OPTICA PURA Y APLICADA - Vol. 26 - 1993

Depth detection on monocular images using active vision

Detección de profundidad en imágenes monoculares mediante visión activa

Xavier Binefa, Jordi Vitrià, Xavier Roca

Grup de Visió per Computador

Dept. Informàtica - Universitat Autònoma de Barcelona

08193 Bellaterra (Barcelona)

Resumen

En esta comunicación se presenta un método para recuperar las profundidades de los puntos de una escena que presentan un cambio de intensidad significativo respecto a su entorno, a partir de sus imágenes obtenidas con un sistema óptico de poca profundidad de campo. El método utilizado, basado en el paradigma de la *visión activa*, permite recuperar esta información manipulando uno de los parámetros del sistema óptico utilizado en la captación de la imagen: el grado de enfoque.

Abstract

In this paper we present a new method to recover the depth of the points in a scene, that show a meaningful change of intensity with regard to their environment, from images obtained using an optical system with a small depth of field. The method used, based on *active vision* paradigm, allows to recover this information by manipulating one of the parameters of the optical system used in the image sensing: the gradient of focus.

1 Introducción

La *visión por computador* pretende construir descripciones de las escenas reales a partir de sus imágenes y simular el sistema de visión humano. En la visión temprana (early vision) el objetivo consiste en recuperar las propiedades geométricas de la escena. Es por ello que también acostumbra a definirse como *óptica inversa*. La recuperación de los parámetros físicos de la escena puede realizarse a partir de varias fuentes de evidencia. Las cinco más establecidas son estéreo, movimiento, variaciones de intensidad del nivel de gris (contornos) , textura y sombreado. En todas ellas, debido a la pérdida de información y al ruido introducido en el proceso de digitalización, la recuperación de los parámetros es un proceso que no posee una solución única. Esto obliga a considerar asunciones sobre la geometría de la escena para alcanzar la unicidad de la solución. Por otra parte la mayoría de soluciones se alcanzan con un uso intensivo de la diferenciación. Es bien conocido que esta operación amplifica el ruido y convierte los procesos en altamente inestables. Resulta pues que una característica común de los problemas de optica inversa es su carácter de problemas *mal puestos*, que se caracterizan por la inestabilidad y la no unicidad de su solución.

Una via ampliamente utilizada para abordar estos problemas consiste en introducir las restricciones por medio de un proceso de regularización. Su objetivo es convertir el problema en *bien puesto*. Un problema es *bien puesto* cuando su solución existe, es única y depende de forma continua de las condiciones iniciales. Ello se consigue con el uso de funcionales estabilizadores que aporten al problema información a priori que se exige a las soluciones (suavidad, rigidez, etc) restringiendo con ello el espacio de soluciones admisibles.

Recientemente ha aparecido otro punto de vista basado en la analogía con los sistemas sensoriales biológicos: la *visión activa* [1]. Esto quiere decir que el observador no adquiere la información visual de forma pasiva, sino que utiliza sus facultades para moverse y modificar las

condiciones de obtención de estas imágenes para conseguir fuentes de información adicionales, y resolver mediante su integración los problemas de la visión de forma discriminante.

En un sistema de visión por computador los parámetros controlables son muchos: vergencia de las cámaras en un sistema estéreo, iluminación, posición de la cámara, etc. Algunos de estos parámetros están directamente relacionados con el sistema óptico utilizado. Este es el caso del enfoque. Experimentos realizados [2] con imágenes reales muestran el papel del enfoque en la percepción de la profundidad del entorno. En particular hay evidencia experimental de que, cuando el objeto más cercano se encuentra enfocado, el sistema visual humano interpreta el grado de desenfoque como aumento de la distáncia. Por otra parte la profundidad de campo de las células de la retina que actuan en el espectro del rojo y del verde es diferente de las que actuan en el azul. También posee interés que la distancia focal del ojo humano varie contínuamente y de forma sincronizada en los dos ojos. Todo ello permite pensar que el grado de enfoque juega un papel muy importante en la determinación de la percepción física del entorno y en particular en la detección de la profundidad.

En esta comunicación se presenta un método basado en la manipulación del enfoque para la obtención de la profundidad de los objetos que aparecen en una imagen monocular, y que ha sido aplicado a un problema en particular: la determinación de la topografía de la superficie de un circuito integrado. En este método partimos de una serie de imágenes tomadas a diferente distancia del circuito (manteniendo constantes los demás parámetros) y determinamos la profundidad de los saltos de intensidad que aparecen en las imágenes.

