin and E. Pekalska, "Open issues in pattern recognition, invited talk," in *International Conference on Computer Recognition Systems*, Rydzyna, Poland, 2005, pp. 27-42.

-
- [20] E. Pekalska, *Dissimilarity representations in pattern recognition. Concepts, theory and applications*, *ASCI Dissertation Series*, 109. Delft: Delft University of Technology, 2005.
- [21] A. Hadjiprocopis, W. Rashid, and P. S. Tofts, "Unbiased segmentation of diffusion-weighted magnetic resonance images of the brain using iterative clustering," *Magnetic Resonance Imaging*, vol. 23, pp. 877–885, 2005.
- [22] D. L. Pham, C. Xu, and J. L. Prince, "A survey of current methods in medical image segmentation," in *Technical Report JHU/ECE 99-01* The Johns Hopkins University, 1998.
- [23] L. Hermes, T. Zoller, and J. M. Buhmann, "Parametric Distributional Clustering for Image Segmentation," in *Computer Vision-CCV, LNCS 2352*, 2002, pp. 577-591.
- [24] L. Goldfarb and J. Hook, "Why classical models for pattern recognition are not pattern recognition models," in *International Conference on Advances in Pattern Recognition*: Springer, 1998, pp. 405-414.
- [25] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed. New York: Wiley, 2000.
- [26] J. M. Buhmann, "Learning and Image Segmentation," in *MSRI Introductory Workshop*, 2005.
- [27] J. Puzicha, T. Hofmann, and J. Buhmann, "Histogram clustering for unsupervised segmentation and image retrieval," *Pattern Recognition Letters*, vol. 20, pp. 899-909, 1999.
- [28] T. Seidl and H. P. Kriegel, "Adaptable Similarity Search In Large Image Databases," in *State-of-the-Art in Content-Based Image and Video Retrieval*, 2001.
- [29] J. Gower, "Measures of similarity, dissimilarity, and distance," in *Encyclopedia of Statistical Sciences*. vol. 5, S. Kotz and N. Johnson, Eds. New York: Wiley, 1985, pp. 397–405.
- [30] C. Chu and J. K. Aggarwal, "The integration of image segmentation using region and edge information," *IEEE Trans. Pattern Anal. Mach. Intelligence*, vol. 15, pp. 1241–1252, 1993.
- [31] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Trans. Patt. Anal. Mach. Intelligence*, vol. 22, pp. 888–905, 2000.
- [32] R. E. Korf, "Search," in *Encyclopedia of Artificial Intelligence*, 1990, pp. 994-998.
- [33] J. J. Fortier and H. Solomon, "Clustering procedures," in *Multivariate Analysis*, New York, 1966, pp. 493-506.
- [34] D. Hand, H. Mannila, and P. Smyth, *Principles of data mining*. Cambridge, MA: The MIT Press, 2001.
- [35] B. Fischer, T. Zoller, and J. M. Buhmann, "Path based pairwise data clustering with application to texture segmentation," in *LNCS energy minimization methods in computer vision and pattern recognition*, Berlin, 2001.
- [36] S. K. Pal and P. P. Wang, *Genetic Algorithms for Pattern Recognition*, ed. Boca Raton: CRC Press, 1996.
- [37] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, "Optimization by simulated annealing," *Science*, vol. 220, pp. 671-680, 1983.
- [38] S. Régnier, "Sur quelques aspects mathématiques des problèmes de classification automatique," *Mathématique et sciences humaines*, vol. 82, pp. 13-29, 1983.

-
- [39] E. Forgy, "Cluster analysis of multivariate data: Efficiency versus interpretability of classification," *Biometrics*, vol. 21, pp. 768-780, 1965.
- [40] P. S. Bradley and U. M. Fayyad, "Refining initial points for k-means clustering," in *Proceedings of Fifteenth International Conf. on Machine Learning* Madison, WI, 1998, pp. 91-99.
- [41] G. Ball and D. Hall, "ISODATA, a novel method of data analysis and classification," in *Technical Report AD-699616*, SRI, Ed. Standford, CA, 1965.
- [42] Z. Huang, "Extensions to the k-means algorithm for clustering large data sets with categorical values," *Data Mining and Knowledge Discovery*, vol. 2, pp. 283-304, 1998.
- [43] L. Kaufman and P. Rousseeuw, *Finding Groups in Data: An Introduction to Cluster Analysis*. New York: John Wiley and Sons, 1990.
- [44] R. Ng and J. Han, "Efficient and effective clustering methods for spatial data mining," in *Proceedings of the 20th Conference on VLDB*, Santiago, Chile, 1994, pp. 144-155.
- [45] G. McLachlan and K. Basford, *Mixture Models: Inference and Applications to Clustering*. New York: Marcel Dekker, 1988.
