# UTILIZACIÓN DE MAPAS DE AUTOORGANIZACIÓN EN LA CLASIFICACIÓN DE COEFICIENTES WAVELET PARA LA CREACIÓN DE ESCENAS DE VIDEO NATURALES

Mauricio Díaz Melo, *M. Sc.*
Corporación Universitaria Unitec

Pedro Raúl Vizcaya Guarín, *Ph. D.*
Pontificia Universidad Javeriana

## Resumen

*Este artículo plantea un nuevo método de parametrización y clasificación de secuencias de video para ser usadas en telefonía visual, basado en el uso de una transformada multiresolución, los conceptos de codificación subbanda y los mapas de autoorganización de características (SOFM) comparado con cuantificadores vectoriales LBG con criterio minimax. Dadas las características de estas escenas, emplear transformaciones de este tipo resulta útil para generar diferentes niveles de detalle en los coeficientes a clasificar y para su posterior utilización durante la transmisión comprimida de un libro de códigos, base para la reconstrucción en el receptor. Igualmente, el clasificador con criterio minimax permite la inclusión de muestras atípicas que, bajo otras circunstancias, pueden clasificarse erróneamente, mientras que el libro de códigos encontrado con el clasificador SOFM presenta características importantes como, por ejemplo, la organización topológica.*

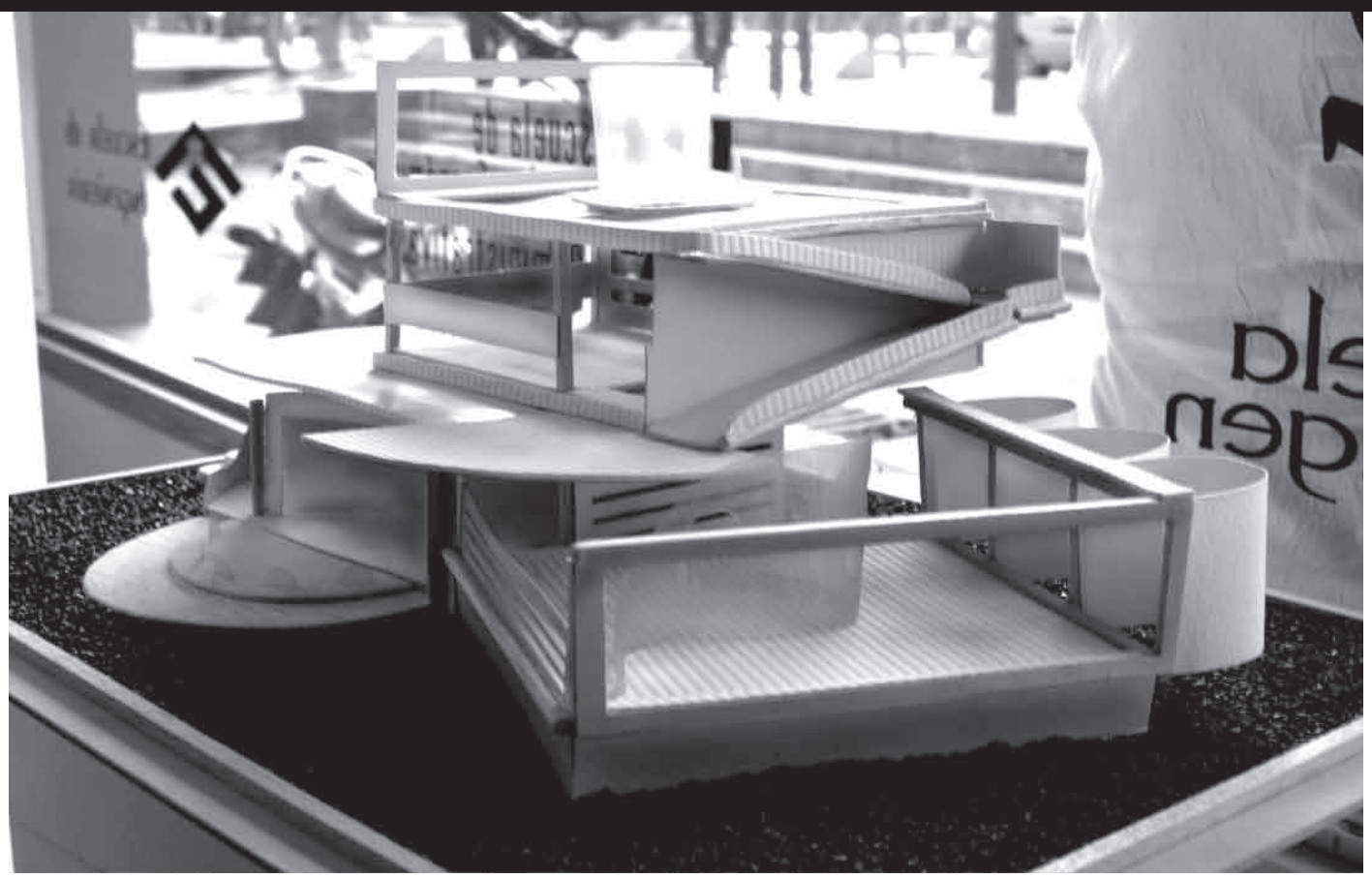
**Palabras claves:** cuantificación vectorial, minimax, SOFM, telefonía visual, wavelet.

## Introducción

Actualmente los sistemas de telefonía visual se basan en estándares (ITU H.100 1998, ITU H320 1999) que no aprovechan óptimamente las cualidades de este tipo de escenas. La información que presentan las secuencias de video comparadas con las de audio en sistemas de telefonía visual, es bastante menor (Essa, 1995). Sin embargo, estudios realizados por diversos investigadores como Green y Kuhl, Sumby y Pollack, O'Neill, citados por Mark y Allen (1994), y el trabajo desarrollado por Bárcenas *et ál*. (2001), concluyen que el video es útil para aumentar o mejorar la inteligibilidad de señales de voz

Mauricio Díaz es ingeniero electrónico de la Pontificia Universidad Javeriana y actualmente se desempeña como coordinador del Centro de Investigación de la Escuela de Ingeniería, Corporación Universitaria Unitec (Bogotá, Colombia); mdiaz@unitec.edu.co. Pedro R. Vizcaya G. es ingeniero electrónico, realizó sus estudios de doctorado en Rensselaer Polytechnic Institute, Troy, (N. Y., EE. UU.) y en la actualidad es profesor asociado del Departamento de Electrónica de la Pontificia Universidad Javeriana (Bogotá, Colombia); pvizcaya@javeriana.edu.co.

*Vestigium,* 3, (1). Febrero de 2007, pp. 34-44.



en ambientes ruidosos, en valores que van desde uno hasta 15 decibeles, pero que con la carencia del audio el mensaje es prácticamente incomprensible, mientras que en el caso contrario no sucede lo mismo.

