

UNA APLICACIÓN DE LA TÉCNICA MULTIEDIT-CONDENSING A LA SEGMENTACIÓN DE ESCENAS NATURALES EN COLOR

FRANCISCO FERRI RABASA, ENRIQUE VIDAL RUÍZ

Se presenta una adaptación de la técnica Multiedit-Condensing (MC), recientemente aplicada a algunos problemas en Reconocimiento de Formas, a un problema de segmentación de imágenes que se encuadra dentro del marco general del Análisis de Imagen en Robótica. Se presenta una revisión de los fundamentos de dicha técnica y se justifica su uso para la solución del problema propuesto. Dicho problema consiste básicamente en la separación de las distintas zonas de interés en una imagen, según su color y ciertos aspectos morfológicos o contextuales, todo ello condicionado por las características de la iluminación natural. El procedimiento es aplicado a un problema concreto consistente en la detección de frutos por visión para su recolección automática.

Keywords: Robótica, Segmentación de escenas, Reconocimiento de Formas, Regla del Vecino más Próximo, Técnica Multiedit-Condensing.

-Francisco Ferri Rabasa. Dpt. d'Informàtica i Electrònica. Universitat de València. Trabajo parcialmente subvencionado por CICYT, ROB89-0285.

-Enrique Vidal Ruíz. Dpt. de Sistemes Informàtics i Computació. Univ. Politècnica de València. Trabajo parcialmente subvencionado por CICYT, TIC-0448/89.

-Article rebut el març de 1990.

1. INTRODUCCIÓN

La Visión Artificial constituye actualmente un área que engloba aproximaciones muy diferentes. El presente trabajo se centra en el análisis de escenas naturales, problema que ha sido ampliamente tratado pero cuyos resultados son difícilmente generalizables debido a los múltiples problemas derivados de la iluminación incontrolada. En este caso, el objetivo final se encuentra dentro del campo de aplicaciones de la Robótica, lo cual va a condicionar fuertemente el desarrollo de la solución al problema propuesto en cuanto a su coste temporal. Más concretamente, dentro de lo que podríamos denominar "Robótica Agrícola", el presente trabajo se centra en la recolección automática dirigida por visión.

El problema de la recolección automática de frutos viene siendo estudiado desde hace mucho tiempo [10] [12], habiéndose llegado recientemente a experimentar robots que consiguen recolectar un porcentaje apreciable de ciertos tipos de fruta en el árbol [7] [8].

En los sistemas de visión propuestos hasta ahora para estos robots se utiliza de alguna manera la información del color, ya sea digitalizando directamente una cierta función de las coordenadas del color, trabajando en alguno de los espacios de color conocidos, o incluso tomando medidas fuera del espectro visible (recolección automática de manzanas, [11]). La mayoría de estos sistemas concentran su potencia en el análisis posterior a la detección del color, utilizando métodos para localizar los objetos que suelen ser costosos, lo cual obliga a incorporar procesadores específicos en algunos de los sistemas presentados en la literatura.

Se propone en este artículo un método para separar los frutos del fondo de una manera más fina, enfocando esta separación como un problema de Reconocimiento de Formas.

Una aproximación similar se ha abordado recientemente en [6] utilizando funciones discriminantes (FD) en el espacio SH (saturación, tinte) e infiriendo estas funciones con un algoritmo genético. Si bien con esta aproximación se obtienen resultados aceptables, queda manifiesta la imposibilidad de separar frutos y fondo sólo con la información de color relativa a un pixel. Esto es debido a la presencia inevitable de brillos, que aparecen aleatoriamente en el fondo (hojas) y prácticamente siempre sobre los frutos debido a su simetría esférica.

Se aporta en este artículo una solución que, en este sentido, mejora sensiblemente la aproximación puntual, sin llegar a la complejidad que representan otros métodos de localización. Esta aproximación consiste en medir el color no sólo para el punto considerado sino también para otros puntos vecinos. Con

estas medidas se construye el vector de características que será utilizado para representar las muestras (pixels). Se clasificarán estas muestras mediante la regla del vecino más próximo (NN, 1-NN), utilizando una serie de prototipos bien clasificados. Obviamente, el problema ahora vendrá dado por la selección de estos prototipos, para cuya solución se propone el uso del método conocido como *Multiedit-Condensing* (MC)[4].

