UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

FACULTAD DE INGENIERÍA

DIVISIÓN DE ESTUDIOS DE POSGRADO

TESIS

"CODIFICACIÓN DE IMÁGENES DIGITALES UTILIZANDO CUANTIFICACIÓN VECTORIAL CLASIFICADA EN EL DOMINIO DE LA DCT-VECTORIAL"

PRESENTADA POR:

EDUARDO ERNESTO CORDÓÑEZ LÓPEZ

PARA OBTENER EL GRADO DE:

MAESTRO EN INGENIERÍA ELÉCTRICA
(COMUNICACIONES)

DIRIGIDA POR:

Dr. FRANCISCO GARCÍA UGALDE

Cd. Universitaria, Junio de 1996

TESIS CON
FALLA DE ORIGEN

A mis padres

Contenido

Capítulo 1

Introducción

La palabra *señal* usualmente se refiere a una onda continua en tiempo y amplitud. Una señal puede verse de una forma más general, como una función del tiempo, el cual puede ser discreto o continuo, y donde los valores de la amplitud pueden ser continuos o discretos, escalares o vectoriales. Algunas veces una señal se refiere a una imagen, en la cual la amplitud depende de dos coordenadas espaciales en lugar de una variable de tiempo; o puede referirse a una imagen en movimiento en donde la amplitud es una función de dos variables espaciales y una temporal.

La palabra *datos* es algunas veces utilizada como un sinónimo para *señal*, pero mas frecuentemente se refiere a una secuencia de números o vectores. Así, la palabra *datos* puede interpretarse como una señal de tiempo discreto.

La palabra *fuente* significa un mecanismo o dispositivo el cual produce señales.

Frecuentemente las señales deben ser transportadas sobre un canal de comunicaciones digital, o deben ser almacenadas en un medio digital. Sin embargo, puede suceder que la señal original no es apropiada para la transmisión o almacenamiento en un medio dado. En este caso la señal digital debe ser transformada o codificada en una forma adecuada para el canal. Esta operación general, en el caso en que no existe distorsión, es conocida como *codificación de señal* y *compresión de señal* cuando la tasa de transmisión se reduce a expensas de una cierta degradación de la señal original.

Una regla o mapeo que especifica como los símbolos de la fuente, o grupos de ellos, son transformados en un nuevo conjunto de símbolos, es llamado *sistema de*

codificación. La palabra *código* es utilizada comúnmente en la teoría de la información, para definir un diccionario, el cual lista los pares de símbolos de entrada y salida de un sistema de codificación.

En la teoría de la información y de las comunicaciones, la codificación de la señal y la compresión de señal juntas forman lo que se conoce como *codificación de fuente*, una terminología debida originalmente a Shannon en su desarrollo clásico de la teoría de la información, la teoría matemática de las comunicaciones. Shannon originó la distinción entre la *codificación de fuente*, la representación eficiente de una señal, y la *codificación de canal*, el control de errores en canales con perturbaciones.

La codificación de fuente trata con la tarea de formar representaciones eficientes de las fuentes de información. Para fuentes discretas, la habilidad de formar descripciones de tasas de datos reducidas está relacionada con el contenido de información y la correlación estadística entre los símbolos de la fuente. Para fuentes analógicas, la habilidad de formar descripciones de tasas de datos reducidas, sujeta a un criterio de fidelidad, está relacionada con la distribución de amplitud y la correlación temporal de la forma de onda de la fuente. El objetivo de la codificación de fuente es, ya sea mejorar la calidad de la descripción para una tasa de datos dada, o reducir la tasa de datos para una calidad de la descripción, dada. En la construcción de un sistema real para codificación de fuente, este objetivo se puede plantear en base a las siguientes condiciones

- El desempeño del sistema debe ser tan bueno como sea posible
- El sistema debe ser tan simple como sea posible.

La primera condición siempre es la más fácil de manejar matemáticamente, y ha ganado una importancia relativa sobre la segunda debido a que la tecnología moderna ha hecho práctica la construcción de algoritmos sofisticados. Las condiciones están claramente contrapuestas ya que el desempeño del sistema, el cual se evalúa en base a una medida de la calidad alcanzada a una tasa dada, usualmente se obtiene a expensas de costo y complejidad adicionales.

Cualquier sistema de codificación de fuente utilizado para compresión, puede ser modelado como un proceso de tres etapas: La primera etapa tiene la función de trasladar la señal original a un nuevo dominio, en el que esté mejor preparada para las operaciones que le aplicará la segunda etapa. La segunda etapa es un bloque de cuantificación. Es aquí en donde se obtiene la mayor parte de la compresión. La tercera etapa es un bloque de codificación sin pérdidas, el cual se utiliza para obtener una compactación adicional de la señal.

En la primera etapa, conocida como procesamiento de la señal, usualmente no se introducen pérdidas de información. A partir de 1990 empiezan a aparecer estudios sobre etapas de procesamiento de la señal orientadas a la optimización de la cuantificación vectorial [VQ28]. Como resultado de estos estudios se ha determinado que existen dos atributos que una etapa de procesamiento debe cumplir

para obtener el máximo desempeño de la cuantificación vectorial: El primero consiste en reducir al máximo la correlación entre los vectores que van a ser cuantificados independientemente. El segundo consiste en preservar al máximo la correlación entre las componentes del vector. Con esta filosofía, se han propuesto esquemas de procesamiento tales como: transformaciones vectoriales, bancos de filtros vectoriales, etc.

En los sistemas de compresión, propuestos hasta ahora, basados en este tipo de etapas de procesamiento, se utiliza la cuantificación vectorial del vecino más cercano con esquemas de asignación fija o dinámica de información. Sin embargo, la mayoría de los trabajos sobre cuantificación vectorial se han dedicado al estudio de fuentes en las que no se cumplen simultáneamente los dos atributos para la optimización de la VQ. Entonces, en base a las características particulares de las fuentes procesadas vectorialmente, surge la motivación de determinar si existen algunas formas de aplicar la cuantificación vectorial, que produzcan mejores resultados que los que hasta ahora se han obtenido con el VQ ordinario.

En este trabajo se utiliza una etapa de procesamiento vectorial, específicamente la transformada coseno discreta vectorial, 2D-VDCT, como punto de inicio a lo que puede ser un estudio más profundo sobre nuevas formas de aplicación de la VQ en sistemas de compresión con procesamiento orientado a vectores. Se eligió esta transformación debido a que, como es una extensión de la 2D-DCT, los esquemas de compresión basados en esa transformación pueden ser tomados como punto de partida para proponer nuevos esquemas de compresión.

Teóricamente, los coeficientes en el dominio transformado deben ser independientes unos de otros. Esto quiere decir que no existe dependencia estadística entre ellos. Esto es válido únicamente cuando la señal de entrada tiene características estadísticas parecidas a las del modelo de señal que dio origen a la DCT como una aproximación de la KLT (la versión tridiagonal del proceso Markov-1) [VQ61]. Sin embargo, frecuentemente la señal, en este caso imágenes, no cumplirá con ese requisito ya que difícilmente puede ser modelada como un proceso estacionario.

Observaciones hechas sobre imágenes en el dominio de la transformada 2D-VDCT revelaron que existe una dependencia estadística entre los coeficientes, debida a las características de la imagen original. Esta dependencia se refleja en patrones de distribución de energía con ciertas configuraciones bien definidas. Generalmente estos patrones están relacionados con características de la señal original tales como orientación de los bordes, texturas, etc.

Estas observaciones indican que existe una cierta cantidad de correlación en el dominio de la 2D-VDCT que no puede ser explotada por la cuantificación independiente de los coeficientes, esto es, mediante el uso de cuantificadores vectoriales sin memoria en cada uno de ellos. La cuantificación con memoria puede desempeñarse mejor que la cuantificación sin memoria, en vectores con cierta dependencia estadística [VQ52]. Aquí debe enfatizarse que los resultados de la

teoría de la información implican que un VQ con memoria no puede desempeñarse mejor que un VQ sin memoria en el sentido de minimizar la distorsión promedio para una restricción de la tasa, dada. De hecho, en la teoría de la información, el modelo matemático básico para un sistema de compresión de datos es exactamente un VQ sin memoria, y tales códigos se pueden desempeñar arbitrariamente cerca del desempeño óptimo que se puede obtener usando cualquier sistema de compresión. Sin embargo, el crecimiento exponencial de la cantidad de cálculos y del espacio de almacenamiento, con el incremento en la tasa, pueden resultar en cuantificadores vectoriales imposibles de construir. Un cuantificador vectorial con memoria puede producir la distorsión deseada con una complejidad practicable. Además, la libertad de utilizar diferentes cuantificadores para cada vector, sin incrementar la tasa puede permitir al código desempeñarse mejor que un cuantificador vectorial sin memoria de la misma dimensión y tasa. Esto se debe a que, en algunos casos, una fuente no estacionaria puede ser modelada por un conjunto de fuentes con algún tipo de estacionariedad, por lo que la aplicación de un VQ para cada una de esas fuentes puede ser mejor que la aplicación de un cuantificador vectorial único[VQ1, VQ52].

La memoria puede ser incorporada de una forma simple, en un cuantificador vectorial, usando libros diferentes para cada vector de entrada, donde los libros son elegidos en base a los vectores de entrada pasados. El decodificador debe saber qué libro se usó por el codificador, para poder reconstruir la señal a partir de los símbolos del canal. Esto puede lograrse de dos formas. En primer lugar, el codificador puede usar un procedimiento de selección del libro de códigos, el cual se base exclusivamente en sus salidas pasadas, y por lo tanto la secuencia de libros de código puede ser seguida en el decodificador. En segundo lugar, el decodificador es informado del libro de códigos seleccionado, vía un canal especial de tasa baja. El primer método es llamado VQ retroalimentada (PVQ, FSVQ, etc.), mientras que el segundo se conoce como VQ adaptable [VQ1].

En el dominio de la 2D-VDCT cada coeficiente tiene diferentes características (nivel de energía, factor de acoplamiento intrínseco, etc.), y por lo tanto no es práctico diseñar un sólo cuantificador para todos ellos. Debido a esto, si se busca utilizar un cuantificador vectorial predictivo, como una forma de introducir memoria en la etapa de cuantificación, los vectores sucesivos de entrada al cuantificador coeficientes con igual índice, pertenecientes a bloques transformados adyacentes, tendrían poca dependencia estadística, entonces este esquema no proporcionaría una ganancia de desempeño en comparación con la cuantificación sin memoria.

Otra forma de introducir memoria en la etapa de cuantificación, consiste en utilizar un cuantificador vectorial clasificado en el cual el modo de operación se elija en base a las características de un conjunto de vectores. En el caso específico de la 2D-VDCT, si la introducción de memoria en la etapa de cuantificación, se hace por medio de un cuantificador vectorial clasificado, entonces los modos de operación de

los cuantificadores de cada uno de los coeficientes, podrían relacionarse fácilmente con la estructura del patrón de la distribución de energía. Además, al añadir un conjunto de plantillas de eliminación de coeficientes y un esquema de asignación de bits al conjunto de cuantificadores clasificados, se obtendría un esquema de codificación zonal, similar al esquema de codificación de transformada basado en la 2D-DCT.

En el diseño de un cuantificador vectorial clasificado, existen dos puntos importantes: El primero es encontrar un parámetro de clasificación, adecuado. El segundo consiste en repartir, de la manera más eficiente, la tasa total del código entre los códigos de los distintos modos de operación del cuantificador. En este trabajo se presentan dos esquemas de clasificación relacionados con la norma de los vectores, en los que la asignación de bits se realiza en base a las matrices de covarianza de los vectores de cada uno de los modos de operación del cuantificador. Ambos clasificadores utilizan un conjunto de umbrales para determinar, en base a la norma del vector, en que modo de operación del cuantificador debe realizarse la codificación. El diseño del primer clasificador se basa en la correlación entre las componentes del vector y en la concentración de la energía, en los diferentes modos de operación, para optimizar un conjunto de umbrales. Aquí lo que se busca es, en base a los umbrales clasificar los vectores de tal forma que los grupos con más energía tengan los mejores atributos para la aplicación de la VQ. El diseño del segundo clasificador se basa en los resultados del teorema Shannon-MacMillan del cual se puede deducir que, conforme el número de dimensiones de un vector aleatorio aumenta, su función de densidad de probabilidad se concentra en una superficie de probabilidad constante. Entonces, una estrategia eficiente de codificación de fuente es distribuir las palabras del código cerca de esa superficie.

Debido a que las fuentes que representan a los coeficientes de la 2D-VDCT muestran una gran dependencia entre sus coeficientes, se espera que la introducción de cuantificación vectorial con memoria, por medio del uso del CVQ, proporcione mejores resultados que la cuantificación vectorial sin memoria. Sin embargo, los métodos de diseño del CVQ que aquí se presentan, son algoritmos de optimización que no garantizan que esa hipótesis se cumpla.

Para realizar el diseño de los cuantificadores vectoriales, tanto clasificado como ordinario, fue necesaria la justificación de la estacionariedad y ergodicidad en base a las longitudes de las secuencias de entrenamiento. Una vez que se diseñaron estos cuantificadores, se procedió a la comparación del desempeño en base a curvas *tasa-distorsión*, obtenidas sobre secuencias de entrenamiento de los distintos coeficientes. Finalmente, los dos esquemas de cuantificación se compararon dentro de un esquema de compresión de imágenes.

En principió se esperaba que el desempeño *tasa-distorsión* del CVQ fuera superior al de la VQ sin memoria, únicamente cuando la información de la clasificación fuera transmitida, de tal forma, que la tasa requerida por ella fuera

despreciable, en comparación con la tasa del código. Sin embargo, los resultados experimentales muestran que, incluso al añadir el índice de clasificación de cada vector a la tasa de codificación, el desempeño del CVQ es superior al del VQ sin memoria. Aunque la brecha entre el desempeño de los dos esquemas disminuye, estos resultados permanecen válidos incluso cuando se utiliza codificación entrópica.

El sistema de compresión que se presenta en este trabajo es un sistema muy sencillo. En un trabajo posterior este sistema podría refinarse en base a los siguientes puntos:

- En el grupo de cuantificadores vectoriales de los coeficientes de la 2D-VDCT, la asignación de bits se podría hacer dinámicamente en base a las características particulares de cada uno de los bloques transformados.
- En los bloques transformados, la información de clasificación de cada uno de los cuantificadores vectoriales podría refinarse en base a una señal de error entre los índices de clasificación óptimos y los predichos por la plantilla asignada a cada bloque.

Capítulo 2

Cuantificación Vectorial

2.1 Introducción

En la sociedad actual, el flujo de información crece notablemente. Esto causa que cada día sea necesario tener más capacidad de transportar y almacenar datos. Para tener un uso más eficiente de estos medios de transporte y almacenamiento, lo más lógico es utilizar únicamente la cantidad indispensable de información para lograr los fines deseados, y discriminar el resto. A esta discriminación de la parte irrelevante de la información se le conoce como compresión. Bajo esta definición, la compresión se puede interpretar como la conversión de una corriente de datos analógicos o discretos de tasa alta a un flujo de datos de tasa más baja. Esta operación debe producir la más alta fidelidad de reproducción posible, cuando es sujeta a las restricciones de la tasa de información y de la complejidad del sistema.

El estudio formal de la compresión de datos se halla comprendido en lo que es la *teoría de codificación de fuente*, Esta teoría se encarga de estudiar la forma de convertir cualquier señal a una representación digital eficiente.

La compresión de datos se encuentra estrechamente ligada con un proceso que se conoce como cuantificación. La cuantificación no es más que la representación de un dato, por medio del elemento más parecido a él dentro de un conjunto finito de elementos. La compresión se alcanza porque la cantidad de información necesaria para identificar al elemento del conjunto es menor que la que se necesita para representar al dato original. Cuando a cada dato original se le asigna un elemento del conjunto, se dice que la cuantificación es escalar (SQ). Cuando a un

bloque de los datos originales se le asigna un elemento del conjunto, se dice que la cuantificación es vectorial (VQ).

2.2 Definiciones y propiedades

La cuantificación vectorial es un mapeo de un conjunto de objetos (en este caso bloques de datos) a un elemento representativo de dicho conjunto.

La VQ es comúnmente utilizada para la compresión de datos, sin embargo, esta no es la única aplicación que tiene. La VQ se utiliza en reconocimiento de patrones, clasificación, transformaciones lineales, etc. [VQ52].

El interés en el estudio de la VQ se debe a que, cuando la dimensión de los vectores se acerca a infinito, el desempeño de codificación se aproxima al límite teórico impuesto por la teoría tasa-distorsión de Shannon. Sin embargo, la complejidad del cuantificador, relacionada directamente con el número de dimensiones del vector, hace que en las aplicaciones prácticas el número de dimensiones se mantenga en un valor pequeño [VQ22]. A continuación se presentan los conceptos teóricos que fundamentan las técnicas de diseño de los cuantificadores vectoriales.

Cuantificador Vectorial

Sea $\boldsymbol{x} \in \boldsymbol{R}^n$ un vector de entrada al cuantificador, y sea $C=\{y_i \in \boldsymbol{R}^n: i=1 \ldots L\}$ un conjunto finito conteniendo L vectores. Un cuantificador vectorial n-dimensional $\boldsymbol{Q}$ con L posibles salidas es un mapeo

$$\boldsymbol{Q}: \boldsymbol{R}^n \rightarrow C$$

tal que

$$\boldsymbol{Q}(\boldsymbol{x})=\boldsymbol{y}_i \qquad \text{si } \boldsymbol{x} \in C_i$$

donde

$$C_i=\{ \boldsymbol{x} \in \boldsymbol{R}^n \mid \boldsymbol{Q}(\boldsymbol{x})=\boldsymbol{y}_i \}, \qquad i=1,2 \ldots L$$

con

$$\cup C_i=\boldsymbol{R}^n \qquad \text{y } C_i \cap C_j=\varnothing \qquad \text{si } i \neq j$$

Cada C_i es llamada una partición, una celda, o la imagen inversa de $\boldsymbol{y}_i$ bajo el mapeo $\boldsymbol{Q}$. C es conocido como libro de código, mientras que las $\boldsymbol{y}_i$ son las palabras del código.

En algunas ocasiones (principalmente en aplicaciones prácticas), es conveniente considerar al cuantificador en dos partes: un codificador α y un decodificador β. El primero asigna a cada vector x_n en un conjunto de entrada X, un símbolo de canal $U_n=g(x_n)$ en un conjunto M. En la mayoría de los casos, U_n es una representación binaria de la posición en una tabla del y_i que mejor representa a x_n. El decodificador β desempeña la lectura del libro de códigos C para generar la reproducción y_i, asignada al vector de entrada x_n.

Tasa del cuantificador

La tasa del cuantificador, la cual indica la cantidad de información por componente requerida para representar un vector, está definida como:

$$r = \frac{\log_2 L}{n} \quad \text{[bits por componente]}$$

Si para buscar una representación más eficiente de los datos, se utiliza codificación de longitud variable (b_i bits para el vector y_i), entonces la tasa promedio utilizada para cada componente está dada por [VQ41]

$$\mathrm{R} = \frac{1}{n}\sum_{i=1}^{\mathrm{L}} b_i \mathrm{P}[X \in C_i] \qquad \text{en donde} \qquad \frac{H(Y)}{n} \le R \le r$$

aquí $H(Y)$ es la entropía del vector aleatorio Y, definida como

$$H(Y) = -\sum_{i=1}^{L} p(Y_i)\log_2 p(Y_i)$$

Distorsión

Para medir el desempeño de un cuantificador es necesario tener una indicación del costo de reproducir el vector aleatorio X con el vector aleatorio Y. La distorsión puede ser, generalmente pero no en todos los casos, utilizada como una medida de ese desempeño [VQ52].

Sea X una fuente aleatoria vectorial con función de densidad de probabilidad conjunta $f_x(x)$. Entonces, la distorsión promedio por dimensión, D, para un vector aleatorio de salida Y y una medida de distorsión $d(\,,\,)$ puede ser expresada como

$$D = E\{d(\mathbf{X},\mathbf{Y})\}/n$$

Matemáticamente, una medida de distorsión es un mapeo de un espacio de producto cartesiano (por ejemplo $R^n \times R^n$) al intervalo $[0,\infty)$. La selección de una medida de distorsión para un problema en particular puede ser un problema difícil y controvertido. Idealmente la distorsión debe cuantificar la calidad subjetiva, debe ser tratable matemáticamente y debe ser calculable. Desafortunadamente, rara vez una medida de distorsión cumple con estas tres cualidades, y frecuentemente la primera está en oposición con las dos últimas ya que una medida de distorsión puede necesitar tener una forma muy complicada para obtener algún grado de calidad subjetiva [VQ49]. Existen muchas medidas de distorsión (Itakura-Saito, Error medio absoluto, etc.) pero la de uso más difundido es el *error cuadrático medio* (MSE), la cual está definida como

$$d(x,y) = \|x - y\|^2 = \sum_{i=0}^{n-1} (x_i - y_i)^2$$

y da una distorsión promedio por dimensión de

$$D = \frac{1}{n} \sum_{i=1}^{L} \int_{Ci} \|\mathbf{x} - \mathbf{y}i\|^2 f_{\mathbf{x}}(\mathbf{x}|\mathbf{x} \in C_i) d\mathbf{x}$$

el error cuadrático medio es una medida de distorsión de tipo *diferencia*. Ésta es una clase muy importante de medidas de distorsión, y deben su nombre a que dependen de sus argumentos únicamente a través de su diferencia [VQ49].

Cuantificadores del vecino más cercano

Cualquier VQ es definido completamente por un libro de códigos y un conjunto de particiones. En algunas casos el conjunto de particiones puede adquirir formas geométricas muy complicadas las cuales sean difíciles de describir geométricamente. Existe una clase particularmente importante de cuantificadores vectoriales la cual es llamada *cuantificadores del vecino más cercano o de Voronoi*. Estos tienen la característica de que el conjunto de particiones está completamente determinado por el libro de códigos y una medida de distorsión, con lo cual se evita tener una descripción explícita del conjunto de particiones. Estos cuantificadores son óptimos, en el sentido de minimizar la distorsión promedio dado un libro de códigos [VQ52].

Formalmente, un cuantificador del vecino más cercano se define como aquel cuyo conjunto de particiones está dado por

$$C_i = \left\{ \mathbf{x} : d(\mathbf{x}, \mathbf{y_i}) \le d(\mathbf{x}, \mathbf{y_j}) \forall j \in (1, \cdots, L) \right\}$$

en donde las ambiguedades se solucionan asignando a x en un punto frontera al vector de código con índice menor.

2.3 Condiciones para optimización

Para realizar un buen diseño de un sistema es indispensable tener un criterio de optimización. La optimización no es mas que la maximización de una medida de desempeño.

Cuando un cuantificador va a ser utilizado como un medio de compresión de datos, los criterios de optimización más comúnmente utilizados son la minimización del valor esperado de la distorsión, o la minimización de la distorsión máxima permitida.

Para que un cuantificador sea óptimo, éste debe de tener un libro de códigos y un conjunto de particiones tales que no existan otro libro y otras particiones que produzcan un valor de distorsión menor al que aquellos producen. Se ha demostrado que, para que un cuantificador sea óptimo, debe cumplir con la condición del vecino más cercano. Cuando el cuantificador cumple con esta condición, entonces la distorsión depende únicamente del libro de códigos, es decir, es una función de nL variables aleatorias. Aún suponiendo que se conoce la función densidad de probabilidad conjunta de las componentes del vector X, la solución analítica para obtener el libro de códigos que minimice a D puede ser imposible de alcanzar. Es por esto que, generalmente, los libros se construyen con métodos numéricos basados en algunas condiciones que aseguren la obtención de al menos un óptimo local.

Aunque no existe una derivación general, en algunos casos particulares (por ejemplo en distribuciones discretas) se ha demostrado que libros localmente óptimos pueden construirse al encontrar un conjunto de vectores C que satisfaga las tres condiciones siguientes [VQ52]:

Condición del vecino más cercano

Para un libro de códigos, C, el conjunto de particiones óptimo satisface la siguiente condición:

$$C_i = \left\{ \mathbf{x} : d(\mathbf{x}, \mathbf{y}_i) \le d(\mathbf{x}, \mathbf{y}_j) \forall j \in (1, \cdots, L) \right\}$$

esto es

$$Q(x)=y_i \qquad \text{si y sólo si } d(x,y_i) \le d(x,y_j) \text{ para toda } j$$

Condición del centroide

Para un conjunto de particiones $\{C_i: i=1, \ldots, L\}$, el libro de códigos óptimo , C, satisface la siguiente condición:

y_i=Cent(C_i)

Condición de frontera con probabilidad cero

Sea B_j la superficie que delimita a la partición C_j. Para que un libro de códigos sea óptimo para una distribución dada, la probabilidad de los puntos en la unión de todas las fronteras del conjunto de particiones debe ser igual a cero. Esto es

$$P(\bigcup_{j=1}^{L} B_j) = 0$$

2.4 Diseño y comparación

Cuando se construye un cuantificador vectorial, el objetivo es encontrar un libro de códigos que minimice la distorsión promedio. En la práctica, cuando la fuente que se desea cuantificar se deriva de imágenes, voz u otro tipo de datos de la vida real, la función densidad de probabilidad del vector aleatorio no es conocida con exactitud, entonces no se puede encontrar el valor esperado de la distorsión. Sin embargo, si el proceso aleatorio es estacionario y ergódico entonces el promedio muestra de término largo es igual a la esperanza matemática, por lo que la distorsión puede ser calculada por un promedio en el tiempo. En base a esta consideración, un método para diseñar el sistema de cuantificación es tomar una secuencia suficientemente larga de datos de la fuente de interés (secuencia de entrenamiento), estimar el valor esperado de la distorsión por medio del promedio muestra de término largo, e intentar diseñar un libro de códigos que minimice este promedio para la secuencia de datos.

Cuando la fuente no es estacionaria ni ergódica, el diseño basado en una secuencia de entrenamiento puede tener sentido siempre que la fuente sea, al menos, asintóticamente estacionaria en la media, y la secuencia de entrenamiento sea lo suficientemente larga [VQ1]. Un método frecuentemente utilizado para diseñar libros de códigos para fuentes asintóticamente estacionarias en la media, consiste en tomar una secuencia de entrenamiento muy larga, diseñar el libro de códigos que minimice la distorsión muestra promedio y codificar una secuencia de prueba generada por la misma fuente pero que no esté en la secuencia de entrenamiento. Si el desempeño en la secuencia de prueba es razonablemente cercano al desempeño de diseño, entonces se puede confiar en que el desempeño del código estará cerca a su valor de diseño en futuras secuencias de prueba. Si este no es el caso, entonces la secuencia de entrenamiento no es lo suficientemente larga, por lo que se debe repetir el diseño con una secuencia mayor.

El uso directo de la cuantificación vectorial está limitado por una seria barrera debida a la complejidad que imponen, tanto el tamaño del libro como la dimensión de los vectores. Reducir la dimensión de los vectores a valores pequeños o reducir los tamaños de los libros, frecuentemente implica sacrificar la posibilidad de hacer una explotación eficiente de la dependencia estadística existente entre un grupo de muestras. Varias técnicas han sido desarrolladas para obtener un compromiso favorable entre desempeño y complejidad. Estas aplican un conjunto de restricciones a la estructura del cuantificador, y producen algoritmos alterados de codificación y diseño. De esta forma se pueden utilizar dimensiones grandes y tasas muy altas sin un incremento prohibitivo en la complejidad, y por lo tanto, se pueden alcanzar calidades que con la VQ ordinaria son impracticables. Algunos de los métodos de cuantificación vectorial con restricciones, más conocidos, son: cuantificación por remoción de la media, forma-ganancia, de enrejado, geométrica, piramidal, etc.

A continuación se presentan los métodos, basados en secuencias de entrenamiento, más comúnmente utilizados para la construcción de cuantificadores vectoriales. El primero en presentarse es el algoritmo LBG, pues éste es utilizado en la construcción de algunos otros cuantificadores más estructurados.

Algoritmo LBG.

Tal vez el método más utilizado para la construcción de cuantificadores vectoriales es el algoritmo LBG (propuesto por Linde, Buzo y Gray [VQ2]). Éste es un método iterativo el cual se basa en las condiciones necesarias para que un libro de códigos sea óptimo. El algoritmo comienza con un libro de códigos inicial de tasa cero, es decir un libro de códigos con sólo un elemento (el centroide de la secuencia de entrenamiento). En cada iteración, el libro de códigos inicial es perturbado de tal forma que se obtiene un nuevo libro con más elementos, el cual es optimizado por medio de las condiciones: del vecino más cercano, del centroide y de fronteras con probabilidad cero. Las iteraciones finalizan cuando el libro de códigos alcanza el tamaño deseado, o el decremento en la distorsión promedio cae por debajo de un umbral preestablecido. A continuación se resume este algoritmo:

1) Dada una secuencia de entrenamiento $\boldsymbol{X}=\{x_0,...,x_{M-1}\}$, m=número de elementos del código, umbral de distorsión $\varepsilon\geq 0$, un libro de códigos de m elementos $C_m=\{y_i: i=1,...,m\}$ y D_{m-1}, la distorsión de la iteración previa, hacer $C_m=\{\text{Cent}(\boldsymbol{X})\}$, $m=0$ y $D_{m-1}=\infty$.
2) Dado C_m, dividir cada vector y_i en dos vectores cercanos $y_i+\xi$ e $y_i-\xi$ en donde ξ es un vector de perturbación fijo. Reemplazar m por $2m$.
3) Por medio de la condición del vecino más cercano, encontrar la partición de mínima distorsión $P(C_m)=\{C_i: i=1,..,m\}$. Calcular la distorsión promedio D_m por medio de

$$D_m = \frac{1}{M} \sum_{i=0}^{M-1} \min_{y \in C_m} d(x_i, y)$$

4) Si $\varepsilon \geq (D_{m-1}-D_m)/D_m$, entonces ir a (6). En caso contrario, continuar.
5) Por medio de la condición del centroide, encontrar el libro de códigos óptimo, C_{m+1}, para $P(C_m)$. Sustituir C_m con C_{m+1}, ir a (3).
6) Si m=número deseado de vectores, entonces finalizar el algoritmo con C_{m+1} como el libro de códigos final. En caso contrario ir a (2).

Cuantificador con remoción de la media

En algunos tipos de fuentes, la amplitud promedio de las componentes del vector $\boldsymbol{X}$, es decir la media muestra m, puede ser considerada como estadísticamente independiente de la variación que las componentes presentan alrededor de ese valor. Esa variación es conocida como residual $\boldsymbol{R}$. Bajo esta consideración, el vector original puede ser descompuesto en dos nuevas variables aleatorias de la siguiente forma:

$$\boldsymbol{X}=m\boldsymbol{1}+\boldsymbol{R}$$

en donde $\boldsymbol{1}$ es el vector n-dimensional cuyas componentes son iguales a la unidad, m es una variable aleatoria escalar y $\boldsymbol{R}$ es un vector aleatorio de dimensión n. A la cuantificación de m y $\boldsymbol{R}$ usando libros de código diferentes se le conoce como cuantificación vectorial por remoción de la media (Figura 2.4.1(a)). Este es un ejemplo de cuantificador vectorial con restricciones.

La representación del vector $\boldsymbol{X}$ por medio de $\boldsymbol{R}$ y m puede ser muy útil por dos razones. En primer lugar se tiene que, el código producto obtenido con los dos cuantificadores óptimos independientes de $\boldsymbol{R}$ y m es óptimo para $\boldsymbol{X}$, en el sentido del vecino más cercano [VQ52]. En segundo lugar se tiene que esta descomposición estructural puede ser una alternativa para simplificar la tarea de codificación de $\boldsymbol{X}$, cuando la utilización de un sólo libro de códigos haga este proceso poco práctico, debido al tamaño requerido por el libro para obtener una calidad de codificación de cierto valor.

Aquí, el libro de códigos de $\boldsymbol{R}$ se genera por medio del algoritmo LBG, mientras que el libro de la variable aleatoria escalar m se construye por medio del método Lloyd-Max para el diseño de cuantificadores escalares [VQ52]. Ambos libros son generados independientemente uno de otro.

Existe una variación de este método en la cual, para obtener el residual, se le sustrae la media cuantificada al vector X. Este método produce mejores resultados en términos de relación señal a ruido pero tiene dos desventajas. En primer lugar, tiene el inconveniente de que el diseño de los dos cuantificadores ya no es independiente uno del otro. En segundo lugar, el codificador no tiene un código producto independiente sino que ahora es un código producto secuencial, lo cual

implica que, para cuantificar el residuo hay que haber cuantificado antes la media, por lo que, en detrimento del tiempo de cálculo, estas dos operaciones no pueden ser realizadas en paralelo [VQ52].

