

## TÉCNICA DE CODIFICACIÓN PIRAMIDAL UTILIZANDO LA CUANTIZACIÓN VECTORIAL PARA LA TRANSMISIÓN PROGRESIVA DE IMÁGENES

Pedro Freddy Huamaní Navarrete

### Resumen

En este artículo se propone una técnica para transmisión progresiva de imágenes. Esta técnica está basada en el modelo piramidal de Burt y Adelson, haciendo uso de la codificación por Cuantización Vectorial en imágenes representadas por una estructura piramidal, denominada pirámide de imágenes diferencia, que son exactamente las imágenes a transmitirse progresivamente del nivel más alto al nivel más bajo, para ir formando las denominadas imágenes media. La técnica es comparada a otros métodos conocidos en la literatura en términos de relación señal-ruido de pico, por su calidad subjetiva y por el esfuerzo computacional asociado.

**Palabras clave:** Cuantización Vectorial, Técnica Piramidal, codificación de imágenes.

### Abstract

In this paper is proposed a technique for progressive images transmission. This technique is based on the pyramidal model of Burt and Adelson, making use of the codification by Vectorial Quantization in images represented by a pyramidal structure, designated images pyramid difference, that they are exactly the images that they will be transmitted progressively of the highest level at bottom most level, to go forming the designated mean images. The technique is compared to other methods known in the literature in terms of relationship sign-peak noise, for your subjective quality and for the effort computational associate.

**Keywords:** Vectorial Quantization, pyramid technique, Image coding.

## **INTRODUCCIÓN**

Los avances de la tecnología digital en la década pasada, principalmente en la transmisión y almacenamiento de datos, ha traído muchas aplicaciones al procesamiento digital de imágenes. Así tenemos: transmisión de imágenes en redes de distribución de TV, monitoreo remoto vía satélite, comunicaciones militares vía radar o sonar, teleconferencia, transmisión de fax, almacenamiento de documentos educacionales, de negocios, de imágenes médicas usadas en sistemas de monitoreo de pacientes, mapas climáticos, etc. (Jain, 1989).

En todas estas aplicaciones, el principal obstáculo radica en la enorme cantidad de datos requeridos para representar digitalmente la imagen. Por lo tanto, el uso frecuente de imágenes digitales implica altos costos de transmisión y/o almacenamiento, sumándose a esto las limitaciones de memoria, canal y velocidad. Es así que se torna muy necesario el uso de técnicas de compresión para la reducción de la tasa de bits, minimizando así el número de bits requeridos para su representación.

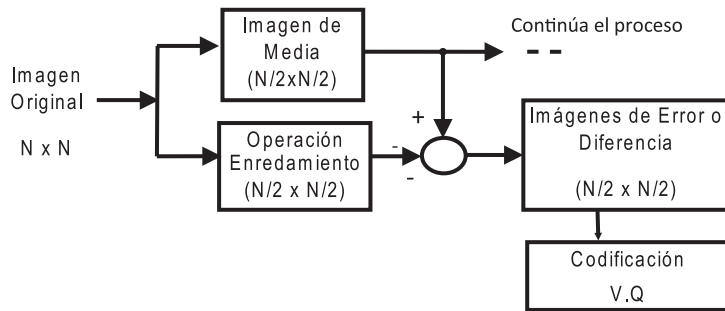
El enfoque principal de este trabajo es transmitir progresivamente imágenes comprimidas. Para esto, una imagen de menor resolución es trans-

mitida, para luego mejorar progresivamente esta imagen con transmisiones complementarias. De esta manera, la transmisión progresiva de imágenes facilita al usuario ver la imagen y tomar una decisión sobre si es necesario o no toda su definición. Todo esto, en un corto tiempo. Este trabajo utiliza una Técnica de Codificación Piramidal (Burt, & Adelson, 1983), conformado por pirámides de medias y diferencias. La pirámide de medias es formada por sucesivas medias sobre bloques de 2x2 pixels iniciando de la imagen original, y la pirámide de diferencia se obtiene con la diferencia de cada imagen media menos una imagen enredada, previamente formada.

