

CODIFICACION DE IMAGENES: UN METODO DE SEGUNDA GENERACION

Ferran MARQUÉS, Josep SAULEDA, Antoni GASULL
Dept. Teoría de la Señal y Comunicaciones
Telf.: 93 - 401 64 57 Fax: 93 - 401 64 47
E.T.S.E.T.B. - U.P.C. Apdo. 30002
08080 Barcelona SPAIN
E-mail: ferran@tsc.upc.es

ABSTRACT

This paper deals with a new technique for coding the shape and position of regions in a segmented image. The shape information is coded by means of a constrained chain code. On its turn, the position of regions within the image is coded by taking advantage of the triple points that appear in the contour image. By using this technique, compression ratios going from 20:1 to 40:1 have been obtained, depending on the initial complexity of the image. In all cases, the quality of the decoded images is fairly better than using classic coding schemes for the same compression ratios.

1.- INTRODUCCION

Las técnicas clásicas de codificación de imágenes se basan principalmente en una división de la imagen en bloques fijos y en el tratamiento individualizado de cada uno de estos bloques [1]. Dependiendo del tipo de tratamiento que se efectúe se tiene un método de codificación u otro. De esta manera, si la codificación de la información contenida en los bloques se realiza tras transformar dicha información a otro dominio, se obtiene el conjunto de técnicas denominadas transformadas. Las transformaciones más usuales son la Transformada Discreta de Coseno (DCT) [1] o la de Karhunen-Loeve (KL) [1]. Otro método de codificación usual representa los bloques obtenidos al hacer la partición de la imagen como vectores de un espacio n -dimensional (siendo n el número de puntos de la imagen en el bloque). Sobre este conjunto de vectores se realiza una Cuantificación Vectorial (VQ) [1].

La idea básica en estos métodos es que la estadística de una imagen es estacionaria a nivel local. O sea, que si se toma áreas pequeñas de la imagen, la información contenida en ellas será fácilmente caracterizable. De este modo se supone que dividiendo la imagen en bloques suficientemente pequeños (típicamente 4×4 ó 5×5 elementos) se podrá conseguir esta situación. En la práctica esta situación no se da, y los bloques que se obtienen no presentan estadísticas estacionarias. Además, el hecho de tratar separadamente los distintos bloques, lleva a un efecto nocivo en la descodificación. Pequeñas variaciones entre la media de los bloques se traduce en la aparición de fronteras ficticias en la imagen descodificada (efecto de bloque) que son visualmente muy molestas para el observador. De hecho, gran parte de la información contenida en una imagen está en los contornos de las figuras. Por tanto, el hecho de introducir contornos erróneos altera substancialmente la interpretación de la misma.

Para intentar paliar estos problemas se han propuesto los llamados métodos de Segunda Generación [2]. Estas técnicas pretenden introducir el conocimiento adquirido sobre el funcionamiento del sistema visual humano en el proceso de codificación. Entre las distintas posibilidades propuestas se encuentra la de realizar una codificación basada en los objetos presentes en la escena. De este modo, se segmenta inicialmente la imagen, obteniéndose así un

conjunto de regiones con características homogéneas, para luego tratar separadamente cada una de estas regiones. Cabe destacar que en los métodos anteriormente descritos también se realiza una segmentación de la imagen (división en bloques), pero en este caso, se pretende obtener una segmentación particular para cada imagen, y no utilizar una partición fija como en los casos anteriores.

Partiendo del supuesto que la segmentación es correcta, la codificación del interior de las regiones es en este caso sencillo ya que su estadística es estacionaria. Además, el tratar separadamente cada región no presenta ahora ningún problema ya que en caso de producir un efecto análogo al anterior efecto de bloque, éste se produciría sobre las fronteras reales de las regiones, reforzando de este modo su presencia.

El precio que se paga al intentar llevar a cabo este esquema de codificación es que en este caso, la ubicación y forma de las regiones (sus contornos) no son conocidas. De esta forma, es necesario introducir una información adicional a la utilizada en los esquemas anteriores. Además, cabe destacar que en los pocos esquemas básicos de codificación basada en regiones que se encuentra en la literatura, la mayor parte de la información necesaria para codificar imágenes se dedica a la codificación de contornos.