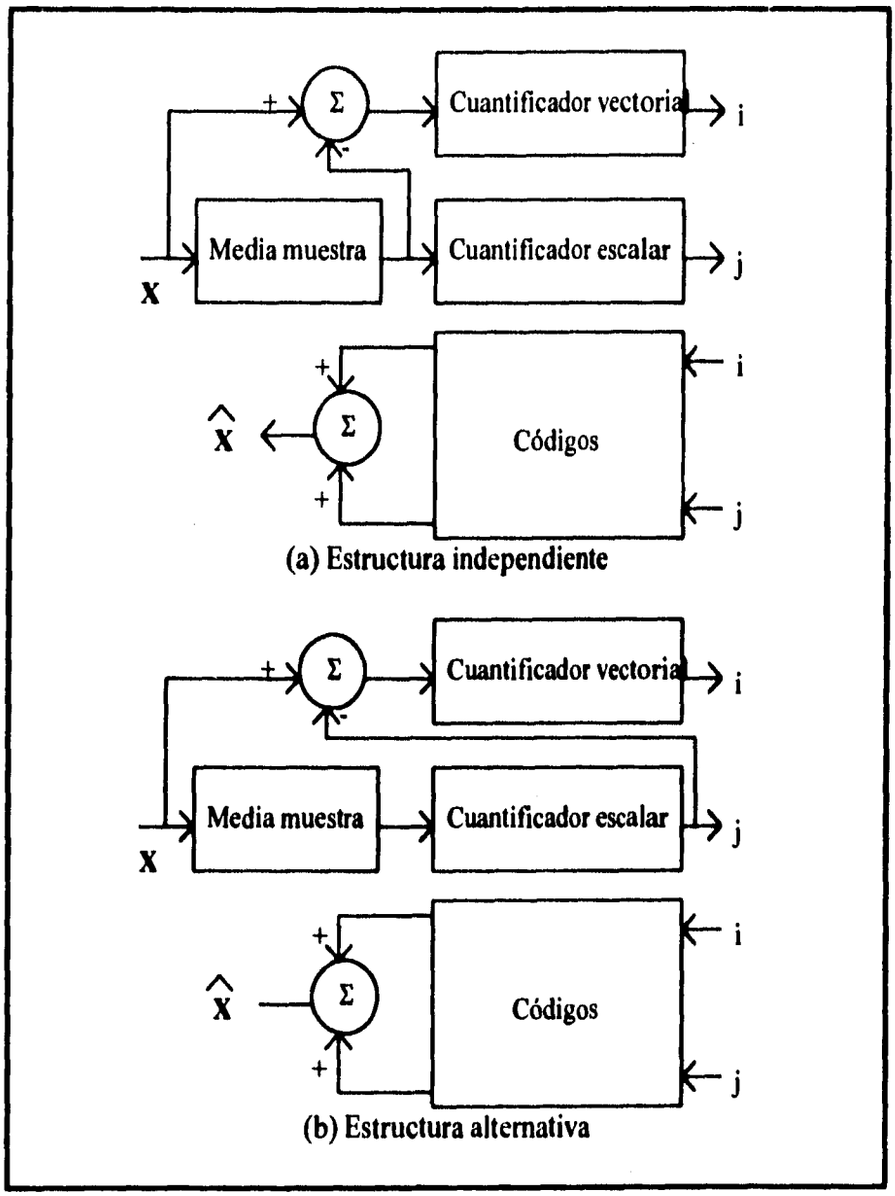


Figura 2.4.1. Cuantificador vectorial con remoción de la media.

Cuantificador *forma-ganancia*

En algunos tipos de fuentes reales (por ejemplo voz y otros tipos de sonidos) sucede que algunos patrones de variación se repiten a diferentes escalas. En estos casos, la distribución de probabilidad de la forma de los vectores puede ser independiente de la escala a la que estén representados. Aquí, un código producto, el cual actúe independientemente en el rango dinámico y la forma del vector, puede simplificar significativamente la tarea de codificación a expensas de una pequeña pérdida en la fidelidad de reproducción [VQ53, VQ54]. En la codificación forma-ganancia, el vector aleatorio $\boldsymbol{X}$ es normalizado por su norma Euclidiana; de esta manera se obtiene un vector aleatorio $\boldsymbol{S}$, el cual está distribuido en la superficie de la hiperesfera de radio unitario en el espacio n-dimensional, y por lo tanto es más susceptible de ser codificado eficientemente que el vector aleatorio original. Sin embargo, la codificación *forma-ganancia* tiene la desventaja de que un código producto con cuantificadores independientes no es óptimo para la codificación. Aquí, la regla óptima de codificación es un proceso de dos pasos (figura 2.4.2), en donde el primer paso involucra exclusivamente al parámetro de forma y un libro de código. El segundo paso depende del primero para el cálculo del vecino más cercano para obtener el valor de ganancia óptimo [VQ52].

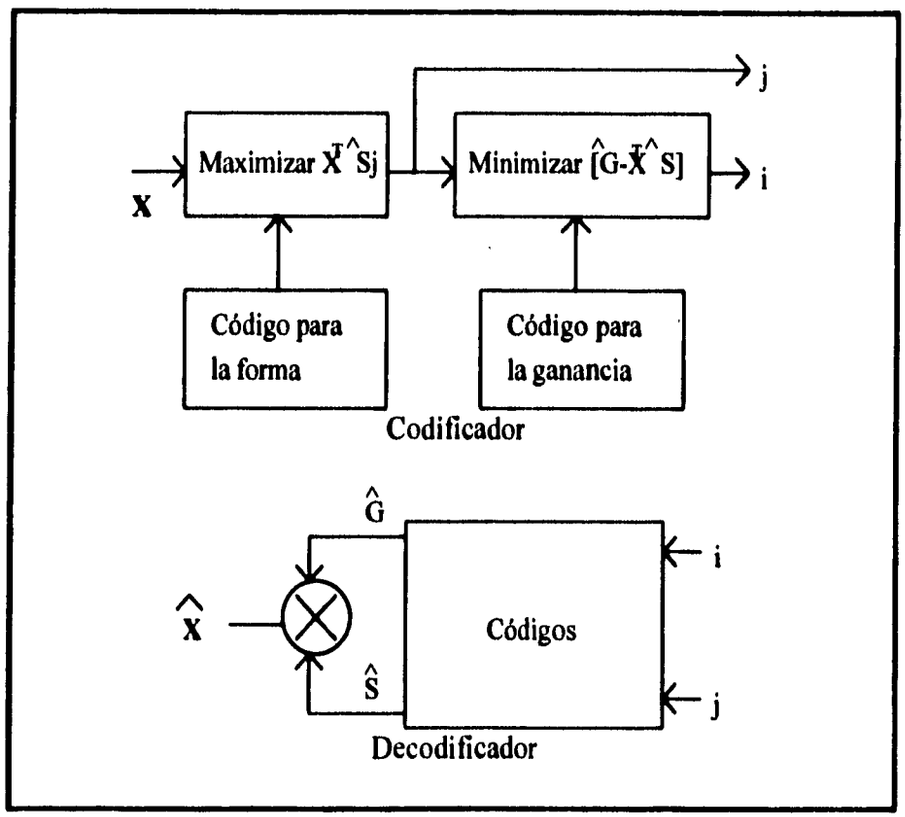


Figura 2.4.2. Cuantificador vectorial de *forma-ganancia*.

Cuantificador de enrejado

Un enrejado n-dimensional, Λ, se define como el conjunto de puntos $\boldsymbol{Y}$ pertenecientes a $\boldsymbol{R}^n$ tales que

$$\Lambda = \left\{ \mathbf{Y} \in \mathbf{R}^n \middle| \exists (u_1, \cdots, u_n) \in \mathbf{Z}^n : \mathbf{Y} = \sum_{i=1}^{n} u_i \mathbf{a}_i \right\}$$

en donde $\{\boldsymbol{a}_i\}$ es un conjunto de vectores linealmente independientes pertenecientes al espacio $\boldsymbol{R}^m$, en donde $m \geq n$. La matriz generadora G, es una matriz cuyos renglones son los vectores base $\boldsymbol{a}_i$. El volumen de la celda de voronoi del enrejado esta dado por

$$vol(\Lambda) = \left| det\left(GG^T\right) \right|^{1/2}$$

Un cuantificador de enrejado es aquel cuyo libro de códigos está formado por un subconjunto de un enrejado. Tal subconjunto es elegido entre los puntos del enrejado que cubren la región de mayor probabilidad de la fuente. Cuando se conoce la región que contiene la mayor probabilidad de la fuente, y se ha elegido un enrejado, un parámetro clave es la densidad del enrejado, esto es, cuantos puntos por unidad de volumen del espacio son contenidos en el libro de códigos. Entre más grande sea la densidad, mayor será la tasa del código y menor la distorsión promedio.

Al utilizar el enrejado, la distribución predefinida de los vectores produce un libro de códigos cuya estructura es periódica y ordenada, permitiendo la construcción de algoritmos de codificación, simples y rápidos. A diferencia de los métodos basados en el algoritmo LBG, en el cuantificador de enrejado no hay necesidad de calcular la norma para buscar el vector más cercano entre todos los del código. En consecuencia, el costo computacional no depende del tamaño del libro de códigos. Por ejemplo, la cuantificación con el enrejado D_n involucra

$f(\boldsymbol{X})$: n redondeos
n-1 sumas
1 prueba de paridad

$g(\boldsymbol{X})$: n diferencias
1 redondeo
1 búsqueda de un máximo.

esto es, entre $2n$ y $3n$+2 operaciones. En comparación con las $(3n\text{-}1)L$ requeridas por un cuantificador vectorial ordinario del vecino más cercano, con L palabras de código y búsqueda exhaustiva.

En algunas aplicaciones, utilizando un enrejado adecuado, un cuantificador vectorial de este tipo combinado con codificación entrópica puede proporcionar un código con desempeño cercano al óptimo, y complejidad de construcción razonable.

Cuantificador geométrico

Shannon observó en 1948 (en *A Mathematical Theory of Communication*) que, con el número de dimensiones, n, suficientemente grande, un vector aleatorio $\boldsymbol{X}$ estacionario ergódico se concentra en una región particular del espacio Euclidiano n-dimensional. La localización adecuada de las palabras de código en esta región de alta probabilidad, es conocida como cuantificación geométrica. Se puede demostrar que la región de alta probabilidad en la que $\boldsymbol{X}$ se concentra, está alrededor de un contorno de densidad de probabilidad constante [VQ30] que, para fuentes uniforme, Gaussiana y Laplaciana, independientes idénticamente distribuidas, es un cubo, una esfera y una pirámide respectivamente. La importancia de este tipo de codificación radica en que se pueden derivar códigos que, asintóticamente en tasa y dimensión, alcanzan una distorsión arbitrariamente cercana al límite teórico. Además, los elementos del código pueden ser un subconjunto de un enrejado localizados en el contorno de probabilidad constante, con lo cual se pueden construir algoritmos rápidos de codificación [VQ30, VQ55, VQ56].

Cuantificador vectorial con estructura de árbol

Una de las técnicas más efectivas para reducir la complejidad de codificación en la cuantificación vectorial, es el uso de un libro de códigos con estructura de árbol. En este tipo de cuantificadores, la búsqueda es realizada en etapas. En cada etapa una parte sustancial del código se elimina de la búsqueda usando un pequeño número de operaciones. En un código de árbol balanceado, el vector de entrada es comparado con m vectores de prueba preestablecidos. El vector de prueba más cercano al vector de entrada determina cual de las m trayectorias del árbol hay que seleccionar para alcanzar la siguiente etapa de prueba.

Cuantificador clasificado

La cuantificación vectorial clasificada es similar al uso de un cuantificador vectorial de árbol en donde el criterio para seleccionar alguna de las m ramas se basa en una característica heurística que el diseñador adopta para identificar el modo particular del vector de entrada. Aquí un clasificador arbitrario se utiliza para seleccionar un conjunto particular del libro de códigos que será utilizado (figura 2.4.3). El libro de códigos global puede estar formado de subconjuntos con diferentes números de vectores. Existen muchas posibilidades para la elección del clasificador. Este puede ser un VQ más simple, el cual únicamente identifique en cuál de las m regiones del espacio de entrada se encuentra el vector que va a ser codificado.

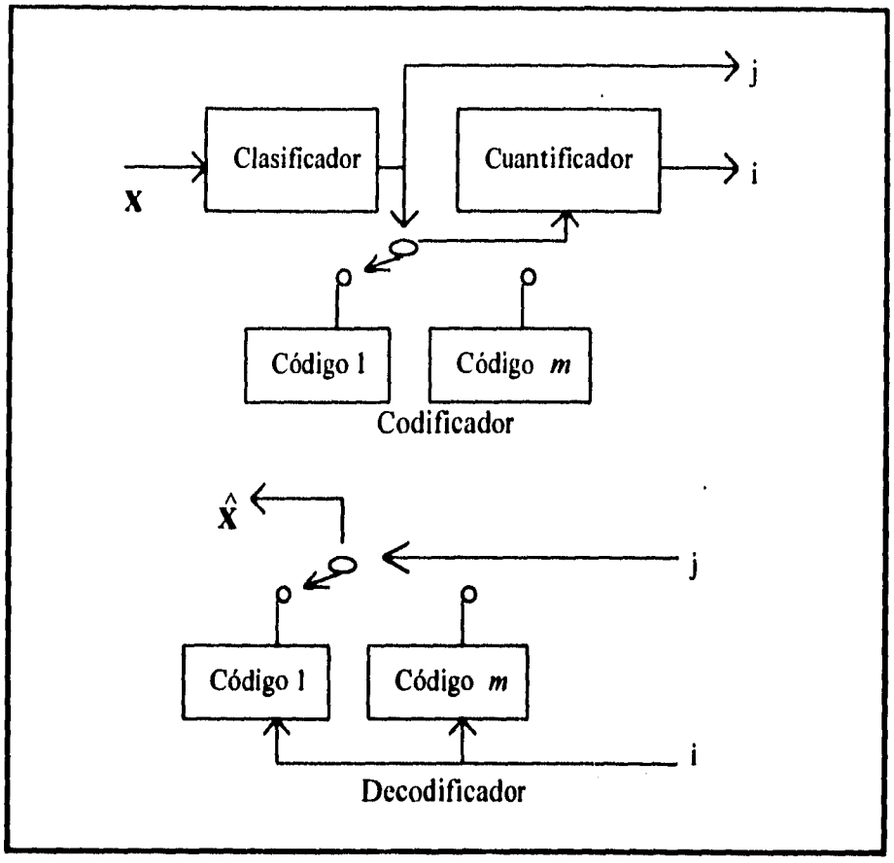


Figura 2.4.3. Cuantificador vectorial clasificado.

Cuantificación vectorial de transformada

En lugar de aplicar directamente un vector a la entrada de un cuantificador vectorial, una transformación lineal ortogonal, ***T***, se usa para procesar el vector. Entonces, el vector transformado puede ser cuantificado y la salida del cuantificador puede ser inversamente transformada para producir la aproximación del vector original (figura 2.4.4). Se puede demostrar que los vectores de código óptimos para realizar la cuantificación en el dominio transformado, son la transformación de los vectores de código obtenidos en dominio original de la señal. Además, la distorsión promedio en el dominio transformado, es la misma que la que se obtiene con la aplicación directa de la cuantificación vectorial a la señal original. Sin embargo, la ventaja de usar la transformación radica en que ésta compacta la energía de los vectores de la señal original en un número pequeño de componentes del vector. Esta propiedad implica que una parte sustancial de las componentes del vector pueden ser descartadas, con lo cual se reduce la dimensión de los vectores que van a ser cuantificados, y la complejidad del codificador disminuye. Por otro lado, en lugar de discriminar los coeficientes de baja energía, estos pueden ser agrupados para formar

un vector, el cual puede ser codificado con una tasa muy baja. En contraste con la aplicación directa de la cuantificación vectorial, en la cuantificación por transformada, el dividir en grupos el conjunto de coeficientes, puede hacer factible el manejo de cuantificadores vectoriales con un número grande de componentes [VQ52].

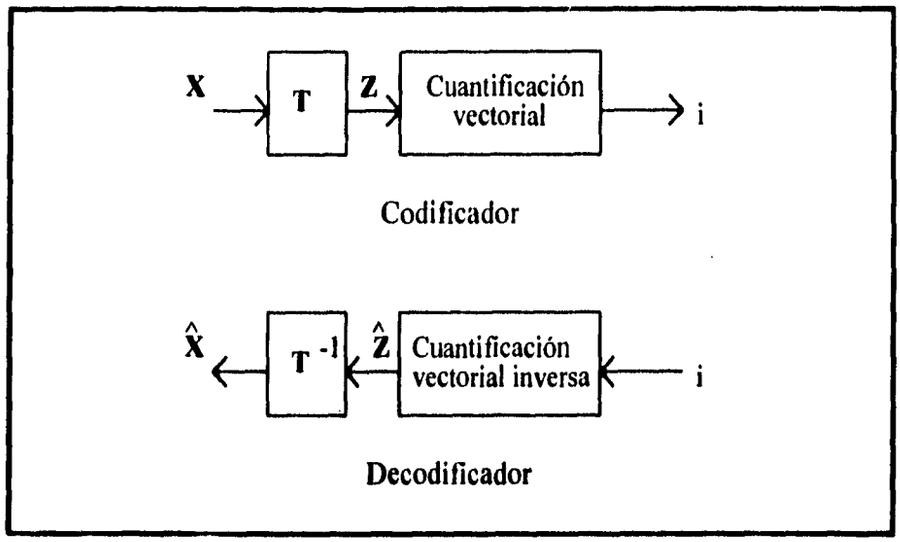


Figura 2.4.4. Cuantificador vectorial de transformada.

Cuantificación vectorial de múltiples etapas

Esta técnica, conocida también como cuantificación vectorial residual, está basada en la idea de dividir la tarea de codificación en diversas etapas sucesivas. En la primera etapa realiza una cuantificación relativamente burda, la segunda etapa opera sobre el error entre el vector original y la salida de la primera etapa. El error obtenido en la segunda etapa lleva a una representación más precisa del vector original. Etapas adicionales pueden ser utilizadas para refinar la cuantificación. El cuantificador vectorial en cada etapa opera en el error de cuantificación de la etapa previa y genera un índice que es enviado al decodificador. En el decodificador, el vector de entrada es representado por medio de la suma de las aproximaciones hechas en cada una de las etapas del codificador.

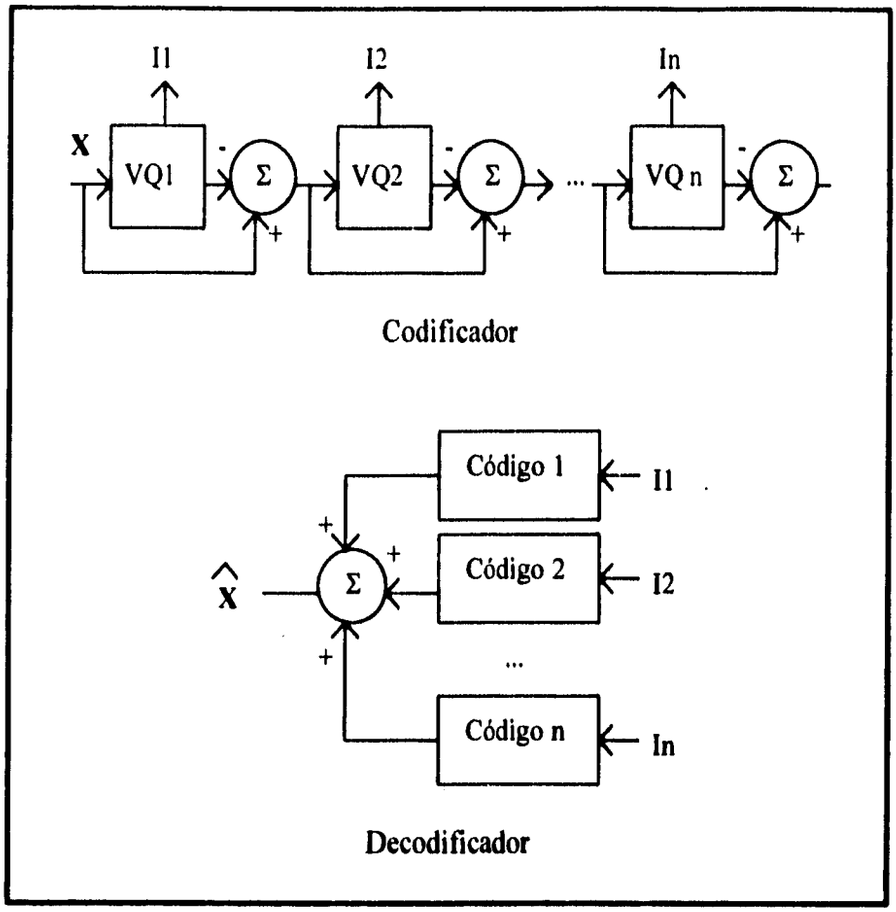


Figura 2.4.5. Cuantificador vectorial de múltiples etapas.

Cuantificación vectorial predictiva

Usualmente se necesita codificar una secuencia de vectores, donde se puede asumir que ellos tiene la misma función de densidad de probabilidad, sin embargo, vectores sucesivos pueden tener cierta dependencia estadística. La aplicación independiente de la VQ a cada uno de los vectores no toma en cuenta esta dependencia. Los codificadores con memoria, al tomar en cuenta la dependencia entre bloques, pueden ser más eficientes, en el sentido de proporcionar mejor desempeño para una tasa y complejidad dadas, que los codificadores sin memoria [VQ49, VQ52]. Una de las formas de incorporar memoria en un codificador, es utilizar un predictor. El diseño de cuantificadores vectoriales predictivos (PVQ) involucra la construcción de un predictor y un cuantificador. El método más simple para resolver este problema de diseño, es usar una metodología de lazo abierto. Esto es, el predictor se diseña en base a las estadísticas de la señal. Una vez que el predictor ha sido diseñado, se construye el libro de códigos en base a una secuencia

de entrenamiento obtenida por procesar la secuencia de entrenamiento original, con el predictor. Una vez que se ha diseñado este par de bloques, la operación de lazo cerrado se logra cuando el predictor opera en la señal reconstruida para predecir la siguiente entrada. Este predictor no es óptimo para la señal cuantificada, pero puede producir un desempeño cercano al óptimo si la reproducción del sistema es suficientemente buena. El cuantificador también opera en vectores cuyas estadísticas pueden diferir de las de la secuencia de entrenamiento para lo que fue diseñado. De nuevo, si la reproducción de la secuencia es lo suficientemente buena, el desempeño del cuantificador será cercano al óptimo.

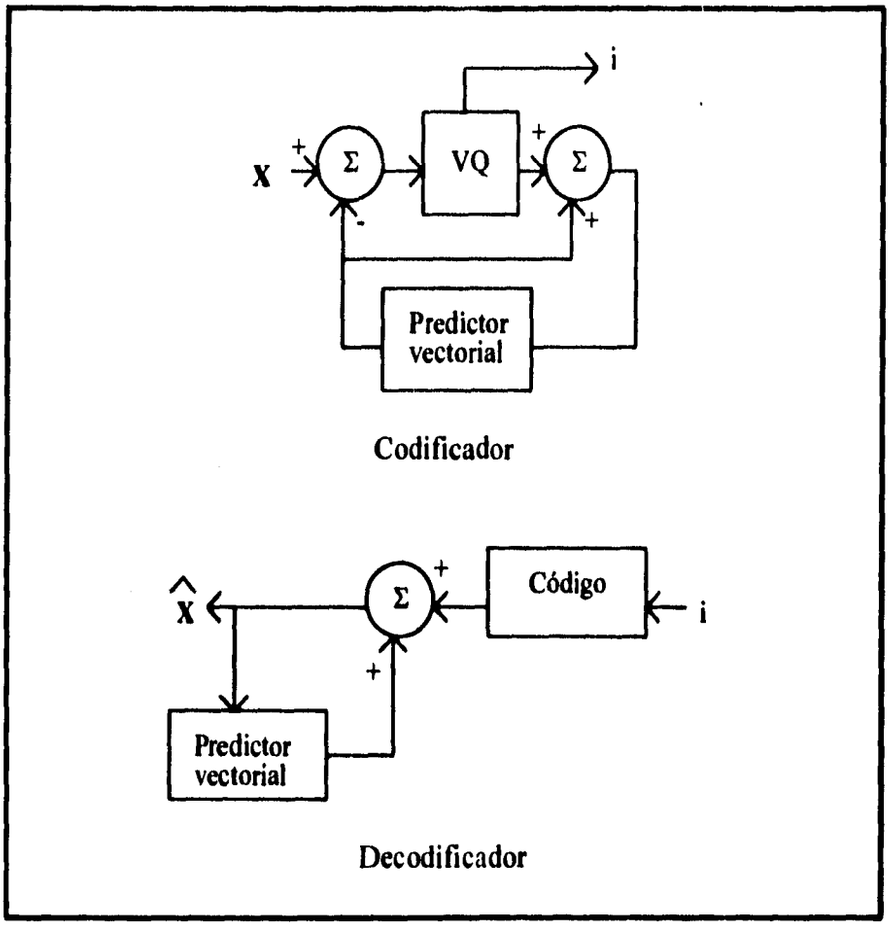


Figura 2.4.6. Cuantificador vectorial predictivo.

Métodos de codificación rápida para cuantificadores del vecino más cercano.

La principal carga de procesamiento en algoritmo LBG y en el codificador del cuantificador del vecino más cercano, está concentrada en la búsqueda del vector de código que minimice $d(x_i,y)$. Si se realiza una búsqueda exhaustiva, es decir, calcular

la distorsión entre x_i y cada uno de los elementos del libro para encontrar el óptimo, la operación de minimización puede ser prohibitiva en tiempo de cálculo, ya que, para un libro de códigos de L elementos de dimensión n , el algoritmo LBG requiere en cada iteración de, al menos, la siguiente cantidad de operaciones:

nL diferencias
nL cuadrados
$(n\text{-}1)L$ sumas

Para evitar este inconveniente, se propusieron esquemas tales como: búsqueda de árbol, VQ-multietapas, enrejados, etc. Sin embargo, muchos de los algoritmos de codificación rápida alcanzan su objetivo de decrementar el tiempo de cálculo a expensas de la calidad de codificación. En 1992 Huang *et al.* encontraron tres métodos de codificación rápida, los cuales producen los mismos resultados que una búsqueda exhaustiva pero reducen el tiempo de codificación entre 80 a 97% [VQ4]. Estos métodos están basados en consideraciones geométricas que permiten limitar la búsqueda a un pequeño subconjunto de vectores del código.

Método I

Dada una métrica $d(x,y)$ en el espacio n-dimensional, sea h_i la distancia entre el vector de entrada x, y el vector del código y_i tal que

$$|d(y_i,0)-d(x,0)| \le |d(y_j,0)-d(x,0)| \forall j = 1...L \qquad [1.4.1]$$

entonces el vector del código, que bajo la regla del vecino más cercano es el más parecido a x, debe estar localizado dentro de la región limitada por la hiperesfera de radio $2h_i$ centrada en y_i (ver figura 2.4.7). Aquellos vectores fuera de esta región pueden ser excluidos de la búsqueda sin calcular $d(x,y)$. Esto reduce el número de operaciones y por lo mismo el tiempo de codificación. El algoritmo se reduce a los siguientes pasos

a) Ordenar los vectores del libro de códigos de acuerdo a $d(y,0)$ en forma decreciente.
b) Para el vector x, calcular $r_x=d(x,0)$
c) Identificar el vector de referencia y_i por medio de (1.4.1) y calcular $h_i=d(x,y_i)$.
d) Identificar un subconjunto S del libro de códigos el cual contenga todos los vectores del código y_k que satisfagan la condición $d(y_k,y_i) \le 2h_i$
e) Por medio de la regla del vecino más cercano, buscar en S al vector y_k más cercano a x. Si $h_k = d(y_k,x) \le h_i$ entonces designar a y_k como y_i y regresar al paso *(d)*. En caso contrario y_i es el vector más parecido a x.

Para mejorar la velocidad de este algoritmo, puede calcularse la tabla de distancias entre los diferentes elementos del código $d(y_k, y_i)$, y almacenarse junto con el libro. Hay que mencionar que esta tabla formada por una matriz de LxL elementos es simétrica, por lo que sólo es necesario almacenar $L(L\text{-}1)/2$ de ellos.

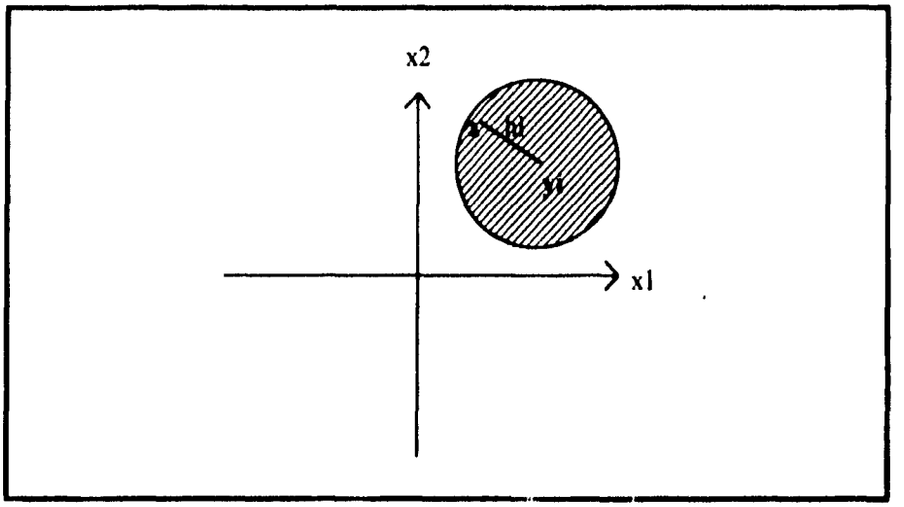


Figura 2.4.7. La región sombreada representa la parte del espacio que es examinada con el método I, para un vector de entrada x y una palabra de código inicial y_i.

Método II

Dada una métrica $d(x,y)$ en el espacio n-dimensional, sea h_i la distancia entre el vector de entrada x, y el vector del código y_i dado por 1.4.1, y sea r_x la distancia entre x y el origen. El mejor vector de código y_i, debe estar localizado dentro de la región limitada por las hiperesferas n-dimensionales de radios r_x-h_i y r_x+h_i.centradas en el origen (ver figura 2.4.8). Este algoritmo está comprendido por lo siguiente

a) Ordenar los vectores del libro de códigos de acuerdo a $d(y,0)$ en forma decreciente.
b) Para el vector x calcular $r_x=d(x,0)$
c) Identificar el vector de referencia y_i por medio de (1.4.1) y calcular $h_i=d(x,y_i)$.
d) Identificar un subconjunto S del libro de códigos el cual contenga todos los vectores del código y_k que satisfagan la condición $r_x - h_i \le r_k \le r_x + h_i$
e) Por medio de la regla del vecino más cercano, buscar en S al vector y_k más cercano a x. Si $h_k = d(y_k, x) \le h_i$ entonces designar a y_k como y_i y regresar al paso *(d)*. En caso contrario y_i es el vector más parecido a x.

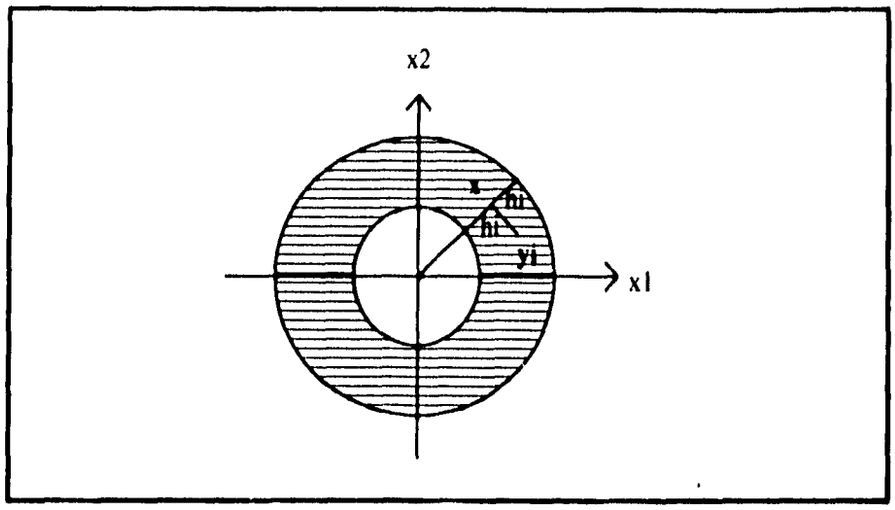


Figura 2.4.8. La región sombreada representa la parte del espacio que es examinada con el método II, para un vector de entrada x y una palabra de código inicial y_i.