Finalmente, la Cuantización Vectorial (V.Q) es la utilizada para comprimir o codificar las imágenes diferencias de cada nivel de la pirámide, iniciándose en la imagen de media de menor dimensión y finalizando en la de dimensión total.

## **DESARROLLO DEL TRABAJO**

Esta técnica utilizada representa una imagen en diferentes dimensiones y formando una pirámide de imágenes de dimensiones cada vez menores. La Figura 1 muestra el diagrama de bloques del análisis de esta técnica.



**Figura 1. Diagrama de Bloques del Análisis**

### PIRÁMIDE DE IMÁGENES MEDIAS

Esta pirámide se forma de una media aritmética sobre cada uno de los bloques de  $2 \times 2$  pixels no superpuestos, de una secuencia de imágenes de resoluciones cada vez menores, donde cada nueva imagen ocupará siempre un cuarto de la dimensión de la imagen de la cual fue obtenida.

### PIRÁMIDE DE IMÁGENES ENREDADAS

La operación de enredamiento es aplicada para cada nivel de la pirámide de imágenes media, partiendo de la imagen original (considerada como “la primera imagen media”). Esta operación consiste en hacer una decimación de una imagen media, sin eliminar ninguna información de línea y columna. La decimación es primero realizada en las líneas y después en las columnas, de forma a reducir a un cuarto la dimensión de la imagen decimada. Como la operación es realizada en cada nivel de la pirámide de imágenes media, se tendrá como consecuen-

cia la creación de una nueva pirámide denominada de enredamiento, la cual contribuirá al cálculo de la pirámide de imágenes diferencia.

### PIRÁMIDE DE IMÁGENES DIFERENCIA

Se consigue con la diferencia entre la pirámide de imágenes media con la pirámide de imágenes enredadas. Por lo tanto, esta diferencia recibe el nombre de Pirámide de Imágenes de Errores, la cual será codificada más adelante.

Los resultados por cada nivel de esta pirámide corresponden a cuatro imágenes diferencia con la apariencia de haber sido filtradas por un filtro pasa-alto (Jain, 1989). Estas cuatro imágenes también presentan una propiedad de dependencia lineal; por tanto, solo se necesita codificar tres sub-imágenes diferencia, porque la cuarta surge de la combinación de las otras tres. De esta manera se mejora la compresión en relación a otros métodos (Garrido, 1989)

## RECONSTRUCCIÓN DE LA PIRÁMI- DE DE IMÁGENES MEDIA

Se realiza a partir de la diferencia entre las imágenes diferencia del último nivel de la pirámide con la imagen de media de menor dimensión. Seguidamente, se realiza la operación de

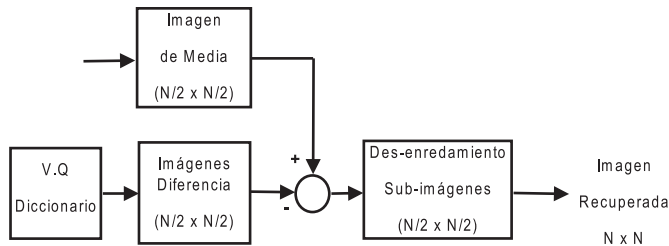


Figura 2. Diagrama de Bloques de la Síntesis

## CODIFICACIÓN POR CUANTIZA- CIÓN VECTORIAL EN LA ESTRUC- TURA PIRAMIDAL

Esta sección trata del método de codificación por Cuantización Vectorial (V.Q), aplicado a la pirámide de imágenes diferencia. Así, sea una imagen dividida en bloques de  $p \times p$  pixels, un V.Q consiste en encontrar un conjunto fijo de vectores (denominado Diccionario de Códigos) que representen todos los bloques de esta imagen. El bloque será codificado escogiendo el vector del diccionario que mejor se aproxime, teniendo como base de medida la Distancia Euclidiana, la cual es mostrada en la siguiente ecuación [11],

$$d(x, y) = \sum_{i=0}^{i=p*p-1} \left| X_i - Y_i \right|^2 \quad (01)$$

des-enredamiento de imagen desde el último nivel hacia el primer nivel, reconstruyendo finalmente la imagen de tamaño original.