La finalidad de este trabajo es aportar un nuevo método de codificación de contornos que reduzca en lo posible la cantidad de información necesaria para representarlos. Para ello se ha partido de un conjunto de segmentaciones obtenidas mediante técnicas basadas en una descomposición multiresolución de la imagen [3]. Como herramienta básica para la codificación de contornos se ha utilizado la técnica de "chain code" (CC) [4] que, pese a su simplicidad, parece ofrecer las mejores prestaciones.

Tras esta introducción, la estructura de la comunicación es la siguiente. En el segundo apartado se discute el tipo de representación que se ha tomado para las imágenes de contornos. En el apartado siguiente se resume el método de codificación, introduciendo el concepto de puntos triples. Finalmente, en el cuarto apartado, se comenta los resultados, presentando dos ejemplos.

2.- DEFINICION DE IMAGEN DE CONTORNOS

Los contornos representados en una imagen segmentada son los cambios de etiqueta que se producen dentro de la imagen. Por tanto, un elemento de contorno viene determinado por una variación entre dos elementos de la imagen segmentada, y no por un solo elemento básico de la imagen. De este modo, si se pretende crear una imagen de contornos que contenga información binaria (discriminar únicamente entre las posiciones en las cuales hay contorno y en las que no), es necesario utilizar una estructura de datos distinta a la de la propia imagen. Para ello, se ha utilizado una representación de imagen en la cual se tiene en cuenta elementos propios de la imagen (pixels), y el conjunto de elementos de contorno que rodean a cada pixel. Esta estructura de datos se representa en la figura 1. En ella se representa mediante cuadrados los elementos de la malla básica (pixels) y mediante líneas los elementos de contorno. Los círculos son elementos que no se utilizan en la representación.

Mediante esta estructura de imagen es posible representar binariamente una imagen de contornos, con lo cual se puede aplicar para su codificación CC. Cabe destacar que en este caso, la discusión sobre el tipo de conexión que tiene el contorno (contornos 4-conectados u 8-conectados [4]) no se tiene sentido. Esto es debido a que la malla sobre la que se definen los elementos de contornos es hexagonal, por lo que todo elemento de contorno tiene un único vecindario posible 6-conectado. Este cambio de tipo de conexión podría parecer que genera un aumento del número necesario de códigos para codificar el contorno mediante CC (se pasa de

un número mínimo de 4 posibilidades a un número fijo de 6). De hecho no es así, ya que si se toma "chain code" derivativo (DCC) [4] en ambos casos se obtiene que tan sólo es menester utilizar 3 códigos diferentes.

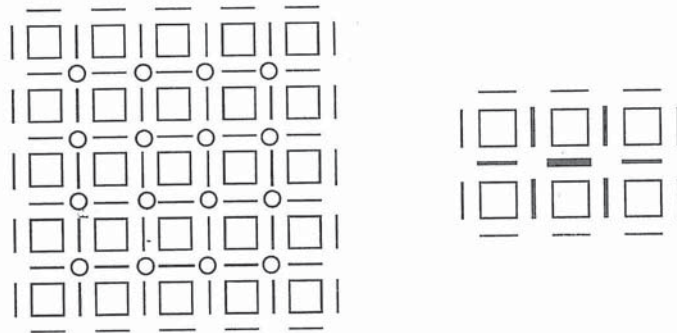


Fig. 1.- Esquema de la estructura de datos utilizada y ejemplo de elemento de contorno 6-conectado.

3.- CODIFICACION MEDIANTE DCC Y PUNTOS TRIPLES

Como se ha comentado anteriormente, la codificación de contornos implica dos tareas distintas: la codificación de su forma y la de su ubicación en la imagen. Aunque ninguno de los dos temas ha sido demasiado tratado en la literatura, sobre la codificación de la forma se ha trabajado más bajo el aspecto de digitalización de curvas y reconocimiento de formas. Por su parte, en el tema de la ubicación de las regiones en la imagen no existen casi trabajos anteriores, pese a que es una de las partes del esquema de codificación que más cantidad de información requiere.