Método III

Ya que el vector de códigos que mejor representa a x, en términos de la métrica $d(x,y)$, al mismo tiempo debe de estar dentro de la esfera de radio h_i centrada en y_i (donde y_i se obtiene por medio de 1.4.1), y en la región limitada por las esferas de radios r_x-h_i y r_x+h_i.centradas en el origen, entonces al limitar a S a la intersección de estas dos regiones (ver figura 2.4.9), se obtiene un conjunto de vectores de código más compacto lo que limita el número de comparaciones para la búsqueda del vector mejor representativo de x.

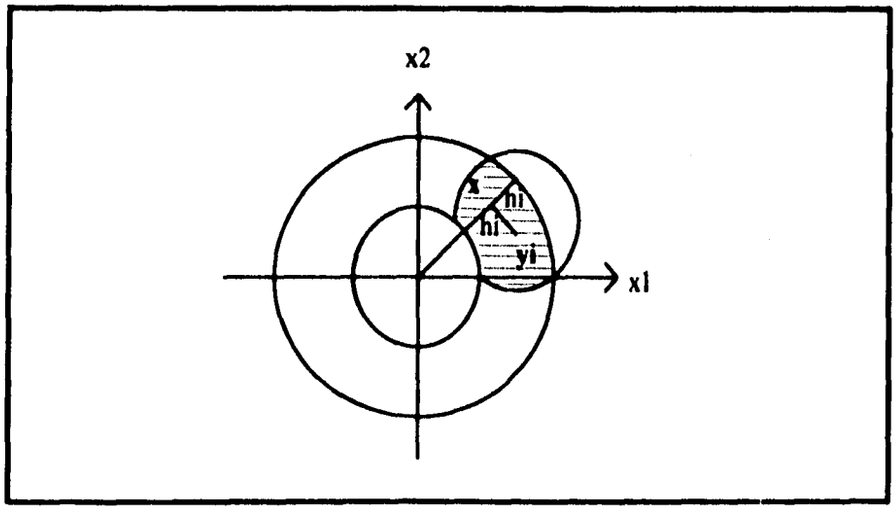


Figura 2.4.9: La región sombreada representa la parte del espacio que es examinada con el método III, para un vector de entrada x y una palabra de código inicial y_i.

Evaluación comparativa o condiciones para aplicabilidad

Todos los métodos de cuantificación vectorial: por remoción de la media, forma-ganancia, de enrejado, geométrico, de estructura de árbol, de transformada, y de múltiples etapas, pertenecientes a la clase de codificadores sin memoria, son formas de cuantificación utilizadas con el fin de evitar el inconveniente de la alta complejidad del cuantificador ordinario del vecino más cercano. Todos estos métodos tienen un desempeño, en términos de la distorsión, para un mismo valor de la tasa del código, inferior al del cuantificador del vecino más cercano ordinario. Sin embargo, como se comenta a continuación, algunas características particulares hacen que, bajo algunas condiciones, estos métodos representen una alternativa efectiva para la codificación de señales.

- Si el vector aleatorio $\boldsymbol{X}$ es tal que éste tiene una amplía varlación de los valores de ganancia o de la media, la cual es casi independiente de la forma de los vectores, entonces los métodos de cuantificación vectorial de *forma-ganancia* o por remoción de la media pueden representar una alternativa de reducción de complejidad del sistema de cuantificación con un ligero decremento en el desempeño del sistema.

- Los cuantificadores vectoriales de enrejado y geométrico tiene una estructura que permite una codificación rápida y uso eficiente de memoria. Los códigos de enrejado tienen un buen desempeño sólo en fuentes que son aproximadamente uniformes en una región limitada del espacio. Por otro lado, los cuantificadores geométricos únicamente son fáciles de diseñar en fuentes con distribuciones independientes, idénticamente distribuidas. La desventaja de estos cuantificadores es que no pueden ser mejorados por una variación del algoritmo LBG sin perder su estructura, y los buenos cuantificadores producidos por el algoritmo de LBG generalmente no pueden ser aproximados por enrejados. Además, debido a la consideración implícita de una resolución muy grande, los códigos de enrejado únicamente son adecuados para aplicaciones con altas tasas de bits y distorsiones pequeñas, con la limitación adicional de que, para que un código de enrejado pueda producir un buen desempeño, debe ser utilizado en conjunto con un cuantificador entrópico [VQ30, VQ41, VQ44, VQ45, VQ56, VQ67].

- El cuantificador vectorial de árbol puede representar una alternativa efectiva en sistemas de codificación en los que importa más la velocidad de búsqueda que el espacio de almacenamiento.

- El punto clave de la codificación de transformada es la compactación de información en unos pocos coeficientes. Esta propiedad implica que una fracción

sustancial de los componentes del vector en el dominio transformado tienen valores cercanos a cero, y por lo tanto pueden ser despreciadas. Esto reduce la dimensión del vector que se codifica, y por lo tanto la complejidad del cuantificador decrece. Otra de las características importantes de la codificación de transformada, es que su estructura permite introducir de una forma directa, condiciones de ponderación relacionadas con la calidad subjetiva de las señales. La codificación de transformada ha demostrado ser muy eficiente en la codificación de imágenes y voz. Es por eso que muchos de los esquemas de compresión actuales están basados en este tipo de técnicas. Sin embargo, el cuantificador vectorial de transformada puede ser de limitada ayuda si el grado de compactación que se puede alcanzar resulta en vectores de dimensiones muy grandes.

- En comparación con un cuantificador simple con la misma tasa de código, un cuantificador de múltiples etapas tiene la ventaja de que el tamaño del libro de código en cada una de ellas, se reduce considerablemente, por lo que el espacio de almacenamiento y la complejidad de la codificación son reducidas substancialmente. Además, su conformación es estructuralmente adecuada para la transmisión progresiva de señales. El precio que se paga por estas ventajas, es una inevitable reducción de la calidad de codificación global alcanzada con las varias etapas [VQ52].

Los métodos de cuantificación vectorial: clasificada y predictiva, pertenecientes a la clase de codificadores con memoria o retroalimentados, son formas de cuantificación utilizadas con el fin de explotar la dependencia estadística entre vectores sucesivos. Bajo algunas condiciones específicas, como se comenta a continuación, estos métodos pueden tener un desempeño, en términos de la distorsión, para un mismo valor de la tasa del código, superior al del cuantificador del vecino más cercano ordinario. En el caso del cuantificador predictivo, este puede dar mejores resultados que el VQ ordinario cuando los vectores tienen una dependencia estadística fuerte, por ejemplo una fuente Gauss-Markov de primer orden con un factor de regresión alto [VQ52]. El cuantificador clasificado puede ser útil en la codificación de señales con modelos estadísticos no estacionarios. Por ejemplo, en [VQ61] se presenta un cuantificador clasificado para la codificación de imágenes en el dominio de la 2D-DCT. Los cuantificadores con memoria tienen la importante desventaja de que la acumulación de errores en el canal puede causar errores de reconstrucción desastrosos. Esto puede ser controlado introduciendo un restablecimiento periódigo del sistema codificación como parte del sistema de corrección de errores.

Capítulo 3

Transformaciones Vectoriales

3.1 Introducción

En los sistemas de compresión de datos la mayor reducción en la tasa de codificación se logra en la etapa de cuantificación. Para optimizar el desempeño de esta etapa, se utiliza un módulo de procesamiento de señal, el cual no introduce pérdidas de información y cuyo propósito es trasladar la señal original a otro dominio en el que los datos estén más estructurados y en consecuencia sean más susceptibles de ser comprimidos eficientemente. Entre los varios atributos deseables de dicha etapa de procesamiento, uno muy importante es el grado de descorrelación que pueda introducir entre las variables involucradas. Se ha demostrado que entre menos correlacionadas estén un grupo de variables, más eficiente será su cuantificación independiente [VQ52].

En este contexto, muchas técnicas tales como predicción lineal y transformación lineal escalar han sido ámpliamente estudiadas como formas de reducción de correlación en sistemas de compresión. Debido a la capacidad de decorrelacionar las componentes de la señal y simultaneamente compactar la mayor parte de la energía en unos pocos elementos, las transformaciones lineales escalares han sido extensamente utilizadas en los sistemas de compresión de imágenes y video actualmente en uso. La transformada coseno discreta (DCT) es quizás, la transformación lineal más ampliamente utilizada en los sistemas de compresión en uso actualmente. En este capítulo se explican los atributos a los que la DCT debe su amplia difusión. Recientemente, las transformaciones lineales vectoriales y otros

esquemas de procesamiento vectorial, han sido propuestos como una nueva técnica para el procesamiento de la señal, previos a la cuantificación, en sistemas de compresión de imágenes [VQ5, VQ22, VQ23, VQ28].

La idea de utilizar etapas de procesamiento vectorial basadas en transformaciones, en la compresión de imágenes fue propuesta originalmente por Weiping Li en [VQ28]. En este trabajo se presenta un esquema de compresión por codificación de transformada, en el que la transformación opera sobre conjuntos de vectores. Posteriormente algunos investigadores [VQ5, VQ23] han utilizado procesamiento vectorial en sistemas multirresolución, descomposición en bandas, etc, con buenos resultados.

El objetivo de los sistemas de procesamiento vectorial, es potenciar de una forma *natural*, el desempeño de la cuantificación vectorial mientras que, como en el caso escalar, se mantiene la filosofía de descorrelación entre variables que van a ser cuantificadas independientemente. Aquí el sentido de la palabra *natural*, se explicará en base a un par de atributos de la señal, que hacen que su cuantificación vectorial sea más eficiente.

El objetivo del presente trabajo es ahondar en el estudio de la codificación por medio de transformaciones vectoriales. Específicamente, se busca explorar nuevas formas de realizar la cuantificación vectorial en un dominio en el que las componentes de los vectores están altamente correlacionadas. La transformada que se utilizará a lo largo de este trabajo es la transformada vectorial coseno discreta bidimensional (2D-VDCT). Esta transformación en especial se eligió en base a que, como se verá en la sexta parte de este capítulo, es la extensión a múltiples dimensiones de la 2D-DCT y por lo tanto se espera que tenga muchas de las virtudes que la 2D-DCT tiene para la codificación de fuente. Para comparar el efecto que, sobre los atributos de optimización de la VQ, tiene el procesamiento orientado a vectores, también se presentará un esquema de transformación escalar orientado a bloques.

3.2 Atributos óptimos de una etapa de procesamiento de la señal para la optimización de la cuantificación vectorial

La experiencia ha demostrado que cada vez que alguna dependencia estadística existe entre las componentes de un vector, alguna ganancia de desempeño puede esperarse al utilizar cuantificación vectorial en lugar de cuantificación escalar. Además, entre mayor sea la dependencia mayor será la ganancia [VQ52]. La dependencia estadística entre un conjunto de variables aleatorias incluye dependencias lineales (correlación), sin embargo no está limitada a este tipo de dependencias. La cuantificación vectorial tiene la característica interesante de que codifica eficientemente a vectores, cuyas componentes tienen dependencias estadísticas no lineales entre sí.

Uno de los resultados fundamentales de la teoría de la codificación de Shannon, es que es más eficiente codificar bloques de datos que codificar datos aislados. Además, entre más dependencia estadística exista entre los datos del bloque más eficiente será este tipo de codificación. La cuantificación vectorial es un tipo de codificación de bloque y esto justifica los resultados experimentales que se han obtenido en trabajos de investigación previos.

Tambien como un resultado de la experiencia, se ha demostrado que entre menos correlacionadas estén un grupo de variables, más eficiente será su cuantificación independiente [VQ52]. No obstante, en este caso aún no existe un teorema general el cual establezca formalemente tal afirmación. Sin embargo, esto se puede aclarar por medio de un ejemplo muy sencillo.

Supongase que una señal sinusoidal está siendo transmitida sobre algún medio. La señal puede ser transmitida como un señal muestreada, en la que las muestras son enviadas secuencialmente. El número de muestras transmitidas depende de la precisión con la que se busca reconstruir la señal. Intuitivamente, entre más muestras se transmitan, mejor será la reconstrucción de la señal. Sin embargo, es bien conocido que todo lo que se requiere para construir una señal senoidal determinística es la magnitud, la fase, la frecuencia, el tiempo de inicio y el hecho de que es una señal senoidal. Esto implica que cinco piezas de información son todo lo que se necesita para reconstruir la señal senoidal exactamente. Desde el punto de vista de la Teoría de la Información, los valores de las muestras de la señal están altamente correlacionados y el contenido de información de ellas es bajo. Por el otro lado, las cinco piezas de información: magnitud, fase, frecuencia, punto de inicio y forma, están completamente descorrelacionadas y tienen exactamente la misma cantidad de información que el número total de muestras. Entonces, la representación de la senoidal en el espacio de los cinco atributos, es más eficiente que la representación en el espacio de las muestras.

Estas consideraciónes acerca de las condiciones bajo las cuales el desempeño de la cuantificación vectorial es más eficiente que el de la cuantificación escalar y que al descorrelacionar un conjunto de variables su cuantificación independiente es más eficiente, lleva a la formulación de dos atributos que hacen que una etapa de procesamiento de señal optimice el desempeño de la cuantificación vectorial. Estos atributos son: [VQ5, VQ22, VQ28].

Atributo 1. *Reducción de la correlación Intervector.*

Las operaciones de procesamiento de señal que reducen la correlación entre los vectores al mínimo nivel permiten a la VQ alcanzar su máximo desempeño mientras se mantienen fijas la dimensión de los vectores y la tasa del código.

Atributo 2. *Preservación de la correlación Intravector.*

Las operaciones que preservan al máximo la correlación de las componentes de cada vector permiten a la VQ alcanzar su máximo desempeño mientras se mantienen fijas la dimensión de los vectores y la tasa del código.

En el momento de elegir entre un conjunto de etapas de procesamiento de la señal la que mejor desempeño pueda proporcionar a la etapa de cuantificación, se puede utilizar como medio de comparación el par de atributos mencionados arriba. Aunque la información necesaria para comparar dos sistemas de procesamiento de la señal, en función del par de atributos, puede obtenerse de la matriz de correlación que indique la dependencia entre todas las variables aleatorias involucradas, el cálculo de tal matriz puede ser muy complicado, además en muchos casos las funciones de densidad de probabilidad conjunta son desconocidas por lo que el cálculo de tal matriz es imposible. Por esta causa es necesario tener una manera más práctica de medir ese par de atributos. En 1993 W. Li [VQ22] definió el factor de acoplamiento intrínseco (ICF) como un indicador del grado de correlación intravector. Este factor se define como

$$F(\boldsymbol{X}) = 1 - \frac{[Det(\boldsymbol{R}_{\boldsymbol{X}})]^{1/D}}{Tr(\boldsymbol{R}_{\boldsymbol{X}})/D}$$

en donde D es la dimensión del vector $\boldsymbol{X}$ y $\boldsymbol{R}_{\boldsymbol{X}}$ es la matriz de covarianza de las componentes del vector. En [VQ52] se muestra que el desempeño de la cuantificación vectorial de un vector $\boldsymbol{X}$ es independiente del sistema coordenado utilizado para describirlo, pero depende de que tan dependientes son los eventos que dan lugar a las componentes del vector. La matriz de covarianza de las componentes del vector es un indicador de esa dependencia, sin embargo, depende del sistema coordenado usado para la representación de los vectores. En cambio, la D-ésima raíz del determinante de dicha matriz carece de ese problema. En diferentes transformaciones la distribución de energía entre los diferentes coeficientes puede no ser igual, es por eso que el ICF se normaliza al valor promedio de la energía de las componentes, esto es, se divide a la raíz del determinante por $\mathrm{Tr}(\boldsymbol{R}_{\boldsymbol{X}})/D$. El ICF es una función que varia entre cero y uno. Cuando es cero indica que las componentes de $\boldsymbol{X}$ son totalmente independientes, cuando es igual a uno las componentes estan totalmente acopladas. Estos hechos se pueden entender de la siguiente forma. Cuando las componentes del vector están muy correlacionadas entre si, todos los elementos en la matriz de covarianza serán muy parecidos, con lo cual el determinante tenderá a cero, y el factor de acoplamiento intrínseco tenderá a uno. En el caso de que las componentes del vector sean independientes idénticamente distribuidas, los elementos de la matriz de covarianza que están fuera de la diagonal principal serán iguales a cero y los que están en ella serán muy parecidos, en consecuencia, la raiz D-ésima del determinante de dicha matriz será muy semejante al valor medio de los elementos de la diagonal, esto es, la traza dividida por D. Esto causará que el factor de acoplamiento intrínseco sea muy cercano a cero.

Bajo estas consideraciones, un esquema de procesamiento de la señal será mejor que otro, en la preservación de la correlación intravector, si el promedio de los ICF de sus coeficientes es mayor.

En el caso de compresión de imágenes, los vectores considerados para la cuantificación vectorial, son arreglos bidimensionales de puntos, esto es, son matrices.

Aunque una caracterización precisa de la correlación intervector para vectores cuadrados permanece en tela de juicio [VQ5], el nivel de compactación de energía entre los diferentes vectores de una señal puede ser utilizado como un indicador de ese atributo [VQ22].

Una vez que se ha determinado cuales son los atributos de una etapa de procesamiento de la señal y como medirlos, en base a ellos se puede hacer una derivación de la forma de la transformación que realice tal etapa.

3.3 Definiciones y propiedades

Matemáticamente, una etapa de procesamiento de la señal puede ser considerada como un mapeo entre dos espacios de referencia diferentes. Bajo este punto de vista es que se fundamenta el uso de esquemas de procesamiento basado en tranformaciones lineales para cumplir con los atributos de optimización de la VQ, establecidos en el capítulo anterior. Además, se explicará la diferencia que existe entre algunas transformaciones lineales, para que en la literatura de procesamiento de señales sean llamadas *transformaciones escalares* o *transformaciones vectoriales.*

Transformación de coordenadas

Sean $(x_1, \ldots, x_N)$ y $(y_1, \ldots, y_N)$ las coordenadas de un mismo punto en dos sistemas de referencia distintos. Supongamos que exiten N relaciones independientes entre las coordenadas anteriores de la forma

$$\begin{aligned} y_1 &= f_1(x_1, \cdots, x_N) \\ &\vdots \\ y_N &= f_N(x_1, \cdots, x_N) \end{aligned}$$

para el caso en que estas relaciones sean lineales, el sistema se puede escribir de la siguiente forma

$$\begin{aligned} y_1 &= a_{11}x_1 + \cdots + a_{1N}x_N \\ &\vdots \qquad \text{lo que en forma matricial es } \mathbf{Y} = \mathbf{AX} \\ y_N &= a_{n1}x_1 + \cdots + a_{NN}x_N \end{aligned}$$

Transformación discreta unidimensional.

En procesamiento digital de señales, una transformada unidimensional y su inversa, usualmente se escriben de la siguiente forma:

$$y_k = \sum_{n=1}^{N} a_{k,n} x_n$$

$$x_n = \sum_{k=1}^{N} a^*_{k,n} y_k$$

en donde $\{a_{k,n}\}$ es el kernel de la transformación y debe satisfacer la siguiente condición de ortogonalidad

$$\sum_{k=1}^{N} a_{k,m} a^*_{k,n} = \begin{cases} 1 & m = n \\ 0 & m \neq n \end{cases}$$

usando notación matricial, este par de transformadas se puede escribir en una forma más compacta como

$$\mathbf{Y} = \mathbf{AX}$$

por lo tanto una transformación lineal discreta puede ser considerada como el paso de un sistema de referencia a otro distinto.

El siguiente ejemplo se utilizará para explicar el hecho de que una transformación lineal discreta puede resultar efectiva, específicamente, en el atributo de la reducción de la correlación para procesar una señal, con el objeto de potenciar la cuantificación. Supóngase un par de variables aleatorias x_1 y x_2 distribuidas como se puede ver en la figura 3.3.1. Por alguna razón, es necesario cuantificarlas independientemente. Si esta cuantificación es realizada directamente, se desperdiciará un conjunto de niveles de representación (figura 3.3.2(a)). En cambio, si la cuantificación se realiza en el sistema de referencia mostrado en la figura 3.3.2(b), el cual es obtenido al rotar los ejes coordenados, entonces los niveles de cuantificación pueden ser distribuidos más eficientemente en el dominio de estas variables aleatorias, lo cual implica una mejor representación de la señal.

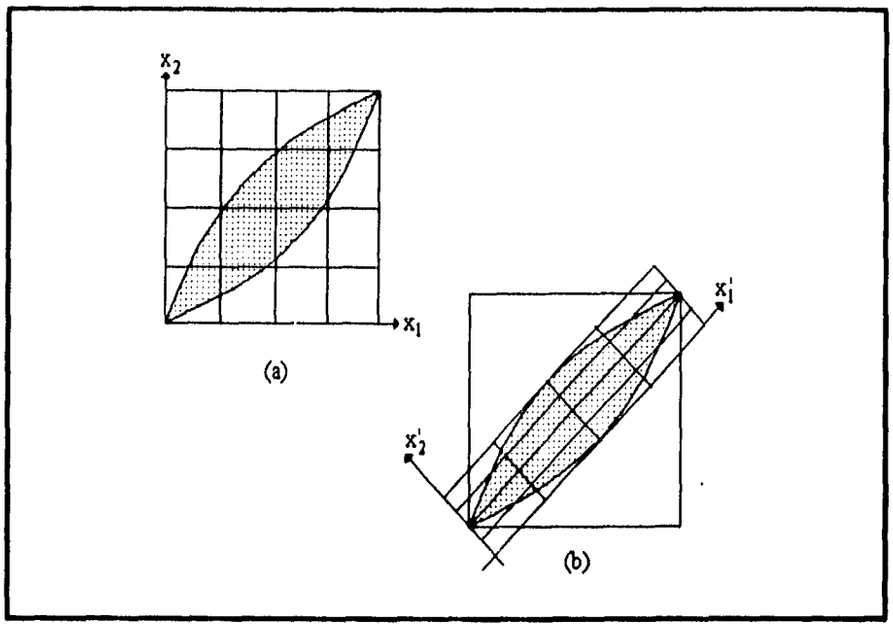


Figura 3.3.1.Cuantificación independiente de las variables aleatorias x_1 y x_2. (a) en el sistema de referencia original. (b) rotando los ejes por medio de una transformación lineal.

Transformación discreta bidimensional.

En procesamiento digital de señales, las señales bidimensionales aparecen muy frecuentemente, entonces, para realizar la operación de descorrelación aparece la necesidad de constuir una transformación lineal bidimensional. Sea $\boldsymbol{X}$ una imagen de (MxN) pixels, una transformada bidimensional y su inversa, usualmente se escriben de la siguiente forma:

$$y_{k,l} = \sum_{m=1}^{M}\sum_{n=1}^{N} V_{m,l} x_{n,m} H_{n,k}$$

$$x_{n,m} = \sum_{l=0}^{M}\sum_{k=0}^{N} V_{m,l} y_{k,l} H_{n,k}$$

en forma matricial $\mathbf{Y} = \mathbf{V}\mathbf{X}\mathbf{H}^T$ $\quad \mathbf{X} = \mathbf{V}^T\mathbf{X}\mathbf{H}$

en donde $\boldsymbol{V}$ y $\boldsymbol{H}$ son matrices reales de dimensión (MxM) y (NxN) respectivamente.

Las transformaciones lineales, unidimensional y bidimensional descritas hasta ahora, son conocidas en el área de procesamiento digital de señales como *transformaciones escalares*. Esto se debe a que las funciones de descorrelación y compactación de la energía, que se busca realizar con este tipo de transformaciones, operan entre cada uno de los puntos de la señal original. Esto es, cada uno de los puntos de la señal original se hace, en el dominio transformado, independiente de los

restantes. Sin embargo, como se vio anteriormente, en el caso de sistemas de codificación en los que se utilizará la cuantificación vectorial, la descorrelación de las componentes de los vectores no es deseable. Esto es, la descorrelación de los puntos de la imagen se debe realizar únicamente entre puntos pertenecientes a distintos vectores de la etapa de cuantificación. Es bajo esta premisa que aparece un tipo de transformaciones lineales con propiedades de descorrelación, orientada a bloques de puntos, en lugar de estar dirigida en una forma punto a punto [VQ28]. Este tipo especial de transformaciones lineales escalares es lo que se conoce, en el contexto del área de procesamiento de señales digitales, como *transformaciones vectoriales*. Las transformaciones vectoriales tienen una estructura especial en el kernel que les permite lograr la descorrelación entre las muestras de la señal pertenecientes a diferentes vectores, mientras que se preserva la correlación entre las muestras dentro de un mismo vector. La construcción de una señal vectorial a partir de un conjunto de muestras, se logra particionando la señal en bloques de tamaño uniforme. En estos bloques, las diferentes muestras pueden ser vistas como las componentes del vector. Para imágenes, los vectores pueden ser formados en base a bloques rectangulares de muestras, los cuales corresponden a subimágenes (comúnmente llamadas vectores). Así, agrupando muestras a partir de una señal escalar, se puede obtener una señal en la cual las muestras son vectores.

Transformación Vectorial 1D

Para aclarar el concepto de las transformaciones vectoriales se utilizará un ejemplo sencillo. Supóngase que se tiene un conjunto de muestras, $(x_1, \ldots, x_{2N})$, el cual va a ser procesado por un sistema de compresión en el que se usa una etapa de cuantificación vectorial con vectores de dimensión dos. Al segmentar la señal en bloques de dos elementos, se obtiene la señal vectorial $(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_N)$. Para que la cuantificación vectorial se desempeñe lo mejor posible, es deseable preservar la correlación entre las componentes de los vectores, sin embargo, como los distintos vectores en los que se particiona la señal original serán cuantificados independientemente, es deseable descorrelacionar entre sí las muestras pertenecientes a distintos vectores. La transformación lineal que operará sobre el conjunto de muestras puede escribirse como

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_{2N-1} \\ y_{2N} \end{bmatrix} = \begin{bmatrix} a_{1,1} & \cdots & a_{1,2N} \\ a_{2,1} & \cdots & a_{2,2N} \\ \vdots & \cdots & \vdots \\ a_{2N-1,1} & \cdots & a_{2N-1,2N} \\ a_{2N,1} & \cdots & a_{2N,2N} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_{2N-1} \\ x_{2N} \end{bmatrix}$$

pero considerando la partición en vectores, la transformación de arriba se puede poner de la siguiente forma

$$\begin{bmatrix} \boldsymbol{Y}_1 \\ \vdots \\ \boldsymbol{Y}_N \end{bmatrix} = \begin{bmatrix} \boldsymbol{A}_{1,1} & \cdots & \boldsymbol{A}_{1,N} \\ \vdots & \cdots & \vdots \\ \boldsymbol{A}_{N,1} & \cdots & \boldsymbol{A}_{N,N} \end{bmatrix} \begin{bmatrix} \boldsymbol{X}_1 \\ \vdots \\ \boldsymbol{X}_N \end{bmatrix} \qquad [3.3.1]$$

en donde las matrices $\boldsymbol{A}_{i,j}$ están dadas por

$$\boldsymbol{A}_{i,j} = \begin{bmatrix} a_{2i-1,2j-1} & a_{2i,2j-1} \\ a_{2i-1,2j} & a_{2i,2j} \end{bmatrix}$$

Al expresar la transformación por medio de (3.3.1) se puede ver más claramente la forma que el kernel debe tener para preservar la correlación entre las componentes de los vectores y eliminar la correlación entre muestras pertenecientes a distintos vectores. Aquí, para lograr la reducción de la correlación entre los $\boldsymbol{Y}_i$, las matrices $\boldsymbol{A}_{i,j}$ deberán tener, entre sí, algún tipo de relación como el que tienen los coeficientes a_{ij} en el caso de una *transformación escalar*. Además, la estructura interna de dichas matrices deberá de ser tal que la correlación entre las componentes de los vectores sea preservada. En resumen, la *transformación vectorial unidimensional* no es más que una transformación lineal, pero con una estructura particular en el kernel que le permite cumplir con los dos atributos considerados como necesarios para la optimización de la cuantificación vectorial.

Para introducir una notación más acorde con la segmentación de la señal en vectores, se redefinirá la transformación lineal y la propiedad de ortogonalidad que el kernel debe cumplir, de la siguiente forma.

Una transformación vectorial unidimensional discreta (1D-VT) de tamaño N y su inversa se definen como [VQ28]

$$\boldsymbol{Y}_k = \sum_{n=1}^{N} \boldsymbol{A}_{k,n} \boldsymbol{X}_n$$

$$\boldsymbol{X}_n = \sum_{k=1}^{N} \boldsymbol{A}^{*}_{k,n} \boldsymbol{Y}_k$$

en donde $\boldsymbol{Y}_k$ y $\boldsymbol{X}_n$ son vectores columna de dimensión M, $\boldsymbol{A}_{k,n}$ son un conjunto de (NxN) matrices de dimensión (MxM) que forman el kernel $\{\boldsymbol{A}_{k,n}\}$ de la transformación, el cual debe satisfacer la siguiente condición de ortogonalidad

$$\sum_{k=1}^{N} A_{k,m} A^{*}_{k,n} = \sum_{k=1}^{N} A^{*}_{k,n} A_{k,m} = \begin{cases} I & m = n \\ 0 & m \neq n \end{cases}$$

en donde I y 0 son las matrices identidad multiplicativa y aditiva, respectivamente. Usando notación matricial, el par de transformadas puede escribirse como

$$Y = AX$$
$$X = A^{*}Y$$

en donde hay que notar que los elementos de A son las matrices $A_{k,n}$.

Transformacion Vectorial 2D

De la misma forma que se utilizó en la formulación de la transformación vectorial unidimensional, se puede llegar a establecer una transformada vectorial para el caso de señales bidimensionales. En este caso, la transformada vectorial es una transformada lineal bidimensional cuyo kernel tiene una estructura tal que le permite preservar la correlación entre las componentes de los vectores, pero elimina la correlación entre muestras pertenecientes a distintos vectores. Una transformada vectorial bidimensional discreta (2D-VT) y su inversa se definen como[VQ22]

$$Y_{k,l} = \sum_{m=1}^{N} \sum_{n=1}^{N} V_{m,l} X_{n,m} H_{n,k}$$

$$X_{n,m} = \sum_{l=1}^{N} \sum_{k=1}^{N} V^{*}_{m,l} Y_{k,l} H^{*}_{n,k}$$

en donde $\{H_{n,k}\}$ y $\{V_{m,l}\}$ son dos conjuntos de NxN matrices, cada una de dimensión MxM y $\{X_{k,l}\}$ y $\{Y_{k,l}\}$ son dos conjuntos de NxN vectores cuadrados, cada uno de dimensión MxM, en el dominio de la señal y transformado respectivamente.

En el caso en que se utilice cuantifición escalar, se puede llegar a la construcción de una *transformación escalar* óptima, en base a una formulación basada en que la matriz de correlación de los coeficientes de la transformación debe ser una matriz diagonal, esto es, $E[x_i x_j]=0$ para $i \neq j$, lo cual lleva a la obtención de la transformada Karhunen-Loève. En el caso de usar cuantificación vectorial, se puede hacer una formulación similar, pero aquí es indispensable introducir dos condiciones en lugar de una. En primer lugar se desea reducir la correlación intervector y en segundo lugar se desea preservar la correlación intravector al nivel original. Esto se escribe como

$$E\left[\boldsymbol{Y}_l\boldsymbol{Y}_k^T\right]=\begin{cases}\boldsymbol{0} & l\neq k\\ \sum_{m=1}^{N}\sum_{n=1}^{N}A_{m,l}E\left\{\boldsymbol{X}_m\boldsymbol{X}_n^T\right\}A_{n,k}^T & l=k\end{cases}$$

[3.2.1]

en el caso de vectores columna y

$$E\left[\boldsymbol{Y}_{k,l}\boldsymbol{Y}_{k',l'}^T\right]=\begin{cases}\boldsymbol{0} & l\neq l' \quad k\neq k'\\ \sum_{m=1}^{N}\sum_{n=1}^{N}\sum_{m'=1}^{N}\sum_{n'=1}^{N}E\left\{V_{m,l}\boldsymbol{X}_{n,m}\boldsymbol{H}_{n,k}V_{m',l'}\boldsymbol{X}_{n',m'}\boldsymbol{H}_{n',k'}\right\} & l=l' \quad k=k'\end{cases}$$

[3.2.2]

cuando se trata de vectores cuadrados.