Diferentes alternativas enfocadas al desarrollo de un sistema de telefonía visual se han desarrollado en los últimos años. Entre las primeras aproximaciones al problema se encuentra la solución planteada por Bárcenas *et ál*. (2001), trabajo en el cual se realizan cambios morfológicos entre diferentes imágenes almacenadas para obtener secuencias de video creíbles; sin embargo, el tiempo de procesamiento que esto implica no permite tener un sistema en tiempo real. Posteriormente se planteó la idea de realizar el procesamiento en un dominio paramétrico (Machado y Santa, 2001), surgiendo así una alternativa que permitió implementar un conversor de texto a voz visual en tiempo real (AVSS). A partir de las necesidades de segmentación de la región de interés (boca), se creó el sistema SPARV (Sistema de segmentación automática de rostros en video) (Baptiste, Sotomayor y Vizcaya, 2002) en el cual se analizaron diferentes métodos de parametrización, incluyendo la transformada discreta de Fourier y un método para la segmentación automática de la boca utilizando las características de movimiento y la detección de la piel por medio del color. Hacia mediados del 2004 se presentó un trabajo que analizó a fondo diferentes transformaciones como métodos de parametrización (Soto y Vizcaya 2004), incluyendo la DCT y el análisis de componentes principales. Igualmente, se desarrolló un algoritmo de interpolación entre imágenes que genera secuencias naturales y se creó un libro de códigos de tamaño reducido. Esta investigación permitió la implementación completa del sistema en tiempo real.

El proyecto "Telefonía visual por canales de muy baja capacidad" (Vizcaya *et ál*., 2004) permitió unir todos los logros desarrollados en trabajos anteriores e implementar un sistema prototipo de videofonía. El proyecto consta de tres módulos claramente definidos: segmentación, codificación y transmisión. La segmentación se realiza utilizando características de la imagen en el plano de luminancia, aplicando sumas acumulativas a lo largo y ancho de la imagen. Esto determina una primera región de interés que sirve de búsqueda para la región de la boca. La codificación de las imágenes se realiza en el dominio de los parámetros (DCT). El diseño del libro de códigos se basa en un criterio que minimiza el error máximo entre las muestras y sus representantes, teniendo en cuenta que se quieren generar secuencias naturales, a diferencia del criterio más ampliamente usado que minimiza la relación señal a distorsión; se utilizó la distancia L1 como métrica. También se desarrolló un algoritmo de interpolación de imágenes basado en la búsqueda de Viterbi.

Este documento plantea una nueva propuesta para un algoritmo de parametrización y codificación de escenas típicas de telefonía visual, basado en la transformada Wavelet y en los conceptos de codificación sub-banda. Dadas las características de este tipo de escenas, el uso de esta transformación es útil para hacer una selección de coeficientes basada en la varianza temporal y su ubicación espacial. Por otra parte, se plantea la creación de un libro de códigos usando el algoritmo SOFM con diferentes

topologías y dimensionalidades. La principal característica que otorga el uso de este clasificador es la organización topológica de la base de datos obtenida.

## Parametrización y selección de características

Dentro del proceso de codificación de una señal cualquiera, es muy común trabajar con una representación de ésta en un espacio que decorrelacione los diferentes datos. Esto se logra encontrando una transformación que genere unos coeficientes (espacio de características) que no posean información redundante, reduciendo de esta forma el tiempo de procesamiento y los requerimientos del sistema. En el área de compresión de imágenes es usual utilizar transformadas como la DCT que implementa el estándar JPEG y, recientemente, se ha incrementado el uso de transformadas multiresolución como las wavelets, en nuevos estándares como JPEG2000 y MPEG4 (González y Woods, 2003, Skodras *et ál*., 2001). Una transformada multiresolución ofrece interesantes características que permiten un análisis más global de este tipo de señales.

## Expansión en series de wavelet

Frecuentemente, una señal o función $f(x)$ puede analizarse de una mejor forma como una combinación lineal de funciones de expansión, cuando estas forman una base del espacio vectorial que contiene a la señal (González y Woods, 2003):

$$f(x)=\sum_k \alpha_k \varphi_k(x) \quad (1)$$

Aquí *k* es un coeficiente entero, finito o infinito, $\alpha_k$ corresponde a valores reales llamados coeficientes de expansión y $\varphi_k(x)$ valores reales llamados funciones de expansión. Si solamente existe un conjunto de coeficientes $\alpha_k$ para un $f(x)$ dado y son linealmente independientes, las funciones $\varphi_k(x)$ se llaman funciones base, mientras que el conjunto de expansión $\{\varphi_k(x)\}$ es denominado una base. Un ejemplo de una base son las funciones exponenciales complejas, que forman un gran espacio de señales (todas aquellas que se pueden representar mediante la serie de Fourier). Estas funciones forman un espacio $V$ de funciones, en homología a un espacio vectorial que se denota de la siguiente manera:

$$V = Span\{\varphi_k(x)\} \quad (2)$$

La expansión en series de wavelets de una función

$f(x) \in L^2(\Re)$ relativa a una función wavelet $\psi(x)$ y una función de escala $\varphi(x)$ se define, de acuerdo a la ecuación 1, de la siguiente forma (Strang y Nguyen, 1997):

$$f(x) = \sum_k c_{j_o}(k)\varphi_{j_o,k}(x) + \sum_{j=j_0}^{\infty}\sum_k d_j(k)\psi_{j,k}(x) \quad (3)$$

Aquí $j_0$ es una escala de inicio arbitraria, los coeficientes $c_{j_o}(k)$ son conocidos como los coeficientes de aproximación o escala y los coeficientes $d_j(k)$ se nombran como coeficientes wavelet o de detalle. Estos nombres se dan porque la primera sumatoria de la ecuación 3 usa funciones de escala que proveen una aproximación de $f(x)$ a una escala $j_0$. Para cada escala $j \geq j_0$ en la segunda sumatoria, una función de resolución más fina (suma de wavelets) es adicionada a la aproximación, aumentando de esta forma el nivel de detalle. Si la función de expansión forma una base ortonormal, los coeficientes de la expansión pueden ser calculados con el producto interno:

$$c_{j_0}(k) = \langle f(x), \varphi_{j_0,k}(x) \rangle = \int f(x) \cdot \varphi_{j_0,k}(x) \cdot dx$$

y

$$d_j(k) = \langle f(x), \psi_{j,k}(x) \rangle = \int f(x) \cdot \psi_{j,k}(x) \cdot dx \quad (4)$$

Para el caso de las bases biortogonales los términos $\varphi$ y $\Psi$ en las ecuaciones anteriores, son reemplazados por sus funciones duales $\tilde{\varphi}$ y $\tilde{\Psi}$, respectivamente (Antonini *et ál*., 1992).

