

## XV

### APLICACIONES DE REDES NEURONALES ARTIFICIALES EN PROCESADO DE IMÁGENES

**D. Cabello; S. Barro; A. Mosquera; M. G. Penedo\*;**

**J. M. Pardo; M. J. Carreira**

*Universidade de Santiago de Compostela*

*(\*) Universidad de A Coruña*

#### 1. INTRODUCCIÓN

La información visual juega un papel muy importante en la comunicación humana, de ahí que no sea de extrañar el gran interés mostrado por el desarrollo de sistemas para el tratamiento computacional de imágenes. Sus aplicaciones son fundamentales e imprescindibles en multitud de campos: ingeniería, física, biología, medicina, etc. Una secuencia típica de procesado de imágenes suele implicar los siguientes pasos: captura de la imagen a partir del proceso físico que la genera mediante el sensor adecuado, restauración y realce en aras a reducir ruidos y mejorar su calidad, codificación, segmentación y extracción de propiedades y posterior interpretación. La forma de abordar estos problemas es muy variada; las soluciones van desde las clásicas, extraídas del campo del procesado de señales, hasta las que integran técnicas de inteligencia artificial en la búsqueda de descripciones simbólicas automáticas de la escena. La literatura disponible en este campo es realmente extensa.

El paralelismo existente tanto en los problemas de procesado de imágenes como en las arquitecturas de redes neuronales artificiales, así como sus características de respuesta (capacidad de aprendizaje, lo que reduce los problemas en muchos casos a un entrenamiento correcto de la red, capacidad de generalización en su

respuesta, potencia y robustez frente al ruido, fácil implementación computacional, posibilidad de salida en tiempo real mediante su implementación física, etc.) ha hecho que muchos autores empleen redes neuronales artificiales en prácticamente todos los procesos involucrados en el tratamiento computacional de imágenes: realce, restauración, codificación, compresión, segmentación, reconocimiento de objetos, interpretación, etc. En este capítulo intentamos mostrar las posibilidades que las redes neuronales artificiales ofrecen para la resolución de dos problemas significativos en visión computacional: la *segmentación* de las imágenes y el *reconocimiento de objetos* sobre ellas. Realizamos para ello una revisión bibliográfica, que en ningún momento pretende ser exhaustiva, pero si ilustrativa de las distintas formas de aproximarse a los problemas planteados y de las soluciones propuestas.

## 2. SEGMENTACIÓN DE IMÁGENES

La segmentación de una imagen es la etapa de procesado más importante en un sistema de visión de bajo nivel. Esta consiste en la división de la imagen en un conjunto de regiones elementales disjuntas caracterizadas por la constancia de alguna propiedad (nivel de gris, textura, color, etc.). En un sistema de interpretación de imágenes siempre existirá como etapa previa un proceso de segmentación. La obtención de una interpretación correcta dependerá en gran medida de la calidad de los resultados del proceso de segmentación.

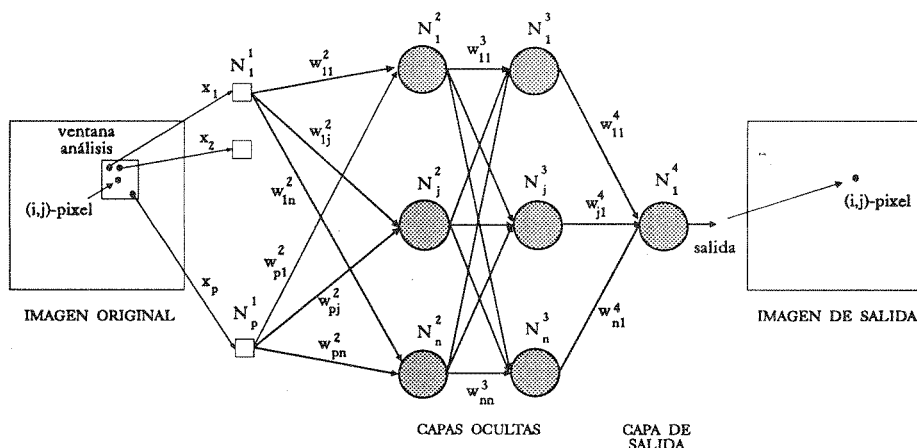
En la bibliografía de procesado de imágenes hay establecidas una gran variedad de estrategias y algoritmos para abordar el problema de la segmentación. El trabajo de Pal y Pal [1993] constituye una buena revisión de estos métodos. Sin embargo, no existe ninguno que proporcione resultados que se puedan considerar óptimos en cualquier tipo de imágenes. De hecho, métodos que pueden segmentar correctamente un tipo de imágenes pueden no ser aplicables a otros tipos de imágenes, usándose entonces el método más adecuado en cada aplicación concreta. Dentro de los procedimientos clásicos de segmentación podemos distinguir entre *procedimientos orientados a contornos* y *procedimientos orientados a regiones*. Los primeros buscan discontinuidades locales en base a operadores de diferencia (Roberts, Sobel, Canny, etc.), mientras que los segundos persiguen la detección de áreas locales con propiedades homogéneas, bien en el propio espacio imagen, con técnicas de crecimiento de regiones, bien en un espacio de propiedades mediante técnicas de clasificación de los pixels por umbralización de histogramas o con algoritmos de detección de agrupamientos (algoritmos de «*clustering*»). En estas últimas técnicas las regiones elementales surgen al transportar la información de clases del espacio de propiedades al espacio imagen; pixels

adyacentes que pertenezcan a la misma clase constituirán una región elemental. Todas estas estrategias se han implementado mediante redes neuronales artificiales; la bibliografía existente es muy extensa. Ilustramos a continuación algunas de las soluciones propuestas.

## 2.1. Soluciones basadas en la detección de contornos

Silverman [1991] propone el uso de una *red perceptron multicapa* entrenada mediante el *algoritmo de «back-propagation»* para extraer contornos en imágenes de ultrasonidos. La red acepta como entradas los valores de los pixels situados sobre ventanas cuadradas en la imagen original, tal como ilustra la figura 1. Para el entrenamiento de la red se necesita una imagen original y los resultados deseados en una nueva imagen transformada. Los pixels asociados a ventanas de análisis sobre la imagen original se proyectan en las entradas de la red y a cada una de las ventanas se le asocia como salida deseada el nivel de gris de su pixel central en la imagen transformada. Una exploración completa de la imagen de entrenamiento y su reiterada proyección sobre la red irá configurando la estructura de pesos hasta que se satisfaga el criterio de convergencia. Durante el proceso de análisis se explora toda la imagen y el valor de la unidad de salida en cada momento se interpreta como la intensidad transformada del pixel central de la ventana de entrada en ese momento, tal como ilustra la figura 1. Silverman muestra que una red consistente en una ventana de entrada de  $2 \times 2$  pixels, dos capas ocultas de cuatro unidades cada una y una sola unidad de salida puede simular correctamente la función del operador gradiente si se usa la propia imagen de gradiente como salida deseada en el proceso de entrenamiento. Este aspecto es importante no por el hecho de simular operadores clásicos, sino por la posibilidad de abordar el problema inverso de crear filtros espaciales para la detección de estructuras específicas. La imagen de resultados para el entrenamiento la generaría un experto delimitando las estructuras de interés. Silverman muestra entonces como una red perceptron con dos capas ocultas de nueve unidades cada una actuando sobre una ventana de análisis de  $9 \times 9$  pixels y entrenada mediante el procedimiento descrito con una imagen de resultados deseados adecuada puede aprender a reconocer contornos de estructuras anatómicas sobre imágenes de ultrasonidos del ojo.