En el caso de vectores columna, con estas condiciones y conociendo el modelo del proceso aleatorio que describe al vector $\boldsymbol{X}$, se puede obtener el kernel $\boldsymbol{V}$, de la transformación óptima para maximizar el desempeño de la VQ. Cuando se trata con vectores cuadrados hay que imponer restricciones adicionales al kernel (matrices $\boldsymbol{V}$ y $\boldsymbol{H}$) para poder poner a $E\{\boldsymbol{X}_{k,l}\boldsymbol{X}_{k',l'}\}$ en función de $E\{\boldsymbol{x}_{n,m}\boldsymbol{x}_{n',m'}\}$. En este caso no siempre es posible encontrar el kernel de la transformación óptima. Por ejemplo, para el modelo de Markov de primer orden y la condición 3.2.2 no es posible encontrar a $\boldsymbol{V}$ y $\boldsymbol{H}$ [VQ5].

Familias de transformaciones Vectoriales

Anteriormente se mencionó que el kernel de una transformación vectorial debe satisfacer la condición de ortogonalidad. Sea $\boldsymbol{T}$ una matriz unitaria, esto es $\boldsymbol{TT^*=T^*T=I}$, entonces

$$\left.\begin{aligned}\boldsymbol{TIT^*}&=\boldsymbol{I}\\ \boldsymbol{T0T^*}&=\boldsymbol{0}\end{aligned}\right\}=\boldsymbol{T}\sum_{k=1}^{N}\boldsymbol{A}_{k,m}\boldsymbol{A}_{k,n}^{*}\boldsymbol{T}^{*}=\boldsymbol{T}\sum_{k=1}^{N}\boldsymbol{A}_{k,m}\boldsymbol{T}^{*}\boldsymbol{T}\boldsymbol{A}_{k,n}^{*}\boldsymbol{T}^{*}=\sum_{k=1}^{N}(\boldsymbol{T}\boldsymbol{A}_{k,m}\boldsymbol{T}^{*})(\boldsymbol{T}\boldsymbol{A}_{k,n}\boldsymbol{T}^{*})^{*}$$

por lo que se puede concluir que, si $\{\boldsymbol{V}_{k,n}\}$ es un kernel y $\boldsymbol{T}$ es una matriz unitaria, entonces $\{\boldsymbol{TV}_{k,n}\boldsymbol{T^*}\}$ también es un kernel. Este hecho puede ser utilizado para construir familias de transformaciones vectoriales, esto quiere decir que usando una VT como semilla y un conjunto de matrices unitarias distintas, se puede obtener un conjuto de transformaciones vectoriales diferentes.

3.4 La transformada coseno discreta

La transformada coseno discreta (DCT) es actualmente la transformación lineal más ámpliamente utilizada en codificación de transformada. Esta transformación tiene un conjunto de funciones base independientes de la señal a transformar.

Además, tiene una buena capacidad de reducción de correlación y de compactación de la energía. La DCT está muy relacionada con la transformada discreta de Fourier (DFT). En particular, una DCT de NxN puede ser expresada en términos de la DFT $2N$x$2N$ de su extensión simétrica par, lo cual lleva a la existencia de algoritmos rápidos de cálculo, basados en la transformada rápida de Fourier (FFT), los cuales hacen posible calcular la DCT de NxN elementos por medio de $O(2N^2\log_2 N)$ operaciones en lugar de $O(N^4)$. La familia de la DCT está constituida por cuatro elementos, cuyas transformadas inversas y directas se definen de la siguiente forma

- DCT-I

$$y_m = \left(\frac{2}{N}\right)^{1/2} k_m \sum_{n=0}^{N} k_n x_n \cos\left(\frac{mn\pi}{N}\right) \quad m = 0 \cdots N$$

$$x_m = \left(\frac{2}{N}\right)^{1/2} k_n \sum_{n=0}^{N} k_m y_m \cos\left(\frac{mn\pi}{N}\right) \quad n = 0 \cdots N$$

$$k_p = \begin{cases} \frac{1}{\sqrt{2}} & p = 0, N \\ 1 & p \neq 0, N \end{cases}$$

- DCT-II

$$y_m = \left(\frac{2}{N}\right)^{1/2} k_m \sum_{n=0}^{N-1} k_n x_n \cos\left[\frac{(2n+1)m\pi}{2N}\right] \quad m = 0 \cdots N-1$$

$$x_m = \left(\frac{2}{N}\right)^{1/2} k_n \sum_{n=0}^{N-1} k_m y_m \cos\left[\frac{(2n+1)m\pi}{2N}\right] \quad n = 0 \cdots N-1$$

$$k_p = \begin{cases} \frac{1}{\sqrt{2}} & p = 0, N-1 \\ 1 & p \neq 0, N-1 \end{cases}$$

- DCT-III

$$y_m = \left(\frac{2}{N}\right)^{1/2} k_m \sum_{n=0}^{N-1} k_n x_n \cos\left[\frac{(2m+1)n\pi}{2N}\right] \quad m = 0 \cdots N-1$$

$$x_m = \left(\frac{2}{N}\right)^{1/2} k_n \sum_{n=0}^{N-1} k_m y_m \cos\left[\frac{(2m+1)n\pi}{2N}\right] \quad n = 0 \cdots N-1$$

$$k_p = \begin{cases} \frac{1}{\sqrt{2}} & p = 0, N-1 \\ 1 & p \neq 0, N-1 \end{cases}$$

- DCT-IV

$$y_m = \left(\frac{2}{N}\right)^{1/2} \sum_{n=0}^{N-1} k_n x_n \cos\left[\frac{(2m+1)(2n+1)n\pi}{4N}\right] \quad m = 0 \cdots N-1$$

$$x_m = \left(\frac{2}{N}\right)^{1/2} \sum_{n=0}^{N-1} k_m y_m \cos\left[\frac{(2m+1)(2n+1)n\pi}{4N}\right] \quad n = 0 \cdots N-1$$

$$k_p = \begin{cases} \frac{1}{\sqrt{2}} & p = 0, N-1 \\ 1 & p \neq 0, N-1 \end{cases}$$

Como es bien conocido, la transformada Karhunen-Loève (KLT) es una transformación óptima debido a las siguientes propiedades [VQ]:

- Descorrelaciona la señal completamente en el dominio transformado.
- Minimiza el error cuadrático medio en la compresión de datos.
- Concentra la mayor varianza (energía) en el menor número de coeficientes transformados.
- Minimiza la entropía total de representación de la secuencia.

La construcción práctica de la KLT involucra la estimación de la matriz de autocovarianza de la secuencia de datos, su diagonalización y la construcción de los vectores base. La incapacidad de predeterminar los vectores base en el dominio transformado, hace que la KLT no sea una herramienta práctia, aún siendo ideal.

Las transformadas coseno discretas I y II aparecen como dos transformaciónes con funciones base fijas, que son buenas aproximaciones para la KLT. De hecho, dado un proceso de Markov de primer orden, la DCT-I y la DCTII son asintóticamente equivalentes a la KLT, conforme N tiende a infinito, o ρ del proceso de Markov tiende a uno, respectivamente. Como la DCT-III es la transpuesta de la DCT-II, esta transformada también tiene el mismo parecido asintótico con la KLT cuando ρ tiende a uno. A continuación se presentarán algunas de las propiedades más importantes de la familia de la DCT.

Ortonormalidad

Aquí, recordaremos la definición de un producto interior de dos vectores reales N-dimensionales g_k y g_m:

$$\langle g_k, g_m \rangle = \sum_{n=1}^{N} g_{n,k}\, g_{n,m}$$

en donde

$$g_k = \left[g_{1,k}, \cdots, g_{N,k}\right]^T \quad g_m = \left[g_{1,m}, \cdots, g_{N,m}\right]^T$$

se dice que g_k y g_m son ortogonales si su producto interior es cero para k distinta de m. Además, si $\langle g_k, g_m \rangle = 1$, se dice que los vectores están normalizados, y la matriz de transformación real correspondiente es unitaria. En base a la definición de producto interno, se puede demostrar que las matrices de las transformaciones que forman la familia de la DCT son ortonormales [VQ61]. La propiedad de ortonormalidad es importante, porque si una transformación tiene esta propiedad entonces la energía y la información de la señal son preservadas bajo la transformación.

Separabilidad

Separabilidad significa que una transformación multidimensional puede ser construida por medio de una serie de transformaciones unidimensionales. Las transformaciones de la familia de la DCT tienen esta propiedad, cuyo beneficio radica en que si se desarrolla un algoritmo rápido para una 1D-DCT, entonces este algoritmo se puede aplicar para construir DCTs multidimensionales.

Escalamiento en tiempo

Ya que las DCTs tratan con puntos de muestreo discretos, un escalamiento en tiempo no tiene efecto en la transformada, excepto en un cambio en las unidades del intervalo de frecuencia en el dominio transformado. Así, si δt cambia a $a\delta t$, entonces δf cambia a $\delta f a$, habiendo considerado que N permanece constante. Así, un escalamiento en tiempo lleva a un escalamiento en frecuencia, sin un escalamiento de la transformación. La siguientes ecuaciones pueden ser utilizadas para interpretar la resolución en frecuencia

$$\delta f \delta t = \left(\frac{1}{2N}\right)$$

$$T = N\delta t$$

$$\delta f = \left(\frac{1}{2T}\right)$$

en donde T es la duración en tiempo para la secuencia de longitud N.

Corrimiento en tiempo

Sea X y $X+$ dos vectores de muestras de dimensión $N+1$, dados por las siguientes expresiones

$$X = [x_0, \cdots, x_N]^T$$

$$X+ = [x_1, \cdots, x_{N+1}]^T$$

entonces la transformada de X+ se puede relacionar con la transfomada de X, por medio de las siguientes expresiones

- DCT-I

$$y+_m = \cos\left(\frac{m\pi}{n}\right) y_m + k_m \operatorname{sen}\left(\frac{m\pi}{N}\right)\left(\frac{2}{N}\right)^{1/2} \sum_{n=1}^{N-1} \operatorname{sen}\left(\frac{mn\pi}{N}\right) x_n + \left(\frac{2}{N}\right)^{1/2} k_m \times$$

$$\times \left\{\left(\frac{-1}{\sqrt{2}}\right)\cos\left(\frac{m\pi}{N}\right) x_0 + \left(\frac{1}{\sqrt{2}} - 1\right) x_1 + (-1)^m \left(1 - \frac{1}{\sqrt{2}}\right)\cos\left(\frac{m\pi}{N}\right) x_N + (-1)^m \frac{x_{N+1}}{\sqrt{2}}\right\}$$

- DCT-II

$$y_{+m} = \cos\left(\frac{m\pi}{n}\right) y_m + k_m \operatorname{sen}\left(\frac{m\pi}{N}\right)\left(\frac{2}{N}\right)^{1/2} k_n \sum_{n=1}^{N-1} \operatorname{sen}\left(\frac{m(n-1/2)\pi}{N}\right) x_n + \left(\frac{2}{N}\right)^{1/2} k_m \times$$

$$\times\left\{(-1)^m x_N - x_0\right\}\cos\left(\frac{m\pi}{2N}\right)$$

- DCT-III

$$y_{+m} = \cos\left(\frac{(2m+1)\pi}{2N}\right) y_m + \operatorname{sen}\left(\frac{(2m+1)\pi}{2N}\right)\left(\frac{2}{N}\right)^{1/2} \sum_{n=1}^{N-1} k_n \operatorname{sen}\left(\frac{(m+1/2)n\pi}{N}\right) x_n +$$

$$+\left(\frac{2}{N}\right)^{1/2}\left\{\left(\frac{1}{\sqrt{2}}-1\right)x_1 - \cos\left(\frac{(2m+1)\pi}{2N}\right)\frac{x_0}{\sqrt{2}} + (-1)^m\left(1-\frac{1}{\sqrt{2}}\right)\operatorname{sen}\left(\frac{(2m+1)\pi}{2N}\right)x_N\right\}$$

- DCT-IV

$$y_{+m} = \cos\left(\frac{(2m+1)\pi}{2N}\right) y_m + \operatorname{sen}\left(\frac{(2m+1)\pi}{2N}\right)\left(\frac{2}{N}\right)^{1/2} \sum_{n=1}^{N-1} \operatorname{sen}\left(\frac{(m+1/2)(n+1/2)\pi}{N}\right) x_n +$$

$$+\left(\frac{2}{N}\right)^{1/2}\left\{-\cos\left(\frac{(2m+1)\pi}{4N}\right)x_0 + (-1)^m \operatorname{sen}\left(\frac{(2m+1)\pi}{4N}\right)x_N\right\}$$

como se puede observar en las ecuaciones anteriores, la propiedad de corrimiento en el tiempo produce en las transformadas de la familia de la DCT, ecuaciones bastante complicadas. Sin embargo, se debe notar que cuando las DCTs instantaneas deben calcularse en un flujo continuo de muestras, estas ecuaciones representan una forma posible de actualización de la transformada, sin tener que calcularla completamente a cada instante.

Convolución

Para secuencias discretas finitas x_n e y_n, se definen dos tipos de convolución. La convolución circular a_n se define para secuencias que son periódicas con periodo de N. a_n está dada por

$$a_n = x_n * y_n = \frac{1}{\sqrt{N}} \sum_{m=0}^{N-1} x_m y_{n-m} \quad n = 0, 1, \cdots, N-1$$

La convolución lineal b_n para secuencias no periódicas x_n e y_n de longitudes L y M, respectivamente, se define como

$$h_n = x_n * y_n = \frac{1}{\sqrt{N}} \sum_{m=0}^{N-1} x_m y_{n-m} \quad n = 0,1,\cdots,N-1 \text{ donde } N \geq L + M - 1$$

Aumentando x_n e y_n con ceros para completar longitudes iguales de tamaño N, bi puede obtenerse como una porción de una convolución circular de las secuencias aumentadas.

El producto de las transformadas DCT de dos secuencias es la transformada DCT de la convolución circular de la versión extendida par de esas dos secuencias con una tercera h_n dada por

$$h_n = 2\left\{\left(\frac{1}{2\sqrt{2}} - 1\right) + exp\left[\frac{j(2n-1)(N-1)\pi}{4N}\right] \frac{\text{sen}\left[\frac{(2n-1)\pi}{4}\right]}{\text{sen}\left[\frac{(2n-1)\pi}{4N}\right]}\right\} / \sqrt{2N}$$

Distribución de varianza

En una señal en el dominio transformado es deseable tener unos pocos coeficientes con varianzas grandes. De esta forma los coeficientes con varianzas pequeñas pueden ser descartados en el proceso de reconstrucción, produciendo un MSE despreciable, entre la señal original y la reconstruida. La distribución de varianza es una característica muy importante cuando una transformada va a ser utilizada para compresión de datos. La distribución de varianza de la DCT-II es muy cercana a la de la KLT, esto ha causado que sea ámpliamete utilizada en sistemas de compresión. La distribución de varianza está muy relacionada con la eficiencia de compactación de la energía, la cual está definida como la porción de energía contenida en los primeros M de N coeficientes transformados y que puede ser calculada por medio de [VQ61]

$$EPE(m) = \frac{\sum_{p=0}^{M-1} E\left[y_p^2\right]}{\sum_{p=0}^{N-1} E\left[y_p^2\right]}$$

Correlación Residual

Una medida del desempeño de una transformación discreta, es el grado de reducción de la correlación que puede lograr entre los elementos de una secuencia dada. Como se ha mencionado, la KLT es una transformación óptima para un proceso de Markov de primer orden. La DCT-II tiene un factor de reducción de

correlación muy cercano al de la KLT para una ámplia variedad de señales. Sin embargo, para secuencias poco correlacionadas la DCT-I se desempeña mejor que la DCT-II.

Máxima reducción de bits

Otra medida del desempeño para una transformación es en términos de la reducción de la cantidad de bits requeridos para representar los coeficientes en el dominio transformado, dada por

$$mrb = -\frac{1}{2N}\sum_{j=0}^{N-1}\log_2\left[E\left\{\boldsymbol{Y}\boldsymbol{Y}^T\right\}_{j,j}\right]$$

en donde la notación $E\{\boldsymbol{Y}\boldsymbol{Y}^T\}_{i,j}$ se utiliza para indicar el elemento (i,j) de la matriz de covarianza del vector $\boldsymbol{Y}$.

Entre más grande sea el *mrb*, mejor será el desempeño de la transformación, en el sentido de la reducción de bits. Bajo esta medida de comparación, la DCT-II tiene un desempeño muy cercano al que tiene la KLT [VQ61].

3.5 La transfomada coseno discreta bidimensional

En codificación de imágenes por medio de transformadas, una imagen de *LxL* puntos generalmente se divide en bloques de tamaño *NxN*. En general, tamaños de bloque de 8x8 y 16x16 son frecuentemente utilizados en codificación de imágenes. Esos tamaños de bloque pequeños permiten introducir características adaptivas basadas en las características particulares del bloque (nivel de detalle, energía, etc). La complejidad, cantidad de memoria y tamaño de los circuitos, también se reduce considerablemente en comparación con la transformación completa de la imagen. Aún cuando existe algun grado de correlación entre bloques vecinos, generalmente está no es lo suficientemente grande como para justificar el incremento de complejidad causado por utilizar bloques de tamaño mayor a 16x16 pixels. Sin embargo, este tipo de partición en bloques tiene la desventaja de que en la codificación de imágenes a tasas muy bajas, puede presentar perturbaciones muy notables por la diferencia entre el tono promedio o el nivel de ruido en bloques vecinos.

Después del mapeo del bloque de *NxN* puntos en el dominio transformado, se puede descartar en base a las características de la imagen, un conjunto de coeficientes. Esto puede ser hecho de forma fija o adaptable. Aunque en el caso de que los coeficientes sean descartados de forma adaptable, se requiere de una cantidad adicional de información colateral, globalmente se puede obtener una

correlación muy cercano al de la KLT para una ámplia variedad de señales. Sin embargo, para secuencias poco correlacionadas la DCT-I se desempeña mejor que la DCT-II.

Máxima reducción de bits

Otra medida del desempeño para una transformación es en términos de la reducción de la cantidad de bits requeridos para representar los coeficientes en el dominio transformado, dada por

$$mrb = -\frac{1}{2N}\sum_{j=0}^{N-1}\log_2\left[E\left\{\boldsymbol{Y}\boldsymbol{Y}^T\right\}_{j,j}\right]$$

en donde la notación $E\{\boldsymbol{Y}\boldsymbol{Y}^T\}_{i,j}$ se utiliza para indicar el elemento (i,j) de la matriz de covarianza del vector $\boldsymbol{Y}$.

Entre más grande sea el *mrb*, mejor será el desempeño de la transformación, en el sentido de la reducción de bits. Bajo esta medida de comparación, la DCT-II tiene un desempeño muy cercano al que tiene la KLT [VQ61].

3.5 La transfomada coseno discreta bidimensional

En codificación de imágenes por medio de transformadas, una imagen de *L*x*L* puntos generalmente se divide en bloques de tamaño *N*x*N*. En general, tamaños de bloque de 8x8 y 16x16 son frecuentemente utilizados en codificación de imágenes. Esos tamaños de bloque pequeños permiten introducir características adaptivas basadas en las características particulares del bloque (nivel de detalle, energía, etc). La complejidad, cantidad de memoria y tamaño de los circuitos, también se reduce considerablemente en comparación con la transformación completa de la imagen. Aún cuando existe algun grado de correlación entre bloques vecinos, generalmente está no es lo suficientemente grande como para justificar el incremento de complejidad causado por utilizar bloques de tamaño mayor a 16x16 pixels. Sin embargo, este tipo de partición en bloques tiene la desventaja de que en la codificación de imágenes a tasas muy bajas, puede presentar perturbaciones muy notables por la diferencia entre el tono promedio o el nivel de ruido en bloques vecinos.

Después del mapeo del bloque de *N*x*N* puntos en el dominio transformado, se puede descartar en base a las características de la imagen, un conjunto de coeficientes. Esto puede ser hecho de forma fija o adaptable. Aunque en el caso de que los coeficientes sean descartados de forma adaptable, se requiere de una cantidad adicional de información colateral, globalmente se puede obtener una

reducción de la tasa, además de que la calidad subjetiva de la imagen recuperada puede aumentar.

Como se mencionó anteriormente, la DCT se aproxima, en base a un conjunto de criterios, a la transformación estadística óptima, la KLT. Esto ha causado que la 2D-DCT-II sea una de las transformaciones más ámpliamente utilizadas, en procesamiento de señal, incluyendo codificación de imágenes. La 2D-DCT-II está definida de la siguiente forma

$$y_{u,v} = \frac{4C(u)C(v)}{N^2} \sum_{m=0}^{N-1} \sum_{n=0}^{N-1} x_{m,n} \cos\left[\frac{(2m+1)u\pi}{2N}\right] \cos\left[\frac{(2n+1)v\pi}{2N}\right] \quad \begin{matrix} u = 0,\ldots,N-1 \\ v = 0,\ldots,N-1 \end{matrix}$$

$$x_{m,n} = \sum_{u=0}^{N-1} \sum_{v=0}^{N-1} C(u)C(v) y_{u,v} \cos\left[\frac{(2m+1)u\pi}{2N}\right] \cos\left[\frac{(2n+1)v\pi}{2N}\right]$$

en donde

$$C(k) = \begin{cases} \frac{1}{\sqrt{2}} & k = 0 \\ 1 & k \neq 0 \end{cases}$$

en estas ecuaciones, el factor de normalización $4/N^2$ aparece completamente en la transformada directa. Aunque este factor puede ser dividido equitativamente entre las dos transformaciones, o movido enteramente a la transformada inversa, factores relacionados con la realización en *hardware*, tales como escalamiento, desbordamiento, etc. son los que determinan en donde es conveniente localizarlo.

En varios sistemas de codificación de transformada basados en la 2D-DCT, se han seguido distintas estrategias en la etapa de cuantificación. Generalmente esta etapa esta dividida en dos partes. La primera, en la cual se utiliza cuantificación escalar, aplicada al coeficiente de *DC* y a los coeficientes de *AC* más relevantes, complementada por modulación por codificación diferencial de pulsos (DPCM). En la segunda se utiliza cuantificación vectorial, en la mayoría de los casos de tipo clasificado, para codificar los coeficientes menos relevantes de la imagen. Por otro lado, se han propuesto esquemas en los cuales se pretende utilizar un conjunto de cuantificadores vectoriales de tamaño uniforme para construir la etapa de cuantificación. Un esquema propuesto en [VQ70] consiste en transformar la imagen en bloques de 8x8 elementos y entonces agrupar los coeficientes de igual índice juntos para formar vectores de tamaño uniforme (transformada bloque coseno). Este esquema ha demostrado ser poco práctico, ya que para obtener una buena calidad de reconstrucción, se requiere que para algunos de los coeficientes los libros de código tengan tamaños muy grandes. Recientemente se han propuesto nuevos esquemas de codificación por transformada, los cuales utilizan transformaciones vectoriales en el

sentido que se explicó en la parte anterior de este capítulo, para dar un procesamiento de la señal más adecuado a las necesidades de una etapa de cuantificación vectorial, la cual opera sobre vectores de tamaño uniforme sobre el bloque de coeficientes. Esto da lugar a la aparición de la transformada coseno vectorial que se describe en la siguiente sección.

3.6 La transformada coseno discreta vectorial.

La transformada coseno discreta vectorial (2D-VDCT) propuesta por W. Li en [VQ22], aparece como una etapa de procesamiento que pretende emular algunas de las propiedades de la transformada coseno, mientras que al mismo tiempo trata de cumplir con los dos atributos que una etapa de procesamiento de la señal debe tener para potenciar al máximo la cuantificación vectorial. La 2D-VDCT está definida de la siguiente forma:

Sea $\{x_{n,m}\}$ un conjunto de NxN vectores cuadrados (matrices), cada uno de dimensión MxM, los cuales forman la imagen original y sea $\{X_{k,l}\}$ el conjunto de NxN coeficientes, cada uno de dimensión MxM, que forman la imagen en el dominio transformado. Estos dos conjuntos están relacionados por medio de

$$X_{k,l} = \sum_{m=0}^{N-1}\sum_{n=0}^{N-1} V_{m,l}\, x_{n,m}\, H_{n,k}$$

$$x_{n,m} = \sum_{l=0}^{N-1}\sum_{k=0}^{N-1} V_{m,l}^{T}\, X_{k,l}\, H_{n,k}^{T}$$

en donde los conjuntos de NxN matrices, cada una de dimensión MxM, $\{V_{m,l}\}$ y $\{H_{n,k}\}$, están dados por

$$V_{m,l} = \sqrt{\frac{2}{N}}\begin{bmatrix} \cos\left(\frac{(2m+1)l\pi}{2N}\right) & 0 & \cdots & 0 \\ 0 & \cos\left(\frac{(2m+1)l\pi}{2N}\right) & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & \cos\left(\frac{(2m+1)l\pi}{2N}\right) \end{bmatrix} = \sqrt{\frac{2}{N}}\cos\left(\frac{(2m+1)l\pi}{2N}\right) I$$

$$H_{m,l} = \sqrt{\frac{2}{N}}\begin{bmatrix} \cos\left(\frac{(2n+1)k\pi}{2N}\right) & 0 & \cdots & 0 \\ 0 & \cos\left(\frac{(2n+1)k\pi}{2N}\right) & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & \cos\left(\frac{(2n+1)k\pi}{2N}\right) \end{bmatrix} = \sqrt{\frac{2}{N}}\cos\left(\frac{(2n+1)k\pi}{2N}\right) I$$

en donde

$$V_{m,0} = H_{m,0} = \sqrt{\frac{1}{N}} I$$

En los resultados mostrados en [VQ5, VQ22, VQ23, VQ28, VQ58]. se puede comprobar que la 2D-VDCT, es superior en el atributo de la preservación de la correlación entre las componentes de los vectores comparada con la 2D-DCT, mientras que tiene un comportamiento similar al de esta última transformación, en el atributo de compactación de la energía. El cual está relacionado con el nivel de reducción de la correlación residual.

El uso de la 2D-VDCT en esquemas de codificación por transformada se reporta en [VQ5], en donde se presenta un sistema de compresión por asignación dinámica de bits, el cual da mejores resultados en términos tasa-SNR, que el esquema de compresión JPEG. En [VQ22] se presenta un sistema por asignación fija de bits el cual sirve como marco de comparación entre la 2D-VDCT y dos esquemas basados en la 2D-DCT. De nuevo, el esquema basado en la 2D-VDCT presenta el mejor desempeño.

Capítulo 4

Cuantificación Vectorial de fuentes Gaussianas y Laplacianas en el dominio de la 2D-VDCT

4.1 Introducción

Los vectores en el dominio de la 2D-VDCT tienen distribuciones que pueden ser modeladas por procesos: Gaussiano para el coeficiente de *DC* y Laplaciano para los demás coeficientes. Además estos vectores tienen la peculiaridad de tener componentes altamente correlacionadas. La calidad en su codificación, usando el cuantificador vectorial del vecino más cercano, generalmente dista mucho del límite superior impuesto por la teoría de la información y particularmente esta distancia se acentúa cuando la codificación se realiza a tasas bajas o moderadas (menos de 1 bit por dimensión) [VQ22, VQ28, VQ58]. Esto motiva la búsqueda de nuevas formas de codificación para vectores con estas características. De una manera algo informal se observó que si se divide el espacio vectorial de uno de los coeficientes de la 2D-VDCT, se pueden obtener regiones en las cuales los atributos para la VQ (estimación del factor de acoplamiento intrínseco y compactación de energía [VQ5]) del coeficiente, son mucho mejores que los que tiene el espacio en su conjunto. Esto sugiere que puede existir un cuantificador vectorial clasificado, que bajo ciertas condiciones podría dar mejores resultados que un cuantificador del vecinos más cercano ordinario. Por otro lado, el hecho de que las componentes de los coeficientes

de la 2D-VDCT estén bastante correlacionadas, aunado a que la distribución de energía entre los diferentes coeficientesde presenta ciertos patrones típicos, indica que despues de pasar por la etapa de procesamiento, en la señal todavía existe cierto grado de correlación que puede ser explotada por un cuantificador vectorial con memoria. Aquí los puntos importantes son, asentar bajo que condiciones puede introducirse memoria en la etapa de cuantificación y además proponer un método de optimización de dicha etapa. En la primera parte de este capítulo se presenta un planteamiento de este problema, el cual es atacado por el lado de los atributos óptimos de la etapa de procesamiento de la señal que precede a la cuantificación vectorial, en un sistema de codificación por transformada. Este planteamiento da como resultado un algoritmo iterativo para el diseño de un clasificador, el cual va a ser utilizado para construir un cuantificador vectorial.

En el contexto de codificación de la fuente, el principio de equipartición asintótica (teorema Shannon-McMillan-Brieman), sugiere que casi todas las palabras del código caen en una región de alta probabilidad, especificada por la entropía de la fuente. Esta región de alta probabilidad tiene una geometría que depende de la distribución particular de cada fuente, por ejemplo, la esfera para el proceso Gaussiano independiente idénticamente distribuido, el hipercubo para la fuente uniforme, etc. Estos resultados de la teoría de la información han sido utilizados por algunos investigadores para construir métodos de codificación para fuentes no uniformes, utilizando cuantificadores de enrejado. En el presente trabajo se utilizan estos resultados para fundamentar la obtención de un segundo método de optimización del clasificador que será utilizado para construir el cuantificador clasificado, el cual es un medio para introducir memoria en la etapa de cuantificación vectorial.

Finalmente, en la última parte de este capítulo se presenta un esquema de compresión de imágenes en base a la transformada vectorial, el cuantificador clasificado propuesto y un esquema de codificación zonal, basado en un conjunto predeterminado de plantillas.

4.2 Codificación de fuente y atributos óptimos para la VQ

En 1990 W. Li, en la búsqueda de las razones por las cuales la cuantificación vectorial no produce ventajas muy notables en la codificación de transformada, en comparación con la cuantificación escalar, encuentra un par de atributos que una etapa de procesamiento debe cumplir para adecuar la señal lo mejor posible para la cuantificación vectorial. Estos atributos son:

- *Preservación de la correlación entre las componentes de los vectores.* La cual está directamente relacionada con el factor de acoplamiento intrínseco ICF.
- *Reducción de la correlación entre los vectores.* Generalmente asociada con el nivel de compactación de energía.

Se ha observado experimentalmente que una secuencia de longitud suficientemente grande de alguno de los coeficientes en el dominio de la 2D-VDCT, puede ser clasificada en términos de la magnitud de la norma de los vectores para producir subconjuntos, los cuales pueden llegar a tener ICFs mucho mayores que el ICF estimado sobre toda la secuencia. Si se toma en cuenta que el ICF está relacionado directamente con el error normalizado de codificación, entonces se puede pensar que bajo algunas condiciones es posible construir un cuantificador vectorial clasificado que produzca mejores resultados en la codificación de los coeficientes de la 2D-VDCT, que el cuantificador del vecino más cercano ordinario. A continuación se presenta una formulación de esta hipótesis.