2 Modelización del problema

El modelo que se ha escogido para la formación de las imágenes captadas es el de la óptica geométrica. En este modelo la imagen de un punto p situado a distancia u del centro óptico es

un punto p' localizado detrás de la lente y a distancia w del centro. La relación entre estos valores y la distancia focal f del sistema óptico viene dada por la conocida ley de Gauss: $\frac{1}{f} = \frac{1}{u} + \frac{1}{w}$

El conocimiento de la distancia w permetiría conocer la distancia en la que se encontraría el objeto. Si el plano detector no estuviera a una distancia w de la lente, el brillo del punto p' se dispersaría según una función gausiana de parámetro σ proporcional a la distancia entre p' y el plano detector.

Consideremos el caso más simple en el que la imagen de la escena aparece como uniformemente desenfocada. Esto puede ser debido a que la escena sea plana o porque la profundidad de campo del sistema óptico provoque que la diferencia de enfoque entre los objetos de la escena no pueda ser apreciada. En el caso de una *escena plana* la imagen totalmente enfocada estará en un plano con un brillo en cada punto $f(x, y)$ proporcional a la radiáncia del punto correspondiente de la escena. La imagen detectada puede expresarse como:

$$I(x, y, z) = f(x, y) \otimes G(x, y, z) \tag{1}$$

donde $I(x, y, z)$ es el brillo del plano imagen situado a distancia z del plano detector y $G(x, y, z)$ es la función gausiana con σ proporcional a z.

El desenfoque puede también modelizarse con ecuaciones diferenciales. Consideremos la ecuación diferencial de la calor:

$$\begin{cases} \frac{\partial I(x,y,z)}{\partial z} = \Delta I(x, y, z) \\ I(x, y, 0) = f(x, y) \end{cases} \tag{2}$$

La solución de esta ecuación es:

$$I(x, y, z) = \int\int G(x - s, y - r, z) f(s, r, z) ds dr \tag{3}$$

donde $G(x, y, z) = \frac{1}{4\pi z} \exp^{-\frac{x^2+y^2}{4z}}$ corresponde a una gausiana de $\sigma^2 = 2z$. Esta solución justifica el hecho de que el desenfoque pueda ser analizado como solución de la ecuación del

calor. Podemos pues identificar nivel de gris con temperatura y evolución por desenfoque con dispersión de temperatura a lo largo del tiempo.

En el caso de *escenas no planas*[3] al variar el punto de vista a lo largo del eje z obtendremos una secuencia de imágenes parcialmente desenfocadas. En este caso la imagen enfocada se encontrará en una superficie no plana: $S = \{(x, y, h(x, y))\} \subset \Re^3$. Tal es el caso de las imágenes de un circuito integrado tomadas con un microscopio óptico. La resolución lateral necesaria conlleva una profundidad de campo más pequeña que las diferencias de altura de los elementos del circuito. En tales imágenes la σ varia en cada punto. Para extender el modelo anterior a estas imágenes supongamos que el plano sensor siga la topografía $3D$ de la escena. Por una parametrización diferente (r, s) de la superficie S anterior podemos expresar el brillo de tal superficie como $f(r, s)$, proporcional a la radiancia del punto correspondiente de la escena. Denotemos la imagen como $I(r, s, 0) = f(r, s)$. Cuando separamos el plano detector una distancia t de la superficie enfocada a lo largo del eje z registraremos una imagen con un desenfoque uniforme expresado como:

$$I(r, s, t) = f(r, s) \otimes G(r, s, t) \tag{4}$$

Resulta pues que la ecuación 2 en coordenadas (r, s) proporciona también el modelo diferencial de comportamiento del desenfoque del brillo de las imágenes que obtendríamos en una cámara con un "plano" sensor adaptado a la superficie del circuito.