En las siguientes secciones se presenta formalmente el problema considerado. Asimismo se revisará la técnica MC y se expondrá la adaptación de esta técnica al problema propuesto. Concluye este artículo presentando los resultados obtenidos y sugiriendo posibles mejoras en la conclusión.

2. PROCEDIMIENTOS DE CLASIFICACIÓN

El problema que se va a abordar consiste en la clasificación de los pixels de imágenes representadas en un cierto espacio que incluye información de color. Para ello se va a proponer el uso de la regla del vecino más próximo (NN - "Nearest Neighbour"). No obstante, el uso directo de esta regla de clasificación presenta ciertas dificultades derivadas de algunas particularidades del problema abordado. Por una parte será necesario un número de referencias o prototipos elevado si se pretende caracterizar toda la tonalidad que pueden presentar los frutos en todas las situaciones (grado de madurez, nivel de luminosidad, etc.) así como la frontera que separa frutos de lo que no lo son. Una consecuencia claramente negativa de este hecho es un excesivo coste computacional de la clasificación, cuya velocidad depende en principio del número de prototipos. Por otra parte, dada la imposibilidad de eliminar el ruido en las referencias, no se podrá aspirar a obtener conjuntos de prototipos libres de puntos incontrolados (sin sentido, etiquetados erróneamente) los cuales, en general, tenderán a aumentar el error de clasificación.

Se propone una solución a estas dificultades consistente en preprocesar las referencias primarias con la técnica *Multiedit-Condensing* [4], que permitirá eliminar prototipos erróneos y reducir drásticamente el tamaño del conjunto final de prototipos. Este método ha sido recientemente utilizado dentro del área del Reconocimiento Geométrico de Formas para ciertos sistemas que tienen que trabajar con patrones con un índice alto de variabilidad dentro de cada clase. Algunos ejemplos son el reconocimiento automático de caracteres impresos con varios tipos de letra y baja calidad [13], la decodificación acústico-fonética [1] o el reconocimiento de palabras aisladas multilocutor [14].

Con esta aproximación se obtiene un conjunto de prototipos con una relación tamaño-prestaciones cuasióptima, que conducirá a una binarización de la imagen lo suficientemente buena como para que con un algoritmo sencillo y rápido (sobre imágenes binarias) se localice un porcentaje elevado de frutos presentes.

3. MULTIEDIT-CONDENSING (MC)

La técnica MC [4] consiste en la combinación de dos procedimientos que apuntan a la selección de un número reducido y consistente de referencias a partir de un conjunto dado de ellas. El objetivo es que las fronteras de decisión inducidas por la regla NN con estas referencias se aproximen a las del clasificador de Bayes para un error de clasificación mínimo.

La técnica *Multiedit* [4] permite seleccionar de un conjunto de referencias dado, aquellas que pertenecen a la región de aceptación de Bayes para cada clase, permitiendo así buenos resultados incluso para distribuciones multimodales de muestras en cada clase.

La técnica *Condensing* [9] [4], por otra parte, intenta seleccionar de un conjunto de referencias previamente multieditado, sólo aquellas que contribuyen realmente a formar las fronteras de decisión; es decir, aquellas referencias que se encuentran en "las periferias" de la clase correspondiente. Este procedimiento da lugar a una reducción drástica del tamaño del conjunto de referencias considerado.

Algoritmo Multiedit

Datos:

T : conjunto de muestras que se pretende multieditar.

$I \in N$: criterio de terminación.

$M \in N$: valor cualquiera, $M > 2$.

Resultados:

T : conjunto de muestras multieditadas.

Método:

1. Dividir aleatoriamente T en M subconjuntos $T(i)$, $i = 1 \dots M$.
2. Para $i = 1 \dots M$, clasificar las muestras de $T(i)$ con la regla NN utilizando como conjunto de referencias $T((i + 1) \bmod M)$.