- [46] A. Dempster, N. Laird, and D. Rubin, "Maximun likelihood from incomplete data via the EM algorithm," *Journal of Royal Statistical Society*, vol. 39, pp. 1-22, 1977.
- [47] C. Fraley and A. E. Raftery, "Model-based Clustering, Discriminant Analysis and Density Estimation," in *Technical Report No. 380* Department of Statistics: University of Washington, 2000.
- [48] J.-M. Bouroche and G. Saporta, *L'analyse des donnees, Que sais-je?* Paris: Presses Universitaires de France, 1980.
- [49] M. Jambu and M.-O. Lebeaux, *Cluster analysis and data analysis*. North-Holland, Amsterdam, 1983.
- [50] G. N. Lance and W. T. Williams, "A general theory of classificatory sorting strategies. I. hierarchical systems," *Computer Journal*, vol. 9, pp. 373-380, 1967.
- [51] S. M. Savaresi, D. L. Boley, S. Bittanti, and G. Gazzaniga, "Cluster Selection in divisive clustering algorithms," in *Proceedings of the 2nd SIAM ICDM*, Arlington, VA, 2002, pp. 299-314.
- [52] S. Guha, R. Rastogi, and K. Shim, "CURE: An efficient clustering algorithm for large databases," in *Proceedings of the ACM SIGMOD International Conference on Management of Data* New York, 1998, pp. 73-84.
- [53] G. Karypis, E. H. Han, and V. Kumar, "CHAMELEON: A hierarchical clustering algorithm using dynamic modeling," *IEEE Computer, Special Issue on Data Analysis and Mining*, vol. 32, pp. 68-75, 1999.
- [54] F. Murtagh, "Complexities of hierarchic clustering algorithms: state of the art," *Computational Statistic Quarterly*, vol. 1, pp. 101-113, 1984.
- [55] F. Murtagh, "Clustering in massive data sets," in *Handbook of massive data sets*, P. Pardalos, J. Abello, and M. Reisende, Eds. Norwell, MA, USA: Kluwer Academic Publishers, 2002, pp. 501-543.
- [56] J. P. Benzécri, "Construction d'une classification ascendante hiérarchique par la recherche en chaîne des voisins réciproques," *Les Cahiers de l'Anayse des Données*, vol. VII, pp. 209-218, 1982.

-
- [57] M. Bruynooghe, "Méthodes Nouvelles en Classification Automatique des Données Taxinomiques Nombreuses," *Statistique et Analyse des Données*, vol. 3, pp. 24-42, 1977.
- [58] E. Diday, "Inversions en classification hiérarchique: application à la construction adaptative d'indices d'agrégation," *Revue de statistique appliquée*, vol. 31, pp. 45-62, 1983.
- [59] C. Ding and X. He, "Cluster Aggregate Inequality and Multi-Level Hierarchical Clustering," in *European Conf. Principles of Data Mining and Knowledge Discovery (PKDD 2005)*, Porto, 2005.
- [60] A. Y. Ng, M. I. Jordan, and Y. Weiss, "On spectral clustering: Analysis and an algorithm," in *Advances in Neural Information Processing Systems*, 2002.
- [61] I. T. Jolliffe, *Principal Component Analysis*. New York: Springer-Verlag, 1986.
- [62] T. Cox and M. Cox, *Multidimensional Scaling*. London: Chapman & Hall, 1994.
- [63] C. J. C. Burges, "Geometric methods for feature extraction and dimensional reduction," in *Data Mining and Knowledge Discovery Handbook: A Complete Guide for Practitioners and Researchers*, 2005.
- [64] S. Singh and K. J. Bovis, "Medical image segmentation in digital mammography," in *Advanced Algorithmic Approaches to Medical Image Segmentation*, London, 2001.
- [65] N. R. Pal and S. K. Pal, "A review on image segmentation techniques," *Pattern Recognition*, vol. 26, pp. 1277-1294, 1993.
- [66] J. K. Udupa, "Three-dimensional Visualization and Analysis Methodologies: A Current Perspective," *Radiographics*, vol. 19, pp. 783-806, 1999.
- [67] Y. J. Zhang, "A survey on evaluation methods for image segmentation," *Patt. Rec. Let.*, vol. 29, pp. 1335-1346, 1996.
- [68] Y. Zhuge, J. K. Udupa, and P. K. Saha, "Vectorial scale-based fuzzy-connected image segmentation," *Computer Vision and Image Understanding*, vol. 101, pp. 177-193, 2006.
- [69] D. Morgenthaler and A. Rosenfeld, "Multidimensional edge detection by hypersurface fitting," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 3, pp. 482-486, 1981.
- [70] T. McInerney and D. Terzopoulos, "Deformable models in medical image analysis: a survey," *Med. Image Anal.*, vol. 1, pp. 91-108, 1996.