Otros autores proponen otras estrategias para abordar la detección de contornos mediante arquitecturas de redes neuronales artificiales. Así, Shah [1990] formula el problema en el contexto de la minimización de una función de energía y Lu y Szeto [1991] proponen una red neuronal jerárquica para mejorar la detección de contornos en imágenes ruidosas, eliminando elementos de contornos falsos y recuperando otros perdidos.



**Figura 1.** Esquema del proceso de análisis de una imagen mediante una red perceptron multicapa (en la figura, dos capas ocultas de  $n$  neuronas cada una). ( $w_{ij}^k$ ; peso de la conexión entre la neurona  $i$  de la capa  $k-1$  y la neurona  $j$  de la capa  $k$ ).

## 2.2. Soluciones basadas en umbralización de histogramas

Babaguchi y col. [1990] usan una red multicapa entrenada con el *algoritmo de «back-propagation»* para umbralizar una imagen. La entrada a la red es el histograma de la imagen y la salida el valor de umbral adecuado para segmentarla. Para el proceso de entrenamiento requieren un conjunto suficientemente grande de imágenes con umbrales conocidos que conduzcan a segmentaciones correctas.

## 2.3. Soluciones basadas en clasificación de pixels

En este tipo de soluciones cada pixel de la imagen queda caracterizado por un vector de propiedades. Cada una de sus componentes pueden, por ejemplo, aportar información del pixel en las distintas bandas espectrales en el caso de imágenes de satélite, surgir de distintas imágenes de resonancia magnética nuclear con diferentes parámetros, o bien, ser los parámetros de un modelo de imagen estimados en la vecindad de ese pixel sobre la imagen original. En estos casos el problema de la segmentación se transforma en un problema de clasificación/partición del espacio de propiedades, problema en el que las redes neuronales artificiales se han mostrado idóneas. Así, por ejemplo, Blanz y Gish [1990] usan una red neuronal de tres capas con un entrenamiento supervisado para la segmentación de imágenes en base a la clasificación de los pixels. El número de neuronas en la capa

de entrada depende del número de propiedades que caracterizan al pixel y el número de neuronas en la capa de salida es igual al número de clases.

Otros esquemas clásicos de uso frecuente son los algoritmos de agrupamiento no supervisados, que intentan detectar agrupamientos naturales homogéneos («clusters») en el conjunto de vectores de entrada en base a una medida de disimilaridad. Los algoritmos de detección de agrupamientos más ampliamente usados son los algoritmos «*c-medias*», bien su versión todo o nada (clasificación «*hard*»), también referenciado como algoritmo «*K-medias*» (KM), bien en su versión borrosa, conocida como algoritmo «*c-medias borroso*» (FCM), o algunas de sus variantes. Estos algoritmos asignan los  $N$  vectores del conjunto de entradas en  $C$  clases de forma que se minimice una función objetivo. Sea  $\mathbf{X} = \{ \mathbf{x}_1, \dots, \mathbf{x}_n \}$  el conjunto de los  $N$  vectores de propiedades en  $\mathbf{R}^p$ . El vector  $\mathbf{x}_j = (x_{j1}, \dots, x_{jp})$  representa las  $p$  propiedades que caracterizan al pixel  $j$  en una ordenación lineal de la imagen. Una  $c$ -partición de  $\mathbf{X}$  viene dada por una matriz  $\mathbf{U} = \{ u_{ik}, 1 \leq k \leq n, 1 \leq i \leq c \}$  en la que sus elementos satisfacen las relaciones

$$\begin{aligned} 0 \leq u_{ik} &\leq 1 && \text{para todo } i, k \\ \sum_i u_{ik} &= 1 && \text{para todo } k \\ \sum_k u_{ik} &> 0 && \text{para todo } i \end{aligned} \tag{1}$$

Los elementos  $u_{ik}$  representan el grado de pertenencia del pixel  $k$  a la clase  $i$ . Si los valores de las  $u_{ik}$ 's están en el conjunto  $\{0,1\}$ , la partición  $\mathbf{U}$  es *hard*; en caso contrario, hablamos de una partición borrosa. Los algoritmos *c-medias* realizan una partición  $\mathbf{U}$  en el conjunto  $\mathbf{X}$  de forma que se minimice la función objetivo dada por la ecuación [Bezdeck, 1981]:

$$J_m(\mathbf{U}, \mathbf{v}; \mathbf{X}) = \sum_i \sum_k u_{ik}^m (\|\mathbf{x}_k - \mathbf{v}_i\|_A)^2 \tag{2}$$

donde  $m \in [1, \infty)$  es un exponente que incide en la borrosidad de la clasificación,  $\mathbf{v} = (\mathbf{v}_1, \dots, \mathbf{v}_c)$  representa los  $c$ -vectores de centros de las  $c$ -clases en  $\mathbf{R}^p$ ,  $\mathbf{A}$  es una matriz  $p \times p$  definida positiva que induce la norma y  $\|\mathbf{x}_k - \mathbf{v}_i\|_A = (\mathbf{x}_k - \mathbf{v}_i)^t \mathbf{A} (\mathbf{x}_k - \mathbf{v}_i)$  es la distancia de  $\mathbf{x}_k$  a  $\mathbf{v}_i$ .

Dado que estos algoritmos se plantean como un problema de minimización de una función, algunos autores proponen formularlo en términos de un modelo de red neuronal de Hopfield [Kamgar-Parsi y col., 1990; Amatur y col., 1992]. Así, éstos últimos realizan una clasificación no supervisada en imágenes de resonancia magnética nuclear de cabeza en base a la minimización de la función objetivo del algoritmo *K-medias* mediante una red de Hopfield modificada. La arquitectura que proponen responde a una red bidimensional de  $N \times C$  elementos de procesamiento, siendo

N el número de pixels en la imagen y C el número de clases. Su conectividad así como los valores de los pesos se determinan al proyectar la función cuyo mínimo resuelve el problema sobre la expresión de la energía que minimiza dicha arquitectura neuronal. La evolución temporal de la red conduce a un estado estable en el que aparece activada una sola neurona por fila y columna, indicando la asignación de ese pixel (índice fila) a esa clase (índice columna). La determinación del número de clases así como la interpretación (análisis) de las regiones resultantes queda encomendada a un experto clínico. La novedad más importante que aporta esta nueva forma de implementar un algoritmo clásico mediante redes neuronales artificiales es la posibilidad de su realización en paralelo empleando procesadores analógicos, por lo que serían de esperar soluciones rápidas al problema de la segmentación.