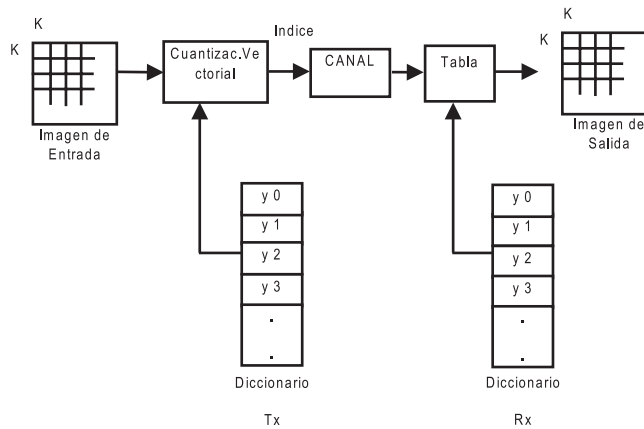
La Figura 2 muestra el diagrama de bloques de la síntesis o reconstrucción de esta técnica.

Donde:

X indica el vector de entrada

Y el vector-código perteneciente al diccionario.

De esta manera, la V.Q es aplicada a cada nivel de la pirámide, realizándose así una codificación basada en tres diccionarios diferentes por nivel, cada uno de estos asociado a cada imagen diferencia de determinado nivel. Solo las imágenes diferencia son codificadas por poseer gran proporción de pixels con niveles de gris próximos a cero, correspondientes a las regiones de baja frecuencia de la imagen original (González, Woods & Eddins, 2004). En la Figura 3, se muestra el diagrama de bloques básico de un Cuantizador Vectorial.



**Figura 3. Esquema Básico de la Codificación Vectorial**

### PROYECTO DEL DICCIONARIO

Existen diversos algoritmos iterativos que, dado el criterio de distorsión, otorgan un diccionario óptimo. Como datos de entrada, estos algoritmos necesitan el tamaño del diccionario ( $N$  niveles) y el tamaño del vector ( $pxp$ ). Otro dato importante es el diccionario inicial que puede, por ejemplo, ser generado con bloques aleatoriamente escogidos de un grupo de imágenes diferentes de la que se desea codificar (Gersho & Gray, 1992). Entre algunos algoritmos iterativos generalmente utilizados, encontramos: el algoritmo LBG y el llamado “Nearest Neighbor” o vecino más próximo (González, & Woods, 1996), que es significativamente más rápido que el primero pero en consecuencia presenta mayor degradación (Garrido, 1989).

Para la formación de los diccionarios, se escogió el algoritmo LBG (González, & Woods, 1996), que toma como datos de entrada:

1. La secuencia de entrenamiento formada, en este caso, por cuatro imágenes diferentes de aquellas que serán codificadas.
2. El tamaño del diccionario.
3. El diccionario inicial generado a partir de bloques aleatorios dentro del margen dinámico de la imagen.
4. El umbral de distorsión, a ser definido.

Los vectores de la secuencia de entrenamiento son mapeados en los vectores-código, según el criterio del menor error cuadrático. Si el error cuadrático medio de todos los vectores fuera menor que el umbral de distorsión, se llega al diccionario final. Caso contrario, se sustituye cada vector-código por el centroide de los vectores que en él se mapearán.

### CUANTIZACIÓN VECTORIAL (VQ) EN LAS IMÁGENES DIFERENCIA

La codificación por VQ es espacial y realizada directamente sobre cada

una de las imágenes diferencia, codificando del nivel más alto (imagen de menor dimensión) al nivel más bajo.

### a) Codificación del Tercer Nivel

La última imagen de media, al igual que las otras, tienen las mismas características de la imagen original. Por tanto, cada uno de sus pixels es un número real positivo que se encuentra dentro de la franja dinámica de imágenes monocromáticas con 256 niveles de gris (0 hasta 255).

En este nivel, son dos los tipos de codificación a utilizar. La primera hace uso de la VQ para el grupo de imágenes diferencia, dividiendo en bloques de 2x2 pixels cada imagen; y, la segunda, utiliza una codificación pixel a pixel.