- [71] A. X. Falcao, J. K. Udupa, S. Samarasekera, S. Sharma, B. E. Hirsch, and R. A. Lotufo, "User-steered image segmentation paradigms: live wire and live lane," *Graph. Models Image Process.*, vol. 60, pp. 233-260, 1998.
- [72] E. Mortensen and W. Barrett, "Intelligent scissors for image composition," in *Proc. SIGGRAPH_95*, Los Angeles, 1995, pp. 191-198.
- [73] T. Cootes, G. Edwards, and C. Taylor, "Active appearance models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, pp. 681-685, 2001.
- [74] T. Cootes, C. Taylor, and D. Cooper, "Active shape models - their training and application," *Computer Vision and Image Understanding*, vol. 61, pp. 38-59, 1995.
- [75] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Trans. Systems, Man, and Cybernetics*, vol. SMC-9, pp. 62-66, 1979.

-
- [76] A. K. C. Wong and P. K. Sahoo, "A gray-level threshold selection method based on maximum entropy principle," *IEEE Trans. Systems, Man, Cybernetics*, vol. 19, pp. 866–871, 1989.
- [77] Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, pp. 1222–1239, 2001.
- [78] Z. Wu and Z. Leahy, "An optimal graph theoretic approach to data clustering: theory and its application to image segmentation," *IEEE Trans. Pattern Anal. Mach. Intelligence*, vol. 15, pp. 1101–1113, 1993.
- [79] D. L. Collins, A. P. Zijdenbos, V. Kollokian, J. G. Sled, and N. J. Kabani, "Design and construction of a realistic digital brain phantom," *IEEE T. Med. Imag.*, vol. 17, pp. 463–468, 1998.
- [80] P. Meer and B. Georgescu, "Edge Detection with Embedded Confidence," *IEEE Transactions On Pattern Analysis And Machine Intelligence*, vol. 23, pp. 1351–1365, 2001.
- [81] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, 2nd ed. Upper Saddle River, NJ: Prentice Hall, 2002.
- [82] J. Canny, "A computational approach to edge detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 8, pp. 269–285, 1986.
- [83] P. V. C. Hough, *Methods and Means for Recognizing Complex Patterns*. USA: US Patent No. 3,069,654, 1962.
- [84] N. Kiryati and A. M. Bruckstein, "Heteroscedastic Hough transform (HtHT): An efficient method for robust line fitting in the 'errors in the variables' problem," *Comput. Vision Image Understand.*, vol. 78, pp. 69–83, 2000.
- [85] J. Matas, C. Galambos, and J. Kittler, "Robust detection of lines using progressive probabilistic Hough transform," *Comput. Vision Image Understand.*, vol. 78, pp. 119–137, 2000.
- [86] A. Singh, D. Goldgof, and D. Terzopoulos, *Deformable models in medical image analysis*: IEEE Computer Society, 1998.
- [87] M. Sonka, "Digital Image Processing, Image Analysis and Understanding," vol. 2005, 1998.
- [88] P. K. Sahoo, S. Soltani, A. K. C. Wong, and Y. C. Chen, "A survey of thresholding techniques," *Comput. Vision Graphics Image Processing*, vol. 41, pp. 233–260, 1988.
- [89] M. Sezgin and B. Sankur, "Survey over image thresholding techniques and quantitative performance evaluation," *Journal of Electronic Imaging*, vol. 13, pp. 146–165, 2004.
- [90] T. W. Ridler and S. Calvard, "Picture thresholding using an iterative selection method," *IEEE Trans. Syst. Man Cybernetics*, vol. SMC-8, pp. 630–632, 1978.
- [91] H. Lee and R. H. Park, "Comments on an optimal threshold scheme for image segmentation," *IEEE Trans. Syst. Man Cybern. SMC*, vol. 20, pp. 741–742, 1990.
- [92] J. Z. Liu and W. Q. Li, "The automatic thresholding of gray-level pictures via two-dimensional Otsu method," *Acta Automatica Sin.*, vol. 19, pp. 101–105, 1993.
- [93] C. V. Jawahar, P. K. Biswas, and A. K. Ray, "Investigations on fuzzy thresholding based on fuzzy clustering," *Pattern Recognition*, vol. 30, pp. 1605–1613, 1997.

- [94] R. L. Kirby and A. Rosenfeld, "A note on the use of gray level, local average gray level space as an aid in threshold selection," *IEEE Trans. Syst. Man Cybern. SMC*, vol. 9, pp. 860–864, 1979.
- [95] B. Chanda, B. B. Chaudhuri, and D. D. Majumder, "On image enhancement and threshold selection using the gray level co-occurrence matrix," *Pattern Recognition Letters*, vol. 3, pp. 243-251, 1985.