Sea $\boldsymbol{X}$ un vector aleatorio n-dimensional distribuido de forma discreta en un subconjunto R de R^n y sea $C=\{y_i;i=1...L\}$ un cuantificador vectorial del vecino más cercano utilizado para representar al vector $\boldsymbol{X}$. Suponiendo que el proceso sea al menos ergódico asintóticamente estacionario en la media, entonces en el límite se puede aproximar a los valores esperados por medio de promedios muestra de término largo [VQ49, VQ52]. Bajo estas suposiciones, el error de representar a $\boldsymbol{X}$ por medio del conjunto de vectores C está dado por

$$Err(X)=\sum_{i=1}^{L}\sum_{X\in R_i} d(X,y_i)$$

donde

$$R=\bigcup_{i=1}^{L} R_i \quad donde \quad R_i=\left\{X: d(X,y_i)\le d(X,y_j) \quad j=1...L\right\}$$
$$R_i\bigcap R_j=\varnothing \quad si \quad i\neq j$$

con una tasa promedio r [bits por vector], que debe cumplir con

$$H=-\sum_{i=1}^{L} p(R_i)\log_2 p(R_i)\le r\le \log_2 L$$

Sea D una partición mutuamente exclusiva de $\boldsymbol{R}$, esto es

$$R=\bigcup_{i=1}^{N} D_i$$
$$D_i\bigcap D_j=\varnothing si \quad i\neq j$$

en donde la energía total de X se divide entre las diferentes regiones D_i de acuerdo a un conjunto de coeficientes A_i, de la siguiente manera:

$$E_X = \sum_{X \in R} \|X\|^2 = \sum_{i=1}^{N} \sum_{X \in D_i} \|X\|^2 = \sum_{i=1}^{N} E_{D_i} = \sum_{i=1}^{N} A_i E_X \qquad [4.2.1]$$

y en donde el grupo de coeficientes debe cumplir con

$$\sum_{i=1}^{N} A_i = 1$$

Si se supone que la pertenencia de un vector a una de las N regiones de la partición D puede especificarse por un índice, entonces la cantidad de información necesaria para representar a la partición D es r_D [bits por vector], la cual debe cumplir con

$$H_D = -\sum_{i=1}^{N} p(D_i) \log_2 p(D_i) \le r_D \le \log_2 N$$

Si a cada una de las regiones D_i de R se le asocia un cuantificador

$$C_i = \{y_{ij} : j = 1 \ldots L_i\} \quad \text{en donde} \quad \bigcup_{j=1}^{L_i} R_{ij} = D_i$$

$$R_{ij} = \{X : d(X, y_{ij}) \le d(X, y_{ik}) \quad k = 1 \ldots L_i\}$$

$$\text{y } R_{ik} \bigcap R_{ij} = \varnothing \quad si \quad k \ne j$$

con tasa de código r_{Di} [bits por vector], la cual cumple con

$$H_{Di} = -\sum_{j=1}^{L_i} p(R_{ij}) \log_2 p(R_{ij}) \le r_{Di} \le \log_2 L_i$$

entonces el error en el que se incurre al representar a X por medio del conjunto de vectores y_{ij} está dado por

$$Err(X, D) = \sum_{i=1}^{N} Err(X : X \in D_i) = \sum_{i=1}^{N} \sum_{j=1}^{L_i} Err(X : X \in R_{ij}) = \sum_{i=1}^{N} \sum_{j=1}^{L_i} \sum_{X \in R_{ij}} d(X, y_{ij}) \qquad [4.2.2]$$

con una tasa promedio

$$\bar{r}_{Di} = \sum_{i=1}^{N} p(D_i) r_{Di}$$

Entonces el problema consiste en encontrar una partición (si es que existe alguna) D de R tal que

$$Err(\boldsymbol{X}) = \sum_{i=1}^{L} \sum_{X \in R_i} d(X, y_i) \geq Err(\boldsymbol{X}, D) = \sum_{i=1}^{N} \sum_{j=1}^{L_i} \sum_{X \in R_{ij}} d(\boldsymbol{X}, y_{ij})$$

sujeta a la restricción en la tasa del código

$$\bar{r}_{Di} + r_D \leq r \qquad [4.2.3]$$

Hasta aquí sólo se ha llegado a la formulación del problema de una manera muy general, ya que no se ha especificado nada acerca de la forma de la partición D, ni se ha asociado a ésta con el factor de acoplamiento intrínseco.

De forma experimental se ha determinado que la relación entre el error normalizado y el factor de acoplamiento intrínseco (ICF) en un conjunto de vectores, está dada por una función del tipo (ver gráficas 5.3.14, 15 y 16 en el capítulo 5)

$$\frac{Err(D_i)}{E_X} = -\alpha \cdot ICF(D_i) + \beta$$

en donde las constantes α y β dependen de la tasa del cuantificador (ver capítulo 5 parte 2). Introduciendo esta condición en la ecuación (4.2.1) se puede obtener una ecuación alternativa a (4.2.2), la cual relaciona el error de representación de $\boldsymbol{X}$ con la distribución de energía y los ICFs de la partición D.

$$Err(\boldsymbol{X}, D) = \sum_{i=1}^{N} \frac{Err(D_i)}{E_{Di}} E_{Di} = \sum_{i=1}^{N} \frac{Err(D_i)}{E_{Di}} A_i E_X$$

$$Err(\boldsymbol{X}, D) = \sum_{i=1}^{N} \left[\left(-\alpha \cdot ICF(D_i) + \beta \right) \left(A_i E_X \right) \right] = E_X \sum_{i=1}^{N} \left[\left(-\alpha \cdot ICF(D_i) + \beta \right) A_i \right] \qquad [4.2.4]$$

Hasta aquí el planteamiento se ha enfocado principalmente en el primer atributo para el desempeño óptimo de la VQ, es decir la correlación intravector

modelada por el ICF. Sin embargo la influencia del segundo atributo también ha sido tomada en cuenta. Esta se encuentra presente desde que se introdujo el conjunto de coeficientes de compactación de energía A_i, esto es, entre más concentrada esté la energía en unos pocos A_i, menos correlación intervector estará presente en los subconjuntos de la secuencia de entrenamiento de $\boldsymbol{X}$ y por lo tanto más eficiente será la cuantificación independiente de los subconjuntos en función de la partición.

Ahora quedan dos puntos por resolver. El primero es, cómo especificar una partición de R lo suficientemente simple para permitir al cuantificador vectorial clasificado ser competitivo en términos de complejidad, con el cuantificador del vecino más cercano. El segundo punto es encontrar una forma de introducir en esta ecuación la restricción en la tasa del codificador clasificado (desigualdad 4.2.3). Considerando que los factores α y β en la ecuación 4.2.4 están relacionados con el número de vectores del código en cada región, se puede introducir la restricción de tasa total del código, con el error de representación al hacer una asignación de bits entre las diferentes regiones de la partición. Esto se puede expresar como

$$Err(\boldsymbol{X}, D) = E_{\boldsymbol{X}} \sum_{i=1}^{N} \left[\left(-\alpha(r_i) \cdot ICF(D_i) + \beta(r_i) \right) A_i \right] \qquad [4.2.5]$$

en donde las funciones α y β tienen la siguiente forma

$$\alpha(r) = a_\alpha r^2 + b_\alpha r + c_\alpha$$
$$\beta(r) = a_\beta r^2 + b_\beta r + c_\beta$$

El criterio que se puede utilizar para repartir los bits entre los cuantificadores de las regiones D_i, es el siguiente [VQ5, VQ50]. Debido a que el rango dinámico de cada región es diferente, el número de bits dedicados para representar los vectores de ellas, debe ser proporcional al rango de variación observado en cada una de las regiones. La estrategia más simple es elegir el número de bits proporcional a la variación de cada coeficiente. Además, como la correlación entre las componentes de los vectores determinan el desempeño de codificación, también se tiene que tomar en cuenta a este factor para la asignación de bits. El determinante de la matriz de covarianza es una buena medida de que tan grande es la variación de las componentes y de que tan cercanamente están correlacionadas [VQ5]. Por eso, para una partición D de R, se puede utilizar a los determinantes de los estimados de las matrices de correlación de las diferentes D_i, para hacer una asignación de bits en base a las siguientes ecuaciones

$$r_i n = r_T + \frac{1}{2}\log_2|\boldsymbol{R}_{Di}| - \frac{1}{2N}\sum_{j=1}^{N}\log_2|\boldsymbol{R}_{Dj}|$$

$$L_i = 2^{r_i} \qquad [4.2.6]$$

en donde $\boldsymbol{R}_{Di}$ es el estimado de la matriz de covarianza del vector $\boldsymbol{X}$ en la región D_i y L_i es el número de vectores del cuantificador asociado a esa región.

4.3 Codificación de fuente y geometrías implícitas

Una geometría implícita útil para la codificación de fuente está asociada con cada fuente estacionaria ergódica [VQ30, VQ55]. Esta geometría puede ser caracterizada en función de superficies de probabilidad constante. Sea $f(\boldsymbol{X})$ la función de densidad de probabilidad del vector aleatorio $\boldsymbol{X}$ n-dimensional, Shannon observó [VQ71] que para L suficientemente grande y ε y Δ arbitrariamente pequeñas

$$\left|\frac{\log f(\boldsymbol{X})}{n} + h(\boldsymbol{X})\right| < \varepsilon$$

para todos los vectores $\boldsymbol{X}$, excepto aquellos en un conjunto de probabilidad total menor que Δ. En donde $h(\boldsymbol{X})$ es la entropía diferencial de $\boldsymbol{X}$, definida como

$$h(\boldsymbol{X}) = -\int_{-\infty}^{\infty} f(\boldsymbol{X})\log f(\boldsymbol{X})d\boldsymbol{X}$$

Esto es, para n suficientemente grande $\boldsymbol{X}$ se localiza en una región particular del espacio n-dimensional. Entonces, la codificación de fuentes aleatorias continuas se resume en localizar los puntos de representación en esta región de alta probabilidad. Esta afirmación se puede fundamentar formalmente de la siguiente manera:
Sea $\boldsymbol{X}$ un vector aleatorio n-dimensional con función de probabilidad conjunta $f(\boldsymbol{x})$, asumiendo que la fuente satisface la propiedad ergódica, la entropía diferencial por grado de libertad es

$$h_n = -\frac{1}{n}\int f(\boldsymbol{x})\log f(\boldsymbol{x})d\boldsymbol{x} \xrightarrow{p} h$$

$$para \quad n \to \infty$$

esto, conocido como el teorema Shannon-McMillan-Brieman [VQ49], implica que la región especificada por

$$S(n,h_n) = \left\{ x : -\frac{1}{n}\log f(x) = h_n \right\}$$

es una geometría suficiente para la codificación óptima de fuente, esto es, $\boldsymbol{X}$ se localiza en la región $S(n,h_n)$ en el sentido de que para cada $\boldsymbol{X}$ y $\alpha>0$ existe una $\hat{\boldsymbol{X}} \in S(n,h_n)$ tal que

$$\frac{1}{n}\left\|\boldsymbol{X}-\hat{\boldsymbol{X}}\right\|_v^{\alpha} \xrightarrow{p} 0$$

para $n \to \infty$

en donde

$$\left\|\boldsymbol{X}-\hat{\boldsymbol{X}}\right\|_v = \left(\sum_{i=1}^{n}\left|X_i-\hat{X}_i\right|^v\right)^{1/v}$$

es la distancia norma por dimensión.

En el caso de vectores con componentes independientes idénticamente distribuidas, la región $S(n,h_n)$ tiene cierta simetría en el espacio n-dimensional y está alineada con el sistema coordenado natural de la fuente. Desafortunadamente, para vectores con componentes correlacionadas, la región $S(n,h_n)$ puede tener una forma irregular y puede no estar alineada con el sistema coordenado natural de la fuente. Esto dificulta la aplicación de los resultados derivados del teorema Shannon-McMillan-Brieman en la construcción de cuantificadores vectoriales simples, para vectores con componentes no independientes. Además, dicho teorema sólo asegura una convergencia en probabilidad cuando n es suficientemente grande. Para el caso en que la dimensión de los vectores es moderada, o la tasa del código es baja, Fisher propone [VQ30] la construcción de un cuantificador vectorial usando varias regiones concéntricas de probabilidad constante en lugar de una sola. Con esto se logra distribuir los puntos de representación en varias superficies de probabilidad constante, lo cual asegura una mejor representación de la fuente. El procedimiento consiste en multiplicar los vectores con un factor de ponderación tal que sean trasladados a una región cercana a la superficie $S(n,h_n)$. De esta manera ese autor obtiene un cuantificador vectorial tipo producto para la codificación de fuentes aleatorias Gaussianas y Laplacianas independientes idénticamente distribuidas.

Si bien los cuantificadores obtenidos por este método son subóptimos, la utilización de los enrejados permite la construcción de códigos para vectores de dimensiones tales que, no serían prácticas de manejar con cuantificadores del vecino más cercano ordinarios.

La filosofía del uso de geometrías implícitas para la codificación de fuentes puede ser utilizada como punto de partida para esbozar una manera de hacer la

división D del espacio R^n planteada en la sección 3.2. Esta división se podría hacer por medio de un conjunto de N regiones de probabilidad constante concéntricas, las cuales concentren a la mayoría de la energía de $\boldsymbol{X}$ en el sentido de que para cada $\boldsymbol{X}$ y $\alpha>0$ existe una $\hat{\boldsymbol{X}}_{ij} \in S_i(n,h_n)$ con $i=1,..,N$ tal que

$$\frac{1}{n}\left\|\boldsymbol{X}-\hat{\boldsymbol{X}}_{ij}\right\|_v^{\alpha} \xrightarrow{p} 0 \qquad \text{donde} \qquad d(\hat{\boldsymbol{X}}_{ij},\boldsymbol{X}) \le d(\hat{\boldsymbol{X}}_{ik},\boldsymbol{X}) \quad k=1...N$$

Aquí habría que hacer las siguientes consideraciones. Las fuentes con las que se está trabajando son fuentes con componentes muy correlacionadas en las que se desconoce el modelo matemático de la dependencia, por lo que es necesario encontrar un tipo de superficie que en promedio aproxime a la superficie de probabilidad constante. Además, como se desea hacer una repartición de bits óptima para cada región, entonces el método de mapear los vectores por medio de un factor de ponderación no es adecuado.

4.4 Cuantificador vectorial clasificado CVQ.

En este tipo de cuantificador, un clasificador se utiliza para seleccionar un subconjunto particular del libro de códigos. La clasificación se basa en una peculiaridad, generalmente elegida de manera heurística, que permite caracterizar a cada vector de entrada. En el caso general, el libro de códigos está dividido en M subconjuntos (modos de operación), no necesariamente del mismo tamaño. En este esquema, el clasificador genera el índice del subconjunto que por medio de un cuantificador del vecino más cercano, será utilizado para representar al vector de entrada. Por lo tanto la palabra del código que genera el cuantificador vectorial clasificado consta de dos partes. La primera es el índice del subconjunto del código y la segunda es el índice de la palabra en tal subconjunto.

Existen muchas posibilidades para la elección del clasificador. Este puede ser un identificador que permite obtener cual de las N regiones del espacio pertenece el presente vector de entrada. En este caso el CVQ es muy similar a un cuantificador vectorial estructurado de dos etapas.

El significado físico particular del vector puede relacionarse con el clasificador, de tal forma que se preserven ciertas características particulares de la señal. Por ejemplo, un detector de bordes puede ser utilizado como clasificador y de esta forma se puede utilizar un código para las texturas y otro para los bordes.

Una vez que se conocen las características de la señal de entrada, hay que determinar el número de modos de operación adecuado para el CVQ. Cuando se ha determinado la cantidad de modos de operación se puede diseñar el clasificador basándose en algún criterio de optimización.

Cuando se ha diseñado el clasificador de N-modos, el libro de códigos puede ser diseñado en una de las siguientes formas. El primer método consiste en pasar el conjunto de entrenamiento original a través del clasificador y así generar N subconjuntos de entrenamiento. De esta forma, el libro de códigos se forma por la concatenación de los libros de códigos generados para cada uno de los subconjuntos de la secuencia de entrenamiento. El tamaño de cada uno de los subcódigos puede determinarse usando un algoritmo de optimización para la asignación de bits. El segundo método difiere del primero en que el tamaño de cada subcódigo se determina en base a un significado perceptual o subjetivo en el cual sea ventajoso hacer una asignación de bits no proporcional a la frecuencia de ocurrencia de los diferentes modos.

En la mayoría de los casos, el cuantificador clasificado es utilizado en sistemas de compresión como un medio de reducir la complejidad de la etapa de cuantificación a expensas de una ligera reducción en el desempeño *tasa-distorsión* del sistema. En este trabajo de tesis se utiliza la cuantificación vectorial clasificada de una manera algo diferente. Aquí se utiliza como un medio para la introducción de memoria en la cuantificación vectorial de los coeficientes de la 2D-VDCT. La introducción de memoria en la etapa de cuantificación consiste en seleccionar la forma de cuantificar un vector en base a las características de un cierto conjunto de vectores. En el caso específico de los coeficientes de la 2D-VDCT, una forma para la introducción de memoria se puede llevar a cabo al elegir el modo de operación de los cuantificadores clasificados de cada uno de los coeficientes, en base a la forma en que la energía de la señal se distribuye entre ellos. Aplicado de esta forma y suponiendo que exista una forma eficiente de codificación de los índices de clasificación, se puede esperar que el cuantificador clasificado explote la cantidad de correlación que la transformación no pudo eliminar de la señal, lo cual puede llevar a que este tipo de cuantificación produzca un mejor desempeño *tasa-distorsión* que el del cuantificador ordinario del vecino más cercano.

4.5 Diseño del clasificador en base a los atributos óptimos para VQ

En la parte 4.2 se derivó una ecuación para el cálculo del error normalizado, el cual resulta de representar a un vector aleatorio $\boldsymbol{X}$ por medio de un conjunto de N cuantificadores vectoriales. Esos cuantificadores estaban asociados a una partición del espacio basada en la magnitud de la norma de los vectores. Esta partición se puede especificar por medio de un conjunto de umbrales como se muestra a continuación.

Sea $\boldsymbol{X}$ un vector aleatorio n-dimensional distribuido de forma discreta en un subconjunto R de R^n y sea D una partición mutuamente exclusiva de R basada en la norma del vector $\boldsymbol{X}$, esto es

$$R = \bigcup_{i=1}^{N} D_i$$

$$D_i \bigcap D_j = \varnothing si \quad i \neq j$$

en donde las regiones D_i están completamente determinadas por un conjunto de N+1 umbrales γ_i, de acuerdo a

$$D_i = \{ \boldsymbol{X} \in R : \gamma_i \leq \|\boldsymbol{X}\| < \gamma_{i+1} \}$$

[4.5.1]

Este conjunto de umbrales, al dividir el espacio n-dimensional en N regiones, constituyen la parte de clasificación del cuantificador vectorial que se está buscando construir. Ahora, hay que determinar como elegir los umbrales que optimicen el desempeño del cuantificador. Como no se conoce una expresión matemática de la función de densidad de probabilidad conjunta del vector $\boldsymbol{X}$, entonces el diseño del cuantificador se realizará por medio del procesamiento de secuencias de entrenamiento. Después, en lugar de expresar la ecuación 4.2.5 en función de la partición D y f_X y luego analíticamente encontrar los γ_i que la minimicen, se propone un método iterativo basado en la optimización independiente de cada uno de los γ_i. Este método consiste de los siguientes puntos.

1) Sean: R un subconjunto de R^n la secuencia de entrenamiento, r la tasa global del cuantificador clasificado, N el número de modos de operación del cuantificador y ε>0 el umbral de finalización del algoritmo.

2) Ordenar los vectores de R en forma creciente de acuerdo a la magnitud de su norma.

3) Elegir un conjunto inicial de γ_i.

4) En función de los γ_i y la ecuación 4.5.1, formar la partición inicial $D=\{D_i : i=1...N\}$ de R.

5) Para los elementos de la partición inicial, encontrar

$$ICF_i = 1 - \frac{|\boldsymbol{R}_{Di}|^{1/n}}{Tr(\boldsymbol{R}_{Di})/n}$$

$$r_i = r + \frac{1}{2n}\log_2|\boldsymbol{R}_{Di}| - \frac{1}{2nN}\sum_{j=1}^{N}\log_2|\boldsymbol{R}_{Dj}|$$

$$A_i = \frac{\sum_{X \in Di} \|X\|^2}{\sum_{X \in R} \|X\|^2}$$

en donde $\boldsymbol{R}_{Di}$ es la matriz de covarianza de $\boldsymbol{X}$ estimada sobre la región D_i.

6) Calcular el error normalizado inicial por medio de

$$Err_0(\boldsymbol{X}, D) = E_{\boldsymbol{X}} \sum_{i=1}^{N} \left[\left(-\alpha(r_i) \cdot ICF_i + \beta(r_i) \right) A_i \right]$$

7) Optimizar secuencialmente cada uno de los umbrales y calcular el error normalizado final como

$$Err_1(\boldsymbol{X}, D') = E_{\boldsymbol{X}} \sum_{i=1}^{N} \left[\left(-\alpha(r_i') \cdot ICF_i' + \beta(r_i') \right) A_i' \right]$$

8) Si el decremento en el error normalizado es menor que el umbral de finalización, esto es

$$\frac{Err_0 - Err_1}{Err_0} < \varepsilon$$

entonces detener el algoritmo con γ_i como los umbrales óptimos. En otro caso hacer

$Err_0 = Err_1$
$ICF_i = ICF_i'$
$r_i = r_i'$
$A_i = A_i'$
y regresar al paso (7).

Optimización del umbral γ_i

Durante el planteamiento del algoritmo se consideró que para un clasificador con N modos de operación se necesitan N+1 umbrales. Sin embargo, como γ_0=0 y γ_{N+1}=∞ entonces únicamente es necesario optimizar N-1 umbrales. El concepto de la optimización del umbral γ_i , el cual divide a las regiones D_{i-1} y D_i, es muy simple y se basa en encontrar la división γ_m, entre las dos regiones que minimice al error normalizado dado por la ecuación 4.2.5. El algoritmo para esta optimización está constituido por los siguientes puntos

1) Calcular el error normalizado local en la región D_T dada por

$$D_T = D_{i-1} \bigcup D_i$$

por medio de

$$Err_T(X, D_T) = E_X\left[\left(-\alpha(r_{i-1}) \cdot ICF_{i-1} + \beta(r_{i-1})\right) A_{i-1}\right]$$
$$+E_X\left[\left(-\alpha(r_i) \cdot ICF_i + \beta(r_i)\right) A_i\right]$$

2) Sea ν una cantidad entera elegida experimentalmente, el número de pasos para la optimización de γ_i. Entonces el tamaño de paso en la optimización del umbral está dado por

$$\Delta\gamma = \frac{\gamma_{i+1} - \gamma_{i-1}}{\nu + 1}$$

3) Moviendo el umbral γ desde $\gamma_{i-1} + \Delta\gamma$ hasta $\gamma_{i+1} - \Delta\gamma$ en pasos de $\Delta\gamma$,obtener el valor de γ_m para el cual Err_{T1} es mínimo, en donde

$$Err_{T1}(X, D_{T1}) = E_X\left[\left(-\alpha(r'_{i-1}) \cdot ICF'_{i-1} + \beta(r'_{i-1})\right) A'_{i-1}\right]$$
$$+E_X\left[\left(-\alpha(r'_i) \cdot ICF'_i + \beta(r'_i)\right) A'_i\right]$$

$$ICF'_i = 1 - \frac{|\boldsymbol{R}_{D'i}|^{1/n}}{Tr(\boldsymbol{R}_{D'i})/n} \qquad r_i = r + \frac{1}{2n}\log_2|\boldsymbol{R}_{D'i}| - \frac{1}{2nN}\sum_{j=1}^{N}\log_2|\boldsymbol{R}_{D'j}|$$

$$A'_i = \frac{\sum_{X \in D'i} \|X\|^2}{\sum_{X \in R} \|X\|^2}$$

están determinados en base a las regiones D'_{i-1} y D'_i definidas como

$$D'_{i-1} = \{X \in D_T : \gamma_{i-1} \le \|X\| < \gamma\}$$

$$D'_i = \{X \in D_T : \gamma \le \|X\| < \gamma_{i+1}\}$$

y las cuales cumplen con la siguientes condiciones

$$D_T = D'_{i-1} \bigcup D'_i$$
$$D'_{i-1} \bigcap D'_i = \varnothing$$

4) Si $Err_{Tl} < Err_T$ entonces hacer $\gamma_i = \gamma_m$. En caso contrario dejar a γ_i sin cambio.

Elección de los umbrales iniciales y convergencia.

El algoritmo anterior al ser un método basado en la optimización individual de cada uno de los umbrales, no garantiza que se alcance un mínimo global en la función del error. Sin embargo, el hecho de que en cada iteración se modifiquen las regiones únicamente si hay un decremento en la función de error local, garantiza que la función de error global es al menos no creciente. Bajo estas condiciones, el conjunto de umbrales utilizado para representar la partición inicial, será determinante en la convergencia hacia un buen mínimo local. Por el momento no se tiene una forma óptima para elegir estos umbrales, pero al elegirlos uniformemente espaciados en el intervalo $[0,\sigma_n]$, en donde σ_n es la varianza de la norma de los vectores, el algoritmo converge en unas pocas iteraciones.

4.6 Diseño del clasificador en base a las geometrías implícitas

Anteriormente se mostró como la geometría de la función de densidad de probabilidad de un vector $\boldsymbol{X}$ puede ser utilizada para construir un método que permita codificarlo. En este método, se desean obtener las superficies $S_i(n,h_n)$ de probabilidad constante que concentran la mayor parte de la energía de $\boldsymbol{X}$, para distribuir el código del cuantificador vectorial en torno a ellas. Como no se conoce la expresión matemática de la función densidad de probabilidad conjunto del vector aleatorio, no es posible determinar la expresión de las superficies $S_i(n,h_n)$ las cuales concentran a X en el sentido de que para cada vector en el dominio de $\boldsymbol{X}$ y $\alpha>0$ existe una $\hat{\boldsymbol{X}}_{ij} \in S_i(n,h_n)$ con $i=1,..,N$ tal que

$$\frac{1}{n}\left\|\boldsymbol{X}-\hat{\boldsymbol{X}}_{ij}\right\|_v^{\alpha} \xrightarrow{p} 0 \qquad \text{donde} \qquad d(\hat{\boldsymbol{X}}_{ij},\boldsymbol{X}) \le d(\hat{\boldsymbol{X}}_{ik},\boldsymbol{X}) \quad k=1...N$$

para evadir esta dificultad, se propone aproximar a las $S_i(n,h_n)$ verdaderas por medio de superficies concéntricas esféricas $Se_i(n,h_n)$. Sin embargo, si no se conocen las superficies $S_i(n,h_n)$ verdaderas ¿cómo determinar cuales son las superficies esféricas que mejor las aproximan? Para resolver este problema hay que recordar que $\boldsymbol{X}$ se encuentra concentrado en las superficies $S_i(n,h_n)$ bajo el criterio de que para cada punto en el dominio de $\boldsymbol{X}$ existe un punto en una de las superficies, tal que la distancia entre estos dos puntos tiende a cero en probabilidad, conforme la dimensión de los vectores aumenta. Entonces bajo este criterio, las superficies esféricas $Se_i(n,h_n)$ que mejor aproximen a las superficies de probabilidad constante,

son aquellas que minimizan el error en el que se incurre al representar a cada punto en el dominio de X por medio del punto más cercano a él, en la esfera más cercana a él. Estas condiciones se pueden expresar matemáticamente de las siguiente forma

$$\mathrm{Se_i}(n,h_n) = \min_{S(n,h_n)} E\left\{\left\|\boldsymbol{X}-\hat{\boldsymbol{X}}\right\|_{\nu}^{\alpha}\right\}$$

si la medida de distorsión utilizada es el error cuadrático medio, esta expresión se transforma en

$$\mathrm{Se_i}(n,h_n) = \min_{S(n,h_n)} E\left\{\left\|\boldsymbol{X}-\hat{\boldsymbol{X}}\right\|^{2}\right\}$$

en donde $\hat{\boldsymbol{X}}$ es un vector sobre la superficie esférica de radio

$$\hat{n}_i = \min_{i=1,\ldots,N}\left(\|\boldsymbol{X}\|-\hat{n}_i\right)^2$$

y donde

$$\hat{\boldsymbol{X}} = \min_{\boldsymbol{X}\in Se_i(\hat{n}_i,h_{\hat{n}i})}\left(\left\|\boldsymbol{X}-\hat{\boldsymbol{X}}\right\|^{2}\right) \qquad [4.6.1]$$

En una superficie esférica de radio $\hat{n}$ centrada en el origen, el punto de la superficie $\hat{\boldsymbol{X}}$, más cercano a un punto $\boldsymbol{X}$, en la vecindad de la superficie, es aquel que está en la intersección entre las superficie y la recta que une al origen con $\boldsymbol{X}$ y el cual está dado por

$$\hat{\boldsymbol{X}} = \hat{n}\frac{\boldsymbol{X}}{\|\boldsymbol{X}\|}$$

introduciendo esta condición en la ecuación 4.6.1 se obtiene

$$\hat{\boldsymbol{X}} = \min_{\boldsymbol{X}\in Se_i(\hat{n}_i,h_{\hat{n}i})}\left(\left\|\boldsymbol{X}-\frac{\boldsymbol{X}}{\|\boldsymbol{X}\|}\hat{n}_i\right\|^{2}\right)$$

$$\hat{\boldsymbol{X}} = \min_{\boldsymbol{X}\in Se_i(\hat{n}_i,h_{\hat{n}i})}\left(\|\boldsymbol{X}\|^{2}\left(1-\frac{\hat{n}_i}{\|\boldsymbol{X}\|}\right)^{2}\right)$$

$$\hat{X} = \min_{X \in Se_i(\hat{n}_i, h_{\hat{n}i})} (\|X\| - \hat{n}_i)^2 = \min_{i=1,\ldots,N} (\|X\| - \hat{n}_i)^2$$

lo cual implica que

$$\mathrm{Se}_i(n, h_n) = \min_{S(n,h_n)} E\left\{\|X - \hat{X}\|^2\right\} = \min_{S(n,h_n)} E\left\{(\|X\| - \hat{n}_i)^2\right\}$$

entonces el problema se reduce a encontrar el conjunto $\hat{n} = \{\hat{n}_1, \ldots, \hat{n}_N\}$ que minimice a

$$E\left\{(\|X\| - \hat{n}_i)^2\right\}.$$

Como la norma de un vector es una cantidad escalar, entonces la expresión anterior se puede resolver numéricamente por medio del cuantificador escalar óptimo de N niveles sobre el conjunto de normas de los vectores en el dominio de X. En base a estos resultados se puede construir un algoritmo iterativo para el diseño del clasificador en base al criterio de geometrías intrínsecas. Este algoritmo consta de los siguientes pasos:

1) Sean: R un subconjunto de R^n la secuencia de entrenamiento del vector X, N el número de modos de operación del clasificador, un conjunto inicial de centroides $\hat{n}_0 = \{\hat{n}_1, \ldots, \hat{n}_N\}$, $D_0 = \infty$ la distorsión inicial y $\varepsilon > 0$ el umbral de finalización del algoritmo.
2) Calcular la norma de cada uno de los vectores de la secuencia de entrenamiento R y construir el conjunto

   $n = \{n_i = \|X_i\| : X_i \in R\}$.
3) En base al conjunto inicial de centroides $\hat{n}_0 = \{\hat{n}_1, \ldots, \hat{n}_N\}$, encontrar la partición óptima en celdas de cuantificación del conjunto de normas, definida como

   $$R_i = \left\{n : d(n, \hat{n}_i) \le d(n, \hat{n}_j) : para \quad j \ne i\right\} \qquad i=1,\ldots,N$$

4) Usando la condición del centroide, encontrar el conjunto de centroides óptimo

   $\hat{n}_1 = \left\{\hat{n}_i^1 = Cent(R_i) : i = 1,\ldots,N\right\}$, para las celdas R_i.

5) En base al promedio muestra de término largo, calcular

$$D_1 = E\left\{\left(n - \hat{n}_i^1\right)^2\right\}. \text{ Si } \frac{D_0 - D_1}{D_0} < \varepsilon$$

entonces tomar a

$$\hat{n}_1 = \left\{\hat{n}_1^1, \ldots . \hat{n}_N^1\right\}$$

como el conjunto de centroides final $\hat{n} = \{\hat{n}_1, \ldots, \hat{n}_N\}$, e ir al paso (6). En caso contrario hacer

$$D_0 = D_1$$

$$\hat{n}i = \hat{n}_i^1 \quad para \quad i = 1, .., N$$

y regresar al paso 3.