La ecuación sobre variedades resultante es de tipo parabólico y por lo tanto verifica el *Principio del Máximo.* Este principio establece que la secuencia de imágenes sobre las variedades $I(r, s, t)$ tomará el máximo valor (aplicado a $-I$ también el mínimo valor) en la superficie $t = 0$ o bien en la frontera lateral del dominio de la imagen f. El Principio del Máximo puede ser usado [4] para probar que cuando t aumenta, los puntos (r, s, t) donde una de las derivadas direccionales $\frac{\partial I}{\partial r}$ o $\frac{\partial I}{\partial s}$ es cero, cumplen el principio de causalidad: forman superficies tridimensionales que empiezan en $I(r, s, 0)$. Los puntos en que una derivada direccional es cero corresponden a las

líneas de cresta del gradiente. Esto significa que hay continuidad de los puntos parabólicos del gradiente en la secuencia de imágenes. Por otro lado, es bien conocido que el desenfoque atenúa las altas frecuencias y por lo tanto las lineas de cresta del gradiente alcanzarán el mayor valor en la superficie en que la imagen se encuentra totalmente enfocada.

En nuestro caso, el plano detector es perpenticular al eje del sistema óptico y no sigue la superficie de la escena. Las imágenes que obtendremos serán cortes de diferentes superficies y en consecuencia obtendremos diferentes grados de enfoque en cada imagen. En las imágenes gradiente , la continuidad de las líneas de cresta en la secuencia de imágenes será mantenida. Podemos detectar la profundidad relativa de cada línea de cresta del gradiente asignandole la imagen en la que esta más enfocada.

3 Metodología Propuesta

Sea $I(x, y, z)$ el conjunto de imágenes del circuito integrado obtenido variando continuamente el centro del sistema óptico a lo largo del eje z, en el sentido de aproximación al objeto y partiendo de un punto en que la escena esta totalmente desenfocada. Las imágenes digitalizadas $I(x, y, z_i)$ son obtenidas con decrementos pequeños del eje z (en nuestro caso de $0.1\mu m$ y magnificación de $100X$.)

La determinación de la profundidad de los contornos se realiza en tres fases:

- Filtraje y determinación del módulo del gradiente de cada imagen de la secuencia.
- Detección de las superficies tridimensionales formadas por la evolución de las crestas del módulo del gradiente en la secuencia de imágenes .
- Para cada superficie de evolución determinar en que imagen los pixels alcanzan su valor de máximo gradiente. Asignar como profundidad relativa el número de la imagen de la secuencia en que esto sucede.