- [96] R. M. Haralick, "Statistical and structural approaches to texture," *Proceedings of the IEEE*, vol. 67, pp. 786-804, 1979.
- [97] N. R. Pal and S. K. Pal, "Entropic thresholding," *Signal Processing*, vol. 16, pp. 97–108, 1989.
- [98] C. Lee, S. Hun, T. A. Ketter, and M. Unser, "Unsupervised connectivity-based thresholding segmentation of midsagittal brain MR images," *Comput. Biol. Med.*, vol. 28, pp. 309–338, 1998.
- [99] H. Suzuki and J. Toriwaki, "Automatic Segmentation of Head MRI Images by Knowledge Guided Thresholding," *Computerized Medical Imaging and Graphics*, vol. 15, p. 233 240, 1991.
- [100] P. A. Bottomley, T. H. Foster, R. E. Argersinger, and L. M. Pfeifer, "A review of normal tissue hydrogen NMR relaxation times and relaxation mechanisms from 1-100 MHz: Dependence on tissue type, NMR frequency, temperature, species, excision, and age," *Medical Physics*, vol. 11, pp. 425-448, 1984.
- [101] R. Momenan, D. Hommer, R. Rawlings, U. Ruttimann, M. Kerich, and D. Rio, "Intensity-Adaptive Segmentation of Single-Echo T1-Weighted Magnetic Resonance Images," *Human Brain Mapping*, vol. 5, pp. 194-205, 1997.
- [102] M. O'Donnell, J. C. Gore, and W. J. Adams, "Toward an Automated Analysis System for Magnetic Resonance Imaging. II. Initial Segmentation Algorithm," *Medical Physics*, vol. 13, pp. 293-297, 1986.
- [103] A. J. Worth, N. Makris, M. R. Patti, J. M. Goodman, E. A. Hoge, V. S. Caviness Jr., and D. N. Kennedy, "Precise Segmentation of the Lateral Ventricles and Caudate Nucleus in MR Brain Images Using Anatomically Driven Histograms," *IEEE Transactions on Medical Imaging*, vol. 17, pp. 303-310, 1998.
- [104] S. W. Zucker, "Region growing: Childhood and adolescence," *Computer Graphics and Image Processing*, vol. 5, pp. 382-399, 1977.
- [105] P. K. Saha and J. K. Udupa, "Fuzzy Connected Object Delineation: Axiomatic Path Strength Definition and the case of Multiple Seeds," *Computer Vision and Image Understanding*, vol. 83, pp. 275-295, 2001.
- [106] J. K. Udupa and S. Samarasekera, "Fuzzy Connectedness and Object Definition: Theory, Algorithms, and Applications in Image Segmentation," *Graphical Models and Image Processing*, vol. 58, pp. 246-261, 1996.
- [107] P. K. Saha and J. K. Udupa, "Relative Fuzzy Connectedness among Multiple Objects: Theory, Algorithms, and Applications in Image Segmentation," *Computer Vision and Image Understanding*, vol. 82, pp. 42-56, 2001.
- [108] J. K. Udupa, P. K. Saha, and R. A. Lotufo, "Relative Fuzzy Connectedness and Object Definition: Theory, Algorithms, and Applications in Image Segmentation," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 24, pp. 1485-1500, 2002.

-
- [109] S. Beucher and F. Meyer, "The morphological approach to segmentation: The watershed transform," in *Mathematical Morphology in Image Processing*, 1993, pp. 433-481.
- [110] J. B. T. M. Roerdink and A. Meijster, "The watershed transform: Definitions, algorithms and parallelization strategies," *Fundamenta Informaticae*, vol. 228, pp. 141-187, 2000.
- [111] A. X. Falcao, J. Stolfi, and R. A. Lotufo, "The Image Foresting Transform: Theory, Algorithms, and Applications," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 26, pp. 19-29, 2004.
- [112] S. L. Horowitz and T. Pavlidis, "Picture Segmentation by a Directed Split-and-Merge Procedure," in *Proc. 2nd Int'l Conf. on Pattern Recognition*, 1974, pp. 424-433.
- [113] Y. Lu, T. Jiang, and Y. Zanga, "Region growing method for the analysis of functional MRI data," *NeuroImage* vol. 20, pp. 455-465, 2003.
- [114] N. Malpica, J. E. Ortuño, and A. Santos, "A multichannel watershed-based algorithm for supervised texture segmentation," *Pattern Recognition Letters*, vol. 24, pp. 1545-1554, 2003.
- [115] Y. L. Chang and X. Li, "Adaptive image region growing," *IEEE Transactions on Image Processing*, vol. 3, pp. 868-873, 1994.
- [116] R. Adams and L. Bischof, "Seeded region growing," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 16, pp. 641-647, 1994.
- [117] A. Mehnert and P. Jackway, "An improved seeded region growing algorithm," *Pattern Recognition Letters*, vol. 8, pp. 1065-1071, 1997.