3. Descartar las muestras que han sido mal clasificadas y formar un nuevo conjunto T con las restantes.
4. Si en las últimas I iteraciones no ha habido edición, terminar, si no, ir al paso 1.

Este algoritmo eliminará aquellos puntos situados en zonas fronterizas donde predominen puntos de otras clases, ya que será muy fácil que, en alguna iteración, su vecino más próximo sea de otra clase. Se ha probado que bajo ciertas condiciones el algoritmo converge [4][5] en el sentido de que el número de prototipos descartados por el *Multiedit* está acotado por $2E$, donde E , es el error de clasificación para la regla NN con los prototipos iniciales. En [4] se presentan evidencias teóricas y experimentales de que la regla NN, usada conjuntamente con esta técnica, converge asintóticamente a la regla de decisión de Bayes.

Algoritmo Condensing (CNN)

Datos:

T : conjunto multieditado de muestras.

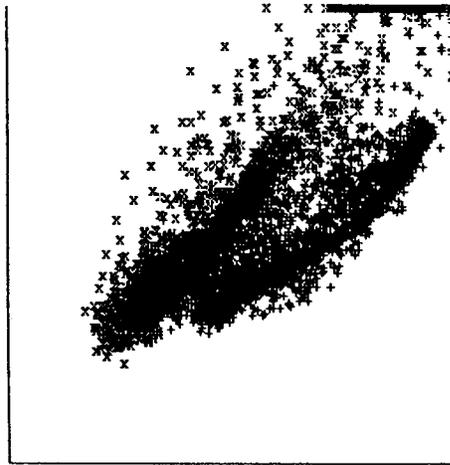
Resultados:

R : conjunto de muestras condensadas.

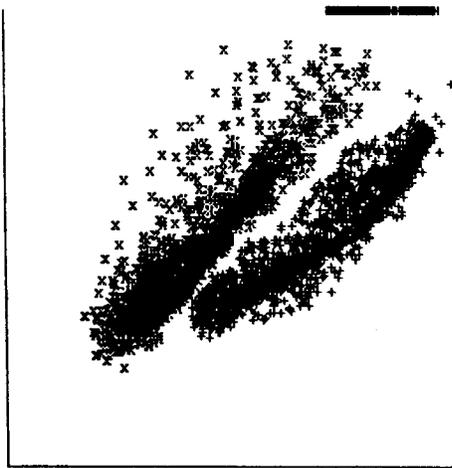
Método:

1. Seleccionar una muestra de T y ponerla en R que estará inicializado vacío.
2. Para todo $p \in T$,
 - clasificar p con el actual contenido de R como conjunto de referencias, si esta muestra es mal clasificada transferirla de T a R .
3. Si se produce una ejecución del paso 2 sin transferencias, o T está vacío, terminar, si no, ir al paso 2.

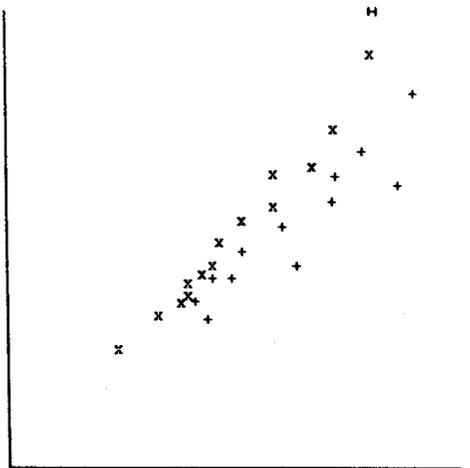
Este algoritmo eliminará aquellas referencias que no son “esenciales” para acertar en la clasificación de todas las demás. Hay que tener en cuenta que el resultado es subóptimo en el sentido de que no se conservan exactamente las fronteras de decisión del conjunto inicial multieditado de referencias. Además el resultado no es único y puede llegar a depender en cierta medida del orden de las muestras en T . Han sido propuestos algunos algoritmos para ordenar el conjunto T de manera que los puntos fronterizos sean considerados en primer lugar. Estos algoritmos suelen llamarse “CNN Modificado” [3] [4].