Otros autores intentan implementar estos algoritmos de agrupamiento mediante el uso de mapas topológicos autoorganizados. Aunque los resultados de la red de Kohonen no son óptimos con respecto a ningún modelo, su terminación es forzada, la convergencia no está asegurada y la salida depende de la secuencia de procesado de los datos, Lippmann [1987] sugiere que la red de Kohonen es en muchos aspectos similar al algoritmo de clasificación *K-medias*. Huntsberger y Ajjimarangsse [1990] sugieren la modificación de la red básica de Kohonen para que proporcione también particiones borrosas. Introducen una nueva capa de salida para la asignación de etiquetas y la sustitución de los factores de ganancia  $\alpha$  encontrados usualmente en las ecuaciones que rigen el entrenamiento de la red de Kohonen por valores de pertenencia borrosos tal como los calculados en el algoritmo *c-medias borroso* (FCM). En un trabajo posterior, Bezdek y col. [1992] proponen una nueva red de Kohonen para la clasificación borrosa (*Fuzzy Kohonen Clustering Network*, FKCN) que integra el modelo FCM en los factores de ganancia y estrategias de actualización de pesos, de forma que FKCN puede considerarse una implementación mediante un proceso autoorganizado del algoritmo FCM.

Hall y col. [1992] y Bezdek y col. [1993] consideran el uso del algoritmo FCM para la segmentación de imágenes de resonancia magnética del cerebro. El número de clases existentes en la imagen los fija un experto en términos del número de tejidos distintos que espera encontrar en ella. Otros autores [Springub y col., 1991] abordan el mismo problema usando directamente un mapa autoorganizado de Kohonen de  $K \times K$  neuronas con ligeras modificaciones en las adaptaciones de pesos y criterios de aprendizaje. El algoritmo genera un mapa topológico correcto en el sentido de que vectores de propiedades correspondientes a una misma unidad del mapa pertenecen a la misma clase en la imagen y vectores de entrada similares aparecen en unidades vecinas, por lo que podrían unirse en una única clase, disminuyendo así su número. Justifican entonces que no sea necesario fijar el número de clases como parámetro de entrada al algoritmo, en oposición a los ejemplos anteriores, en los que C es un parámetro de entrada al fijar el número de neuronas en la red.

Mosquera y col. [1993a] avanzan en la idea de usar un mapa de propiedades autoorganizado para extraer la estructura de clases subyacente en el conjunto de datos y la aplican a la segmentación de texturas. En este caso los componentes del vector de propiedades de cada pixel son los parámetros de un Campo Aleatorio de Markov de segundo orden estimados en su entorno. Emplean una red neuronal con una distribución bidimensional de  $K \times K$  neuronas entrenada mediante el esquema básico de Kohonen, pero con métrica, parámetros de ganancia y criterio de convergencia distintos. La distancia entre los vectores de entradas y los vectores de pesos en una iteración del proceso de entrenamiento está ponderada por la activación de las unidades en la iteración anterior. Así, la distancia entre el vector de entradas  $\mathbf{x}_s$  y el vector de pesos  $\mathbf{w}_r$  de la neurona  $k_r$  en la iteración  $t+1$  viene dada por la ecuación

$$d_{\mathbf{x}_s, \mathbf{w}_r}^2 = \frac{\|\mathbf{x}_s - \mathbf{w}_r\|^2}{1 + \beta \frac{n_r(t)}{n}} \quad (3)$$

donde  $n$  representa el número de datos en el conjunto de entrenamiento,  $n_r(t)$  el número de veces que la neurona  $k_r$  fue seleccionada como ganadora en la iteración anterior y  $\beta$  un factor de peso. De esta forma se consigue que las unidades con mayor activación previa tengan mayor tendencia a ser activadas de nuevo, favoreciendo el proceso de convergencia del mapa. En cuanto al factor de ganancia  $\alpha$ , implementan un factor individual para cada unidad ganadora  $k_c$  que depende de su activación en la iteración anterior de la forma

$$\alpha_c(t+1) = \frac{n_c(t)}{n} \quad (4)$$

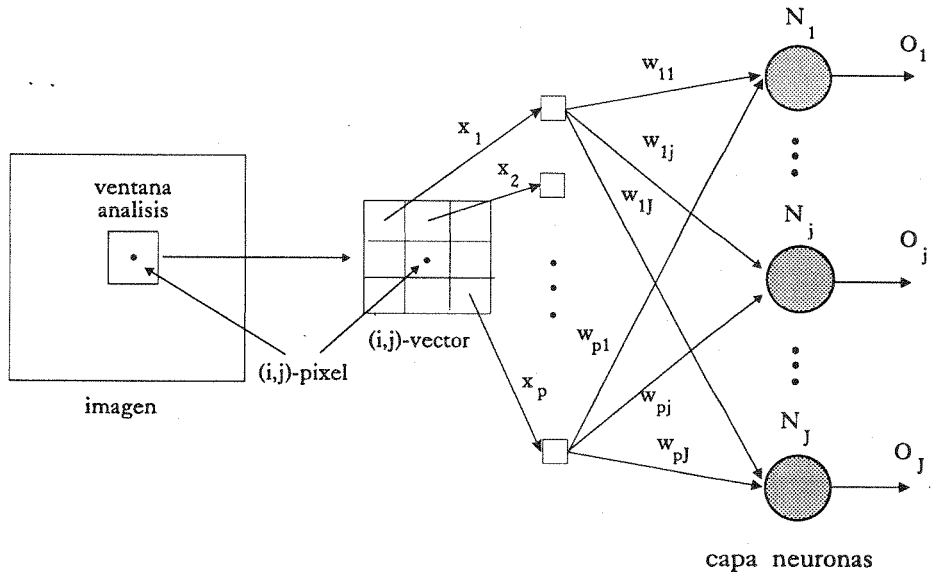
Además, la incidencia que esta ganancia tiene sobre las unidades vecinas en esa iteración viene ponderada por una función decreciente con la distancia a la que se encuentre cada unidad de aquella que presente menor salida y, por tanto, resulte ganadora. Estas elecciones para  $\alpha$  posibilitan que el aprendizaje se concentre en la parte central de la zona de influencia. Finalmente, el criterio de convergencia se establece como la permanencia de un mapa de activación de neuronas en iteraciones sucesivas. Dicho mapa representa el número de veces que cada unidad resulta ganadora en una iteración completa.

Una vez entrenada la red, cada punto del espacio de propiedades se corresponde con el punto más próximo en el mapa generado. De esta forma surge un mapa de activaciones neuronales; la presencia de una clase se asocia a un grupo de neuronas vecinas activadas. Como resultado del proceso extraen información sobre las clases existentes en el espacio de propiedades (número y vectores característicos) así como una segmentación de la imagen en base a información

estadística: píxeles adyacentes cuyos vectores de propiedades se proyecten sobre unidades de un grupo de neuronas formarán una región elemental. En el apartado siguiente mostramos algunos resultados alcanzados con esta técnica.