- [118] J. M. Beaulieu and M. Goldberg, "Hierarchy in picture segmentation: a stepwise optimization approach," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, pp. 150-163, 1989.
- [119] T. Pavlidis and Y. Liow, "Integrating region growing and edge detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, pp. 225-233, 1990.
- [120] J. Freixenet, X. Muñoz, D. Raba, J. Martí, and X. Cufí, "Yet Another Survey on Image Segmentation: Region and Boundary Information Integration," in *European Conference on Computer Vision*, Copenhagen, Denmark, 2002, pp. 408-422.
- [121] B. Taskar, P. Abbeel, and D. Koller, "Discriminative probabilistic models for relational data," in *Proceedings of UAI-2002*, 2002.
- [122] D. Jensen and J. Neville, "Linkage and autocorrelation cause feature selection bias in relational learning," in *Proceedings of the 19th International Conference on Machine Learning*, 2002.
- [123] L. Anselin, *Spatial econometrics: Methods and models*. The Netherlands: Kluwer Academic Publisher, 1998.
- [124] G. J. Upton and B. Fingleton, *Spatial data analysis by example* vol. 1: Point pattern and quantitative data. Toronto, Canada: Wiley, 1985.
- [125] A. D. Cliff and J. K. Ord, *Spatial Autocorrelation*. London: Pion, 1973.
- [126] R. R. Sokal and N. L. Oden, "Spatial autocorrelation in biology 1. Methodology," *Biological Journal of the Linnean Society*, pp. 10-199, 1978.

- [127] L. J. Hubert, R. G. Golledge, and C. M. Constanzo, "Generalized procedures for evaluating spatial autocorrelation," *Geographical Analysis*, pp. 13-225, 1981.
- [128] A. Getis and K. J. Ord, "The analysis of spatial association by use of distance statistics," *Geographical Analysis*, vol. 24, pp. 189-206, 1992.
- [129] L. Lebart, "Analyse statistique de la contiguïté," *Publication de l'Institut de Statistiques de l'Université de Paris*, vol. 28, pp. 81-112, 1969.
- [130] L. Lebart, "Contiguity Analysis and Classification," in *Data Analysis*, Berlin, 2000, pp. 233-244.
- [131] S. Ollier, "Des outils pour l'intégration des contraintes spatiales, temporelles et évolutives en analyse des données écologiques." vol. Phd Thesis: Université Claude Bernard - Lyon 1, 2002.
- [132] P. A. P. Moran, "Notes on continuous phenomena," *Biometrika*, vol. 37, pp. 17-23, 1950.
- [133] P. de Jong, C. Sprenger, and F. van Veen, "On extreme values of Moran's I and Geary's c," *Geogr. Anal.*, vol. 16, pp. 17-24, 1984.
- [134] S. Dray, P. Legendre, and P. R. Peres-Neto, "Spatial modelling: a comprehensive framework for principal coordinate analysis of neighbour matrices (PCNM)," *Ecological Modelling*, vol. 196, pp. 483-493, 2006.
- [135] L. Anselin, "The Moran Scatterplot as an ESDA Tool to Assess Local Instability in Spatial Association," in *Spatial Analytical Perspectives on GIS*, Londres, 1996, pp. 111-125.
- [136] L. Anselin, "Local Indicators of Spatial Spatial Association-LISA," *Geographical Analysis*, vol. 27, pp. 93-115, 1995.
- [137] R. C. Geary, "The contiguity ratio and statistical mapping," *The incorporated Statistician*, vol. 5, pp. 115-145, 1954.
- [138] D. A. Griffith, *Spatial Autocorrelation, A primer*. Washington, D.C.: Association of American Geographers, 1987.
- [139] G. d'Aubigny, "Spatial autocorrelation: from Moran's I to CAR models," in *Fifth French-Danish Workshop on Spatial Statistics and Image Analysis in Biology*, Saint-Pierre de Chartreuse, France, 2004, pp. 17-20.
- [140] W. R. Tobler, "Cellular Geography," in *Philosophy in Geography*, Dordrecht, 1979, pp. 379-386.
- [141] A. D. Gordon, "A survey of constrained classification," *Computational statistics & Data Analysis*, vol. 1, pp. 17-29, 1996.
- [142] S. Zhong and J. Ghosh, "Scalable, model-based balanced clustering," in *SIAM, International Conference on Data Mining (SDM-03)*, San Francisco, CA, 2003, pp. 71-82.
- [143] A. D. Gordon, "Classification and assignment in soil science," *Soil use and management*, vol. 3, pp. 3-8, 1987.
- [144] L. Lebart, "Programme d' Agregation avec Contraintes (CAH Contiguit)," *Les Cahiers d'Analyse des Donnes*, vol. 3, pp. 275-287, 1978.