Durante la codificación por VQ surge un problema –como es de esperar, aparecen errores, los cuales se propagan conforme se avanza en dimensión durante la reconstrucción–. Por lo tanto, estos errores eventualmente pueden replicarse creando un efecto de bloques en la imagen original. Para minimizar este problema, se optó por codificar estas imágenes diferencia con un Cuantizador Vectorial de dos Niveles Jerárquicos, donde el primer nivel codifica directamente las imágenes diferencia y el segundo codifica los errores obtenidos en la primera codificación.

### Cuantizador Vectorial de Dos Etapas

Este modelo codifica la primera etapa con un cuantizador vectorial directo  $q$  ( $\cdot$ ), y un diccionario de códigos “A”, donde el vector de entrada “x” es

cuantizado. Luego, un segundo cuantizador vectorial  $q'$  ( $\cdot$ ), con otro diccionario de códigos “A’”, opera sobre el vector diferencia entre el vector original y la salida cuantizada de la primera etapa (Gersho & Gray, 1992).

Realizada la codificación, se observa la presencia de tenuidad así como información de bordes en cada imagen, lo que obliga a la división del diccionario en dos grupos. Para su construcción, se realizaron entrenamientos de bloques de 2x2 pixel con imágenes tenues y de bordes, tomándose en cuenta la separación de los bloques pertenecientes a la tenuidad (denominados inactivos) y los bloques pertenecientes a los bordes (denominados activos), tal como se muestra a continuación:

Si: Imagen  $\in \pm$  Umbral  $\rightarrow$  Bloque Inactivo

Caso Contrario  $\rightarrow$  Bloque Activo

Donde:

Umbral, corresponde a un valor diferente para cada nivel de la pirámide de imágenes diferencia. La selección de este valor está basada en previa experiencia.

### b) Codificación del Segundo Nivel

En este nivel, la codificación también se lleva a cabo solo para tres imágenes diferencias. Se utilizan palabras-códigos con 4 elementos cada una, ya que estas imágenes presentan muchos detalles que pueden no ser tomados en cuenta si las codificamos con vectores de más elementos. Al igual que en el nivel anterior, se

observó enorme cantidad de regiones tenues y con bordes; por esa misma razón, se optó por una nueva partición del diccionario en dos, para codificar independientemente la parte tenue y de borde, y teniendo en cuenta la expresión de bloques activos e inactivos.

**c) Codificación del Primer Nivel**

Para este nivel, se aumenta el número de elementos para cada vector-código, caso contrario la compactación de datos quedaría reducida. Por tanto, son tomados 16 elementos para cada vector. Nuevamente es utilizada la técnica de división del diccionario para la parte tenue y de bordes, por la enorme cantidad de los mismos.

La Tabla 1 muestra el tamaño del diccionario para cada nivel, así como el número de pixels para cada palabra-código.

Niveles	Bloques	Diccionarios
		Tenue y Borde
3er Nivel	2x2	14 242
2do Nivel	2x2	18 238
1er Nivel	4x4	24 488

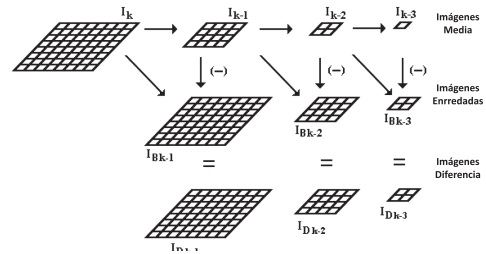
**Tabla 1. Tamaño de Diccionarios para la VQ**

**Reconstrucción de la Imagen**

Iniciando en la imagen media del nivel más alto de la pirámide, se selecciona un vector para realizar una búsqueda en el diccionario de códigos de la primera etapa, de tal manera a localizar al que mejor lo represente. Seguidamente, es calculado el vector de error. Una vez codificado, se transmite

el par de índices de las dos etapas, donde la tarea del decodificador es generar la composición de los dos vectores códigos correspondientes a estos índices de cada diccionario de cada etapa.