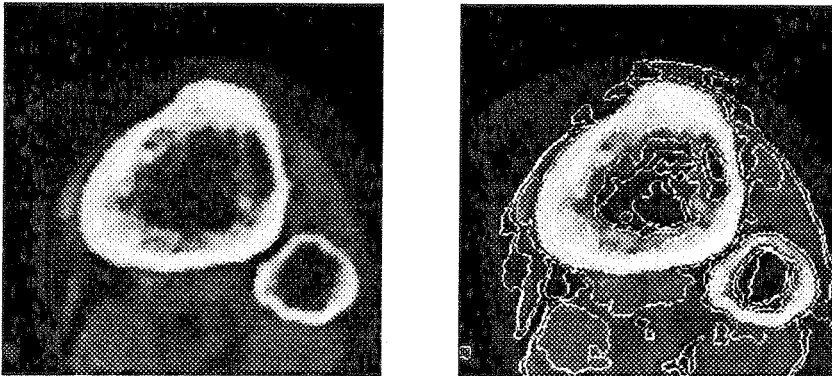
### 2.4. Soluciones basadas en integración de información estadística y espacial

En los esquemas presentados hasta ahora la segmentación se realiza bien en base a información espacial, lo que lleva a segmentaciones incompletas de la imagen, bien en base a información estadística mediante una partición en el espacio de propiedades, que, aunque conduce a segmentaciones completas, no considera la distribución espacial de los píxeles. Un aspecto relevante y reciente consiste en el desarrollo de estrategias que permitan realizar la segmentación considerando ambos tipos de informaciones. Esta idea también se ha plasmado mediante redes neuronales artificiales. Silverman [1991] y Cabello y col. [1993] realizan esa integración para la segmentación de imágenes biomédicas al considerar como vector de propiedades característico de un píxel su propio nivel de gris y el de los píxeles de su entorno. El problema de la segmentación se transforma de nuevo en un problema de clasificación de un conjunto de vectores  $p$ -dimensionales que resuelven mediante una estructura neuronal lineal de  $J$  unidades, conteniendo cada una un vector de pesos  $p$ -dimensional, tal como ilustra la figura 2. Mediante un proceso de aprendizaje autoorganizado, la estructura neuronal genera un mapa topológico que responde a



**Figura 2.** Proceso de segmentación de una imagen mediante una red de Kohonen lineal.  $O = (o_1, \dots, o_j)$ : vector de etiquetas del píxel central de la ventana de entrada en ese instante.

la estructura subyacente en el conjunto de datos. Cada punto del espacio de entradas (vectores  $p$ -dimensionales de los pixels) se corresponden con el punto (vector de pesos de las neuronas) más próximo en el mapa generado. Las regiones elementales surgen de nuevo al trasladar esa información al espacio imagen. A título ilustrativo, la figura 3 muestra las posibilidades de este tipo de redes en la segmentación de imágenes de tomografía computacional de tibia [Cabello y col., 1993]. La figura 3.a muestra una imagen de  $180 \times 180$  pixels y 256 niveles de gris correspondiente a un corte transversal de la tibia en la zona diafisaria alta; la figura 3.b muestra los resultados del proceso de segmentación con una red de 7 neuronas y una ventana de análisis de  $3 \times 3$  pixels. Los contornos de las regiones obtenidas se muestran superpuestos sobre la imagen original. Como podemos observar, la segmentación obtenida responde a la estructura anatómica subyacente; los contornos delimitan las estructuras anatómicas existentes en la imagen, extrayéndose correctamente tanto la geometría del hueso como un mapa de distribución de tejidos de distintas densidades en la imagen. Esta segmentación es difícil de obtener de forma automática mediante los métodos de segmentación clásicos, (Cabello y Col., 1993).



**Figura 3.** a) Imagen de tomografía computacional de una tibia en zona diafisaria alta. b) Contornos obtenidos tras su segmentación mediante la red de Kohonen lineal superpuestos sobre la imagen original.

Siguiendo con la idea de integración de información, otros autores presentan soluciones en base a redes de Hopfield modificadas que realizan clasificaciones válidas minimizando funciones que integran información estadística y espacial. Así, Huang [1992] plantea una red de  $N \times N \times M$  neuronas para segmentar una imagen de  $N \times N$  pixels en  $M$  clases, de forma que asocia a cada pixel  $M$  neuronas, una por cada clase. Las neuronas las modelada como amplificadores no lineales con una relación

entrada/salida monótonica tipo sigmoide. Para resolver el problema la red debe de minimizar una función de energía que formula como suma de tres términos: uno de exclusión, que intenta impedir que dos pixels vecinos pertenezcan a clases distintas, otro de cohesión, que favorece que pixels vecinos pertenezcan a la misma clase y un tercero de inhibición local, que persigue que, de las  $M$  neuronas asociadas a un pixel, sólo una esté activa. La matriz de conexiones surge al proyectar esta función sobre la expresión general de la energía minimizada por la arquitectura de red propuesta. Los niveles de gris de los pixels fijan los potenciales iniciales de las neuronas y la resolución del conjunto de ecuaciones que rigen sus dinámicas lleva a la red a un estado estable en el que cada pixel se asigna a la clase indicada por su neurona activada. Huang muestra la efectividad del método en la segmentación de imágenes en base a sus niveles de gris incluso en presencia de una fuerte componente de ruido gaussiano.

Manjunath y col. [1990] proponen para el análisis de imágenes de texturas un modelo de imagen compuesto: un Campo Aleatorio de Markov Gaussiano para modelar la distribución de niveles de gris de los pixels, supuestos pertenecientes a una misma textura, y un modelo markoviano discreto para la distribución espacial de las etiquetas de los pixels en la imagen. Esas etiquetas indican la clase a la que pertenece cada pixel. Plantean el problema de la segmentación supervisada de imágenes de texturas de  $N \times N$  pixels en  $M$  clases, de las que conocen sus vectores característicos, mediante una red de Hopfield modificada de  $M$  capas con  $N \times N$  neuronas cada una que minimiza una función de energía de Gibbs. De nuevo, la proyección de esta función sobre la expresión general de la energía minimizada por la red determina su esquema de conexiones y entradas. Para su dinámica proponen algoritmos de relajación determinística y estocástica que conducen a un mínimo de la función de Gibbs, que representa una estimación de la distribución de *máxima probabilidad «a posteriori»* de las etiquetas de textura dada la distribución de niveles de gris en la imagen.

Mosquera y col. [1993b] abordan también la segmentación de imágenes de texturas en base a estas ideas, que ilustramos ahora un poco más extensamente. Plantean un problema de segmentación no supervisado, por lo que enlazan con el mapa autoorganizado descrito previamente para extraer la información de clases/texturas existentes en la imagen, así como sus vectores de parámetros característicos, proporcionados por los vectores promedio de los vectores de pesos de las neuronas que integran los distintos grupos de activación existentes en el mapa de activaciones. De nuevo plantean el problema de la segmentación sobre el plano imagen como un problema de optimización; para resolverlo proponen la minimización de una función objetivo que integra información estadística y espacial. Esta función consta de dos términos; el primero corresponde a un error cuadrático de asignación de un punto de la imagen a una textura dada y el segundo favorece la

asignación de puntos vecinos a la misma textura. La función viene dada por la ecuación

$$J \propto \sum_{s \in \Omega} \sum_{l=1}^K \left[ \theta_s - \hat{\theta}_l \right]^2 I_{sl} - \lambda \sum_{s \in \Omega} \sum_{r \in N_s} L_s^t L_r \tag{5}$$

donde  $K$  es el número de texturas en la imagen  $\Omega$ ,  $\hat{\theta}_l$  el vector de parámetros característico de la  $l$ -ésima textura y  $\theta_s$  el vector de parámetros estimado en el pixel  $s$ .  $L_s$  es su vector de etiquetas;  $L_s = (l_{s1}, \dots, l_{sK})^t$ ,  $l_{sl} \in \{0,1\}$  y  $\sum_l l_{sl} = 1$ ; es decir, cada pixel pertenece sólo a una clase.  $N_s$  es el conjunto de vecinos considerado en el modelo de distribución de etiquetas; la etiqueta de un pixel es ahora función tanto de su vector de propiedades como de las etiquetas de sus vecinos.  $\lambda$  es un factor de peso entre ambos términos. Como hemos indicado, tanto el número de clases,  $K$ , como sus vectores característicos,  $\hat{\theta}_l$ , se obtienen como resultado del análisis de la imagen por la red de kohonen descrita previamente.