## Selección de la función wavelet

La influencia de las funciones escogidas como base de la transformación, afecta directamente todos los aspectos de un sistema de codificación con wavelets. La principal característica a tener en cuenta radica en la complejidad computacional de la transformación, mientras que en segundo lugar se haya la habilidad del sistema para comprimir y reconstruir imágenes con un aceptable error.

Para la reconstrucción exacta de una imagen fotorrealista[1] es apropiado usar una base ortonormal con una función wavelet "suave", es decir, que no contenga saltos bruscos, ya que este tipo de bases permite representar la imagen con unos pocos coeficientes. Los filtros por medio de los cuales se implementa esta transformación deben ser cortos (de no muchos coeficientes). Por otra parte, es deseable que los filtros resultantes sean FIR con fase lineal para su fácil implementación en cascada, sin la necesidad de ser compensados en fase. Sin embargo, no hay una solución no trivial para hallar filtros ortonormales FIR de fase lineal con las características de reconstrucción exactas (Antonini, *et ál*., 1992). Es posible preservar la linealidad de fase sacrificando el criterio de ortonormalidad y usando bases biortogonales. Es por esto que las funciones wavelets más ampliamente usadas en sistemas de compresión son las wavelets de Daubechies y las wavelets biortogonales. El estándar de compresión JPEG 2000 (Skodras *et ál.*, 2001) utiliza la función wavelet conocida como Daubechies 7/9. En este trabajo se utiliza esta misma función.

A continuación se muestra la función de escala y la función discreta wavelet para los filtros de análisis y síntesis.

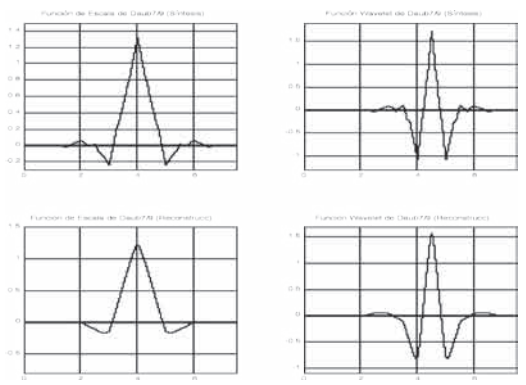


**Figura 1.** Funciones de escala y wavelet.

La función de escala wavelet en dos dimensiones se muestra a continuación.

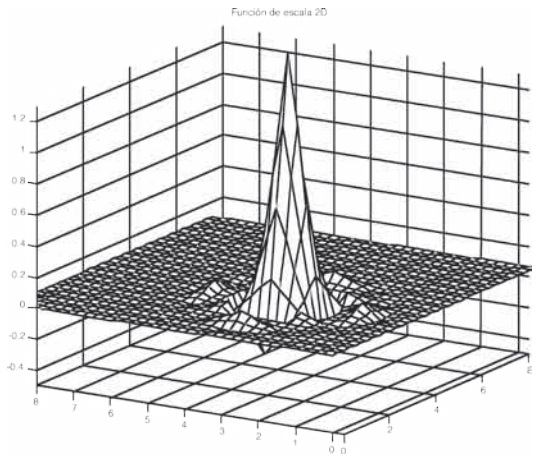



**Figura 2.** Función de escala en 2D.

## Selección de coeficientes

Para la selección de los coeficientes a clasificar, se hace una búsqueda de aquellos que presentan mayor varianza en una determinada región de interés que contiene la cara del sujeto en la escena. Este análisis de movimiento se realiza

utilizando la acumulación de las diferencias absolutas entre píxeles de cuadros consecutivos. Finalmente se obtiene una máscara que indica cuáles de esos coeficientes tienen una mayor varianza temporal. Este análisis evita el almacenamiento de todos los coeficientes del video de entrenamiento, manteniendo solamente el acumulador, los coeficientes actuales y los inmediatamente anteriores.

Una vez encontrada la máscara con los valores de varianza temporal de los coeficientes en el video de entrenamiento, es posible seleccionar cuáles de estos coeficientes son los que se usan para conformar el clasificador. Esta alternativa permitiría ir variando los coeficientes de forma dinámica para crear un clasificador adaptable, sin embargo, este tema no se discutirá en este trabajo.

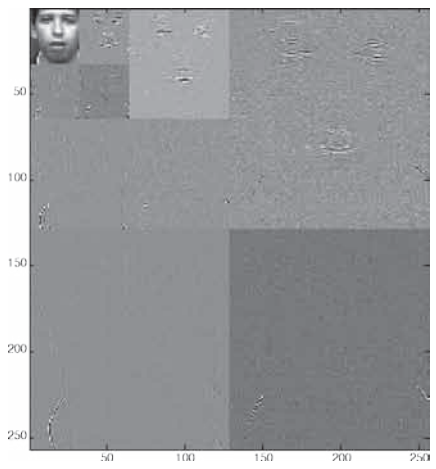



**Figura 3.** Transformación de 3 niveles de 256x256.

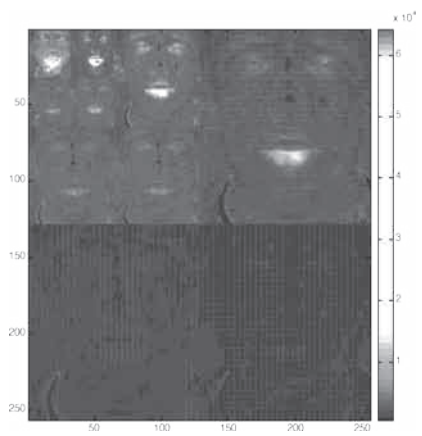



**Figura 4.** Varianza temporal de los coeficientes usados en el entrenamiento.

Inicialmente se aplicó un umbral global para encontrar los coeficientes que se utilizan en el clasificador. Se probaron diferentes valores en el intervalo del valor medio del acumulador temporal hasta 10 veces ese mismo valor medio. Sin embargo, vale la pena señalar, como se muestra en la figura 6, que la aplicación de un umbral determinado para cada nivel de la transformación es más indicada, ya que de otra forma pueden no ser tenidos en cuenta coeficientes que sirven en el proceso de clasificación. Este

umbral para cada nivel está determinado de la siguiente forma:

$$u(x,y)=\begin{cases} f(x,y) & \text{si} \quad f(x,y)\le 0.9\cdot(\max(f(x,y))) \\ 0 & \text{otro caso} \end{cases} \qquad (5)$$