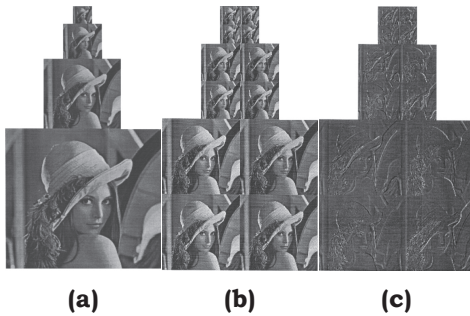
La reconstrucción se realiza progresivamente, reconstruyendo primero las imágenes de media por cada nivel de la pirámide. La Figura 4 muestra una representación de un ejemplo de descomposición piramidal.



**Figura 4. Representación de un ejemplo de descomposición piramidal.**

**PROCESO DE SIMULACIÓN**

Para ejecutar el proceso de simulación, se utilizó el entorno del Software Matlab, con imágenes en tonos de gris y típicas en investigaciones que se relacionan con técnicas de compresión o codificación. A continuación, en la Figura 5a, se muestra el resultado de obtener la pirámide de imágenes media correspondiente a las resoluciones de 32x32, 64x64, 128x128 y 256x256 pixels. De la misma forma, la Figura 5b muestra el resultado del conjunto de imágenes enredadas con las resoluciones (64x64, 128x128 y 256x256), y la Figura 5c, la perteneciente a las imágenes diferencia con las resoluciones (64x64, 128x128 y 256x256).



**Figura 5. Imágenes LENA media, enredada y diferencia.**

### RESULTADOS

En esta sección, se muestran los resultados visuales y numéricos de la codificación por Cuantización Vectorial en la pirámide de imágenes diferencia. Para esta operación, se utilizaron dos grupos de imágenes de 256x256 pixels con 8 bpp. Un primer grupo de entrenamiento para sintetizar los diccionarios teniendo en cuenta la tenuidad y bordes existente en cada imagen de cada nivel de la pirámide; y un segundo grupo para realizar las pruebas y, por tanto, ajenas al grupo de entrenamiento.

Para medir la calidad de la imagen reconstruida (análisis objetivo), se

utilizó la relación señal-ruido de pico, PSNR, definida por (Garrido, 1989):

$$PSNR = 10 * \log_{10} \left[ \frac{255^2}{\frac{1}{MN} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} (x_{ij} - \hat{x}_{ij})^2} \right] \text{ dB} \quad (02)$$

Donde:

255 el valor pico de la señal.

X: la imagen original.

$\hat{X}$  : la imagen reconstruida.

Ambas imágenes con tamaño de M x N pixels.

De esta forma, la simulación fue realizada en el grupo de imágenes ajenas al entrenamiento, mostrándose los resultados numéricos para cada imagen media. Ver Tabla 2. De igual forma, los resultados visuales, pertenecientes a la imagen LENA, son mostrados en la Figura 6.

Imágenes Media	PSNR (dB)	Bpp
LENA <sub>32 x 32</sub>	330.2	7.0
LENA <sub>64 x 64</sub>	41.3	3.7
LENA <sub>128 x 128</sub>	35.4	2.4
LENA <sub>original</sub>	31.5	1.0

**Tabla 2. Resultados de Simulación**





**Imagen Original**



**Imagen Recuperada**

**Figura 6. Imagen LENA codificada**

### CONCLUSIONES

Muchas técnicas requieren operaciones de transformaciones al dominio de la frecuencia, por lo cual utilizan mayor tiempo computacional y, por consiguiente, se tornan más lentos que el método propuesto en este artículo.

La construcción de la pirámide de medias requiere poco esfuerzo computacional, basado en simples cálculos de medias aritméticas sobre bloques de  $2 \times 2$  pixels no sobrepuestos. Los cálculos son ejecutados tres veces consecutivas, originando de esta manera una pirámide media con cuatro niveles. La última imagen media con dimensión más pequeña, no necesitará ser más sub-muestreada. De este modo, se cuantiza y codifica directamente pixel a pixel para posteriormente ser transmitida.