La minimización de esta función la realizan implementando una red neuronal formada por un número de capas igual al de clases y cada una de ellas compuesta por una matriz de  $M \times M$  neuronas, siendo  $M$  la dimensión de la imagen, tal como ilustra la figura 4. Asumen que los elementos de la red son binarios y están indexados como  $(i,j,l)$ , donde  $(i,j)$  indica la posición en la imagen y  $l$  la capa. La  $(i,j,l)$ -ésima neurona está en ON si su salida  $V_{ijl}$  es 1, y esto indicará que el pixel  $(i,j)$  de la imagen pertenece a la clase  $l$ , de acuerdo con la suposición de que cada pixel sólo puede pertenecer a una clase. Sea  $T_{ijl:i'j'l'}$  el peso asociado a la interconexión entre las neuronas  $(i,j,l)$  e  $(i',j',l')$ ,  $I_{ijl}$  la entrada a la neurona  $(i,j,l)$  y consideremos la red con conexiones simétricas, esto es  $T_{ijl:i'j'l'} = T_{i'j'l':ijl}$  y sin realimentación, es decir,  $T_{ijl:ijl} = 0$ . Entonces la expresión de la energía de la red es

$$E = -\frac{1}{2} \sum_{i=1, j=1, l=1, i'=1, j'=1, l'=1}^{M, M, K, M, M, K} T_{ijl:i'j'l'} V_{ijl} V_{i'j'l'} - \frac{1}{2} \sum_{i=1, j=1, l=1}^{M, M, K} I_{ijl} V_{ijl} \tag{6}$$

Insertando el factor multiplicativo  $1/2$  en (5) e identificando con (6) obtienen los parámetros de la red. Así, considerando  $s = (i,j)$ ,  $I_{sl} = V_{ijl}$  y  $r = (i',j')$ , los parámetros serán:

$$I_{ijl} = (\theta_{ij} - \hat{\theta}_l)^2 \tag{7}$$

$$T_{ijl:i'j'l'} = \begin{cases} \lambda & \text{si } (i',j') \in N_s \text{ y } l' = l \\ 0 & \text{en otro caso} \end{cases}$$

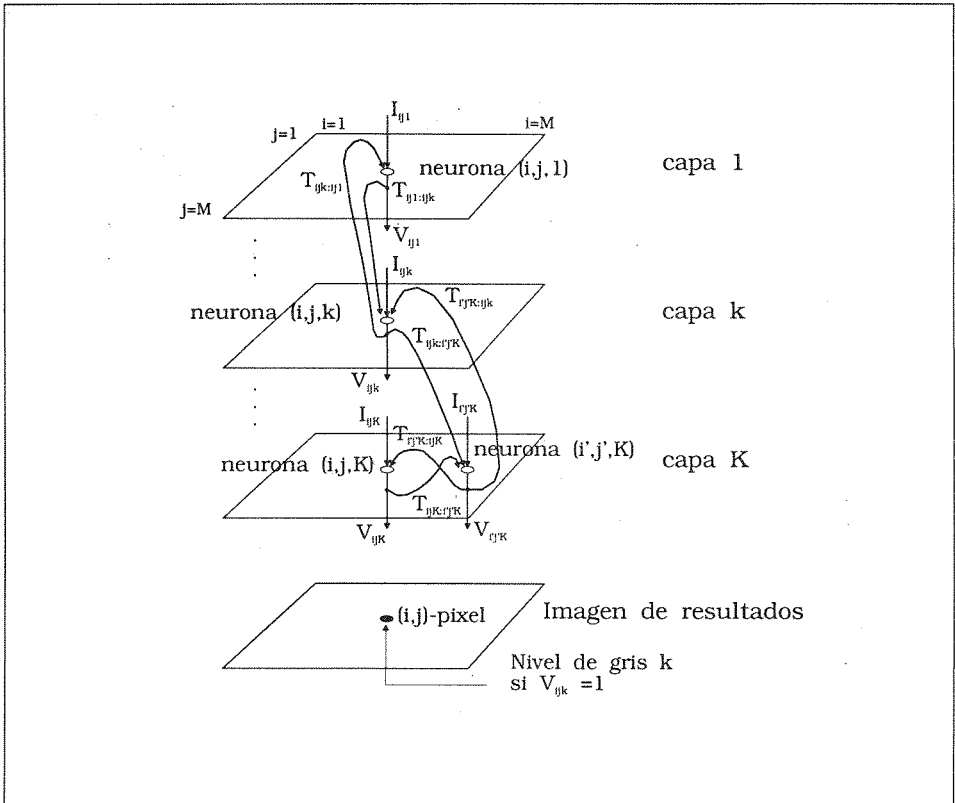
La ecuación (7) indica que no existen conexiones entre neuronas pertenecientes a diferentes capas y, dentro de una capa, las conexiones sólo alcanzan a sus

vecinas en el campo que consideramos. La expresión de la energía de la red puede reescribirse como

$$E = -\frac{1}{2} \sum_{i=1, j=1, l=1}^{M, M, K} U_{ijl} V_{ijl} \quad (8)$$

donde  $U_{ijl}$  representa el potencial de la  $(i,j,l)$ -ésima neurona. Identificando términos con la ecuación (6) y teniendo en cuenta los parámetros dados por (7), este potencial se puede expresar como

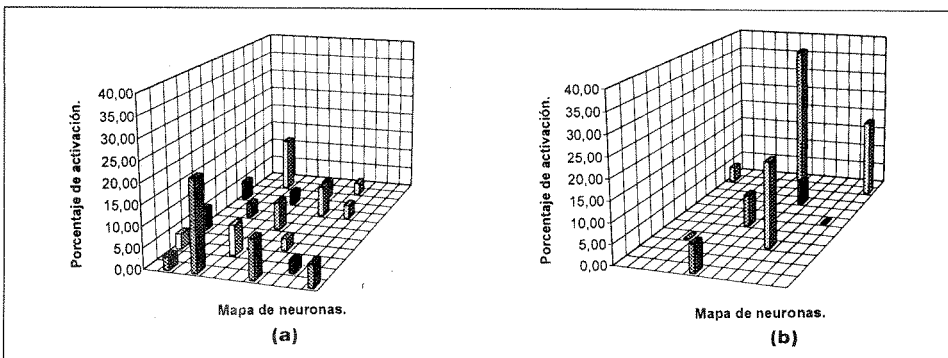
$$U_{ijl} = \sum_{(i',j') \in N_s} V_{i'j'l} - (\theta_{ijl} - \hat{\theta}_l)^2 \quad (9)$$



**Figura 4.** Estructura general de la red de Hopfield multicapa empleada en el proceso de segmentación de imágenes [Mosquera y col., 1993a].