- [145] P. Legendre, "Constrained clustering," in *Developments in Numerical Ecology*, Berlin, 1987, pp. 289-307.
- [146] P. Monestiez, "Méthode de classification automatique sous contraintes spatiales," *Statistique et Analyse des Données*, vol. 3, pp. 75-84, 1977.

-
- [147] F. W. Wharton, "A contextual classification method for recognizing land use patterns in high resolution remotely sensed data," *Pattern Recognition*, vol. 15, pp. 317-324, 1982.
- [148] B. J. L. Berry, "Essay on commodity flows and the spatial structure of the indian economy," in *Research paper 111* Department of Geography: University of Chicago, 1996.
- [149] A. K. Jain and F. Farrokhnia, "Unsupervised texture segmentation using gabor filters," *Pattern Recognition*, vol. 24, pp. 1167-1186, 1991.
- [150] M. A. Oliver and R. Webster, "A geostatistical basis for spatial weighting in multivariate classification," *Mathematical Geology*, vol. 21, pp. 15-35, 1989.
- [151] B. D. Ripley, *Statistical Inference for Spatial Processes*. Cambridge: Cambridge University Press, 1988.
- [152] C. Perruchet, "Constrained agglomerative hierarchical classification," *Pattern Recognition*, vol. 16, pp. 213-217, 1983.
- [153] F. Murtagh, "Contiguity-constrained hierarchical clustering," in *Partitioning Data Sets*, I. J. Cox, P. Hansen, and B. Julesz, Eds.: DIMACS, AMS, 1994, pp. 143-152.
- [154] F. Murtagh, A. Ganz, and S. McKie, "The structure of narrative: The case of film scripts," *Pattern Recognition*, doi:10.1016/j.patcog.2008.05.026, 2008.
- [155] P. Hansen, B. Jaumard, C. Meyer, B. Simeone, and V. Doring, "Maximum Split Clustering Under Connectivity Constraints," *Journal of Classification*, vol. 20, pp. 143-180, 2003.
- [156] J. Theiler and G. Gisler, "A contiguity-enhanced k-means clustering algorithm for unsupervised multispectral image segmentation," in *Proc. SPIE*, 1997, pp. 108-118.
- [157] C. Ambroise, M. Dang, and G. Govaert, "Clustering of spatial data by the EM algorithm," in *geoENV I - Geostatistics for Environmental Applications. Quantitative Geology and Geostatistics*, 1996, pp. 493-504.
- [158] F. Murtagh, "A survey of algorithms for contiguity-constrained clustering and related problems," *The Computer Journal*, vol. 28, pp. 82-88, 1985.
- [159] P. J. Green and R. Sibson, "Computing dirichlet tessellations in the plane," *The Computer Journal*, vol. 21, pp. 168-173, 1977.
- [160] D. F. Watson, "Computing the n-dimensional Delaunay tessellation with application to Voronoi polytopes," *Computer Journal*, vol. 24, pp. 167-172, 1981.
- [161] K. R. Gabriel and R. R. Sokal, "A new statistical approach to geographic variation analysis," *Systematic Zoology*, vol. 18, pp. 258-278, 1969.
- [162] G. Toussaint, "The relative neighborhood graph of a finite planar set," *Pattern Recognition*, vol. 12, pp. 261-268, 1980.
- [163] A. Ferligoj and V. Batagelj, "Some types of clustering with relational constraints," *Psychometrika*, vol. 48, pp. 541-552, 1983.
- [164] K. Wagstaff, C. Cardie, S. Rogers, and S. Schroedl, "Constrained K-means Clustering with Background Knowledge," in *Proceedings of the Eighteenth International Conference on Machine Learning (ICML 2001)*, 2001, pp. 577-584.
- [165] I. Davidson and S. S. Ravi, "Clustering With Constraints: Feasibility Issues and the k-Means Algorithm," in *SIAM Data Mining Conference. Best Paper Award*, 2005.
- [166] S. Basu and I. Davidson, "Clustering with Constraints: Theory and Practice," in *12th ACM SIGKDD International Conference (KDD)*, 2006.

- [167] S. Basu, A. Banerjee, and R. J. Mooney, "Semi-supervised clustering by seeding," in *Proceedings of 19th International Conference on Machine Learning (ICML-2002)*, 2002, pp. 19–26.
- [168] E. P. Xing, A. Y. Ng, M. I. Jordan, and S. Russell, "Distance metric learning, with application to clustering with side-information," *Advances in Neural Information Processing Systems 15*, pp. 505–512, 2003.
- [169] A. Ferligoj and V. Batagelj, "Clustering with relational constraint," *Psychometrika*, vol. 47, pp. 413–426, 1982.