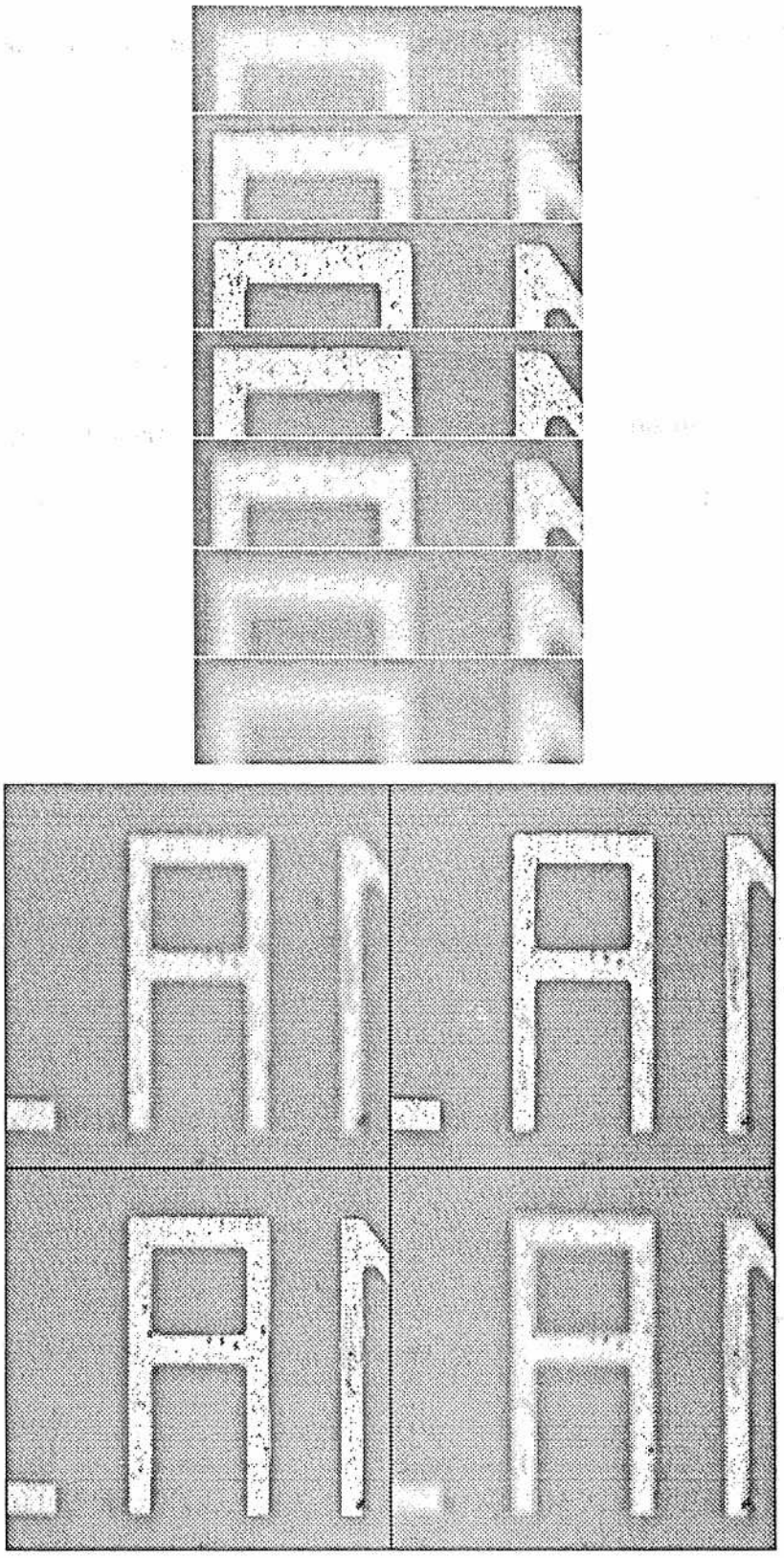

Figura 1: Secuencias de imágenes

3.1 Filtraje y módulo del gradiente

Las pistas de aluminio que aparecen en las imágenes presentan un ruido impulsivo que puede ser eliminado mediante filtraje no lineal. Se ha tratado cada imagen de la secuencia con un Filtro Alternado Secuencial [5] iterando cierres seguido por aperturas de tallas consecutivas. Este filtro morfológico presenta interesantes propiedades: No desplaza los contornos, mantiene su contraste y elimina los extremos poco significativos en área.

Antes de calcular el módulo del gradiente se ha regularizado cada imagen de la secuencia con el fin de evitar la ampliación del ruido que caracteriza la diferenciación . Para ello se ha calculado la convolución de cada imagen con la primera derivada de un spline cúbico (muy similar a una gaussiana). Su justificación a partir de la teoría de regularización estandard se encuentra en [6].

3.2 Evolución de las crestas del gradiente

Por la modelización del problema realizada anteriormente, la evolución de las líneas de cresta del gradiente es contínua a lo largo de la secuencia. Para obtener las paredes por las que evoluciona se ha utilizado un algoritmo bien conocido de la Morfología Matemática: El algoritmo de los "Watersheds" [7]. Su uso es habitual en el proceso de segmentación de imágenes y permite detectar los contornos de los objetos de una imagen a partir del módulo del gradiente de la imagen. Partiendo de que los objetos presentarán un nivel de gris uniforme, su extracción consistirá en obtener las líneas de máximo contraste que los circunden. Se definen los contornos como las líneas de cresta ("watershed lines") de la imagen gradiente (considerada como superficie topográfica) que circundan a los mínimos. En el caso de imágenes bidimensionales, estas líneas pueden ser consideradas como alternativa a los cruces por cero del laplaciano de una imagen. Recientemente se han obtenido versiones muy eficientes del algoritmo de detección de crestas basadas en simular la inmersión del módulo del gradiente a partir de sus mínimos [7].

El algoritmo de inundación anterior permite considerar imágenes tridimensionales. Para ello

necesitamos precisar la noción de mínimo y de cuenca de una imagen $3D$.

Definición 1 *Un* mínimo *de valor h de una imagen $I : \mathcal{Z}^3 \longmapsto [0, N]$ es una componente conectada de puntos $M \subset \mathcal{Z}^3$, todos ellos de valor h, tal que cualquier camino que partiendo de M y con al menos un punto del complementario de M, pase por lo menos por un punto de valor superior a h.*

Esta definición es muy diferente de la definición de mínimo local. En este último caso solo necesitamos conocer los vecinos de un punto para determinar si un punto particular es mínimo. En el caso digital exigimos que todos los puntos exteriores a M tengan un valor superior.