6) En base al conjunto de centroides $\hat{n} = \{\hat{n}_1, \ldots, \hat{n}_N\}$ calcular el conjunto de umbrales del clasificador de acuerdo a

$$\gamma_i = \frac{\hat{n}_{i+1} + \hat{n}_i}{2} \quad i = 1, \ldots, N-1$$

4.7 Esquema de compresión de imágenes por medio de la transformada coseno discreta vectorial y cuantificadores clasificados

La codificación por transformada desarrollada hace mas de dos décadas, ha probado ser un esquema efectivo para la compresión de imágenes y es la base de todos los estándares de codificación con pérdidas en uso actualmente. Un codificador por transformada básico segmenta la imagen en bloques cuadrados pequeños. Cada bloque sufre una transformación ortogonal para producir un bloque de coeficientes. A continuación los coeficientes son cuantificados y codificados individualmente. Los coeficientes del bloque que tienen las energías más altas son cuantificados más finamente y aquellos con menos energía son cuantificados burdamente, o en algunos casos, son truncados. El codificador trata a los coeficientes cuantificados como símbolos, los cuales son codificados en base a su entropía. El decodificador reconstruye las intensidades de los pixels de la imagen a partir del flujo de bits que recibe, siguiendo las operaciones inversas realizadas en el codificador. Sin embargo los codificadores de transformada utilizados en el presente realizan un procesamiento de la señal orientado a escalares, y en consecuencia la cuantificación que utilizan es de ese tipo. Uno de los resultados fundamentales de la teoría de la información, es que es más eficiente la codificación de bloques

(vectores) que la de datos aislados (escalares), lo que implica que la cuantificación de vectores es más eficiente que la cuantificación de escalares. Sin embargo, los resultados obtenidos al emplear cuantificadores vectoriales en esquemas simples de codificación por transformada indican que las mejoras del desempeño obtenidas no son muy significativas y no llegan a justificar el incremento que causan en la complejidad del sistema. En el capítulo 3 se vio que la mejor transformación para el propósito de cuantificación vectorial efectiva, debe cumplir con un atributo adicional a los que cumple la etapa de transformación en sistemas con cuantificación escalar. Este atributo es que debe preservar la correlación entre las componentes de los vectores. Entonces la mejor transformación para el propósito de cuantificación vectorial debe ser una transformación orientada a vectores. En 1991 Weiping Li [VQ58] propuso el uso de transformaciones vectoriales, la 2D-VDCT entre otras, como etapa de procesamiento previo a la cuantificación vectorial en un sistema de compresión. Los sistemas de compresión por transformada propuestos hasta ahora son una generalización de los esquemas de codificación por transformada de muestreo zonal (asignación fija de bits a cada cuantificador) y de codificación de umbral (asignación dinámica de bits a cada cuantificador). En estos esquemas se utiliza el cuantificador vectorial del vecino más cercano en la etapa de cuantificación. En el presente trabajo se ha explorado una nueva forma de realizar la cuantificación de los coeficientes en el dominio de la 2D-VDCT. El esquema que se propone es un cuantificador vectorial con memoria basado en un cuantificador clasificado, los resultados positivos obtenidos a nivel de la codificación individual de los coeficientes, motiva probar el esquema de cuantificación dentro de un sistema de compresión de imágenes. Por motivos de tiempo y simplicidad, únicamente se usará el esquema de codificación por transformada con asignación fija de bits sobre un conjunto de máscaras, es decir, el sistema de codificación zonal con sólo un conjunto de cuantificadores.

La codificación zonal se basa en la premisa de que los coeficientes de máxima varianza, en el sentido de la teoría de la información, transportan la mayor cantidad de información de la imagen y deben ser retenidos. Las localidades de los coeficientes con las k varianzas más grandes son indicados por medio de una máscara o plantilla, en donde los coeficientes retenidos son denotados por uno, y los descartados por cero. Para diseñar máscaras zonales, las varianzas de cada coeficiente deben ser calculadas ya sea sobre un conjunto representativo de subimágenes transformadas, o basados en un modelo global de la imagen. Las máscaras zonales pueden ser afinadas para imágenes individuales y almacenadas con la imagen codificada.

Debido a que el rango dinámico y el factor de acoplamiento intrínseco de cada coeficiente retenido es diferente, es imposible diseñar un cuantificador vectorial único para todos los coeficientes. Además, el objetivo es optimizar el desempeño *tasa-distorsión*. Un método para realizar la cuantificación vectorial en el esquema de

codificación zonal, consiste en establecer la tasa promedio del sistema, de tal forma que el número total de bits para la cuantificación vectorial del conjunto de coeficientes, sea un valor fijo, entonces se reparte la tasa total entre los diferentes cuantificadores, de manera que la distorsión promedio sea minimizada. Intuitivamente, se desea asignar más bits a los vectores que son más dificiles de codificar y que tienen una mayor rango dinámico y menos bits a los coeficientes más fáciles de codificar y con un rango dinámico pequeño. El problema consiste en encontrar una buena medida de que tan susceptible de codificar es un vector. Como las varianzas de las componentes y la correlación entre ellas determinan el desempeño de codificación del vector, hay que tomar ambos factores en cuenta para la asignación de bits. El determinante de la matriz de covarianza es una buena medida de que tan grandes son las varianzas de las componentes y que tan correlacionadas están [VQ5, VQ28, VQ50,]. Por eso se pueden usar los valores de los determinantes para la asignación de bits de acuerdo a:

$$r_i = \frac{r_T}{n} + \frac{1}{2}\log_2\left(|\boldsymbol{R}_i|\right) - \frac{1}{2Nn}\sum_{j=0}^{N-1}\log_2\left(|\boldsymbol{R}_i|\right)$$

en donde r_i es la tasa asignada al cuantificador del i-ésimo coeficiente, N es el número de coeficientes, n es el número de dimensiones de los coeficientes, R_i es la matriz de covarianza del i-ésimo coeficiente y r es la tasa total promedio del codificador.

Una vez que se ha determinado cual es la tasa para la cuantificación de cada uno de los coeficientes, se puede proceder a diseñar los códigos de los cuantificadores clasificados utilizando alguno de los métodos presentados en las secciones cinco y seis de este capítulo.

Otra parte importante en el diseño del sistema de compresión, es determinar cual es la forma de las mascarillas que se utilizarán para la codificación zonal. En este trabajo, se utilizará un enfoque energético. Al determinar un nivel máximo en la relación señal a ruido que se pretende alcanzar con el codificador, se puede determinar que cantidad de energía, E_D, de la señal original puede ser discriminada para producir la SNR deseada. Esto es, si se desea una calidad de codificación de SNR_o decibeles, entonces la cantidad de energía que se puede discriminar eliminando algunos de los coeficientes de la imagen transformada es

$$E_D < E_0 = \frac{Energía(Bloque)}{10^{\frac{SNR_O}{10}}}$$

y la relación, γ_E, entre la energía de los coeficientes suprimidos y la energía total del bloque está dada por

$$\gamma_E = \frac{Energía(Bloque) - Energía(Suprimida)}{Energía(Bloque)} > \gamma_0 = \left(1 - \frac{1}{10^{-\frac{SNR_0}{10}}}\right) \qquad [4.7.1]$$

entre más cercana esté γ_E a γ_o más grande debe ser la calidad de la codificación, SNR_p, de los coeficientes que se han preservado, la cual debe cumplir con

$$SNR_p = 10\log_{10}\left(\frac{\gamma_E}{10^{-\frac{SNR}{10}} - 1 + \gamma_E}\right)$$

por esa razón, se debe elegir un valor de γ_E que logre un balance entre el número de coeficientes suprimidos y la tasa requerida para lograr codificar a los coeficientes preservados con la calidad requerida por la ecuación anterior.

En la figura 4.7.1 se puede ver el conjunto de plantillas de codificación zonal que se utilizarán en este esquema de compresión. Estas plantillas fueron obtenidas al observar las características de un conjunto de imágenes transformadas. Si el número de plantillas es pequeño en comparación con el número de coeficientes del bloque, entonces la cantidad de información lateral necesaria para identificar la plantilla del bloque será despreciable, en comparación con la cantidad de información asociada con los índices de los coeficientes cuantificados. En este trabajo, el tamaño de la transformada es de 8x8 vectores. Como se tiene un conjunto de doce plantillas, se requieren 4 bits para identificar la plantilla de cada bloque. Esto implica una cantidad adicional de 0.004 bits por cada pixel, lo que es prácticamente despreciable.

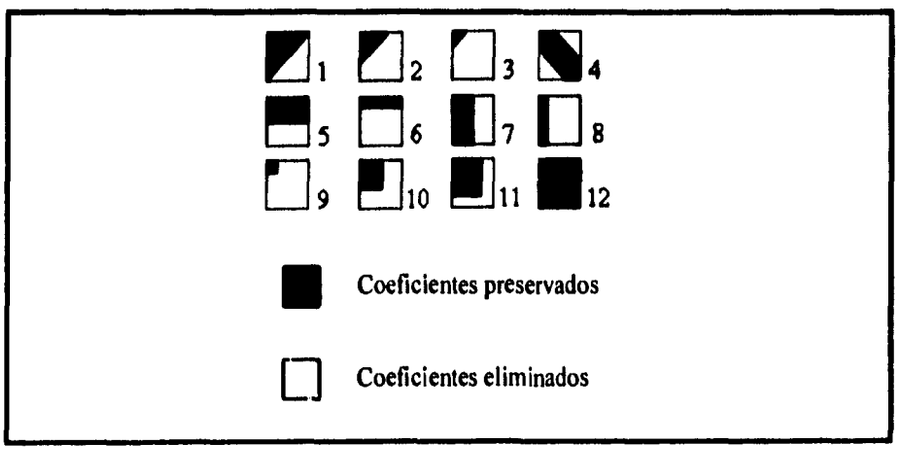


Figura 4.7.1. Conjunto de plantillas de codificación zonal para el esquema de compresión 2D-VDCT-CVQ.

Aquí las plantillas cumplen una doble función. En primer lugar serán utilizadas para la codificación zonal, y en segundo lugar, cada plantilla incluirá los índices de clasificación de los cuantificadores. Estos índices deben ser calculados en base a un enfoque de entrenamiento, esto es, para cada plantilla determinar los índices de clasificación óptimos para un conjunto de bloques transformados (secuencia de entrenamiento) y elegir aquellos índices que mejor representen a los de las secuencia.

En las figuras 4.7.2 y 4.7.3 se pueden ver los diagramas de bloques del codificador y del decodificador respectivamente. En la etapa de codificación, después de que la imagen pasa por la etapa de transformación vectorial, se determina que coeficientes serán eliminados en base al umbral γ_E. Cada uno de los coeficientes preservados es procesado individualmente por un cuantificador vectorial clasificado, el cual ha sido diseñado en base a una secuencia de entrenamiento especial para dicho coeficiente. Después de la etapa de codificación, todos los índices obtenidos son agrupados. Como la estructura de la etapa de cuantificación está controlada por un conjunto de plantillas, es necesario enviar información adicional acerca de la plantilla utilizada en el bloque.

En la etapa de decodificación, los índices procedentes del canal son enviados a los cuantificadores de acuerdo a la estructura determinada por el índice de la plantilla utilizada en el bloque. Cada uno de los cuantificadores genera un coeficiente. Los coeficientes correspondientes a los vectores suprimidos en la etapa del codificador se llenan con ceros. Todos los coeficientes son presentados simultáneamente a la entrada de la etapa de transformación vectorial inversa, la cual genera la imagen de salida.

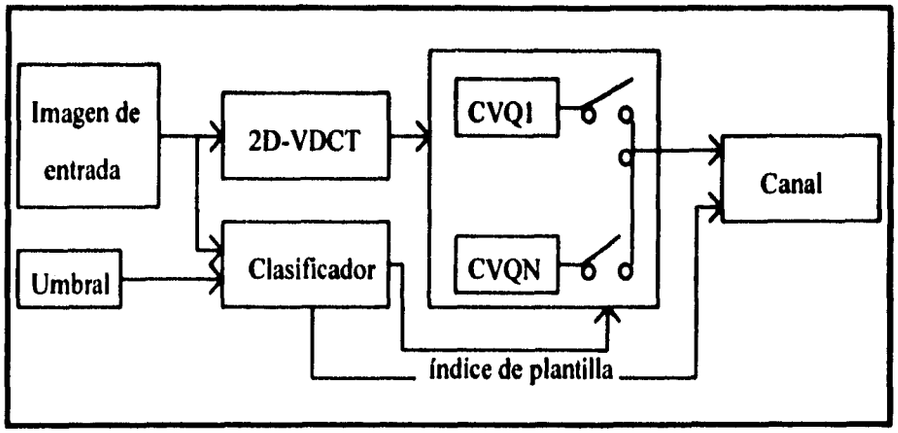


Figura 4.7.2. Diagrama de bloques del codificador del sistema de compresión de imágenes basado en la 2D-VDCT y el cuantificador vectorial clasificado.

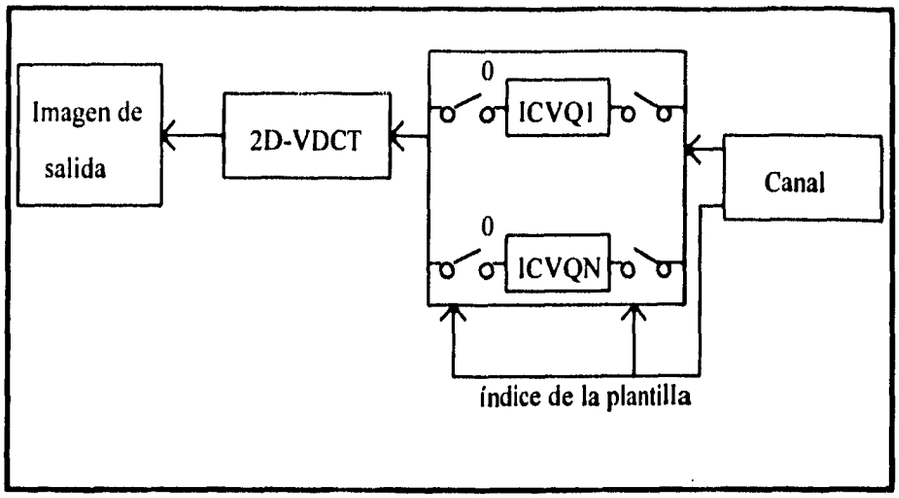


Figura 4.7.3. Diagrama de bloques del decodificador del sistema de compresión de imágenes basado en la 2D-VDCT y el cuantificador vectorial clasificado.

Finalmente cabe señalar que: como el cuantificador vectorial clasificado de cada uno de los coeficientes asigna diferentes tasas a cada uno de los modos de operación y como las plantillas suprimen distintas cantidades de coeficientes, entonces el esquema de compresión 2D-VDCT-CVQ puede ser visto como un esquema simple de asignación dinámica de bits, aún cuando se utilice una asignación fija de información para los diferentes coeficientes.

Capítulo 5

Resultados experimentales

5.1 Introducción

Es bien conocido que los coeficientes llamados de corriente alterna, *AC*, de la transformada coseno discreta bidimensional (2D-DCT) pueden ser modelados como fuentes Laplacianas y el coeficiente llamado de corriente directa, *DC*, como una fuente Gaussiana [VQ34, VQ41, VQ55, VQ61]. Dado que la transformada vectorial coseno discreta bidimensional (2D-VDCT) puede ser construida por medio de un grupo de 2D-DCTs escalares [VQ5, VQ28], es lógico que las distribuciones de los coeficientes tengan distribuciones que puedan ser modeladas por procesos aleatorios Laplacianos y Gaussianos *N*-dimensionales. Sin embargo, debido a la capacidad de la 2D-VDCT de preservar la correlación intravector, es de esperarse que las componentes estén bastante correlacionadas entre sí.

La búsqueda de un método de cuantificación eficiente de fuentes aleatorias con distribuciones Gaussianas y Laplacianas, ha producido variadas investigaciones desde que el uso de la transformada coseno discreta comenzó a tener auge en los sistemas de compresión de señales [VQ30, VQ41, VQ55, VQ56, VQ62]. Sin embargo, todos los estudios realizados se enfocan al caso en el que las componentes de los vectores son variables aleatorias independientes idénticamente distribuidas, o bien son modelos muy simples de correlación tales como el modelo de Markov de primer y segundo orden. Como es bien sabido, el modelo de Markov no es adecuado para representar muestras obtenidas a partir del procesamiento de imágenes [VQ53].

Debido a que el uso de las transformaciones vectoriales en sistemas de compresión de imágenes es bastante reciente, no se han realizado estudios en busca de formas diferentes de cuantificación vectorial de fuentes Gaussianas y Laplacianas fuertemente correlacionadas. Los sistemas de compresión de imágenes, usando la 2D-VDCT, que se han propuesto hasta ahora, utilizan el cuantificador ordinario del vecino más cercano en esquemas de muestreo zonal o en esquemas de asignación dinámica de bits [VQ5, VQ22, VQ28, VQ58].

Considerando que el desempeño de la VQ del vecino más cercano ordinaria, aplicada a los coeficientes de la 2D-DCT, dista mucho de los límites teóricos impuestos por la teoría de la información para fuentes con valores similares de ICF [VQ22, VQ28, VQ68], las características particulares de las distribuciones de estos coeficientes motivan la búsqueda de nuevos métodos de codificación, para ver si es posible encontrar una forma de VQ que permita explotar la geometría de las distribuciones y el modelo de correlación, y así obtener una ganancia de desempeño sobre el cuantificador del vecino más cercano ordinario. Como en principio esta dependencia es desconocida, entonces es adecuado manejarla como un modelo implícito durante el diseño de los cuantificadores vectoriales. O similarmente, hacer el diseño en términos de la filosofía de secuencias de entrenamiento, bajo una justificación de ergodicidad y estacionariedad basada en la longitud de las secuencias. En este trabajo se propusieron dos métodos de diseño para cuantificadores vectoriales clasificados, los cuales se presentan como una alternativa al cuantificador del vecino más cercano ordinario. Debido a que el diseño de uno de los métodos de cuantificación propuestos, requería de una relación explícita entre el error normalizado, el factor de acoplamiento intrínseco y la compactación de la energía, fue necesario hacer caracterizaciones de la relación entre estos atributos y el error normalizado, en función de la tasa de los códigos para los coeficientes de la 2D-VDCT. La comparación entre los métodos de cuantificación propuestos y el cuantificador del vecino más cercano se hace de dos formas. En el primer caso se tiene la comparación del desempeño sobre cada coeficiente como una función de la tasa del código. Esta comparación se realiza con y sin codificación entrópica. En el segundo caso se utiliza un esquema de compresión de imágenes, como el marco de comparación. En la parte final de este capítulo, se presenta una descripción de las herramientas utilizadas para realizar todas las pruebas requeridas durante el trabajo, también se da una ligera descripción de las funciones incluidas en los programas que fue necesario construir.

5.2 Metodología

El presente trabajo está dividido en dos áreas principales. La primera consiste en el estudio de las características que adquieren los vectores después de pasar por la etapa de procesamiento de la señal. En la segunda parte se estudian algunas formas

de realizar la cuantificación de dichos vectores y se hace una comparación de desempeño entre los métodos propuestos y el cuantificador del vecino más cercano ordinario. Para dar al lector un punto de vista global de este trabajo de tesis, se presenta un diagrama esquemático en el cual se destacan las partes más importantes (ver figura 5.2.1).

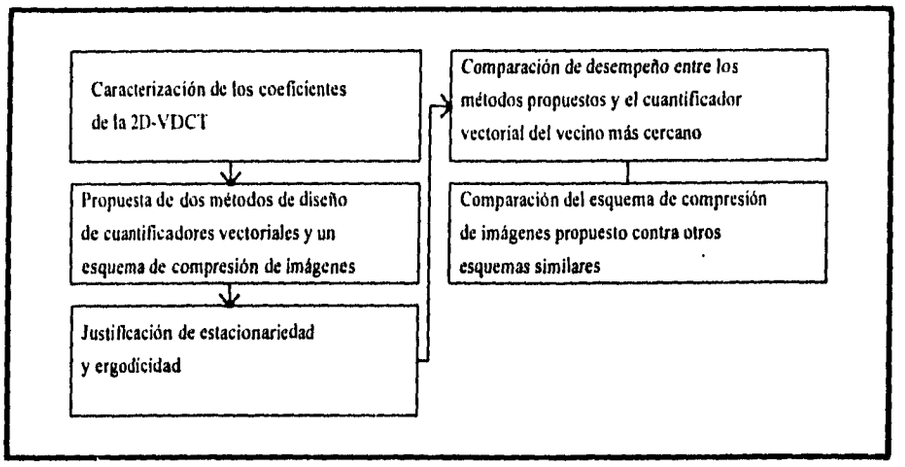


Figura 5.2.1 Puntos fundamentales del trabajo de tesis.

A continuación, como una guía introductoria se esboza la metodología seguida durante la fase de caracterización y comparación de los resultados de los métodos propuestos en este trabajo. Adicionalmente, se presenta una descripción de las herramientas de *software* que fue necesario construir.

- Caracterización de las fuentes aleatorias *n*-dimensionales derivadas del procesamiento de imágenes por medio de la 2D-VDCT y comparación contra otros tipos de fuentes.
    - Caracterización de las funciones de densidad de probabilidad de las componentes de los coeficientes 2D-VDCT en base a histogramas.
    - Medición de ICF y compactación de la energía en los coeficientes de la 2D-VDCT.
    - Comparación por medio de: ICF, compactación de la energía y matriz de auto correlación, de las siguientes fuentes: coeficientes de la 2D-VDCT, 2D-BDCT, Laplaciana independiente idénticamente distribuida, Gaussiana independiente idénticamente distribuida, Gauss-Markov de primer orden a=0.9 y vectores en el dominio imagen.

- Justificación de estacionariedad y ergodicidad en los coeficientes de la 2D-VDCT, por medio del tamaño de las secuencias de entrenamiento.
    - Cuantificador vectorial del vecino más cercano.
    - Cuantificador clasificado.

- Comparación del desempeño entre el cuantificador del vecino más cercano y los cuantificadores clasificados propuestos.
    - Gráficas de relación señal a ruido en función de la tasa del cuantificador *sin* codificación entrópica. Comparaciones realizadas sobre las siguientes fuentes: 2D-VDCT.C0, 2D-VDCT.C1, 2D-VDCT.C24, 2D-VDCT.C28, 2D-VDCT.C63, Laplaciana iid, Gaussiana iid, Gauss-Markov primer orden, e imágenes.
    - Gráficas de relación señal a ruido en función de la tasa del cuantificador *con* codificación entrópica. Comparaciones realizadas sobre las siguientes fuentes: 2D-VDCT.C0, 2D-VDCT.C1, 2D-VDCT.C24, 2D-VDCT.C28, 2D-VDCT.C63, Laplaciana iid, Gaussiana iid, Gauss-Markov primer orden e imágenes.

- Esquema de compresión de imágenes por medio de la 2D-VDCT.
    - Características de las imágenes utilizadas.
    - Características de las secuencias de entrenamiento.
    - Medidas de comparación.
    - Resultados en la bibliografía.
    - Resultados obtenidos.

- Comparaciones en base a tiempos de proceso y espacios de almacenamiento.
    - Cálculo de los umbrales del clasificador por los métodos: atributos óptimos y geometrías intrínsecas.
    - Tiempo de cálculo en el diseño de los cuantificadores vectoriales del vecino más cercano y clasificado.
    - Tiempo de codificación por medio de los cuantificadores vectoriales: del vecino más cercano y clasificado.

- Herramientas de desarrollo y comparación.
    - Cálculo de la 2D-VDCT y la 2D-BDCT.
    - Caracterización de fuentes aleatorias *n*-dimensionales.
    - Diseño de cuantificadores vectoriales: VQ, CVQ_{AO} y CVQ_{GI}.
    - Caracterización y procesamiento de imágenes.
    - Caracterización de secuencias de entrenamiento.
    - Graficación.

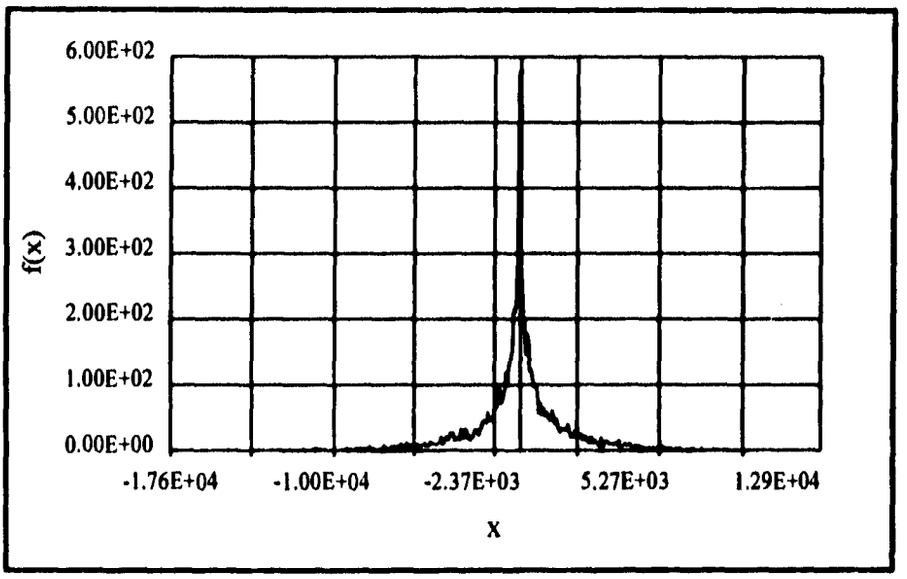
5.3 Caracterización de los coeficientes derivados de la 2D-VDCT

Para poder buscar un método efectivo de cuantificación vectorial de los coeficientes de la 2D-VDCT, es conveniente ratificar la suposición de que la forma de las distribuciones de estos coeficientes derivados del procesamiento de imágenes

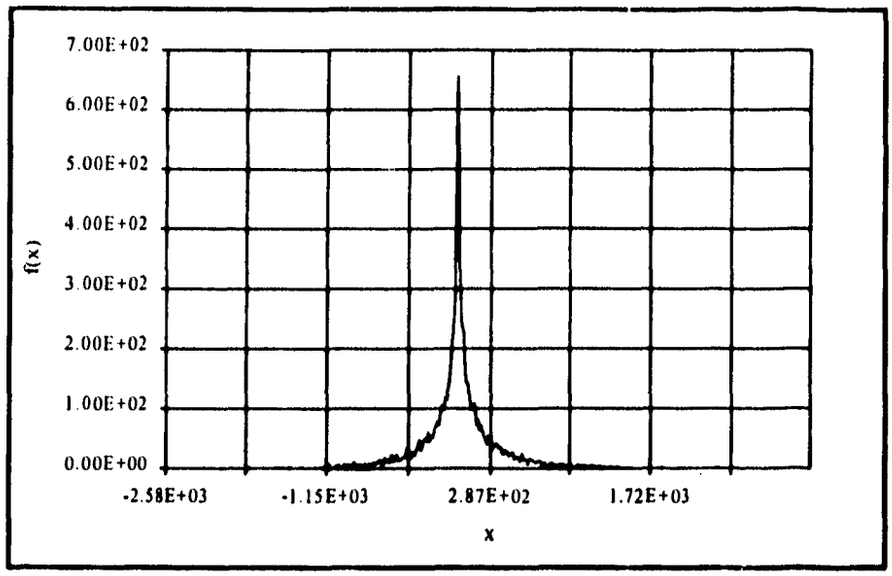
son: Gaussiana en el coeficiente *C0* y Laplacianas en el resto. Para comprobar esta suposición, la herramienta más adecuada es la prueba de Kolmogorov-Smirnov [VQ34, VQ60], la cual implica el cálculo de una distancia entre una función de distribución teórica propuesta y una función distribución muestra (histograma). Como el presente trabajo no es necesariamente de un estudio riguroso de esas distribuciones, únicamente se consideró indispensable hacer una comparación subjetiva entre las distribuciones muestra y las ya bien conocidas curvas de las distribuciones Gaussiana y Laplaciana.

En las gráficas 5.3.1, 5.3.2 y 5.3.3 se pueden observar los histogramas, obtenidos a partir del procesamiento de imágenes, para secuencias de entrenamiento de los coeficientes C1, C24 y C63 de la 2D-VDCT. Estos histogramas están construidos a partir de 512 segmentos distribuidos en el intervalo de variación de cada componente. Por cuestiones de espacio únicamente se muestran los histogramas de la componente (0,0) de los vectores (estos *vectores* realmente son matrices de dimensión 4x4), haciéndose notar que los histogramas de las otras componentes presentan comportamientos muy similares. Como era de esperarse, los histogramas presentan la forma conocida de la distribución Laplaciana.

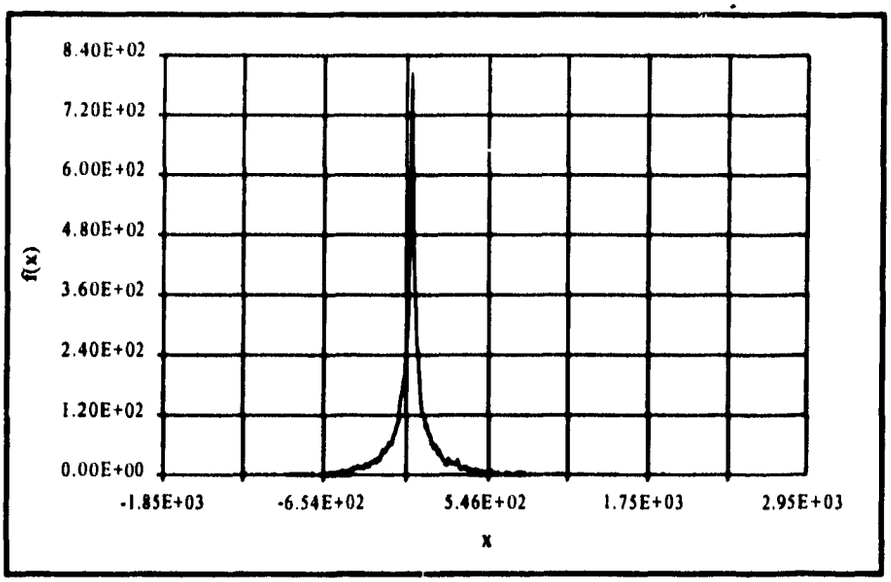
En los coeficientes modelados por distribuciones Laplacianas, se puede observar que la varianza de las componentes disminuye en relación a las frecuencias espaciales asociadas con el coeficiente.



Gráfica 5.3.1 Histograma del coeficiente C1, componente 0, de la 2D-VDCT (8x8,n=16).



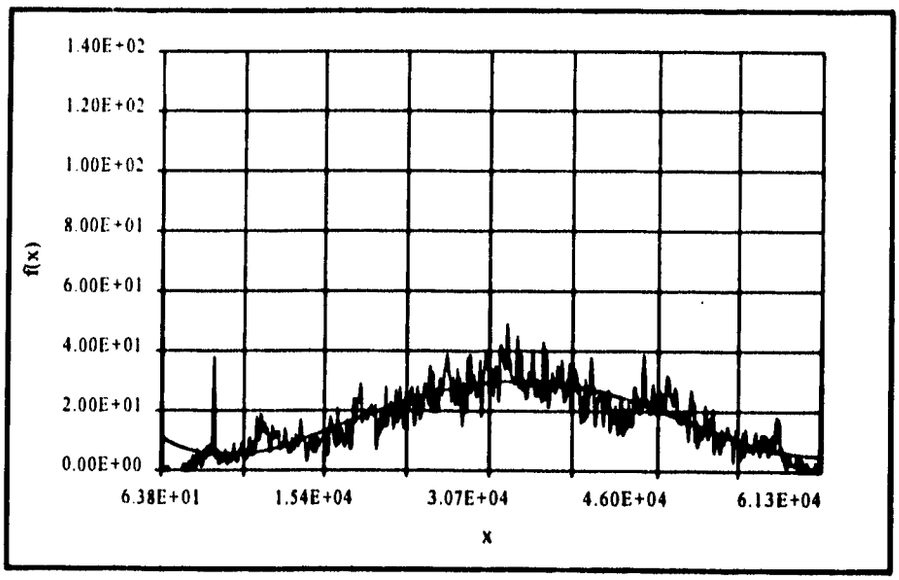
Gráfica 5.3.2. Histograma del coeficiente C24, componente 0, de la 2D-VDCT (8x8,n=16).



Gráfica 5.3.3. Histograma del coeficiente C63, componente 0 de la 2D-VDCT (8x8,n=16).

En el caso del coeficiente *C0* de la 2D-VDCT, supuestamente con componentes con distribuciones Gaussianas, la forma del histograma no concuerda tan claramente con la forma de este tipo de distribución (ver figura 5.3.4). Sin

embargo, cuando se suaviza este histograma, el parecido aumenta notablemente. Aquí, al igual que en el caso de los coeficientes Laplacianos, por cuestiones de espacio, únicamente se presenta la distribución de la componente (0,0) del vector.