- [170] K. L. Wagstaff, "Value, Cost, and Sharing: Open Issues in Constrained Clustering," in *KDID 2006, LNCS 4747*, S. Dzeroski and J. Struyf, Eds. Berlin Heidelberg: Springer-Verlag, 2007, pp. 1–10.
- [171] I. Davidson and S. Basu, "A Survey of Clustering with Instance Level Constraints," *ACM Transactions on Knowledge Discovery from Data*, 2007.
- [172] V. Batagelj and A. Ferligoj, "Constrained clustering problems," in *Advances in Data Science and Classification*, A. Rizzi, M. Vichi, and H. Bock, Eds.: Springer-Verlag, 1998, pp. 137–144.
- [173] L. P. Lefkovich, "Conditional clustering," *Biometrics*, vol. 36, pp. 43–58, 1980.
- [174] V. Batagelj, "Generalized Ward and Related Clustering Problems," in *Classification and Related Methods of Data Analysis*, H. Bock, Ed. North-Holland, Amsterdam, 1988, pp. 67–74.
- [175] F. Murtagh, *Multidimensional Clustering Algorithms*: Physica-Verlag, 1985.
- [176] F. Murtagh, A. Ganz, and S. McKie, "The structure of narrative: The case of film scripts," *Pattern Recognition*, doi:10.1016/j.patcog.2008.05.026, 2008.
- [177] I. C. Lerman and K. Bachar, "Construction et justification d'une méthode de classification ascendante hiérarchique accélérée fondée sur le critère de la vraisemblance du lien en cas de données de contiguïté. Application en imagerie numérique," *Publication interne 1611, IRISIA, Université de Rennes*, 2004.
- [178] T. H. Cormen, C. E. Leiserson, and R. L. Rivest, *Introduction to Algorithms*. Cambridge, MA: MIT Press, 2001.
- [179] E. M. Haacke, R. W. Brown, M. R. Thompson, and R. Venkatesan, *Magnetic Resonance Imaging: Physical Principles and Sequence Design*. New York: Wiley, 1999.
- [180] X. Zeng, L. H. Staib, R. T. Schultz, and J. S. Duncan, "Segmentation and measurement of the cortex from 3D MR images using coupled surfaces propagation," *IEEE Trans. Med. Imag.*, vol. 18, pp. 100–111, 1999.
- [181] F. Barkhof, P. Scheltens, and W. Kamphorst, "Pre- and post-mortem MR imaging of unsuspected multiple sclerosis in a patient with Alzheimer's disease," *J Neurol Sci.*, vol. 8, pp. 117–175, 1993.
- [182] L. van den Hauwe, P. M. Parizel, J. J. Martin, P. Cras, P. De Deyn, and A. M. De Schepper, "Postmortem MRI of the brain with neuropathological correlation," *Neuroradiology*, vol. 9, p. 337:343, 1995.
- [183] S. Warfield, J. Dengler, J. Zaers, C. R. G. Guttmann, W. M. Wells III, G. J. Ettinger, and J. K. Hiller, R., "Automatic Identification of Gray Matter Structures from MRI to Improve the Segmentation of White Matter Lesions," *Journal of Image Guided Surgery*, vol. 1, pp. 326–338, 1995.

-
- [184] S. K. Warfield, K. H. Zou, and W. M. Wells, "Simultaneous Truth and Performance Level Estimation (STAPLE): An Algorithm for the Validation of Image Segmentation," *IEEE Transactions on Medical Imaging*, vol. 23, pp. 903-921, 2004.
- [185] G. W. Milligan and M. C. Cooper, "An examination of procedures for determining the number of clusters in a data set," *Psychometrika*, vol. 50, pp. 159-179, 1985.
- [186] A. Gordon, *Classification*, 2nd ed. London: Chapman and Hall/CRC press, 1999.
- [187] A. D. Gordon, "How many clusters? An investigation of five procedures for detecting nested cluster structure," in *Data Science, Classification, and Related Methods*, C. Hayashi, N. Ohsumi, K. Yajima, Y. Tanaka, H. Bock, and Y. Baba, Eds. Tokio: Springer-Verlag, 1998.
- [188] R. Tibshirani, G. Walther, and T. Hastie, "Estimating the number of clusters via the gap statistic," *J. Royal Statist. Soc. B*, vol. 63, pp. 411-423, 2001.
- [189] H. Lawrence and P. Arabie, "Comparing partitions," *Journal of Classification*, vol. 2, pp. 193-218, 1985.
- [190] W. M. Rand, "Objective criteria for the evaluation of clustering methods," *Journal of the American Statistical Association*, vol. 66, pp. 846-850, 1971.
- [191] R. B. Calinski and J. Harabasz, "A dendrite method for cluster analysis," *Communications in Statistics*, vol. 3, pp. 1-27, 1974.