La conectividad introduce también complicaciones que no se dan en el caso contínuo. En imágenes digitales en trama cuadrada (trama cúbica en imágenes tridimensionales) hemos de considerar que entendemos por puntos vecinos de uno dado. Es muy operacional considerar una matriz 3×3 ($3 \times 3 \times 3$ en el caso tridimensional) en la que los valores uno de la matriz representen los vecinos del punto central. Las conectividades usuales en imágenes bidimensionales son 8-conectividad (todos los valores de la matriz a uno) o bien 4-conectividad (esquinas de la matriz a cero y el resto a uno). Las dos conectividades proporcionan regiones cerradas sin que haya motivos intrínsecos para elegir entre una u otra. En el caso tridimensional podemos alcanzar regiones sólidas de muchas maneras. En nuestro caso hemos considerado conectividad a 6 (conectividad a 4 en el plano que consideramos y a uno los pixels superior e inferior del centro). Esta conectividad garantiza que las fronteras seran cerradas y reseguibles a conectividad a 26.

Definición 2 *Una* componente maximal *(cuenca) de un mínimo M de una imagen I es el mayor conjunto $C(M)$ de puntos del dominio de I que podemos alcanzar a través de caminos, con valores de gris monotonamente crecientes, con origen en los puntos de M.*

Tendremos pues un conjunto de cuencas $\{C(M_i)\}$ cada una asociada a un mínimo de la imagen.

Definición 3 *Las* líneas de cresta *("watersheds") de la imagen I se definen como:*

$$W(I) = \bigcup_{i \neq j} (C(M_i) \cap C(M_j))$$

En el caso de imágenes tridimensionales las líneas de cresta serán las superficies con mayor valor de gris que recubren las componentes conectadas determinadas por cada mínimo.

Por lo dicho anteriormente, las diferentes imágenes gradiente $g(x, y, z_i)$ obtenidas forman una imagen tridimensional. En ella los contornos que circundan los mínimos forman superficies en el espacio que corresponden a la evolución de los contornos por desenfoque.

3.3 Determinación de la profundidad

Una vez determinadas las superficies tridimensionales que envuelven los mínimos, es necesario localizar las líneas de máximo contraste que evolucionan a través de las mencionadas superficies. El hecho de conocer solamente la conectividad en la que podemos recorrer estas superficies asi como la poca robustez de considerar el valor del gradiente pixel a pixel, obliga a utilizar un método global de detección de las líneas de máximo contraste. En [8] asumimos la planaridad de las regiones de los objetos sobre los que aplicar el método. Aqui la robustez la conseguimos aplicando el algoritmo de detección de crestas actuando únicamente sobre el dominio de las superficies 3D detectadas. Los únicos mínimos desde los que empezamos la inundación seran los puntos de estas superficies que pertenezcan al primer y último plano de la secuencia de imágenes (en las que todos los objetos estan desenfocados - sin crestas de interés-). Este método es general al no exigir ninguna asunción sobre planaridad de los objetos. Una vez determinadas las líneas de cresta sobre las superficies detectadas, aquellas se extenderan sobre la secuéncia de imágenes. Tomamos como indicador de profundidad de los pixels de las líneas de cresta la imagen en la que aparece cada punto de la cresta.

4 Resultados Experimentales

En la figura 1 pueden verse dos secuencias (de siete y cuatro imágenes respectivamente) de un circuito integrado tomadas en las condiciones expresadas anteriormente. En la primera secuencia podria asumirse el caracter plano de los objetos que en ella aparecen. Visualmente nos apoyamos no solo en la nitidez de los contornos sino tambien en la claridad de la textura. En la segunda secuencia vemos que las pistas del circuito estan ligeramente inclinadas, disminuyendo la profundidad de izquierda a derecha.

En la figura 2 se aprecian los resultados de profundidad conseguidos con el algoritmo. Vemos que los resultados coinciden con la apreciación visual de las imágenes. En la secuencia de imágenes cuadradas puede observarse que un mismo objeto no tiene que ser forzosamente plano para que el algoritmo funcione. Una misma pista esta localizada a profundidades diferentes.