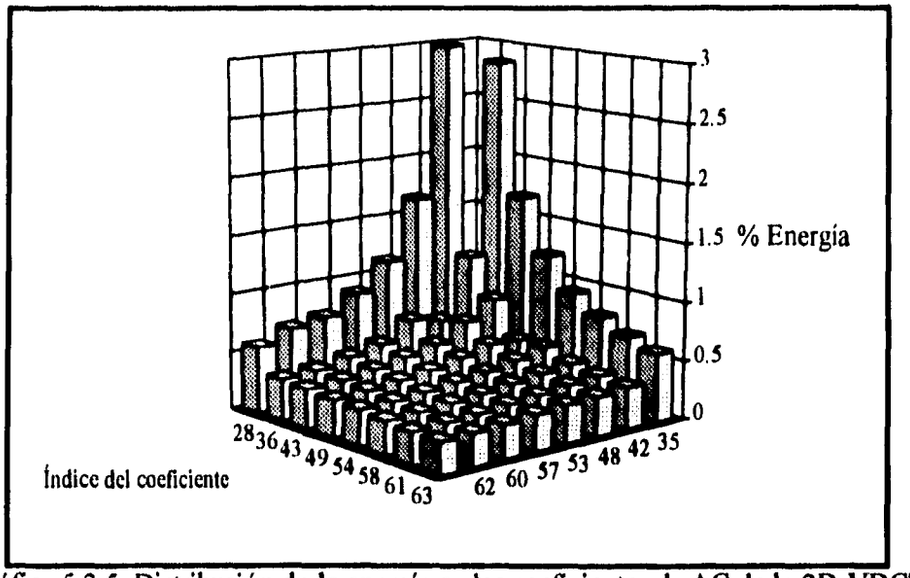


Gráfica 5.3.4. Histograma del coeficiente C0, componente 0, de la 2D-VDCT (8x8,n=16).

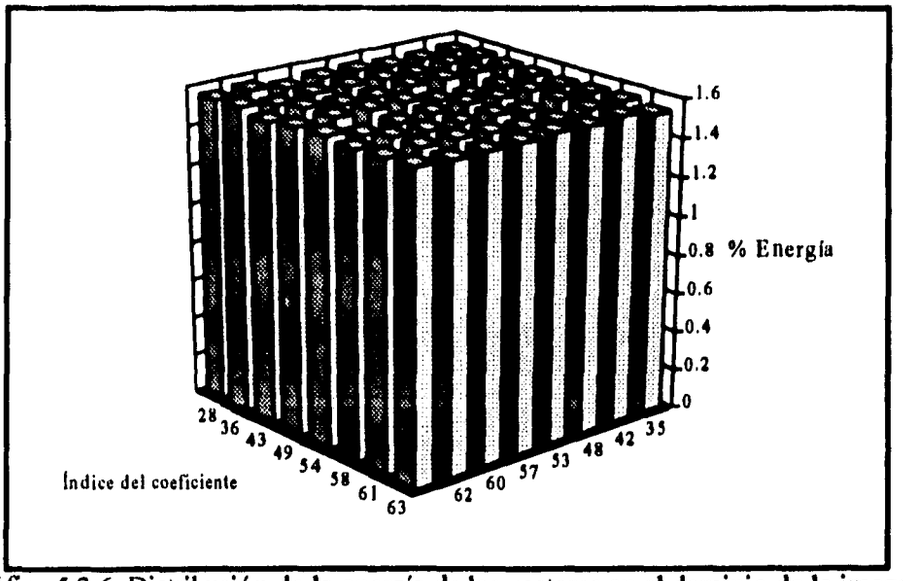
Una vez que se tiene una medida subjetiva de la forma de las distribuciones de los coeficientes, es deseable corroborar que la transformación vectorial cumple con las funciones de preservación de la dependencia entre las componentes de los coeficientes y reducción de la correlación entre los coeficientes. Estos atributos, son los que en un esquema de compresión de datos, la hacen adecuada como etapa de procesamiento previo a la cuantificación vectorial. Los atributos antes mencionados pueden ser medidos por medio del factor de acoplamiento intrínseco y la compactación de la energía respectivamente.

Como se puede ver en la figura 5.3.5, el grado de concentración de la energía en los coeficientes en el dominio de la 2D-VDCT es mucho mayor que el que se tiene entre los vectores en el dominio de la imagen (figura 5.3.6). En el dominio transformado, la mayor parte de la energía, 62.25%, está contenida en el coeficiente de *DC*. Aún cuando este coeficiente no es graficado en la figura 5.3.5, se puede observar que unos pocos coeficientes de *AC*, los correspondientes a los bordes verticales y horizontales puros y entre ellos los de más baja frecuencia, concentran el resto de la energía. En contraste, en el dominio de la imagen, la distribución de la energía es casi uniforme en todos los vectores. Entre más concentrada esté la energía

en unos pocos coeficientes transformados, menor será la correlación entre ellos, y en consecuencia más efectiva será la cuantificación vectorial [VQ5, VQ22, VQ28, VQ58]. Por lo tanto vemos que la 2D-VDCT está cumpliendo con el segundo atributo que debe tener una etapa de procesamiento, para ser adecuada para la VQ.



Gráfica 5.3.5. Distribución de la energía en los coeficientes de AC de la 2D-VDCT (8x8, n=16). El coeficiente de DC, el cual no se muestra, contiene el 62.25% del total de la energía de la secuencia de entrenamiento.



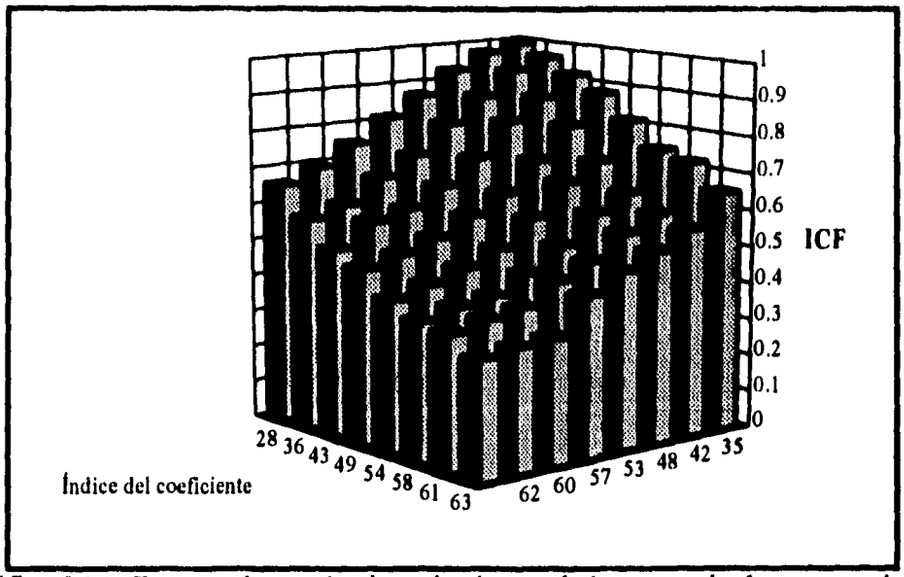
Gráfica 5.3.6. Distribución de la energía de los vectores en el dominio de la imagen.

Aunque la 2D-VDCT tiene un menor grado de compactación de la energía que el que se logra con una 2D-BDCT escalar [VQ5], la preservación de la correlación entre las componentes de los vectores es la que la hace superior como etapa de procesamiento previo a la cuantificación vectorial. La medida que nos permite sopesar la correlación intravector es el factor de acoplamiento intrínseco. En la gráficas 5.3.7 y 5.3.8 se muestran los valores de ICF para los coeficientes de la 2D-VDCT y para la 2D-BDCT respectivamente. Como se puede observar, todos los coeficientes de la 2D-VDCT tienen componentes mucho más fuertemente acopladas que los coeficientes de la 2D-BDCT. El coeficiente con menor grado de correlación entre sus componentes es el coeficiente de más alta frecuencia, el cual tiene un ICF=0.335 en la 2D-VDCT y 0.021 en la 2D-BDCT.

Como se vio en el capítulo 3, el factor de acoplamiento intrínseco es una función del determinante de la matriz de covarianza del vector en cuestión. Como en este caso no se tiene una expresión matemática para el modelo de correlación entre las componentes de los coeficientes de la 2D-VDCT, no es posible encontrar una expresión analítica para el factor de acoplamiento intrínseco, por lo cual es necesario utilizar un estimador sobre una secuencia de entrenamiento. En este caso se utilizó el estimador no sesgado de la matriz de correlación muestra promedio de término largo, dado por la siguiente ecuación

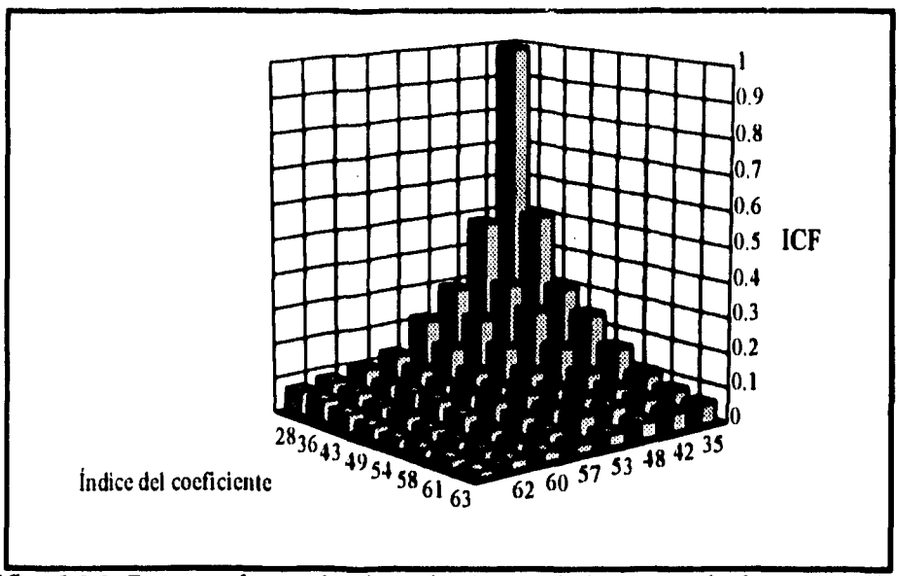
$$\boldsymbol{R}_X = \frac{1}{N-1}\sum_{X \in R} XX^T$$

en donde N es la cardinalidad del conjunto R y $\boldsymbol{R}_X$ es una matriz cuadrada de dimensión nxn.



Gráfica 5.3.7. Factores de acoplamiento intrínseco de la secuencia de entrenamiento en el dominio de la 2D-VDCT (8x8,n=16).

Es conveniente mencionar que aunque los coeficientes en el dominio de la 2D-VDCT tienen un factor de acoplamiento intrínseco menor al que tienen los vectores en el dominio de la imagen [VQ5, VQ28], la compactación de la energía debida a la transformada hace que la VQ sea más eficiente en ese dominio. Aún no se tiene una medida clara de bajo que condiciones cada uno de esos atributos sea más significativo para el desempeño de la VQ sobre un conjunto de vectores [VQ22].

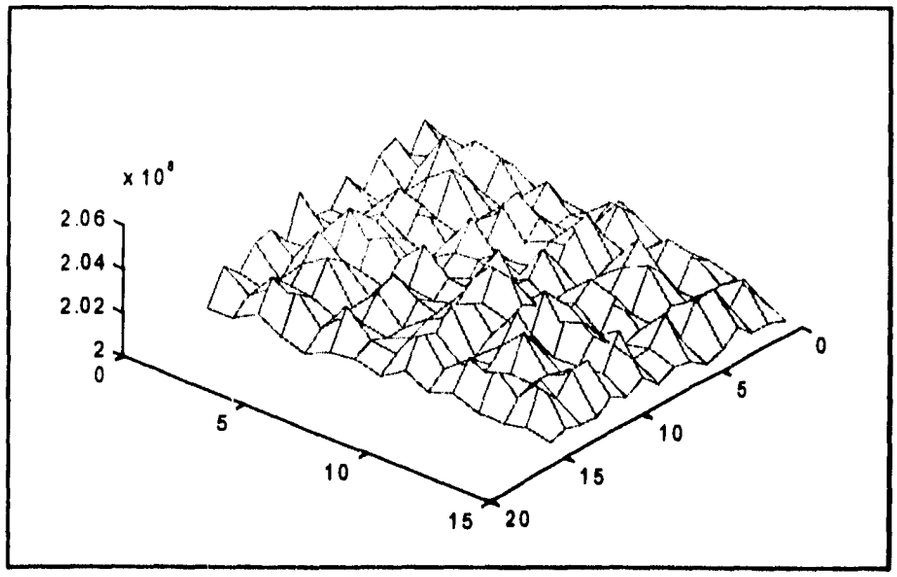


Gráfica 5.3.8. Factores de acoplamiento intrínseco de la secuencia de entrenamiento en el dominio de la 2D-BDCT (32x32).

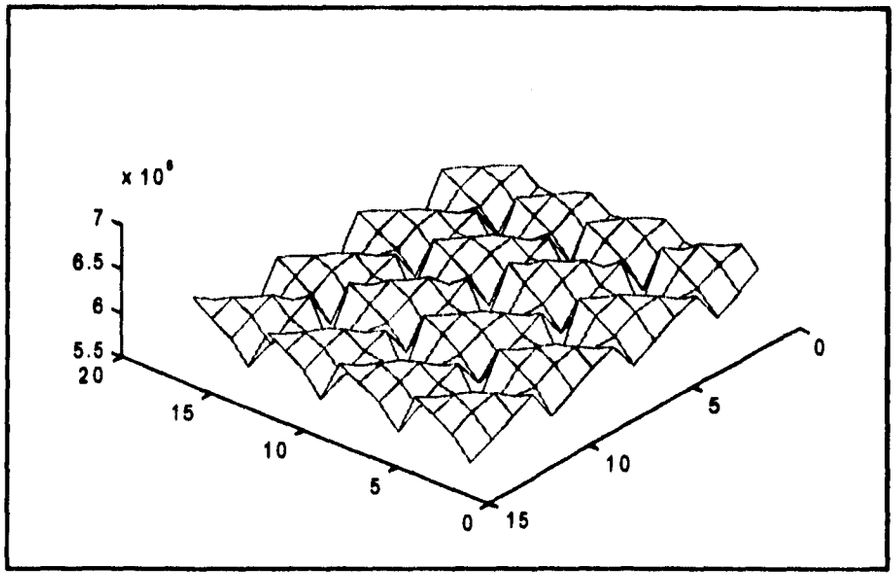
Para continuar con la caracterización de los coeficientes de la transformada 2D-VDCT se presentan las gráficas de las matrices de covarianza de estos coeficientes y se comparan con las matrices obtenidas sobre datos modelados por procesos aleatorios comúnmente encontrados en la bibliografía. Aquí, aunque para algunos de los modelos se tiene la expresión matemática de la función de densidad de probabilidad conjunta, las matrices de autocovarianza se calcularon con el mismo estimador que se utilizó para los coeficientes de la 2D-VDCT.

En las gráficas 5.3.9 a la 5.3.12 se pueden ver los estimados de las matrices de covarianza de los coeficientes de la 2D-VDCT. Como era de esperarse, el coeficiente con mayor factor de acoplamiento intrínseco, el coeficiente C0 con ICF=0.997, presenta una fuerte correlación entre cada uno de los pares posibles de componentes. Esto es, cada componente está fuertemente correlacionada con las restantes. Conforme la frecuencia asociada con el coeficiente crece, se puede observar que la correlación se concentra cada vez más en la diagonal principal,

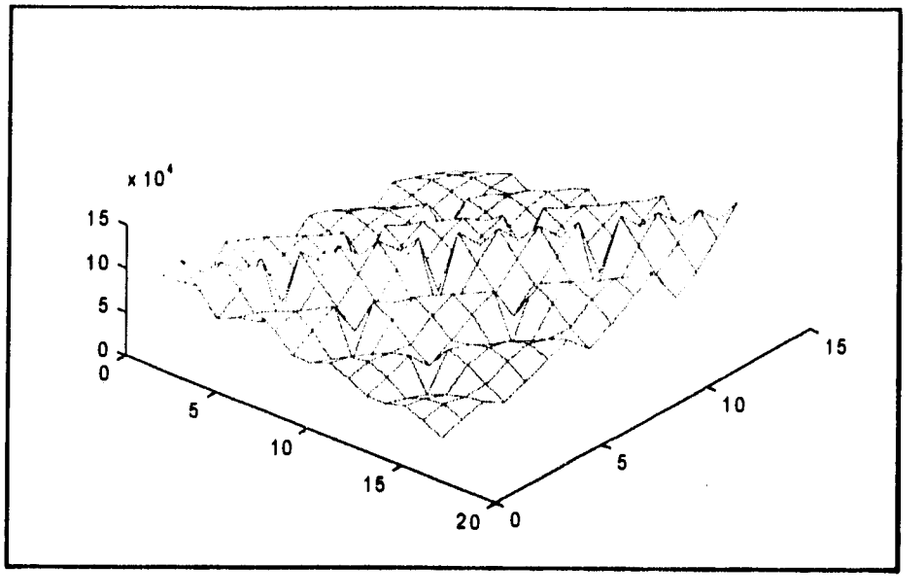
aunque hay que notar que en el caso del coeficiente con componentes menos acopladas, el *C63* con ICF=0.335, todavía hay correlación fuerte fuera de la diagonal principal.



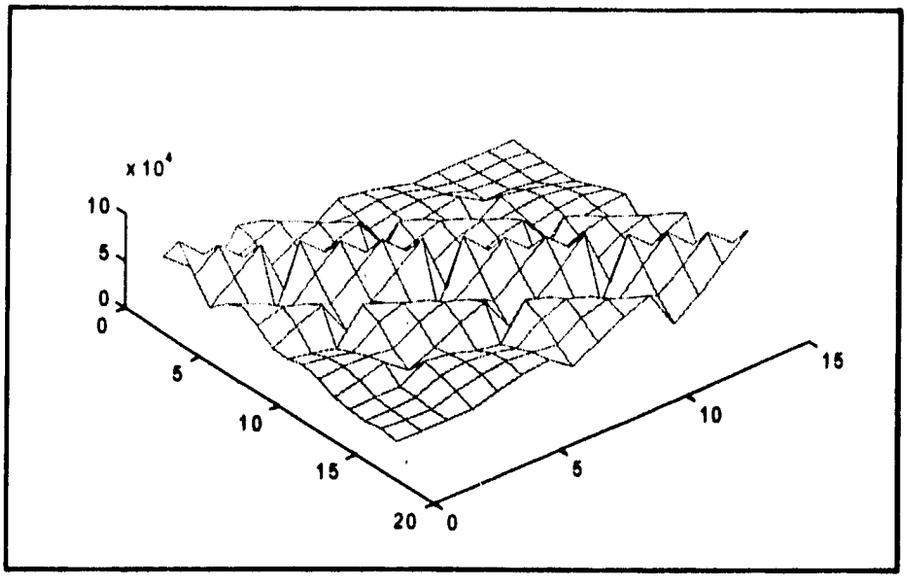
Gráfica 5.3.9. Matriz de covarianza del coeficiente C0 de la secuencia de entrenamiento, en el dominio de la 2D-VDCT (8x8, n=16).



Gráfica 5.3.10. Matriz de covarianza del coeficiente C1 de la secuencia de entrenamiento en el dominio de la 2D-VDCT (8x8,n=16).



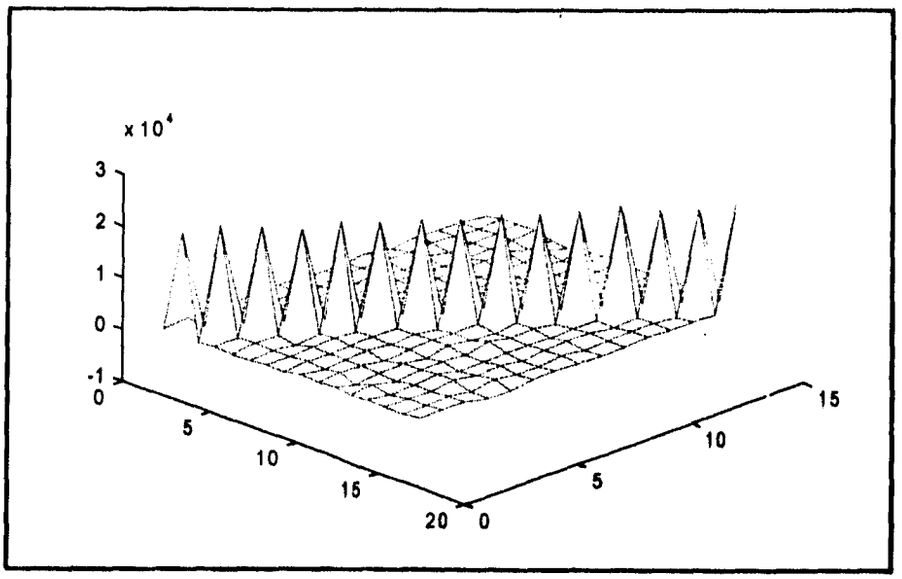
Gráfica 5.3.11 Matriz de covarianza del coeficiente C24 de la secuencia de entrenamiento en el dominio de la 2D-VDCT (8x8,n=16).



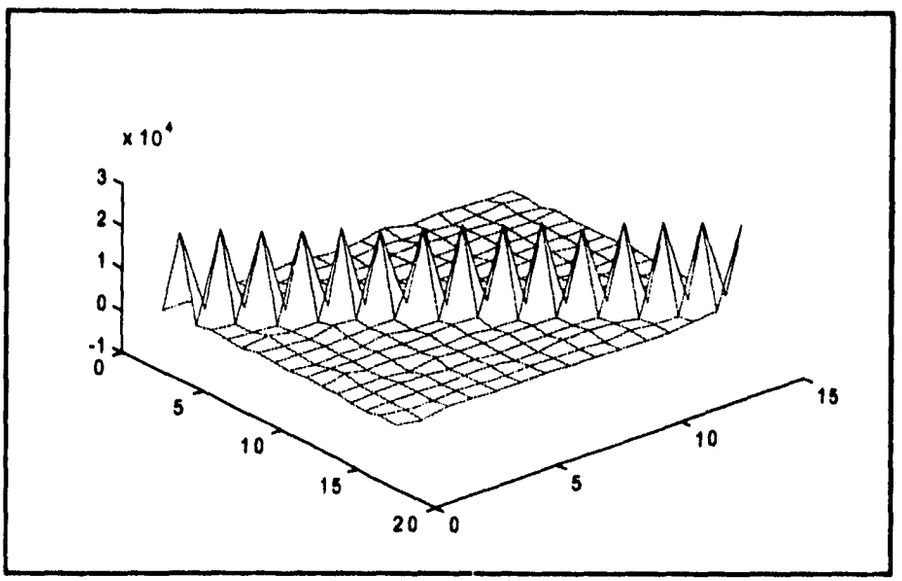
Gráfica 5.3.12 Matriz de covarianza del coeficiente C63 de la secuencia de entrenamiento en el dominio de la 2D-VDCT (8x8,n=16).

En el caso de las fuentes aleatorias Gaussiana y Laplaciana independientes idénticamente distribuidas (ver gráficas 5.3.13 y 5.3.14) se puede observar que los elementos de la diagonal principal en la matriz de covarianza tienen la misma

magnitud. Además, los elementos fuera de esta diagonal tienen correlación despreciable. Esto concuerda con la definición de dichos procesos.



Gráfica 5.3.13. Matriz de correlación del proceso Gaussiano independiente idénticamente distribuido, 320000 muestras.



Gráfica 5.3.14. Matriz de correlación del proceso Laplaciano independiente idénticamente distribuido, 320000 muestras.

Para el caso del proceso Gauss-Markov de primero orden, la matriz de correlación está definida como [VQ52]

$$\boldsymbol{R}_{X=}(i,j) = a^{|i-j|}\sigma^2$$

para generar la secuencia de entrenamiento, en el presente caso se utilizaron los valores a=0.9 y σ^2=5.0E12. Como se puede ver en la gráfica 5.3.15, los valores arrojados por el estimador de la matriz de correlación presentan una buena aproximación a los valores teóricos.

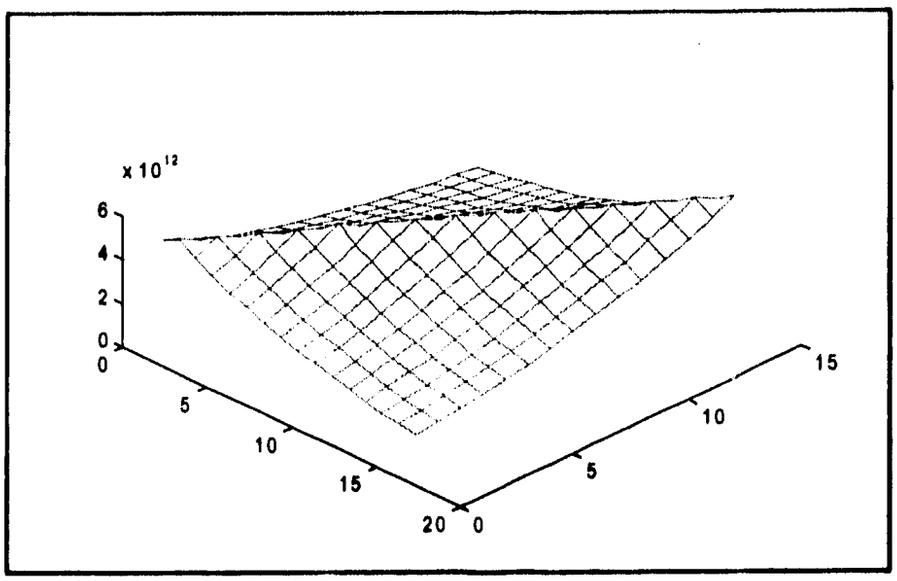


Figura 5.3.15 Matriz de correlación del proceso Gauss-Markov de primero orden, a=0.9,σ^2=5.0E12, 320000 muestras.

En el planteamiento del diseño del clasificador basado en los atributos óptimos, fue necesario relacionar el error normalizado en función de la tasa del código y el factor de acoplamiento intrínseco. Como no se tenía un modelo matemático de las funciones de densidad de probabilidad de los coeficientes de la 2D-VDCT, fue necesario obtener esta relación en base a resultados experimentales. En la gráfica 5.3.14 se muestran las curvas del error normalizado en función del factor de acoplamiento intrínseco para diferentes tasas del libro de códigos. Cada una de estas curvas fue aproximada por medio de un polinomio de primer orden, obteniendose las siguientes expresiones:

$$\overline{Error}_{r=6bpv} = -0.6223 \cdot ICF + 0.7080$$

$$\overline{Error}_{r=5bpv} = -0.8229 \cdot ICF + 0.8197$$

$$\overline{Error}_{r=4bpv} = -0.9178 \cdot ICF + 0.9316$$

$$\overline{Error}_{r=3bpv} = -0.9070 \cdot ICF + 1.0600$$

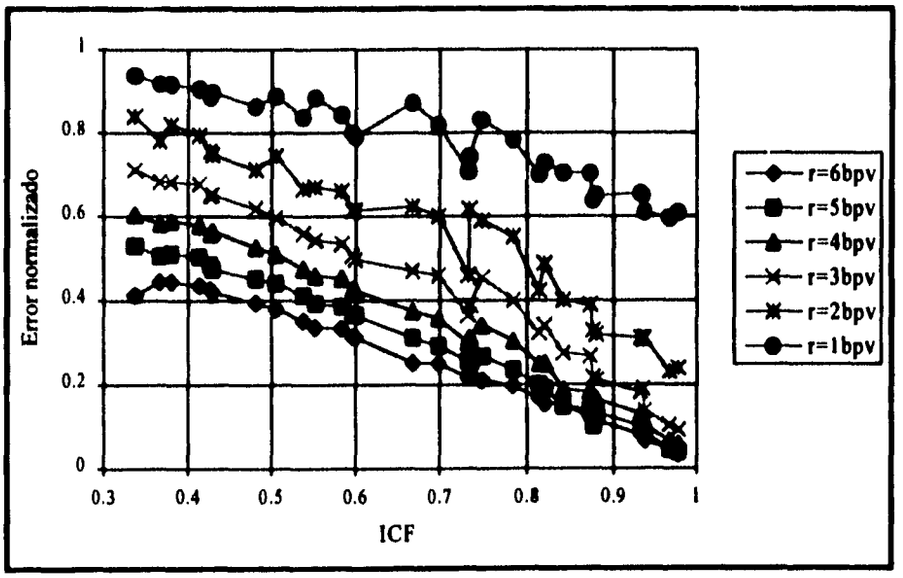
$$\overline{Error}_{r=2bpv} = -0.7907 \cdot ICF + 1.1626$$

$$\overline{Error}_{r=1bpv} = -0.5686 \cdot ICF + 1.1237$$

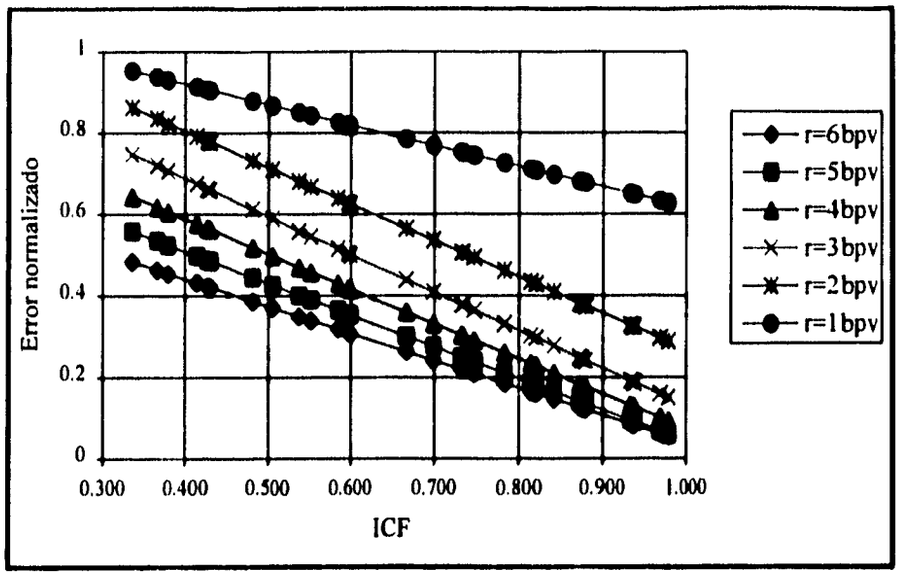
Los polinomios correspondientes a los valores experimentales de la gráfica 5.3.16 se muestran en la gráfica 5.3.17. En base a los valores de α y β de cada uno de estos polinomios se obtuvieron dos funciones que relacionan a esas variables con la tasa del código (ver gráfica 5.3.18). Para ajustar los valores de α y β se utilizaron polinomios de segundo orden, los cuales dan las siguientes relaciones:

$$\alpha(r) = 0.0528 \cdot r^2 - 0.3590 \cdot r - 0.3162$$

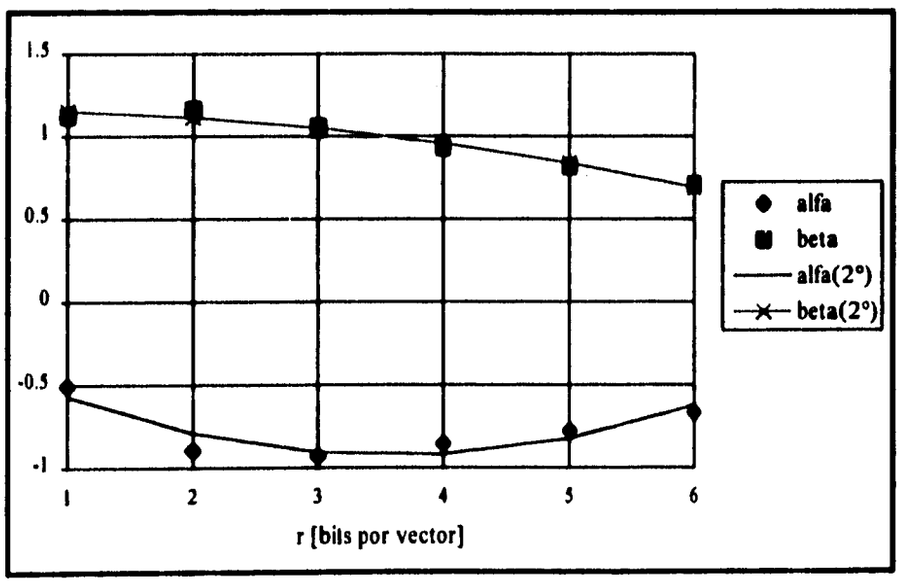
$$\beta(r) = -0.0141 \cdot r^2 + 0.1912 \cdot r + 0.5123$$



Gráfica 5.3.16. Error normalizado vs. factor de acoplamiento intrínseco en fuentes Laplacianas.



Gráfica 5.3.17. Error normalizado vs factor de acoplamiento intrínseco, polinomios de primer orden.



Gráfica 5.3.18. Coeficientes de los polinomios del error normalizado para diferentes tasas del código.