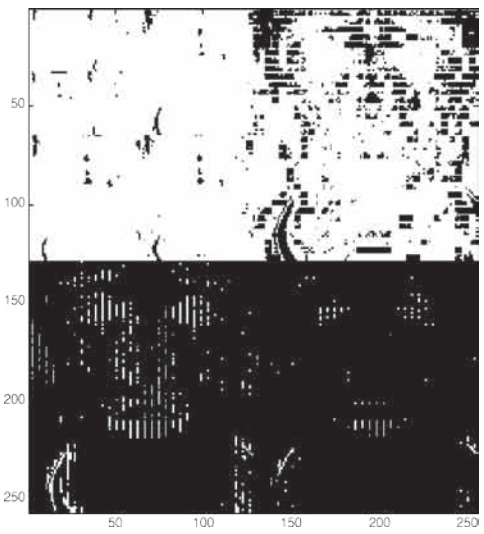



**Figura 5.** Coeficientes seleccionados con un umbral global (en blanco).

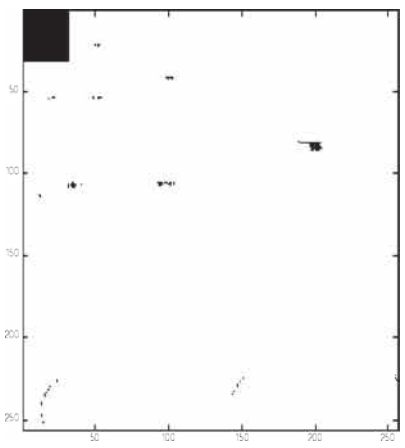


**Figura 6.** Ubicación espacial de los coeficientes seleccionados con un umbral diferente para cada nivel (en negro).

## Obtención del libro de códigos
## Cuantificador Vectorial con criterio minimax

La cuantificación vectorial o cuantificador multidimensional es un concepto desarrollado ampliamente desde comienzos de la década de los ochenta (Gray, 1984, Linde *et ál*., 1980, Nasrabadi, 1985). Es un sistema que asigna una secuencia de vectores continuos o discretos en una secuencia digital, para usar sobre un canal de comunicación o para su almacenamiento digital. Su principal éxito radica en la compresión de los datos. Esta asignación puede o no tener memoria en el sentido de dependencia de las acciones pasadas del codificador, tal y como sucede en el caso escalar, por ejemplo en técnicas como PCM (Lloyd, 1982).

Aunque la teoría de información establece que siempre se conseguirán mejores resultados codificando vectores que escalares, los cuantificadores escalares han permanecido durante años como los sistemas de compresión de datos más comunes, lo cual se debe, en gran parte, a su simplicidad y buen desempeño cuando la capacidad del canal es bastante grande.

Sin embargo, la disminución de las tasas de bits para los canales de comunicación y la necesidad de un mejor uso del canal, han dado paso a diversos algoritmos de diseño que han desarrollado una gran variedad de cuantificadores vectoriales para diversas aplicaciones, entre ellas la telefonía visual.

En el trabajo de Vizcaya *et ál.* (2004), se desarrolla un método de cuantificación que permite encontrar un libro de códigos apto para la transmisión de telefonía visual. Este método se basa en el algoritmo LBG (Linde-Buzo-Gray) (Linde *et ál*., 1980), sin embargo, no utiliza el paradigma del error cuadrático medio (MSE) como medida de distorsión para el cálculo de nuevos centros, sino un criterio que minimiza el máximo error entre el centro y sus muestras (minimax). Esta aproximación permite la construcción de un video resintetizado más natural, ya que este criterio permite incluir imágenes atípicas y descartar la inclusión de imágenes muy parecidas. El algoritmo iterativo para la búsqueda del libro de códigos apropiado (*codebook*), se basa fundamentalmente en dos condiciones que asignan el nuevo centro a una región y que determinan la pertenencia o no de las muestras a una clase, usando el criterio minimax.

## Cuantificador Vectorial con SOFM (mapas de autoorganización de características)

Una de las principales aplicaciones de los mapas de autoorganización de características, desarrollados por Kohonen (Kohonen, 1984, Haykin, 1994), consiste en transformar una señal patrón de dimensiones arbitrarias en un mapa discreto de una o dos dimensiones, y realizar esta transformación de forma que se adopte un ordenamiento topológico. Kohonen basó su desarrollo en el funcionamiento del cerebro humano y la forma en que las capas de la corteza cerebral se organizan de acuerdo a la función fisiológica que desarrollan. Una aproximación inicial se plantea desde el punto de vista de las redes neuronales. El algoritmo se puede resumir de la siguiente manera:

1. Inicialización. Escoger valores aleatorios para el vector inicial de pesos sinápticos $\mathbf{w}_j(0)$. La única restricción que

existe es que $\mathbf{w}_j(0)$ debe ser diferente para $j = 1,2,...N$, donde N es el número de neuronas.
2. Muestreo. Dibujar una muestra $\mathbf{x}$ de la distribución de entrada con una cierta probabilidad; el vector $\mathbf{x}$ representa la señal sensada.
3. Criterio de similaridad. Encontrar la neurona que "mejor se acomoda" (neurona ganadora) $i(\mathbf{x})$ en un tiempo n, usando el criterio de mínima distancia Euclidiana.

$$i(\mathbf{x}) = \arg\min \left\| \mathbf{x}(n) - \mathbf{w}_j \right\|, \quad j = 1,2,...,N \tag{6}$$

4. Actualizar. Ajustar los vectores de pesos sinápticos de todas las neuronas, usando la fórmula de actualización que involucra a los vecinos (regla de Kohonen).

$$\mathbf{w}_j(n+1) = \begin{cases} \mathbf{w}_j(n) + \eta(n)\left[\mathbf{x}(n) - \mathbf{w}_j(n)\right], & j \in \Lambda_{i(x)}(n) \\ \mathbf{w}_j(n) & \text{de otra forma} \end{cases} \tag{7}$$

5. Donde $\eta(n)$ es un parámetro conocido como la tasa de aprendizaje y $\Lambda_{i(x)}(n)$ es la función de aprendizaje alrededor de la neurona ganadora $i(\mathbf{x})$; tanto la tasa como la función de aprendizaje cambian dinámicamente para obtener los mejores resultados.
6. Volver al paso 2 hasta no observar cambios en el mapa resultante.