- [192] S. Dudoit and J. Fridlyand, "A prediction-based resampling method for estimating the number of clusters in a dataset," *Genome Biology*, vol. 3, pp. 1-21, 2002.
- [193] E. Dimitriadou, S. Dolnicar, and A. Weingessel, "An Examination Of Indexes For Determining The Number Of Clusters In Binary Data Sets," *Psychometrika*, vol. 67, pp. 137-160, 2002.
- [194] "BrainWeb: Simulated Brain Database."
- [195] R. K. S. Kwan, A. C. Evans, and G. B. Pike, "MRI simulation-based evaluation of image-processing and classification methods," *IEEE Transactions on Medical Imaging*, vol. 18, pp. 1085-1097, 1999.
- [196] C. A. Cocosco, V. Kollokian, R. K. S. Kwan, and A. C. Evans, "Brainweb: Online interface to a 3D MRI simulated brain database," *Neuroimage (Proceedings of 3-rd International Conference on Functional Mapping of the Human Brain, Copenhagen)*, vol. 5, p. S425, 1997.
- [197] L. Lemieux, A. Hammers, T. Mackinnon, and R. S. N. Liu, "Automatic Segmentation of the Brain and Intracranial Cerebrospinal Fluid in T1-weighted Volume MRI Scans of the Head, and Its Application to Serial Cerebral and Intracranial Volumetry," *Magnetic Resonance in Medicine*, vol. 49, 2003.
- [198] D. W. Shattuck, S. R. Sandor-Leahy, K. A. Schaper, D. A. Rottenberg, and R. M. Leahy, "Magnetic Resonance Image Tissue Classification Using a Partial Volume Model," *NeuroImage*, vol. 13, 2001.
- [199] D. Pham and J. Prince, "Adaptive fuzzy segmentation of magnetic resonance images," *IEEE Trans Med Imaging*, vol. 18, 1999.
- [200] K. V. Leemput, F. Maes, D. Vandermeulen, and P. Suetens, "Automated model-based tissue classification of MR images of the brain," in *Technical Report Nr.KUL/ESAT/PSI/9903* Katholieke Universiteit Leuven, Medical Image Computing (ESAT-Radiology), 1999.

- [201] D. L. Collins, A. P. Zijdenbos, V. Kollokian, J. G. Sled, N. J. Kabani, C. J. Holmes, and A. C. Evans, "Design and construction of a realistic digital brain phantom," *IEEE Trans Med Imaging*, vol. 17, pp. 463–468, 1998.
- [202] "IBSR, Center for Morphometric Analysis at Massachussets General Hospital," <http://www.cma.mgh.harvard.edu/ibsr/>.
- [203] J. Simon, L. Jacobs, and M. Campion, "A longitudinal study of brain atrophy in relapsing multiple sclerosis. The multiple sclerosis collaborative research group (MSCRG)," *Neurology*, vol. 53, pp. 139-148, 1999.
- [204] R. A. Bermel and B. R., "The measurement and clinical relevance of brain atrophy in multiple sclerosis," *The Lancet Neurology*, vol. 5, pp. 158-170, 2006.
- [205] A. Brunetti, A. Postiglione, E. Tedeschi, A. Ciarmiello, M. Quarantelli, E. M. Covelli, G. Milan, M. Larobina, A. Soricelli, A. Sodano, and B. Alfano, "Measurement of Global Brain Atrophy in Alzheimer's Disease With Unsupervised Segmentation of Spin-Echo MRI Studies," *Journal of Magnetic Resonance Imaging*, vol. 11, pp. 260–266, 2000.
- [206] M. P. Sanfilipo, R. H. B. Benedict, R. Zivadinov, and R. Bakshi, "Correction for intracranial volume in analysis of whole brain atrophy in multiple sclerosis: the proportion vs. residual method," *NeuroImage*, vol. 22, pp. 1732– 1743, 2004.
- [207] R. D. Sinelnikov, *Atlas de Anatomía Humana* vol. III. Moscú: Editorial Mir, 1977.
- [208] J. M. Charcot, *Lectures on Diseases of the Nervous System*. London: The New Sydenham Society, 1977.
- [209] L. Lebart, M. Piron, and A. Morineau, *Statistique Exploratoire Multidimensionnelle, [Visualisation et Inférence en Fouille de Données]*, 4ème edition, refondue ed. Paris: Dunod, 2006.
- [210] M. A. Wong, "A hybrid clustering method for identifying high density clusters," *J. of Amer. Statist. Assoc.*, vol. 77, pp. 841-847, 1982.
- [211] R. Sedgewick, *Algorithms in C++, Part 5 Graph Algorithms*: Addison-Wesley, 2002.
- [212] R. E. Tarjan, "Depth first search and linear graph algorithms," *SIAM Journal on Computing*, vol. 1, pp. 146-160, 1972.