La dinámica de la red conduce a un conjunto de estados de las neuronas de forma que su energía sea mínima. Un algoritmo de relajación determinístico proporciona soluciones válidas. En ese algoritmo, para generar un nuevo estado de la red neuronal se asume que cada pixel tiene una única etiqueta, con lo cual sólo una neurona está activa en cada columna de la red. El nuevo estado de la red se calcula mediante el método «winner-takes-all» sobre cada columna de la red, de forma tal que la neurona con mayor potencial se coloca en *ON* y el resto en *OFF*. Este proceso se asume que es simultáneo para todas las columnas de la red. Este esquema hace converger al sistema hacia un estado estable.

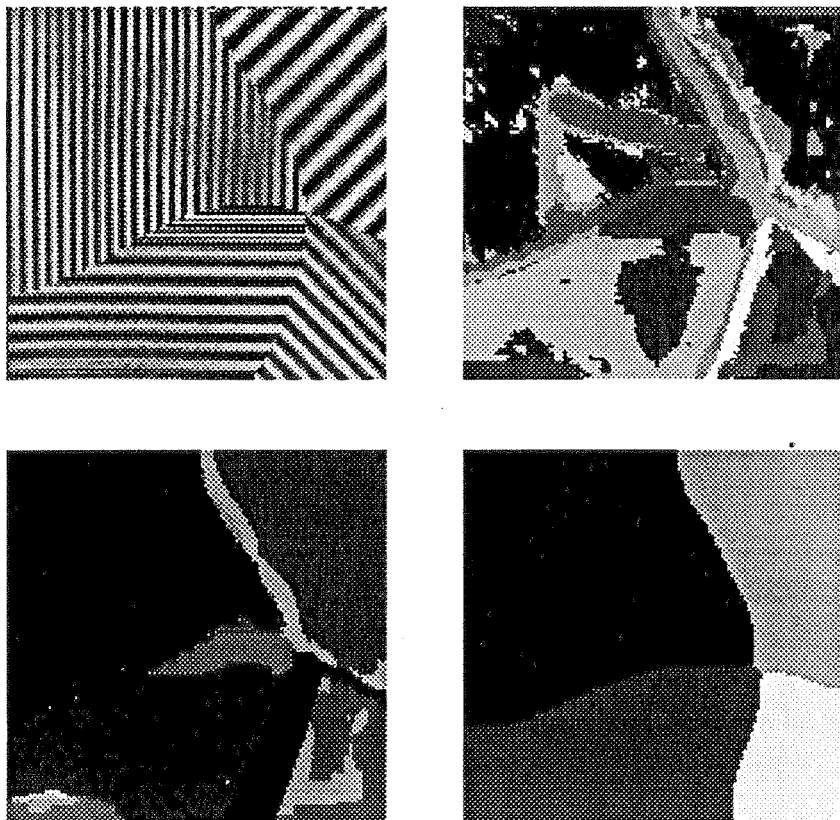
A título ilustrativo, presentamos algunos resultados de este esquema. La figura 6.a muestra una imagen de 128\*128 pixels composición de diferentes texturas sintéticas obtenidas como visualización de un Campo Aleatorio de Markov de segundo orden con distintos vectores de parámetros. Tras un proceso de estimación de los parámetros del campo en la vecindad de cada pixel obtenemos los vectores de propiedades que los caracterizan. Para extraer la estructura de clases empleamos un mapa autoorganizado de 12\*12 neuronas. Con el propósito de comparar resultados lo entrenamos tanto con el esquema clásico de Kohonen como con el esquema modificado y descrito previamente. Una vez realizado el entrenamiento, cada vector del espacio de entradas se clasifica en términos de la unidad que activa por presentar con ella una mayor similitud. De esta forma surge un mapa de activaciones en el que asociamos la existencia de una clase en el espacio de propiedades a un grupo de neuronas vecinas activadas. Las figuras 5.a y 5.b muestran los mapas de activaciones obtenidos tras sendos entrenamientos. Sobre el de la figura 5.a, que responde al esquema de entrenamiento de Kohonen, se detectan 19 grupos



**Figura 5.** Frecuencias de activación de las neuronas en un mapa autoorganizado de 12\*12 elementos de procesado tras el entrenamiento clásico de Kohonen (a), y mediante el esquema modificado propuesto en el texto (b). El espacio de entradas responde a los vectores de propiedades de los pixels de la imagen de texturas mostrada en la figura 6.a.

espaciales de activación aislados, en este caso de una sola neurona cada uno, y sobre el de la figura 5.b, que responde al esquema modificado, sólo 9 [Mosquera y col., 1993a]. Las figuras 6.b y 6.c muestran las segmentaciones obtenidas al transportar esa información de clases al espacio imagen. Las regiones elementales aparecen representadas mediante máscaras. La segmentación proporcionada por la red de Kohonen con el esquema de aprendizaje modificado es mejor que la proporcionada por la red con el aprendizaje original.

Por otra parte, la integración de información espacial en el proceso de segmentación mejora los resultados. Así, la figura 6.d muestra la segmentación proporcionada por una red de Hopfield con 9 capas,  $\lambda = 25$ ,  $N_s$  el conjunto de pixel



**Figura 6.** (de izquierda a derecha y de arriba hacia abajo). a) Imagen composición de distintas texturas sintéticas. b), c) y d) Resultados de la segmentación mediante la red de Kohonen con entrenamiento clásico y modificado y con la red de Hopfield, respectivamente. Las regiones se representan mediante máscaras.

comprendidos en un círculo de radio de 5 pixels y usando como  $\hat{\theta}_i$  los vectores de pesos de las 9 neuronas activadas en el mapa de la figura 6.b. Partiendo de una distribución aleatoria de la activación de las neuronas de la red de Hopfield, mediante el algoritmo de relajación determinístico y tras 40 iteraciones, se alcanza un estado estable en el que sólo hay actividad en 4 capas, detectándose 4 clases de texturas en la imagen. Los contornos de las nuevas regiones generadas, representadas en la figura 5.d mediante máscaras, coinciden de forma apreciable con aquellos percibidos por un observador sobre la imagen original.

### **3. RECONOCIMIENTO DE OBJETOS**

El reconocimiento de objetos es uno de los aspectos del procesado de imágenes que más atención ha recibido por parte de los investigadores, existiendo una bibliografía muy amplia al respecto. Este problema se puede plantear a muy distintos niveles de complejidad, desde el aislamiento y/o reconocimiento de estructuras concretas sobre una imagen hasta la obtención automática de una interpretación global de la misma. El abanico de soluciones propuestas abarca desde técnicas clásicas de reconocimiento de patrones, establecidas en la bibliografía desde hace ya algún tiempo y perfectamente conocidas, hasta técnicas que involucran estrategias y herramientas propias de la inteligencia artificial y que conducen a sistemas basados en conocimiento para la interpretación automática de imágenes.