(a)



(b)



(c)

Figura 1. Ilustración de la técnica *Multiedit-Condensing* para un conjunto de referencias representado en el plano Rojo-Verde.

- (a) Original,
 - (b) multieditado, y
 - (c) condensado.
- (+: clase 1, x: clase 2, H: clase 3).

Como ejemplo ilustrativo se presenta en la figura 1 la proyección sobre los ejes R (rojo) y G (verde) de un conjunto de referencias en el espacio de representación dado por las coordenadas de color RGB. Estas referencias han sido tomadas de una imagen de las que después se utilizarán para evaluar la técnica MC. Puede observarse un considerable solapamiento, sobre todo entre las clases 1 y 2. En la figura 1(b) se muestra el conjunto multieditado correspondiente. Como puede verse han sido eliminadas todas las referencias de la zona de solapamiento, respetando en cierto modo las distribuciones originales de cada clase. Finalmente, en la figura 1(c) se muestran las referencias retenidas por el *Condensing* a partir del conjunto multieditado.

El coste computacional de la técnica MC es relativamente elevado, sobre todo debido al coste del *Multiedit*, lo que puede afectar a la realización de experimentos con un número alto de referencias ($10^5 - 10^6$). Efectivamente, cada iteración de dicho algoritmo requiere en promedio el cálculo de $|T|^2/M$ distancias, lo cual puede ser prohibitivo si el cálculo de la distancia es además apreciablemente costoso. No obstante, en la práctica, esto no constituye en la mayoría de los casos un problema importante, dado que esta técnica se aplica en una fase de preproceso previa a la tarea de reconocimiento. Por otra parte, en [13] se sugieren métodos para reducir el coste computacional del *Multiedit* consistentes en aplicarlo separadamente a grupos de clases que tengan sus distribuciones efectivamente solapadas. Esta información se puede extraer de la primera iteración del algoritmo.

4. APLICACIÓN AL PROBLEMA PROPUESTO

En cualquier sistema de visión para aplicaciones industriales es esencial obtener un adecuado balance entre efectividad y coste computacional. Es por esto que, siguiendo una filosofía similar a la presentada en [2] (aunque modificando sustancialmente el método), el presente problema de segmentación ha sido enfocado como un problema de Reconocimiento de Formas.

En la presente aplicación de la técnica MC se utiliza el color (coordenadas de color RGB) como factor discriminante principal; esto permite una rápida adquisición de las muestras durante el tiempo de clasificación, puesto que sólo hay que digitalizar las señales RGB analógicas entregadas por una cámara de vídeo adecuada. No obstante, esta información puntual conviene extenderla convenientemente con objeto de recoger cierta información contextual que aporte alguna indicación sobre la "forma" o "relieve" local de la imagen considerada. En la extensión propuesta, se consideran como características de un punto las

coordenadas de color de cuatro de sus vecinos además de las suyas propias. Sin embargo, no se considera la componente azul de los vecinos puesto que se ha comprobado mediante pruebas (informales) previas que su poder discriminante frente a las demás es claramente despreciable.

Así pues, si (i, j) son las coordenadas un pixel bajo análisis, el vector de características que asociamos a dicho pixel es:

$$v(i, j) = (R(i, j), G(i, j), B(i, j), \\ R(i + h, j), G(i + h, j), R(i - h, j), G(i - h, j), \\ R(i, j + h), G(i, j + h), R(i, j - h), G(i, j - h));$$

Donde $R(i, j), G(i, j), B(i, j)$ son las componentes Roja, Verde y Azul del pixel (i, j) respectivamente, y h es un parámetro que será elegido en función del tamaño medio de los frutos. Este tamaño será conocido con suficiente aproximación si se fija la distancia de la cámara al árbol y la óptica utilizada.

Se obtiene por tanto un espacio de representación 11-dimensional, lo cual se traduce en una complejidad aceptable del cálculo de distancias y, por tanto, del clasificador NN.