La mayoría de las imágenes naturales tienen alto grado de correlación entre pixels vecinos. Por tanto, la mayoría de los pixels en las imágenes diferencia tienden a concentrarse en valores cerca de cero (próximos del color negro), o sea con menor energía (Gornsztejn, 1993). Esto hace que el conjunto de imágenes diferencia se asemejen al conjunto de imágenes obtenidas en una codificación por sub-bandas (Woods & O'neil, 1986).

El grupo de imágenes diferencia representará las transiciones de la imagen, a pesar de poseer poca energía; por lo tanto, son muy importantes porque determinan la mayor o menor definición en la imagen reconstruida.

Las operaciones de medias se asemejan al tipo de filtrado de media no lineal, localizada y espacial, porque cada bloque de  $2 \times 2$  pixels es tratado independientemente sin ser sobre-

puestos a los bloques vecinos. De esta manera, solo se codifican tres imágenes diferencia por cada nivel de la pirámide, porque estas poseen la propiedad de dependencia lineal.

Sin embargo, la codificación con un único diccionario para las tres imágenes diferencia, por cada nivel de la pirámide, trae una degradación en la reconstrucción de la imagen, ya que los errores se propagan a través de la pirámide.

Por lo tanto, el uso de tres diccionarios diferentes por cada nivel de la estructura piramidal aumenta la tasa de compresión, pero mejora notablemente el aspecto visual y numérico de la imagen reconstruida. Así mismo, se optó por la división en dos de los diccionarios; uno para codificar la parte tenue y otro para la parte de bordes.

De esta manera, se pudo representar la enorme cantidad de vectores tenues en cada imagen diferencia, lo cual fue posible al seleccionar el tamaño de las particiones para tenuidad y bordes a través de la técnica de ensayo y error.

### **Propuestas Futuras**

Organización de los diccionarios para la operación de búsqueda del vector más próximo. Esto quiere decir que se podrá establecer, una búsqueda por energía, dividiendo el diccionario en particiones.

Optimización del tamaño de las particiones en los diccionarios de tenuidad y bordes. Es decir, encontrar una cantidad óptima de vectores, de tal forma a representar adecuadamente las informaciones de tenuidad y bordes en la codificación.

## REFERENCIAS BIBLIOGRÁFICAS

- 1 González, R.; Woods, R. & Eddins, S. (2004). "Digital Image processing using MATLAB". Editorial Dorling Kindersley.
2. González, R. C. & Woods R. E. (1996). "Tratamiento Digital de Imágenes". Addison-Wesley Iberoamericana S.A, U.S.A.
3. Garrido, D. P. (1990). "Laplacian Pyramid Coding Using Scalar and Vector Quantization". Rio de Janeiro ICASSP-90 Technical Program Committee.
- 4 Burt, P.J. & Adelson, E. H. (1983). "The Laplacian Pyramid as a Compact Image Code," Transactions on Communications, vol. 31, pp. 532-540.
- 5 Woods, J.W. & O'neil, S. D., (1986). "Subband Coding of Images," Transactions on Communications, vol. 34, pp. 1278-1288.
- 6 Linde, Y.; Buzo, A. & Gray, R. M. (1980). "An Algorithm for Vector Quantizer Design," Transactions on Communications, vol. 28, pp. 84-95, Janeiro.
- 7 Garrido, D. P. (1989). "Um Sistema de Codificação de Imagem via Dois Canais," Dissertação de Mestrado, Departamento de Engenharia Elétrica - PUC-Rio.
- 8 Smith, M. J. T. & Eddins, S. L. (1990). "Analysis/Synthesis Techniques for Subband Image Coding," IEEE Transactions on Acoustics, Speech and Signal Processing, vol. 38, no 8, pp. 1446-1456.
- 9 Jain, A. K. (1989). "Fundamentals of Digital Image Processing," Englewood Cliffs, New Jersey: Prentice Hall.
- 10 Gornsztejn, J. (1993). "Técnicas de Compressão de Imagens por Sub-Bandas" Dissertação de Mestrado, Departamento de Engenharia Elétrica - PUC-Rio.
- 11 Gersho, A. & Gray, R. M. (1992). "Vector Quantization and Signal Compression", Kluwer Academic Publishers.