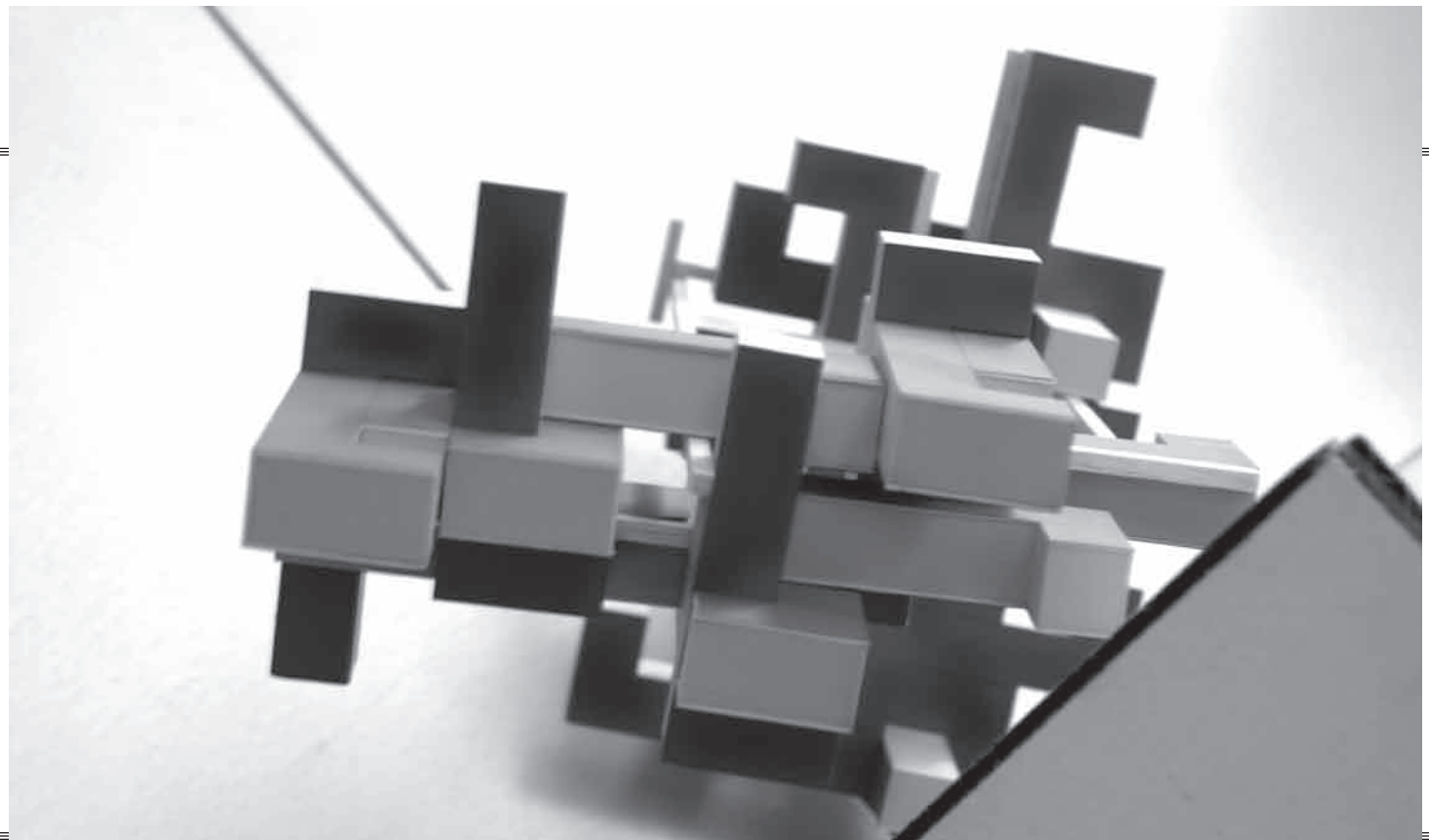
Se puede hacer un paralelo entre el algoritmo de SOFM y el algoritmo LBG. En ambos casos se busca minimizar una función criterio y posteriormente se reasignan las muestras. Desde este punto de vista el algoritmo de mapas de autoorganización es un algoritmo de cuantificación vectorial. Diversas aplicaciones han sido desarrolladas, entre ellas el método conocido como Cuantificación Vectorial de Aprendizaje (*Learning Vector Quantization*, LVQ) (Haykin, 1994).

El método de SOFM posee tres propiedades que lo caracterizan:

1. La aproximación del espacio de entrada. Un mapa de autoorganización de características $\Phi$, representado por el conjunto de vectores de peso sináptico $\{\mathbf{w}_j \mid j = 1,2,...,N\}$, en un espacio de salida A, provee una buena aproximación del espacio de entrada $X$.
2. Ordenamiento topológico. El mapa de características $\Phi$ calculado con el algoritmo SOFM está topológicamente ordenado en el sentido de la localización espacial de los nodos en la estructura, la cual corresponde a un dominio particular de las características de los patrones de entrada.
3. Correspondencia de la función de densidad de la entrada. El mapa de características $\Phi$ refleja las variaciones estadísticas de la distribución de la entrada.

## Resultados
## Simulaciones

El proceso de selección de los coeficientes wavelet con base en la varianza temporal de la secuencia, busca obtener la mínima cantidad de coeficientes que permita hacer una clasificación óptima para resintetizar secuencias creíbles. Inicialmente se tomaron los coeficientes que superaran un valor determinado por la media de su distribución, sin embargo, la cantidad de coeficientes obtenidos en estos casos puede llegar a los 36.000 para regiones de 265x256 píxeles. Por esta razón, se determinó la realización de un umbral diferente para cada nivel de la transformación, lo cual resulta en la determinación de coeficientes más adecuados para realizar la clasificación. Los umbrales para cada nivel se determinan de acuerdo a un porcentaje (90%)



del valor máximo en cada nivel y para cada transformación (horizontal, vertical, diagonal) de la imagen. Con este procedimiento se reduce el número de coeficientes en un 99%, tomando sólo aquellos que realmente son de interés para el análisis y codificación de cada cuadro del video.

Para la creación del libro de códigos se desarrolló el algoritmo de SOFM y se comparó con resultados obtenidos aplicando el algoritmo LBG minimax.

Inicialmente se hicieron simulaciones bidimensionales, para lo cual se tomaron dos variables aleatorias gausianas bidimensionales $\mathbf{X}_1$ y $\mathbf{X}_2$ con 200 y 400 realizaciones cada una. Estas se describen con su media y matriz de covarianza de la siguiente forma:

$$\text{Para } \mathbf{X}_1\text{:}\quad \mu_1 = [-3,-2] \qquad K_1 = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$$

$$\text{Para } \mathbf{X}_2\text{:}\quad \mu_2 = [4,1] \qquad K_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{8}$$

Al aplicar el algoritmo tradicional LBG y el LBG modificado, se obtienen centros mostrados en negro:

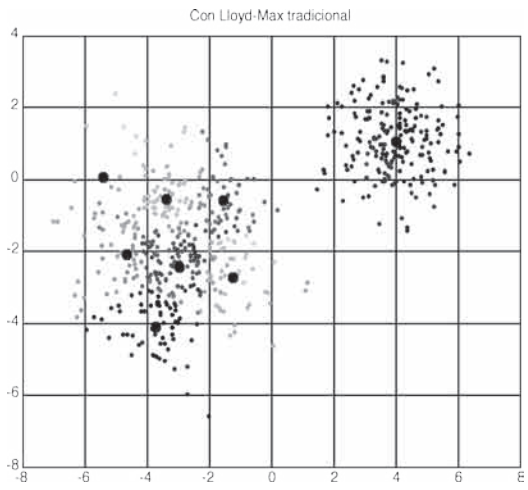



**Figura 7.** Clasificador LBG tradicional.

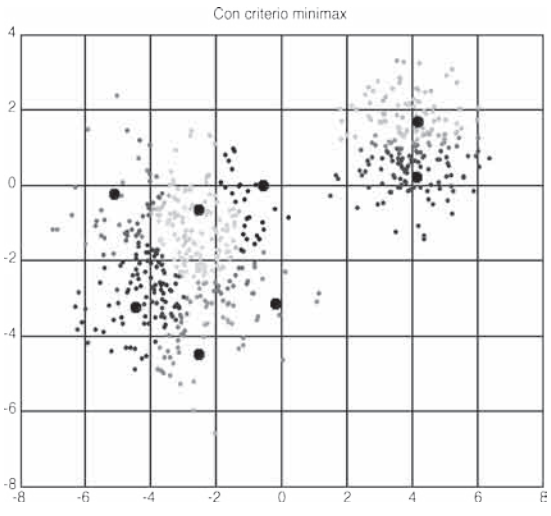



**Figura 8.** Clasificador LBG mínimas.

Al aplicar el algoritmo de SOFM con una topología rectangular de 8 nodos, se obtienen estos centros (en negro) con sus respectivas relaciones topológicas (en gris):

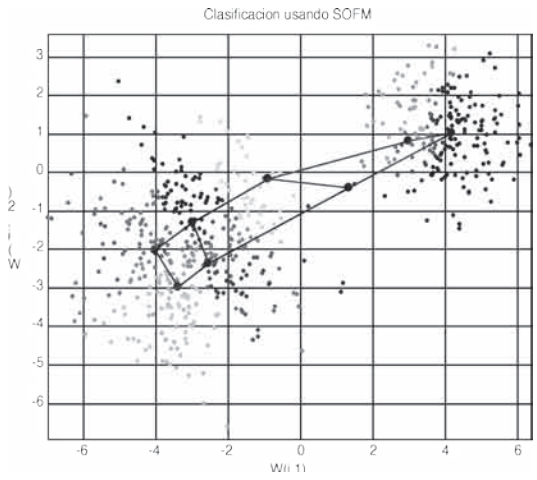



**Figura 9.** Clasificador SOFM.

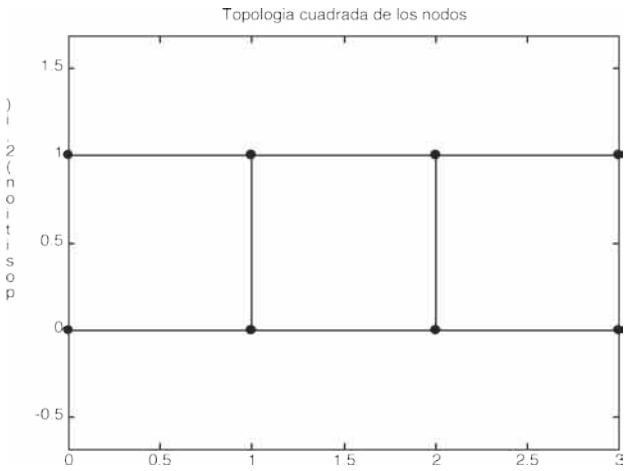



**Figura 10.** Topología cuadrada de 2x4.

En este ejemplo se demuestran claramente las ventajas del diseño del clasificador usando el criterio minimax: la figura anterior ilustra que el uso del algoritmo LBG tradicional encuentra los centros de masa como los representantes de cada clase, mientras que con el criterio modificado los centros encontrados son aquellos que minimizan el máximo error; esto representa un libro de códigos con menos imágenes repetidas (los centros no se acumulan en regiones con alta densidad de muestras).

Al hacer uso de la clasificación SOFM, ésta se basa en un criterio de minimización del error cuadrático medio (como en LBG tradicional), sin embargo, se presenta una ventaja adicional que consiste en la preservación topológica de los centros encontrados con base en una topología previamente establecida. Los centros encontrados, por lo tanto, tienden a comportarse como en el caso de LBG tradicional, restringidos por la preservación de la

continuidad que impone la estructura topológica.

Las pruebas y algoritmos desarrollados para el caso bidimensional fueron aplicados sobre diferentes videos de entrenamiento obteniendo el libro de códigos reducidos de 8 y 16 palabras código. Esta simulación pretende demostrar de forma práctica e intuitiva los beneficios y desventajas en la creación de libro de códigos con los diferentes procedimientos.


**Figura 11.** Libro de códigos obtenido con SOFM topología cuadrada 2x4.

En la figura anterior se observa la relación topológica que se da entre imágenes clasificadas y los parámetros encontrados con la transformada wavelet. La distribución de las bocas se encuentra de acuerdo a lo esperado por la topología. Por ejemplo: la imagen 1 está conectada con la imagen 2 y la imagen 5 en la topología rectangular; la boca de las imágenes 2 y 5 son transiciones suaves de la boca en la imagen 1, etc.

## Pruebas y mediciones

Las evaluaciones subjetivas realizadas están inspiradas en los modelos MOS (Mean Opinion Score) y DMOS (Degradation Mean Opinion Score) (Tobias, 1999). En este trabajo se desarrolló una encuesta que consta de 7 preguntas. Las preguntas invitan al encuestado a dar su concepto personal sobre la "naturalidad" de los videos que se le presentan. En todos los casos se presentan comparaciones entre 2 y hasta 3 videos; la persona debe elegir el que a criterio personal sea "más natural". Una aclaración previa al comienzo de la prueba consistió en limitar el enfoque de naturalidad, como una secuencia que cambia suavemente y no presenta saltos abruptos. A pesar de que los videos disponen de audio para facilitar su comprensión, se explicó en un comienzo que el objetivo de la encuesta no era evaluar la calidad de la señal sonora.

La población que contestó la evaluación corresponde a un amplio intervalo de edades y diferentes sexos; en total se realizaron 30 pruebas. Los resultados obtenidos en dichas pruebas cualitativas se encuentran respaldados por cantidades cuantitativas, como son el error cuadrático medio (MSE) y la correlación entre trayectorias seguidas sobre puntos predefinidos.