En la bibliografía se proponen una gran variedad de soluciones basadas en redes neuronales artificiales para abordar la tarea de extracción o reconocimiento de objetos en imágenes. La complejidad de las redes propuestas depende del nivel de complejidad del problema planteado, existiendo soluciones que van desde las redes clásicas (Kohonen, Hopfield, Perceptron Multicapa, ...) para la identificación de objetos hasta estructuras más sofisticadas y complejas que persiguen obtener una interpretación de la imagen. Vamos a comentar a continuación algunas de las soluciones propuestas que consideramos ilustrativas de las distintas formas de plantear y resolver estos problemas.

Ghosh y Pal [1992] estudian la extracción de objetos sobre un fondo ruidoso. Este problema, que en el procesado clásico de imágenes se puede resolver mediante una umbralización adecuada del histograma de la imagen, lo solucionan mediante un esquema autoorganizado ligeramente distinto a los expuestos hasta ahora. Consideran un número de neuronas igual al de pixels en la imagen. Sus entradas son vectores con información local del entorno de cada pixel, por lo que las entradas difieren de una neurona a otra. En el entrenamiento autoorganizado sólo modifican sus pesos aquellas neuronas en las que el producto escalar de su vector de pesos por su vector de entradas supera un cierto umbral, y las de su entorno. El problema de

elección del umbral lo resuelven en base a criterios estadísticos. Una vez entrenada la red consideran que forman parte del objeto aquellos pixels cuyas neuronas presentan en su salida un producto escalar superior al umbral. El algoritmo paralelo que proponen proporciona resultados satisfactorios incluso en entornos muy ruidosos.

La identificación de tumores sobre imágenes médicas es otro de los problemas de reconocimiento de objetos en los que se han planteado con éxito las redes neuronales artificiales. Silverman y Noetzel [1990] y Penedo y col. [1992] plantean redes perceptron multicapa para resolverlo. Los primeros la aplican a la detección de tumores sobre imágenes de ultrasonidos y los segundos a la detección de zonas de imagen sospechosas de ser nódulos pulmonares sobre imágenes de rayos X de tórax. Para ello deben de enseñar a la red a reconocer esas estructuras sobre un fondo anatómico muy complejo y ruidoso. El esquema de aprendizaje, realizado mediante el algoritmo de *backpropagation*, y el análisis posterior de nuevas imágenes, una vez entrenada la red, responde al ya comentado e ilustrado con la figura 1. Así, para el entrenamiento parten de una imagen original sobre la que un experto clínico delimita mediante herramientas de interacción gráfica los contornos de las estructuras patológicas existentes en ella. Se genera entonces una imagen binaria en la que los tumores delimitados se representan mediante máscaras de nivel de gris '1' sobre fondo '0'. Con ambas imágenes, original y máscara, se realiza el entrenamiento. Éste consiste en proyectar sobre las entradas de la red los pixels asociados a las ventanas de análisis sobre la imagen original, asociándole como salida deseada el nivel de gris del pixel central de la ventana equivalente en la imagen máscara. El tamaño de la ventana lo determina el tamaño de las lesiones a detectar. De nuevo, la reiterada exploración de la imagen original irá configurando la estructura de pesos de la red hasta que se satisfaga el criterio de convergencia. En el proceso de análisis se genera una imagen de resultados, de acuerdo con el esquema de la figura 1, en la que el valor de los pixels, entre '0' y '1', se interpreta como una medida de confianza de la existencia de una lesión en esa posición, obteniéndose además una aproximación a la morfología de las estructuras patológicas. Los resultados que obtienen son bastante satisfactorios, con altas sensibilidades y baja razón de falsos positivos.

Li y col. [1989], Nasrabadi y col. [1990] y Wu [1992] plantean el problema de la identificación de objetos en una escena, ya sean aislados o solapados e invariante a traslaciones y rotaciones, desde una perspectiva totalmente distinta. Su objetivo es realizar un «*matching*» entre un modelo de los objetos y la escena, y lo plantean como un problema de optimización que resuelven mediante redes de Hopfield. Los objetos los modelan mediante grafos en base a una técnica de aproximación poligonal; sus nodos son puntos de propiedades tales como los vértices y guardan en ellos información de ángulos y distancias, con los que establecen medidas de compatibilidad. El proceso de reconocimiento consiste en el *matching* entre el *grafo modelo*

y el *grafo escena*, que realizan mediante una red bidimensional de Hopfield de  $N \times M$  elementos de procesado, siendo  $N$  y  $M$  el número de nodos en los grafos modelo y escena respectivamente. Los pesos de las interconexiones responden a las medidas de compatibilidad entre puntos de propiedades y el estado final de las neuronas representará una medida de igualdad entre dos nodos, uno de cada grafo. Una salida próxima a 1 indicará un *matching* válido entre ambos.

Para finalizar, citar los trabajos de Coppini y col. [1992] y Weng y col. [1992]. Los primeros presentan una arquitectura basada en redes neuronales para el reconocimiento y reconstrucción de órganos sobre imágenes médicas 3-D, obtenidas mediante resonancia magnética nuclear o tomografía computacional. El sistema incluye un bloque de entrada o *retina* y dos módulos de procesado; el primero de ellos es un *sistema focalizador de atención*, que determina en cada fase del proceso el órgano a analizar, y el segundo un *sistema de detección de regiones*, que extrae los pixels asociados a ese órgano. Todos los módulos los implementan mediante redes *feed-forward* entrenadas mediante al algoritmo *back-propagation*. Tras procesar uno a uno todos los cortes, el sistema produce como salida una reconstrucción tridimensional de los órganos de interés. Por otra parte, Weng y col. abordan también el reconocimiento de objetos 3-D en escenas del mundo real. Proponen para ello una nueva red neuronal que llaman «*cresceptron*». Es una red autoorganizada, jerarquizada en niveles, que detecta de forma automática conceptos nuevos en la entrada y en respuesta crea nuevas neuronas y sinapsis para memorizar tanto el concepto como su contexto. Las muestras aprendidas se generalizan en items perceptualmente equivalentes. La red reconoce estos items en las imágenes a analizar al asociar conocimiento con la entrada. Estos dos trabajos ilustran bastante bien el desarrollo de nuevas arquitecturas neuronales y su aplicación en tareas complejas de interpretación de imágenes.

#### 4. CONCLUSIONES

En este trabajo hemos pretendido ilustrar mediante una pequeña revisión bibliográfica las posibilidades que las redes neuronales artificiales brindan para abordar y resolver de forma satisfactoria problemas típicos del procesado digital de imágenes. Aún cuando algunas de las soluciones presentadas son una mera traslación de algoritmos clásicos al dominio de las redes neuronales artificiales, otras proponen nuevas formas de abordar problemas conocidos. En cualquier caso, estas nuevas implementaciones presentan una serie de ventajas asociadas a las características de respuesta de las redes neuronales: capacidad de aprendizaje, lo que puede reducir la formulación del problema a un entrenamiento correcto de la red, fácil implementación computacional, potencia y robustez frente al ruido, posibilidad de salida en tiempo real mediante la implementación física de la red, etc. Por ello, las posibles

aplicaciones de las redes neuronales artificiales en problemas de visión artificial constituye un campo de investigación abierto. Tanto la proyección sobre el dominio de las redes neuronales artificiales de problemas típicos de análisis de imágenes, como pueden ser, por ejemplo, la extracción de propiedades en bajo nivel, el *matching* de imágenes o el reconocimiento de objetos, así como el desarrollo de redes específicas para problemas concretos en base a nuevas arquitecturas y/o neuronas de órdenes elevados son objetivos de investigación importantes para el futuro.