Este método no posee limitaciones en cuanto al número de niveles de profundidad que puede detectar , depende solamente del número de imágenes de la secuencia (en las que no pueden presentarse inversiones de contraste significativas). La obtención de un mapa denso de profundidades requiere el uso de información textural (donde la haya) y está actualmente en estudio. La información que obtengamos fusionada con el método expuesto aumentará la robustez y generalidad del método.

5 Conclusiones

Se ha presentado un método robusto de calculo de profundidad relativa a partir de imágenes ópticas, basado en una extensión de la Optica Geométrica, que intenta simular el sistema de visión humano. El método, al utilizar sólo contornos, puede ser utilizado en metrología microscópica. En ella los objetos aparecen siempre parcialmente desenfocados por lo que la medición manual es dificil de realizar de forma precisa y comparable.

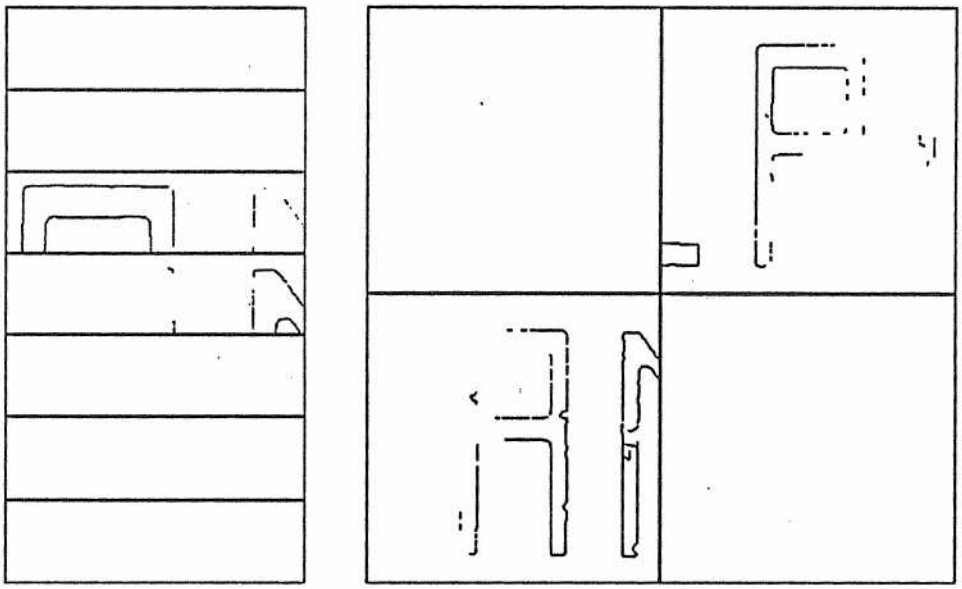

Figura 2: Profundidad detectada.

6 Agradecimientos

La Comisión Interministerial de Ciencia y Tecnología (CICYT) facilitó este trabajo bajo el proyecto ROB90-0435 y la Comissió Interdepartamental per la Recerca i Innovació Tecnològica (CIRIT) dispuso una beca para su realización.

Referencias

[1] J. ALOIMONOS y D. SHULMAN.: *Integration of visual modules.* (Academic Press Inc. , San Diego, 1989).

[2] A. PENTLAND: PAMI 9, Num 4, (1987).

[3] X.BINEFA, J.VITRIA y J.J.VILLANUEVA: *SPIE Conf. on Machine Vision in Microelectronics Manufacturing, SPIE/IS&T's Electronic Imaging*, 9-14 Feb 1992.

[4] R. HUMMEL: *Proc. IEEE Conf on Computer Vision and Pattern Recognition*, pp. 204-209, 1986.

[5] J. SERRA: *Image Analysis and Mathematical Morphology. Volume two: Theoretical Advances.* (Academic Press, London, 1988).

[6] V. TORRE y T. POGGIO: IEEE PAMI8 num. 2 , pp. 147-163, 1986.

[7] L. VINCENT y P. SOILLE: IEEE PAMI 13 num. 6, pp.583-598, 1991.

[8] X.BINEFA, J.VITRIA y M. VANRELL: *V Simposium Nacional de Reconocimiento de Formas y análisis de imágenes* Valencia, pp. 132-139, 21-25 Sbre. 1992

* * *