La elección de este conjunto de características constituye un burdo intento de "capturar" la *variación de color* sobre la superficie de los frutos debido a su forma esférica y a sus propiedades colorimétricas. Obviamente, un estudio más profundo del problema podría conducir a la elección de un conjunto óptimo de características, lo cual desembocaría posiblemente en un espacio de representación con más dimensiones. Comentaremos en la conclusión posibles extensiones al presente trabajo en este sentido.

El procedimiento de clasificación adoptado en este trabajo es la regla del vecino más próximo (1-NN) en un intento de obtener un error de clasificación mínimo. Sin embargo, como se ha comentado en la sección 1, se hace completamente necesario un preproceso de los prototipos seleccionados, en el presente caso mediante la técnica *Multiedit*, para efectivamente aproximarse a este error mínimo. No obstante, aunque esta técnica proporcionaría un buen clasificador, éste sería excesivamente lento, puesto que el número de prototipos retenidos por el *Multiedit* es considerablemente grande. Como se ha visto en la sección anterior, el *Condensing* da una solución: del total de prototipos multieditados se seleccionarán sólo aquellos que son necesarios para obtener una clasificación sin diferencias apreciables con respecto a la proporcionada por el conjunto completo de prototipos multieditados.

En definitiva, se consigue con esto un procedimiento de clasificación de puntos con un coste computacional moderado, que además resultaría fácilmente paralelizable y con el que se obtienen porcentajes de error considerablemente bajos, como se verá a continuación.

5. RESULTADOS EXPERIMENTALES

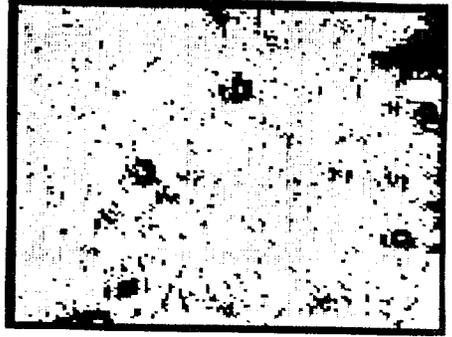
La presente técnica ha sido utilizada como proceso preliminar en varios procedimientos de segmentación y localización de frutos con resultados diferentes que serán comentados brevemente más adelante; sin embargo, en el presente artículo se insistirá en los resultados de la clasificación puntual, es decir, pixel a pixel. Por tanto, se comprobará la eficiencia del método teniendo en cuenta solo el tanto por ciento de puntos bien clasificados.

Los resultados que se van a presentar corresponden a un experimento para el cual se han seleccionado 10 imágenes en color, siguiendo el criterio de que sean representativas (varios niveles de luminosidad), y que en todas ellas aparezcan las diferentes clases consideradas. En el presente experimento estas clases son únicamente tres: *frutos*, *hojas* y *fondo (cielo)*. En la figura 2(a) se puede ver la representación de intensidad de una de estas imágenes que puede ser considerada como representativa puesto que en ella aparecen zonas muy iluminadas (algunas casi saturadas) mientras que otras están prácticamente a oscuras. Utilizaremos esta imagen para comprobar los resultados de la técnica propuesta.

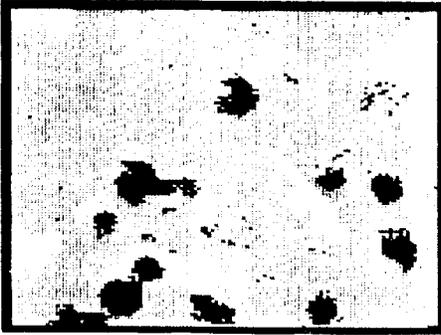
Se han utilizado dos de estas imágenes para construir el conjunto de prototipos con el que serán clasificadas las demás, para ello estas dos imágenes han sido segmentadas y etiquetadas manualmente. Así mismo han sido tratadas de idéntica forma las imágenes restantes con objeto de medir el error de clasificación. Dada la dificultad que entraña este proceso, se ha optado por no considerar los puntos de la imagen en la zona de transición de una región a otra, con lo cual se reduce el tamaño del conjunto de entrenamiento. Con todo ello, el número de prototipos a procesar con el MC es excesivamente grande, por lo que se lleva a cabo además una reducción aleatoria en cada clase, con el objetivo añadido de balancear las poblaciones en las diferentes clases. lo cual resulta particularmente conveniente para el *Multiedit*.