Una de las preguntas de la encuesta buscaba determinar cómo influye el número de niveles de la transformada en el proceso de resíntesis. Los resultados indican una mayoría que selecciona entre los videos resintetizados con 2 y 4 niveles. El número de niveles de la transformada influye directamente en la forma como se representa la imagen. Es decir, la aproximación de una señal, dada por la transformada wavelet, es del tamaño de la imagen original decimada por un factor de 2N, donde N representa el número de niveles. Esto significa que con niveles mayores es posible representar una aproximación (pasabajos) de la imagen con pocos coeficientes. Sin embargo, muy pocos coeficientes no logran una representación óptima. Estos resultados permitieron plantear un método de selección de los coeficientes que tenga en cuenta tanto los de aproximación como los de detalles. En este caso sí se registró un decremento del MSE a medida que el nivel de la transformada aumenta, como podría preverse inicialmente.

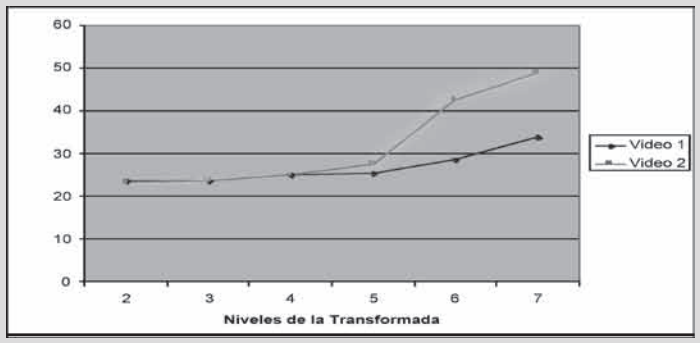
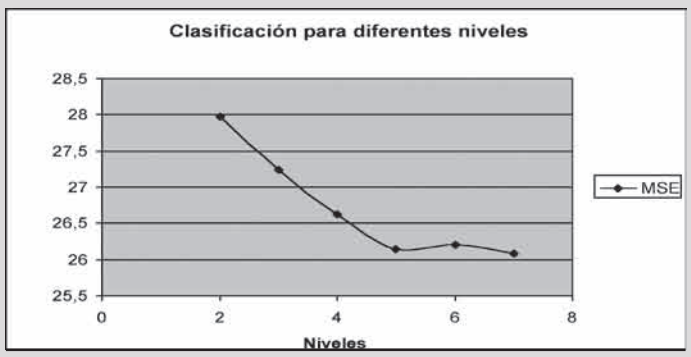



**Figura 12.** Medición MSE con diferentes niveles y manteniendo todos los coeficientes de la aproximación.




**Fifura 13.** Medición MSE con diferentes niveles y selección de coeficientes igual.

Otro interesante resultado muestra la preferencia de los encuestados por los videos resintetizados con los parámetros wavelet de una región de 256x256 píxeles, con aproximadamente 1.000 coeficientes, por encima de los videos resintetizados con la DCT en una región de 100x60 píxeles con 60.000 coeficientes.

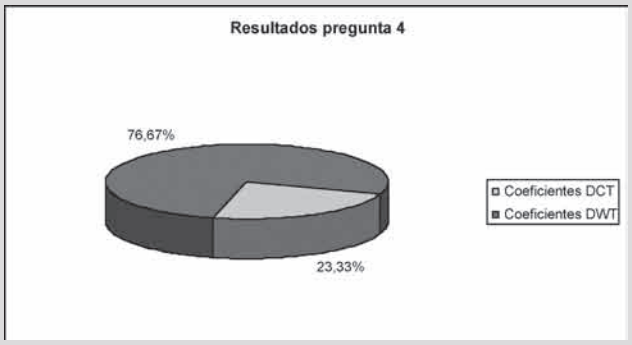



**Figura 14.** Preferencias de videos resintetizados con DCT y con DWT.

Este resultado se sustenta por las mediciones de los coeficientes de correlación de las trayectorias seguidas en dos puntos diferentes, correspondientes a regiones de la boca (punto 1) y de la mejilla (punto 3). Se puede ver en las tablas 1 y 2 que el coeficiente de correlación con respecto a la trayectoria original del punto 1, es un poco mayor en el caso de la DCT; sin embargo, en el punto 3 el coeficiente de correlación de la trayectoria es mayor al tener los parámetros DWT. Al tenerse en cuenta parámetros en una región mayor a la de solamente la boca, el clasificador puede representar mejor estas regiones otorgando más naturalidad al video.

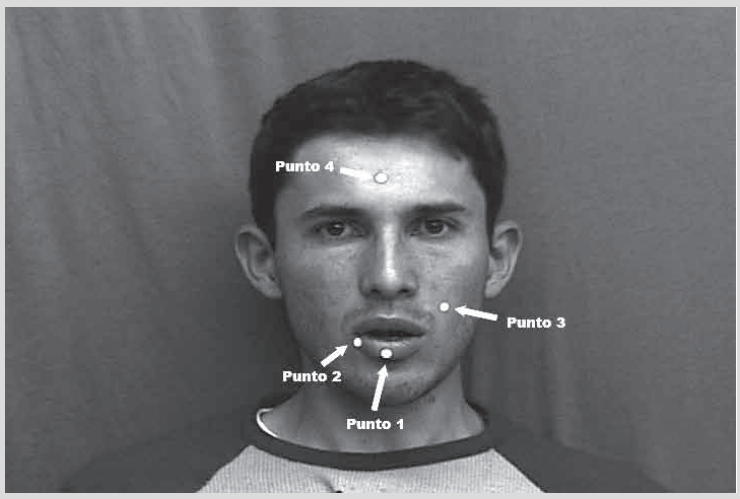



**Figura 15.** Definición de trayectorias de puntos a seguir.

| Transformada | Coeficiente de correlación |
|---|---|
| DCT | 0,9940 |
| DB 9/7 | 0,9573 |

**Tabla 1.** Coeficientes de correlación punto 1 (labio).

| Transformada | Coeficiente de correlación |
|---|---|
| DCT | 0,7846 |
| DB 9/7 | 0,8372 |

**Tabla 2.** Coeficientes de correlación punto 3 (mejilla).

Finalmente, dos de las preguntas de la encuesta se diseñaron con el fin de determinar cómo influye la forma y la dimensionalidad de la topología en las secuencias generadas. Los resultados de las encuestas muestran que se presenta una diferencia estadísticamente significativa entre los resultados con un nivel de significancia de 5%. Sin embargo, desde el punto de vista subjetivo del autor los videos no presentan mayores diferencias. En este caso, se plantea la hipótesis de que el algoritmo de SOFM converge a mínimos locales y no al cuantificador óptimo. Este planteamiento se sustenta con las gráficas de medición del error MSE (Gráfico 15). Se observa que la diferencia en el error es muy pequeña para diferentes formas de la topología y diferentes dimensionalidades. La idea se basa en que el proceso de cuantificación demuestra que la superficie cercana al punto del mínimo global es una superficie bastante rugosa. El clasificador se encarga de encontrar puntos cercanos a este mínimo global, garantizando mantener la estructura topológica.