En este trabajo se ha pretendido realizar un algoritmo que efectúe, en lo posible, ambas codificaciones simultáneamente. De este modo, si se codifica la forma de la región mediante una cadena de movimientos (CC), sólo es necesario para reproducir la imagen de contornos dar la posición de inicio de cada cadena. En el peor de los casos, esta posición se puede codificar mediante las coordenadas directas del punto inicial (2 octetos en caso de imágenes de 256x256). Realmente, se puede utilizar la información contenida en la cadena de una región para marcar los puntos de inicio de otras cadenas. Esto se puede realizar introduciendo en la cadena un nuevo código que marque los puntos triples dentro de los contornos. Como punto triple se entiende aquel punto del contorno que tiene más de dos elementos de contorno vecinos (elementos de contorno 6-conectados al punto en cuestión). Estos puntos marcan la existencia de dos contornos que se tocan y, por tanto, la presencia de dos regiones vecinas. De este modo, estos puntos triples pueden tomarse como inicio de nuevas cadenas de códigos que no se necesitará referenciar explícitamente.

Mediante esta técnica, se consigue ubicar todas las regiones de la imagen cuyos contornos tengan, como mínimo, un punto de contacto con otros contornos. Quedan por tanto excluidas las regiones completamente interiores a otras regiones. En estos casos es necesario codificar su posición explícitamente. De hecho, la reducción en información es drástica ya que sobre el conjunto de segmentaciones que se ha utilizado, el tanto por ciento de regiones totalmente interiores que ha aparecido es menor al 10%.

La codificación de los contornos, como ya se ha comentado, se realiza mediante DCC. Por tanto, el conjunto básico de códigos que representa los movimientos dentro del contorno está formado por tres elementos (*r*: recto, *d*: derecha e *i*:izquierda). Inicialmente, a este conjunto de códigos se debe añadir un cuarto para marcar los puntos triples, *t*. Una vez se obtiene la cadena de movimientos del contorno, ésta se codifica mediante técnicas de Huffman [5]. Para ello se forman agrupaciones de movimientos, y se asigna a cada grupo un código determinado teniendo en cuenta su probabilidad de aparición. Los estudios realizados sobre el orden óptimo de las agrupaciones muestran que, para un conjunto amplio de imágenes, la estadística de las agrupaciones se mantiene casi constante hasta el orden 3. Es por esto que se ha optado por formar palabras de código agrupando los movimientos de tres en tres.

Para aumentar la capacidad de compresión, se ha seguido la idea presentada en [6]. En este trabajo, se propone introducir condiciones en la formación de los contornos para evitar la aparición de algunas agrupaciones concretas de movimientos. De este modo, se ha forzado a que las regiones en la segmentación no presenten puntos aislados. Se entiende como puntos aislados de una región, aquellos puntos que sólo tienen uno de sus vecinos perteneciente a la misma región. Así, se anula la posibilidad de tener los conjuntos *ii* o *dd* en las cadenas. Al anular esta posibilidad, se puede actuar de dos formas distintas: o bien tenerlo en cuenta a la hora de codificar pero seguir con cuatro códigos básicos (*r*, *d*, *i*, *t*), o bien transformar el código *t* en uno de estos conjuntos prohibidos *ii* o *dd*. El estudio del problema ha mostrado que la compactación de la información resulta mayor si se opta por esta última posibilidad. Por tanto, las cadenas de códigos se forman con únicamente tres elementos básicos (*r*, *d*, *i*). La decisión de qué código prohibido debe substituir en la cadena al código *t* viene dada por el entorno en el cual aparece este código. Si su substitución por un conjunto *ii* puede llevar a confusiones, se decide introducir el conjunto *dd* y viceversa. Este tipo de problemas surgen, por ejemplo, en cadenas del tipo *...rdtr...* donde la substitución llevaría a *...rdddr...* y no es posible saber que par *dd* en la cadena implica un punto triple.

Cabe destacar que la codificación de las cadenas no se debe hacer automáticamente mientras se recorre el contorno. Esto es así ya que al recorrer el contorno pueden aparecer en la cadena marcas de puntos triples que posteriormente no deban ser codificadas como tales. Este es el caso de, por ejemplo, dos regiones cuadradas unidas por un lado. Al recorrer el contorno del primer cuadrado, se hallarán dos puntos triples producidos por el segundo cuadrado. De hecho, el segundo punto no es necesario marcarlo como punto triple ya que no da información nueva sobre la posición de ninguna región.