### **AGRADECIMIENTOS**

Los autores del trabajo quieren agradecer a la CICYT el soporte económico recibido a través del proyecto TIC91/0816 en cuyo contexto se han realizado algunas de las estructuras descritas en el trabajo.

## REFERENCIAS

- Amatur, S.C., Piraino, D. y Takefuji, T., «Optimization Neural Networks for the Segmentation of Magnetic Resonance Images», *IEEE Transactions on Medical Imaging*, Vol. 11, N. 2, (1992), 215-220.
- Babaguchi, N., Yamada, K., Kise, K. y Tezuka, T., «Connectionist model binarization», *Proc. 10th ICPR*, (1990), 51-56.
- Bezdek, J.C., *Pattern Recognition with Fuzzy Objective Function Algorithms*, Plenum Press, New York, 1981.
- Bezdek, J.C., Tsao, E. Chen-Kuo y Pal, N.R., «Fuzzy Kohonen Clustering Networks», *Proc. first IEEE Conf. on Fuzzy System*, San Diego, California, USA, (1992), 1035-1043.
- Bezdek, J.C., Hall, L.O. y Clarke, L.P., «Review of MR Image Segmentation Techniques using Pattern Recognition», *Medical Physics*, Vol. 20, N. 4, (1993), 1033-1048.
- Blanz, W.E. y Gish, S.L., «A Connectionist Classifier Architecture Applied to Image Segmentation», *Proc. 10th International Conference on Pattern Recognition*, Atlantic City, NJ, USA, (1990), 272-277.
- Cabello, D., Penedo, M.G., Barro, S., Pardo, J.M. y Heras, J., «CT Image Segmentation by Selforganizing Learning», en Mira, J., Cabestany, J. y Prieto, A. (Eds.), *Lecture Notes in Computer Science: New Trends in Neural Computation*, Vol. 686, Springer-Verlag, Berlin, 1993, 651-656.
- Coppini, G., Poli, R., Rucci, M. y Valli, G., «A Neural Network Architecture for Understanding Discrete Three-Dimensional Scenes in Medical Imaging», *Computers and Biomedical Research*, 25, (1992), 569-585.
- Ghosh, A. y Pal, S.K., «Neural Network, Self-organization and Object Extraction», *Pattern Recognition Letters*, 13, (1992), 387-397.
- Hall, L.O., Bensaid, A.M., Clarke, L.P., Velthuizen, R.P., Silbiger, M.S. y Bezdek, J.C., «A Comparison of Neural Network and Fuzzy Clustering Techniques in Segmenting Magnetic Resonance Images of the Brain», *IEEE Transactions on Neural Networks*, Vol. 3, N. 5, (1992), 672-682.
- Huang, Ch.L., «Parallel Image Segmentation using Modified Hopfield Model», *Pattern Recognition Letters*, Vol. 13, (1992), 345-353.
- Huntsberger, T.L. y Ajjimarangsee, P., «Parallel Self-organizing Feature Maps for Unsupervised Pattern Recognition», *International J. General System*, Vol. 16, (1990), 357-392.

- Kamgar-Parsi, B., Gualtieri, J.A., Devaney, J.E. y Kamgar-Parsi, B., «Clustering with Neural Networks», *Biological Cybernetics*, 63, (1990), 201-208.
- Li, W. y Nasrabadi, N.M., «Object Recognition based on Graph Matching Implemented by a Hopfield-style Neural Network», *IEEE Int. Joint Conf. Neural Network*, Washintong, DC, USA, (1989), 287-290.
- Lippmann, R., «An Introduction to Neural Computing», *IEEE ASSP Magazine*, abril, (1987), 4-22.
- Lu, S. y Szeto, A., «Improving Edge Measurement on Noisy Images by Hierarchical Neural Networks», *Pattern Recognition Letters*, Vol. 12, (1991), 155-164.
- Manjunath, B.S., Simchony, T. y Chellappa, R., «Stochastic and Deterministic Networks for Texture Segmentation», *IEEE Transactions on Acoustic, Speech and Signal Processing*, Vol. 38, N. 6, (1990), 1039-1049.
- Mosquera, A., Cabello, D., Mallo, J.M., Carreira, M.J. y Penedo, M.G., «Self-organizing Feature maps for Texture Image Segmentation», en Pölzleitner, W. y Wenger, E. (Eds.), *Image Analysis and Synthesis*, R. Oldenbourg Verlag, München, 1993a, 329-338.
- Mosquera, A., Cabello, D., Carreira, M.J. y Penedo, M.G., «Texture Image Segmentation using a Modified Hopfield Network», en Mira, J., Cabestany, J. y Prieto, A. (Eds.), *Lecture Notes in Computer Science: New Trends in Neural Computation*, Vol. 686, Springer Verlag, Berlin, 1993b, 657-663.
- Nasrabadi, N.M., Li, W. y Choo, Chang Y., «Object Recognition by a Hopfield Neural Network», *Third International Conference on Computer Vision*, Osaka, Japon, (1990), 325-328.
- Pal, N.Ry Pal, S.K., «A Review on Image Segmentation Techniques», *Pattern Recognition*, Vol. 26, N. 9, (1993), 1277-1294.
- Penedo, M.G., Cabello, D., Barro, S., Carreira, M.J. y Mosquera, A., «Application of a Feedforward Artificial Neural Network for Pattern Recognition in Medical Images», en Brody, W.R. y Johnston, G.S. (Eds.), *Computer Applications to Assist Radiology*, Symposia Foundation, Carlsbad, CA, USA, 1992, 472-477.
- Shah, J., «Parameter Estimation, Multiscale Representation and Algorithms for Energy-minimizing Segmentation», *Proc. 10th International Conference on Pattern Recognition*, Atlantic City, NJ, USA, (1990), 815-819.
- Silverman, R.H., «Segmentation of Ultrasonic Images with Neural Networks», *International Journal of Pattern Recognition and Artificial Intelligence*, Vol. 5, N. 4, (1991), 619-628.

- Silverman, R.H. y Noetzel, A.S., «Image Processing and Pattern Recognition in Ultrasonograms by Backpropagation», *Neural Networks*, Vol. 3, (1990), 593-603.
- Springub, A., Scheppelmann, D. y Meinzer, H.P., «Segmentation of Multisignal Images with Kohonens Selflearning topological map», en Klette, R. (Ed.), *Research in Informatics: Computer Analysis of Images and Patterns*, Vol. 5, Akademie Verlag, Berlin, 1991, 148-152.
- Weng, J.J., Huang, T.S. y Ahuja, N., «3-D Object Recognition from Images: A Neural Network Approach», *Proc. 2nd Singapore International Conference on Image Processing*, World Scientific Publishing Co., (1992), 399-403.
- Wu, J.K., «Invariant Image Recognition Using Neural Networks», *Proc. 2nd Singapore International Conference on Image Processing*, World Scientific Publishing Co., (1992), 533-535.