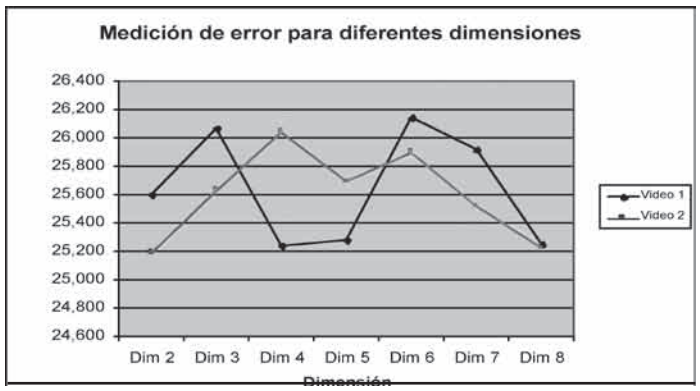



**Gráfico 15.** Medición de MSE para clasificación SOFM con topologías de diferentes dimensionalidades.

## Conclusiones

El uso de una transformada multiresolución permite obtener

parámetros para la generación de un libro de códigos que reproduce secuencias de video naturales. Dentro de las funciones evaluadas, es posible el uso de funciones wavelet como la Haar, Daubechies 2 o una función wavelet biortogonal como la Daubechies 7/9. Esta última otorga una ventaja adicional: la obtención de coeficientes que permitan una adecuada compresión y reconstrucción del libro de códigos.

La selección de coeficientes basados en la varianza temporal permite obtener parámetros para una clasificación eficiente. Sin embargo, el criterio de selección debe estar asociado a todos los coeficientes de igual forma, tanto de aproximación como de detalles, generando así un clasificador que mejora con el número de niveles de la transformación. Por otra parte, el uso de la varianza temporal en la selección de los coeficientes permite plantear una primera aproximación al uso de un libro de códigos, basada en unos parámetros dinámicos, es decir, un clasificador que se adapte a los coeficientes encontrados con los cambios en el tiempo.

La selección de una región de interés más amplia a solamente la boca reduce los saltos presentados en los videos resintetizados en regiones como las mejillas, aumentando de esta forma la naturalidad de la secuencia obtenida.

Es posible obtener un libro de códigos ordenado topológicamente de acuerdo a una estructura con una dimensión determinada y previamente establecida. Las imágenes que componen este libro de códigos, por lo tanto, presentan relaciones que pueden ser explotadas para generar secuencias de video suaves.

El ordenamiento topológico, característica presente en el libro de códigos, permite plantear alternativas en la interpolación de imágenes para la reconstrucción e, igualmente, buscar métodos de compresión de la base de datos aprovechando este ordenamiento.

Los resultados encontrados utilizando el método de cuantificación con el algoritmo LBG modificado con criterio minimax, validan los resultados hallados en otras investigaciones, extendiendo su uso a otro tipo de imágenes parametrizadas con una transformada multiresolución.

**Referencias bibliográficas**


Antonini, Marc *et ál*. "Image coding using wavelet transform". *Transactions on Image Processing.* Vol. 1, No. 2. (Abril, 1992), pp. 205-220.

Baptiste, Carlos, Mauricio Sotomayor y Pedro Vizcaya. *Segmentación y parametrización automática de rostros en video*. Bogotá: Pontificia Universidad Javeriana, 2002.

Bárcenas, Edson *et ál*. "Análisis y síntesis de voz visual en el idioma español". Trabajo de grado (Ingeniería Electrónica). Bogotá: Pontifica Universidad Javeriana, 2001.

Essa, Irfan Aziz. "Analysis, Interpretation, and Synthesis of Facial Expressions". PhD thesis (Arts and Sciences). Massachusetts Institute of Technology, Department of Media Arts and Sciences, 1995.

Gray, Robert. "Vector Quantization". *IEEE ASSP Magazine.* (Abril, 1984), pp. 4-29.

González, Rafael y Richard Woods. "Digital Image Processing". Second edition. *Prentice Hall*. New Jersey. (2003), pp. 349-402.

Haykin, Simon, *Neural Networks. A Comprehensive Foundation*. New Jersey: Englewwood Cliffs, 1994.

Internacional Telecommunication Union, ITU. "Visual Telephone Systems". *ITU-T Recommendation H.100*, 1998.

Internacional Telecommunication Union, ITU. "Narrow-band visual telephone systems and terminal equipment". *ITU-T Recommendation H.320*, 1999.

Kohonen, Teuvo. "Self-Organization and Associative Memory". *Springer Series in Information Sciences*. Helsinki, 1984.

Linde, Y., A. Buzo y Robert M. Gray. "An algorithm for vector quantizer design". *IEEE Transactions on Communications*. Vol. 28 (1980), pp. 84-95.

Lloyd, S. P. "Least squares quantization in PCM". *IEEE Transactions on Information Theory*. Vol. 28. (March, 1982), pp. 129-137.

Machado, Juan Felipe y Diego Santa. "Síntesis paramétrica de voz visual". Trabajo de grado (Ingeniería Eléctrica). Bogotá: Pontificia Universidad Javeriana, 2001.

Mak, M.W. y W.G. Allen. "Lip-motion analysis for speech segmentation in noise". *Speech Communications*. Vol. 14, No. 3. (Jun., 1994), pp. 279-296.

Nasrabadi, N. M. "Use of vector quantizers in image coding". *Proc. IEEE Int. Conf. Acoustic, Speech, Signal Processing.* (Marzo, 1985), pp. 125-128.

Soto, Carolina y Pedro Vizcaya., *Generador de corpus para síntesis de voz visual*. Bogotá: Pontificia Universidad Javeriana, 2004.

Skodras, Athanassios, Charilaos Christopoulos y T. Ebrahimi. "The JPEG200 Still Image Compression Standard". *IEEE Signal Processing Magazine.* (September, 2001), pp. 36-58.

Strang, Gilbert y Truong Nguyen. *Wavelets and Filter Banks*. Massachusetts: Wellesley, 1997.

Tobias, Ö. *How to Meet the User Requirement Of Room-Based Videoconferencing*. Estocolmo: Royal Institute of Technology, 1999.

Vizcaya, Pedro *et ál.* "Codificación y Decodificación de Secuencias de Telefonía Visual". Memorias del IX Simposio de Tratamiento de Señales, Imágenes y Visión Artificial. Manizales (Colombia), Septiembre de 2004.


**Notas**

[1] Fotorrealista en el sentido de tener transiciones suaves entre las diferentes regiones que limitan los elementos presentes en la imagen.