4.- RESULTADOS

En esta comunicación únicamente se ha tratado el problema de codificación de los contornos de una imagen segmentada, sin estudiar el problema de la codificación del interior de las regiones. Para representar esta información se ha utilizado la media del nivel de gris de cada región, lo cual da buenos resultados visuales, ya que la segmentación inicial es de alta calidad. Se debe tener en cuenta que, ya que el método de codificación no presenta pérdidas a partir de la imagen segmentada, los resultados que se presentan son los de segmentación. De este modo, en la figura 2.a se muestra una imagen original junto al resultado tras descodificar. Sobre esta imagen sencilla, la compresión alcanzada es de 40:1, siendo el número de bits por pixel de contorno (bppc) 1.37. Por su parte, en la figura 2.b se muestra el caso de una imagen más complicada, con la cual se consigue una compresión de 21:1 y el número de bppc es de 1.27. Se debe ver que, aunque en el segundo caso la compresión general es menor, el bppc necesario también lo es. Esto es debido al distinto número de regiones que produce la segmentación en cada caso.

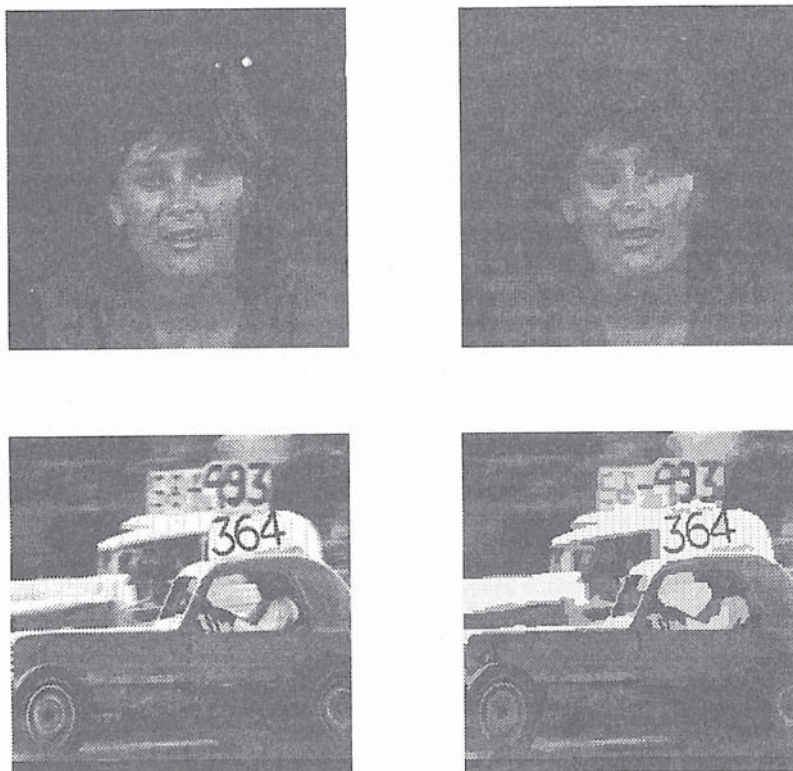


Fig. 2.a.- Imagen sencilla y su codificación a partir de una segmentación con 128 regiones.
 Fig. 2.b.- Imagen compleja y su codificación a partir de una segmentación con 256 regiones.

REFERENCIAS

- [1] *N.S. Jayant and Peter Noll*. "Digital coding of waveforms" Prentice-Hall, inc. Englewood Cliffs, New Jersey 1984.
- [2] *M. Kunt, Ikonopoulou, M. Kocher*, "Second-Generation image coding techniques", Proc. IEEE, vol. 73, num. 4, pp. 549-574, April 1985.
- [3] *F. Marqués, J. Cunillera, A. Gasull*, "Hierarchical segmentation using Compound Gauss-Markov Random Fields", Proceedings de la ICASSP 92 - International Conference on Acoustics, Speech, and Signal Processing, pp. III. 53-56, San Francisco USA, Marzo 1992.
- [4] *H. Freeman*, "On the encoding of arbitrary geometric configurations", IRE Trans. Electron. Comput., vol EC-10, pp. 260-268, June 1961.
- [5] *D. A. Huffman*, "A method for the construction of minimum-redundancy codes", Proceedings of the IRE, pp. 1098-1101, 1952.
- [6] *M. Eden, M. Kocher*, "On the performance of a contour coding algorithm in the context of image coding. Part I: contour segment coding", Signal Processing, vol. 8, pp. 381-386, 1985.