(a)



(b)



(c)



(d)

Figura 2. Resultados de la segmentación de una de las escenas utilizadas en los experimentos.

- (a) Representación en niveles de gris,
- (b) clasificación usando el conjunto de referencias original,
- (c) idem con el conjunto multiresolución, y
- (d) idem con el conjunto condensado.

El conjunto de puntos obtenido, consta de 1177 referencias, a las cuales se les ha aplicado la técnica *Multiedit* tomando el valor de M aleatoriamente entre 3 y 5 en cada iteración y con criterios de convergencia $I = 4$ e $I = 7$. Esto da lugar a dos conjuntos de referencias $M1$ y $M2$, a los cuales se les aplica la técnica *Condensing*, con lo que se obtienen los conjuntos condensados $C1$ y $C2$.

En la tabla 1 se presentan los resultados de clasificar con la regla NN el conjunto test formado por las 8 imágenes restantes. Se incluye además del tamaño de cada conjunto, una medida del coste computacional relativo en tiempo de clasificación (que depende esencialmente del tamaño del conjunto), así como el error de clasificación obtenido para las imágenes test.

Es necesario hacer notar que el error de clasificación se mide en función de la discrepancia entre el etiquetado automático y el manual antes explicado, por lo que debe considerarse únicamente como orientativo (un mismo valor del error puede ser muy bueno para una imagen y muy malo para otra). Un juicio cualitativo de la bondad del resultado obtenido puede realizarse comparando las segmentaciones obtenidas con las imágenes originales (ver figura 2).

Tabla 1

Resultados de clasificación con los conjuntos de referencias obtenidos y con el conjunto original.

	Orig.	$M1$	$C1$	$M2$	$C2$
Tamaño	1077	822	12	807	13
Coste rel.(%)	100	76.3	1.11	74.9	1.21
Error (%)	7.41	6.02	6.25	6.17	6.30

Como puede verse en la tabla, el error se ve sensiblemente disminuido con el *Multiedit* de acuerdo con las predicciones teóricas, mientras que el *Condensing* lo aumenta aunque muy poco. Cabe hacer notar que si bien el *Multiedit* mejora el resultado de la clasificación, es realmente la contribución del *Condensing* lo que hace posible la aplicación de este método al problema considerado. La diferencia abismal existente entre el coste computacional de la tarea de clasificación con el conjunto condensado frente al multieditado (del 1.1% al 75%) es lo que conduce a tiempos de proceso aceptables con un error claramente inferior al producido con el conjunto inicial y sólo ligeramente superior al del obtenido con el conjunto multieditado.

A título comparativo, cabe decir que el error de clasificación obtenido es bueno con respecto a los resultados que se obtienen con funciones discriminantes en [6]. Si se mide el error considerando únicamente dos clases (fruto y no-fruto),

e intentando compatibilizar las medidas con las allí presentadas, se obtiene en nuestro caso para $C1$ un error de 2.88% frente a un error de 3.30% con un clasificador cuadrático [6].

En la figura 2 pueden verse los resultados de clasificación para una determinada imagen con el conjunto inicial de prototipos y con los conjuntos $M1$ y $C1$. Dada la dificultad de esta imagen concreta el resultado puede considerarse satisfactorio. Tras un simple postproceso, este resultado conduciría eventualmente a la detección correcta de prácticamente todos los frutos presentes.

6. CONCLUSIÓN

Se ha presentado en el presente trabajo una aproximación a la segmentación de imágenes a bajo nivel bajo la perspectiva del Reconocimiento de Formas. Es importante hacer notar que esta aproximación es generalizable a cualquier problema de segmentación en el que el objetivo consista en la separación de regiones cuyos puntos se puedan representar mediante un vector de características y cuyas muestras estén manifiestamente afectadas por la presencia de ruido. Además, esta técnica de segmentación a bajo nivel se puede aplicar (si el problema lo requiere) conjuntamente con otras técnicas clásicas de segmentación de imágenes que podríamos denominar de alto nivel.

En cuanto al objetivo final de la aplicación aquí presentada, es decir, la localización de frutos en árbol, se puede decir que en el experimento llevado a cabo, la mayor parte de los frutos son detectados a partir de la clasificación de pixels simplemente etiquetando las componentes conexas de la imagen binaria resultante y discriminando según la relación área/perímetro; si bien es necesario llevar a cabo además una eliminación del ruido, en el sentido de eliminación de puntos aislados y de "agujeros" en los objetos.

En cualquier caso, el objetivo en el presente trabajo consistía en la obtención de una buena tasa de clasificación de pixels, y precisamente el trabajo futuro irá encaminado al diseño de un procedimiento que incorpore esta clasificación pero que trabaje con estructuras de datos más complejas para atacar directamente el problema de la localización de objetos.

7. REFERENCIAS

- [1] **P. Aibar, M.J. Castro, F. Casacuberta y E. Vidal** (1990). "Multiple Template Modelling of Sublexical Units". In *Speech Recognition and Understanding: Recent Advances, Trends and Applications*, P. Laface (ed.) Springer-Verlag.
- [2] **W.E. Blanz, J.L.C. Sanz y D. Petkovic** (1989). "Control-Free Low-Level Image Segmentation: Theory, Architecture, and Experimentation". In *Advances in Machine Vision*, J.L.C. Sanz (ed.) Springer-Verlag.
- [3] **K. Chidananda Gowda y G. Krishna** (1979). "The Condensed NN Rule using the Concept of Mutual Nearest Neighborhood". *IEEE Trans. on Info. and Theory*, Vol. IT-25, 488-490.
- [4] **P.A. Devijver y J. Kittler** (1982). "Pattern Recognition, A Statistical Approach". Prentice Hall, Englewood Cliffs.
- [5] **P.A. Devijver** (1986). "On the editing rate of the multiedit algorithm". *Pattern Recognition Letters*, 4, 9-12.
- [6] **J. Domingo, F. Ferri, J. Albert y V. Cerverón** (1990). "Genetic Algorithm like a Learning Method for Colour Image Segmentation". Proc. of Fourth International Symposium on Knowledge Engineering.
- [7] **A. Grand d'Esnon** (1985). "Robotic Harvesting of Apples". *Proc. of Agrimation I*, 210-214.
- [8] **R.C. Harrell y P. Levi** (1988). "Vision Controlled Robots for Automatic Harvesting of Citrus". *Proceedings ROVISEC-7*.
- [9] **P.E. Hart** (1968). "The Condensed Nearest Neighbor Rule". *IEEE Trans. Info. Theory*, Vol. IT-14, 515-516.
- [10] **E.A. Parrish y A.K. Goksel** (1977). "Pictorial Pattern Recognition Applied to Fruit Harvesting". *Trans. of the ASAE* 20 (5), 822-827.
- [11] **G. Rabatel** (1988). "A Vision System for MAGALI, the Fruit Picking Robot". *Proc. of AG ENG 88*, Paper N 88.293.
- [12] **C.E. Schertz y G.K. Brown** (1968). "Basic Considerations in Mechanizing Citrus Harvest". *Trans. of the ASAE* 11 (2), 343-348.
- [13] **J. Voisin y P.A. Devijver** (1987). "An Application of the Multiedit-Condensing Technique to the Reference Selection Problem in a Print Recognition System". *Pattern Recognition*, Vol. 20, N 5, 465-474.
- [14] **N. Yalabik, Yarman-Vural y A. Mansur** (1988). "Modified Condensed Nearest Neighbor Rule as applied to Speaker Independent Word Recognition". *Speech Communication* 7, 411-415.