5.4 Justificación de estacionariedad y ergodicidad

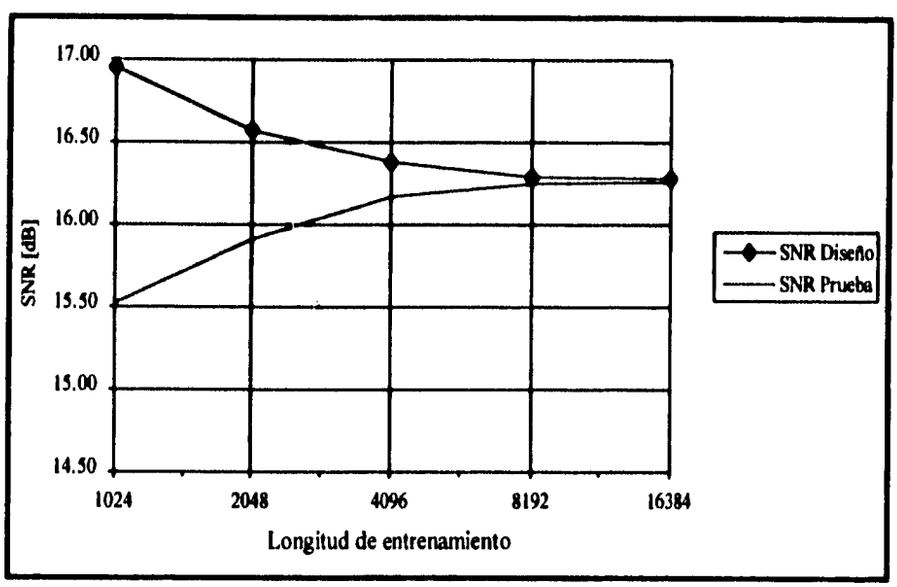
En el diseño de un cuantificador vectorial, se busca encontrar un código que sea óptimo bajo ciertas condiciones. Generalmente, la condición utilizada es la minimización del valor esperado de la distorsión, habiendo fijado previamente un valor para la tasa del código.

En la práctica la distorsión es el promedio del término largo de la distorsión muestra. Si el proceso es estacionario y ergódico, entonces este promedio es igual al valor esperado matemático. El valor esperado de la distorsión es útil para desarrollar límites teóricos en la teoría de la información, pero frecuentemente son imposibles de calcular ya que es necesario tener una expresión matemática para las distribuciones de probabilidad. Por lo tanto, un método para el diseño de sistemas es trabajar con una descripción implícita de la distribución, esto es, tomar secuencias de datos de entrenamiento, estimar la *verdadera* pero desconocida distorsión por medio del promedio del término largo y tratar de diseñar un código que minimice la distorsión promedio muestra para la secuencia de entrenamiento. Si la fuente es en efecto estacionaria y ergódica, el promedio muestra resultante debe ser muy cercano al valor esperado y el mismo código usado en datos futuros debe producir aproximadamente los mismos promedios. Cuando las fuentes no son estacionarias ni ergódicas, la propiedad deseada es que si se diseña un código en una secuencia de entrenamiento suficientemente larga y entonces se usa el código en datos futuros producidos por la misma fuente, entonces el desempeño del código en datos nuevos debe ser aproximado al alcanzado en los datos de entrenamiento. El punto de interés teórico es proporcionar condiciones bajo las cuales esta declaración pueda ser rigurosa. Para medidas de distorsión razonables, una condición suficiente para que esto sea verdadero, para un diseño del cuantificador vectorial sin memoria, es que la fuente sea asintóticamente estacionaria en la media [VQ1, VQ49, VQ52, VQ68]. Las fuentes asintóticamente estacionarias en la media incluyen a todas las fuentes estacionarias, ciclo estacionarias y asintóticamente estacionarias.

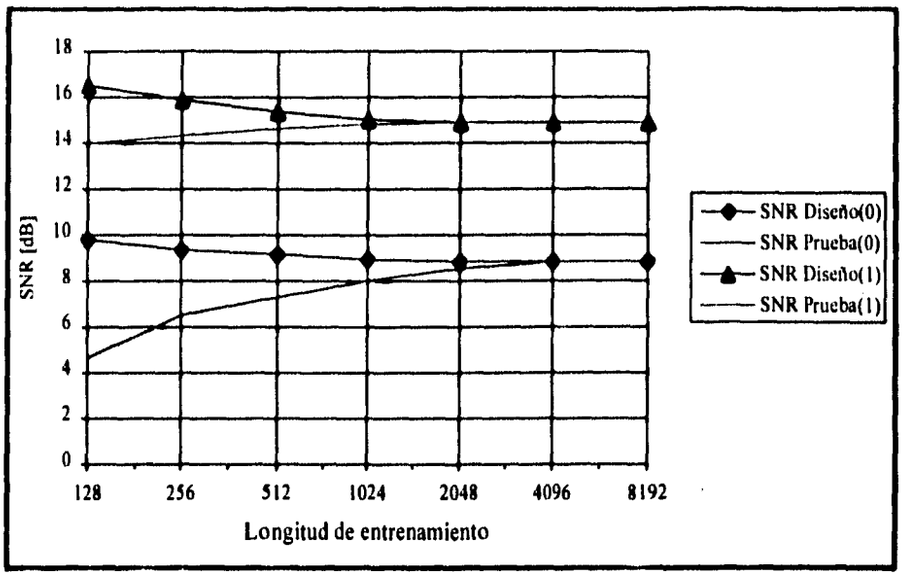
Con este razonamiento se justifica que el método de diseño general usando secuencias de entrenamiento largas, no requiere estacionariedad ni ergodicidad para tener una base teórica sólida. Entonces el método de diseño, basado en secuencias de entrenamiento, para fuentes asintóticamente estacionarias en la media es: tratar de diseñar un código que minimice la distorsión muestra promedio para una secuencia de entrenamiento muy larga. Entonces usar el código en una secuencia de prueba producida por la misma fuente, pero que no esté en la secuencia de entrenamiento. Si el desempeño es razonablemente cercano al valor de diseño, entonces se tiene cierta seguridad de que el código continuará produciendo aproximadamente el mismo desempeño en el futuro. Si el desempeño de diseño y prueba son muy diferentes, entonces probablemente la secuencia de entrenamiento no fue lo suficientemente larga. En otras palabras, no hay que tratar de probar matemáticamente que una

fuente es asintóticamente estacionaria en la media, en lugar de eso hay que tratar de diseñar códigos para ésta y entonces ver si funcionan con datos nuevos [VQ1, VQ5].

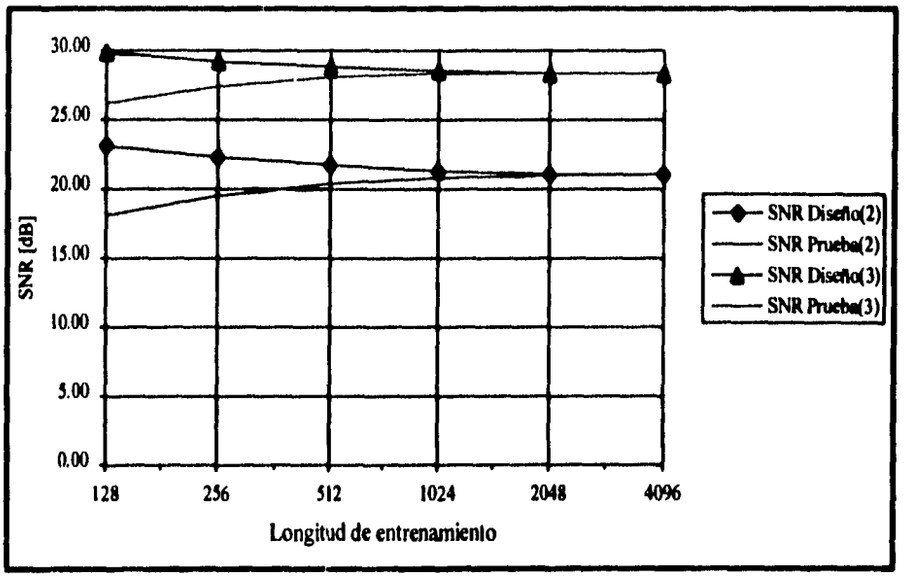
En la elección del tamaño de la secuencia de entrenamiento hay que tomar en cuenta dos factores: el primero es que debe ser lo suficientemente larga como para modelar a la fuente como un proceso, al menos, asintóticamente estacionario en la media. El segundo, es que si la secuencia es demasiado larga, la cantidad de cálculo requerida para el diseño del cuantificador será excesivamente grande. Por lo tanto hay que hallar un equilibrio entre estos dos puntos para obtener un tamaño adecuado. En las gráficas 5.4.1 a la 5.4.3 se compara el desempeño de diseño y de prueba del cuantificador vectorial, contra la longitud de la secuencia, en las secuencias de entrenamiento del VQ del vecino más cercano y del los cuatro modos de operación del VQ clasificado.



Gráfica 5.4.1. SNR (VQ@8bpv) vs. longitud de la secuencia de entrenamiento en el coeficiente C1 de la 2D-VDCT (8x8,n=16).



Gráfica 5.4.2. SNR (CVQ modos 0 y 1 @6bpv) vs longitudes de la secuencia de entrenamiento en el coeficiente C1 de la 2D-VDCT (8x8,n=16).



Gráfica 5.4.3. SNR (CVQ modos 2 y 3 @6bpv) vs. longitud de la secuencia de entrenamiento en el coeficiente C1 de la 2D-VDCT (8x8,n=16).

En base a las gráficas anteriores, el tamaño de la secuencia de entrenamiento que finalmente se eligió para hacer los diseños de los cuantificadores vectoriales fue de 8768 vectores para VQ, 4096 vectores para CVQ modo 0, 2048 vectores para

CVQ modos 1 y 2, y 1024 vectores para CVQ modo 3. Con estas longitudes el desempeño de prueba esta entre 0.1dB y 0.2 dB del desempeño de diseño. Las secuencias de entrenamiento son el resultado del procesamiento de 75 imágenes de la vida real, 8 sintéticas y 4 biomédicas. No todas las imágenes son del mismo tamaño. Las dimensiones más comunes son de 512x512, 256x256 y 512x480. Como secuencias de prueba se utilizaron los vectores resultantes de procesar, por medio de la 2D-VDCT, 4 imágenes no incluidas en las secuencias de entrenamiento. Estas imágenes son:

Tabla 5.4.1 Imágenes en la secuencia de prueba

Imagen	Dimensiones	Vectores/coef.
Girl.raw	512x512	256
Lady.raw	256x256	64
Lena.raw	512x512	256
Peppers.raw	512x512	256

5.5 Comparación del desempeño VQ vs CVQ

Cuando en el capítulo anterior, se planteó la forma de construir los cuantificadores clasificados, en base a las geometrías intrínsecas y a los atributos óptimos, no se obtuvo una expresión que garantizara que el desempeño obtenido con el cuantificador clasificado fuera mejor que el desempeño del cuantificador vectorial del vecino más cercano ordinario. Es más, la motivación que llevo al planteamiento de la hipótesis:

La aplicación de un cuantificador clasificado a las secuencias de entrenamiento en el dominio de la 2D-VDCT, puede producir un mejor desempeño, en el sentido tasa-distorsión, que el cuantificador vectorial del vecino más cercano ordinario.

nace de apreciaciones puramente experimentales, entonces para comprobar si la hipótesis es verdadera o falsa, se debe realizar una comparación del desempeño de los dos tipos de cuantificadores. Esta comparación se hizo de la siguiente forma:

- El número de modos de operación del cuantificador clasificado, se fijo en 4.
- El tamaño de la secuencia de entrenamiento del cuantificador clasificado se eligió de tal forma que los subconjuntos de entrenamiento de cada uno de los 4 modos de operación cumplieran con las longitudes necesarias para garantizar la ergodicidad (gráficas 5.4.2 y 5.4.3).
- Como el tamaño obtenido para la secuencia de entrenamiento del cuantificador clasificado resultó ser mayor que el requerido para justificar la ergodicidad en el caso del cuantificador del vecino más cercano ordinario y

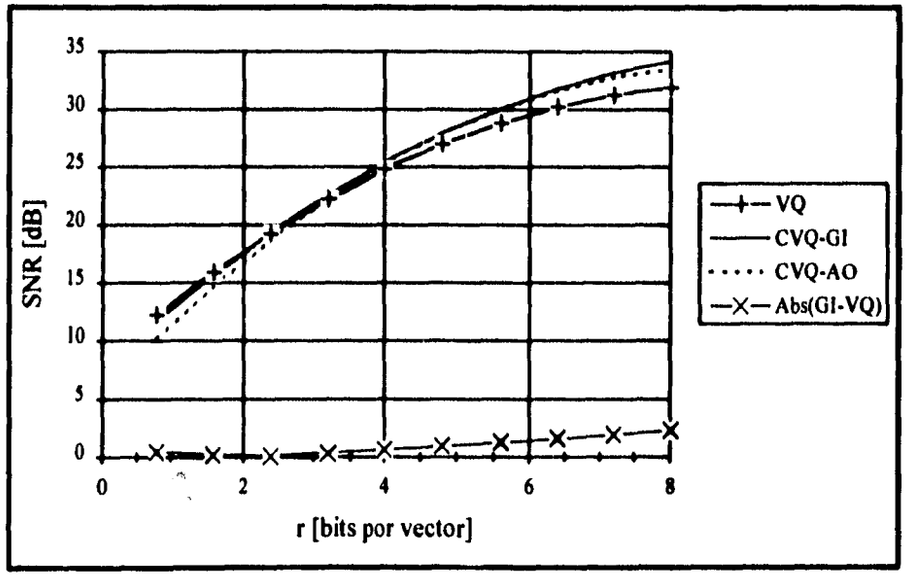
como la comparación se debía hacer bajo las mismas condiciones, entonces la misma secuencia de entrenamiento del cuantificador clasificado se utilizó como secuencia de entrenamiento del cuantificador del vecino más cercano ordinario.

- Para distintas tasas del código, se obtuvo la relación señal a ruido de cuantificación, SNR (en decibeles), dada por

$$SNR = 10\log_{10}\left(\frac{E\{X^T \cdot X\}}{E\{\|X - Q(X)\|^2\}}\right) = 10\log_{10}\left(\frac{\sum_{X\in R} X^T \cdot X}{\sum_{X\in R}\|X - Q(X)\|^2}\right)$$

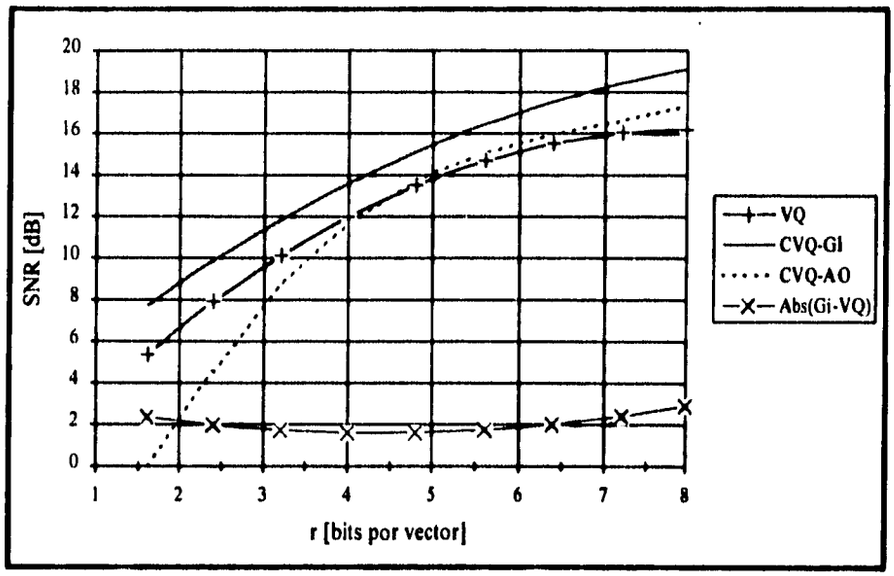
para los dos tipos de cuantificadores. Con los resultados obtenidos se construyeron las curvas SNR contra la tasa del código en bits por vector.

Como en este trabajo se utilizó la 2D-VDCT de 8x8=64 vectores, entonces, por cuestiones de espacio, no es práctico mostrar las gráficas tasa-distorsión de cada uno de ellos. En lugar de mostrar todas las gráficas se determinó que mostrando los resultados para el coeficiente Gaussiano (C0), un Laplaciano de baja frecuencia (C1) y uno de alta frecuencia (C63), se tendría un panorama general de la diferencia del desempeño entre los dos tipos de cuantificadores. En las gráficas 5.5.1 a 5.5.3 se puede observar la SNR de cuantificación, contra la tasa del código del cuantificador clasificado (con los dos tipos de clasificadores), comparado contra el cuantificador del vecino más cercano.



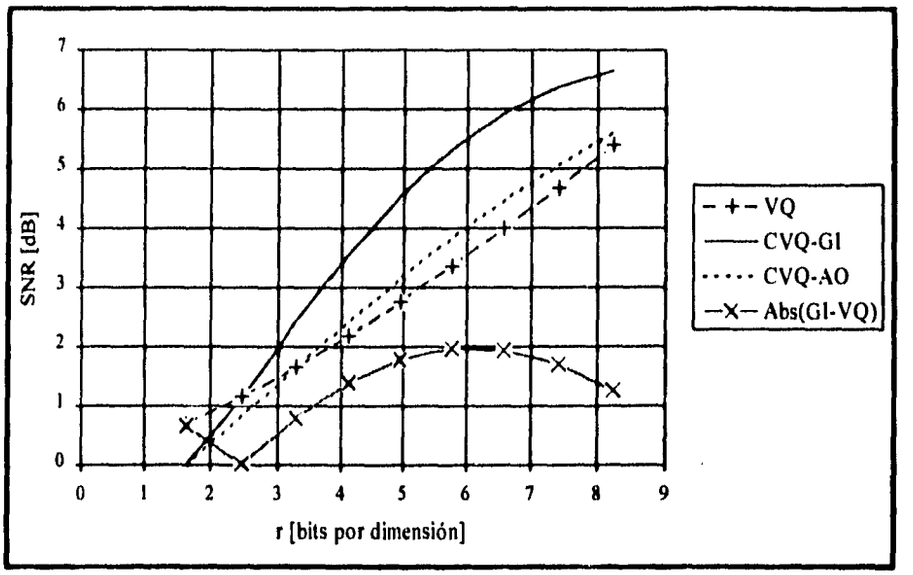
Gráfica 5.5.1 CVQ vs VQ en el coeficiente C0 de la secuencia de entrenamiento en el dominio de la 2D-VDCT (8x8,n=16).

En el caso del coeficiente Gaussiano o de DC (gráfica 5.5.1), se puede observar que el desempeño de los dos cuantificadores clasificados es bastante parecido, aunque el que se basa en las geometrías implícitas (CVQ-GI) mantiene una ligera ganancia sobre el que se basa en el factor de acoplamiento intrínseco y la compactación de energía (CVQ-AO). Los cuantificadores clasificados son una mejor alternativa que el cuantificador del vecino más cercano ordinario (VQ) sólo para tasas del código mayores a 3.5 bits por vector. Esto se puede deber a que a tasas bajas la cantidad de información utilizada para representar la clasificación viene a significar una parte muy importante de la tasa total del código, acentuando el hecho de que el cuantificador clasificado es subóptimo.



Gráfica 5.5.2. CVQ vs VQ en el coeficiente C1 de la secuencia de entrenamiento en el domino 2D-VDCT (8x8,n=16).

En el caso del coeficiente Laplaciano de baja frecuencia C1, el cuantificador clasificado obtenido por el criterio de las geometrías implícitas (CVQ-GI) produce mejores resultados que los otros dos cuantificadores, el del vecino más cercano (VQ) y el clasificador basado en el criterio de los atributos óptimos (CVQ-AO). La distancia entre los dos cuantificadores clasificados es mayor en las tasas bajas, permaneciendo casi constante a partir de r=4.5 bits por vector, punto en el cual el CVQ-AO sobrepasa al VQ. Sin embargo, en la codificación de la fuente aleatoria C1 por medio del cuantificador clasificado, es imposible que con cuatro modos de operación se puedan tener tasas de código inferiores a los 1.5 bits por vector, debido a la tasa dedicada a la representación de la clasificación.



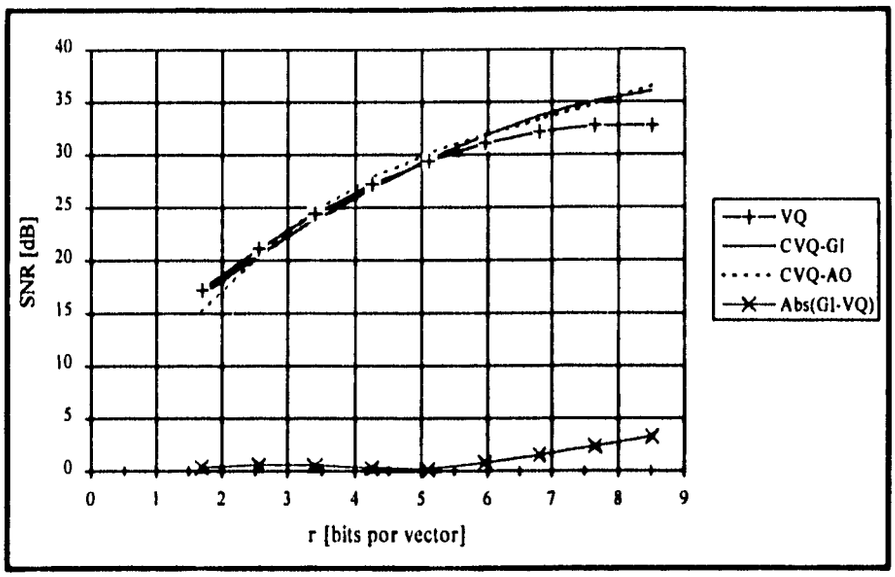
Gráfica 5.5.3. CVQ vs VQ en el coeficiente C63 de la secuencia de entrenamiento en el dominio de la 2D-VDCT (8x8,n=16).

Al igual que en los otros coeficientes Laplacianos, en la gráfica del coeficiente C63 se puede notar que el desempeño del cuantificador clasificado basado en el criterio de las geometrías implícitas produce mejores resultados a partir de cierta tasa (en este caso r=2.5bpv), comparado con el cuantificador de vecino más cercano y con el cuantificador clasificado basado en el criterio de los atributos óptimos. En contraste con los casos analizados anteriormente, la brecha entre las curvas de desempeño de los CVQ no tiende a mantenerse constante.

Como los diferentes códigos que componen al cuantificador clasificado pueden tener distintos tamaños, podría pensarse que en la acción de este cuantificador puede estar incluida cierta cantidad de codificación entrópica. Entonces, podría suceder que al introducir codificación entrópica en los dos esquemas de cuantificación bajo comparación, la ganancia del desempeño obtenida con el cuantificador clasificado podría perderse, lo cual haría injustificable el uso del cuantificador clasificado en lugar del cuantificador del vecino más cercano en las fuentes Gaussiana y Laplaciana. Entonces, para verificar si es que esta pérdida de ganancia ocurre, hay que realizar las pruebas de distorsión en función de la tasa, cuando se incluye codificación entrópica de los índices del código en el caso de VQ y codificación entrópica de los índices de clasificación y de código, en el cuantificador clasificado.

Como esquema de codificación entrópica de los índices de los códigos, se eligió el codificador de Huffman. En las gráficas 5.5.4. a la 5.5.6, se pueden ver las

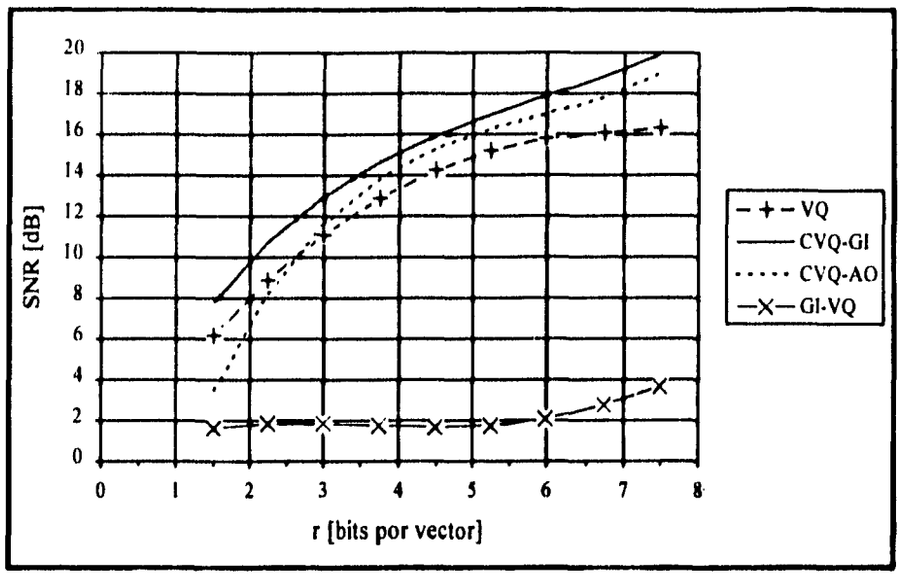
curvas SNR en función de la tasa del código para las mismas fuentes de las comparaciones anteriores.



Gráfica 5.5.4. CVQ+Huffman vs VQ+Huffman en el coeficiente C0 de la secuencia de entrenamiento en el dominio de la 2D-VDCT (8x8,n=16).

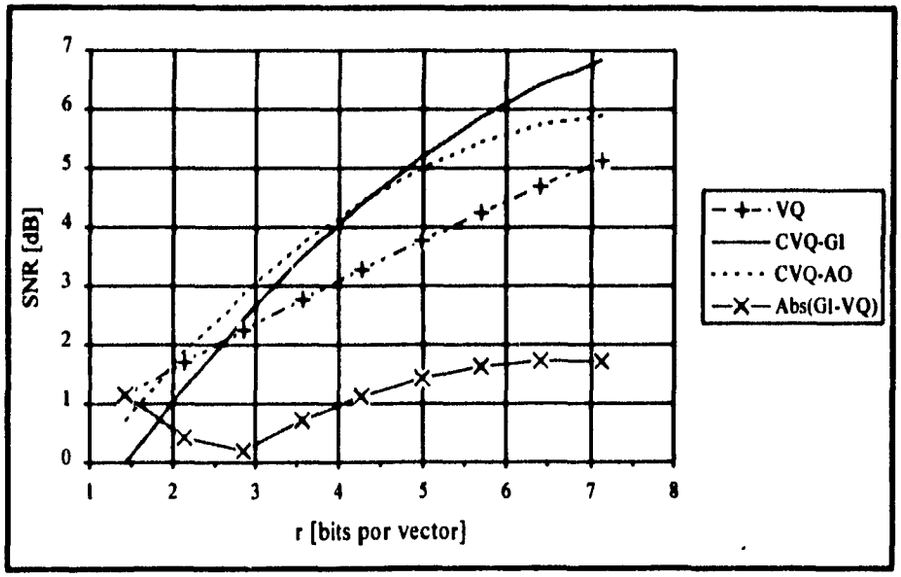
Lo primero que se puede notar en la gráfica del coeficiente C0 es que ahora el valor de la tasa del código para el cual el cuantificador clasificado tiene un mejor desempeño que el cuantificador del vecino más cercano, es mayor que cuando no se incluía la codificación entrópica (5.5 bits en vez de 3.5). Esto quiere decir que el bloque de codificación entrópica no realiza la misma reducción de la tasa en los dos esquemas. Sin embargo, el esquema de CVQ sigue proporcionando una ganancia de desempeño sobre el esquema VQ. Entre los dos esquemas de CVQ también ocurrió un cambio, ahora el CVQ-AO tiene un desempeño más cercano al de CVQ-GI y además, lo llega a sobrepasar en algunos puntos ($2.5< r <6$). Sin embargo, CVQ-AO sigue teniendo un desempeño notablemente más bajo que el de los otros dos esquemas a tasas bajas.

El hecho de que el bloque de codificación entrópica (en este caso Huffman) no logre la misma reducción de la tasa en los esquemas CVQ y VQ, está de acuerdo con la hipótesis de que el CVQ ya incluye cierta codificación de este tipo.



Gráfica 5.5.5. CVQ+Huffman vs VQ+Huffman en el coeficiente C1 de la secuencia de entrenamiento en el dominio de la 2D-VDCT (8x8,n=16).

Al igual que en el caso del coeficiente C0, al incluir un bloque de codificación entrópica en los esquemas de cuantificación del coeficiente C1, la curva de CVQ-AO se hace más parecida a la de CVQ-GI, sólo que en este caso el desempeño de CVQ-AO no llega a ser mayor que el de CVQ-GI. Aquí la brecha entre CVQ-GI y VQ no sufre cambios muy notables.



Gráfica 5.5.6. CVQ+Huffman vs VQ+Huffman en el coeficiente C63 de la secuencia de entrenamiento en el dominio de la 2D-VDCT (8x8,n=16).

Cuando se aplica codificación entrópica a los esquemas CVQ y VQ sobre el coeficiente C63, el cambio más notable es que la curva de desempeño de CVQ-AO ahora se aproxima más a CVQ-GI que a la de VQ, como sucedía cuando no se utilizaba Huffman sobre los índices. Sin embargo, la brecha entre CVQ-GI y VQ no cambia significativamente. El punto para el cual el desempeño de CVQ-GI sobrepasa al de VQ tampoco experimenta cambios significativos.

En la literatura relacionada con la cuantificación vectorial comúnmente se presentan resultados en fuentes Gaussianas y Laplacianas independientes e idénticamente distribuidas, así como en fuentes Gauss-Markov de primer orden. Para tener un punto de comparación del comportamiento de los métodos de cuantificación propuestos en este trabajo, se hicieron pruebas de comparación entre el desempeño de los cuantificadores clasificado y del vecino más cercano en la fuentes mencionadas. En las comparaciones anteriores, el criterio de desempeño que se utilizó fue la curva relación señal a ruido en función de la tasa del código. En este caso, para estar acorde con la literatura, las curvas son la relación entre la SNR, pero en función de la tasa del código normalizada a la dimensión del vector, es decir el número de bits por cada muestra.

Las muestras con función de densidad de probabilidad Laplacianas fueron generadas por el método de la transformación inversa [VQ60], usando como excitación un generador de números para una variable aleatoria uniforme.

Esto es, si u es una variable aleatoria uniformemente distribuida en el intervalo [0,1], la variable aleatoria x resultante de transformar a u por medio de

$$x = F_x^{-1}(u) = \begin{cases} \frac{1}{\lambda} Ln(2u) & u < 1/2 \\ -\frac{1}{\lambda} Ln(2u-2) & u \geq 1/2 \end{cases}$$

en donde

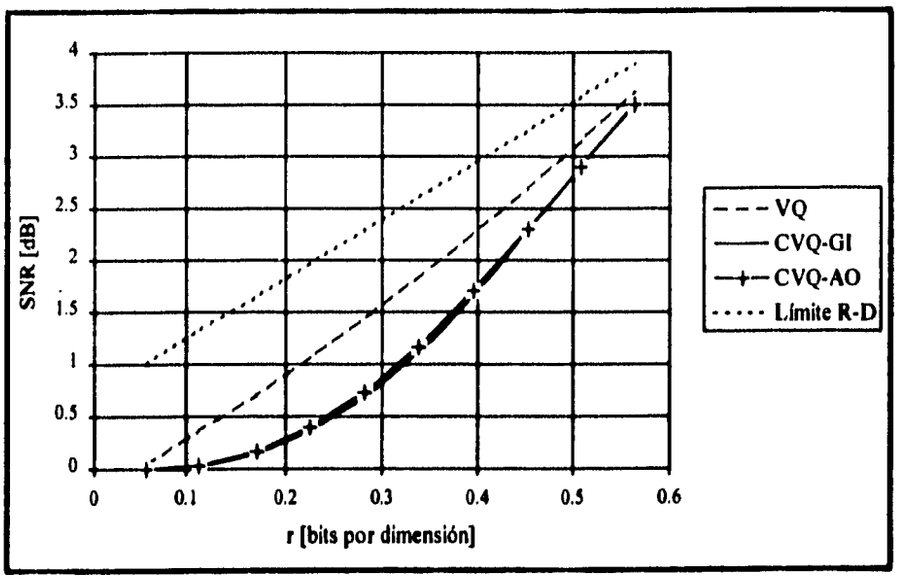
$$F_x^{-1}(x) = \int_{-\infty}^{x} f_x(t)dt$$

es una variable aleatoria con función de densidad de probabilidad Laplaciana dada por

$$f_x(x) = \frac{\lambda}{2} e^{-\lambda|x|}$$

El tamaño de la secuencia de entrenamiento se eligió de nuevo en base al criterio de ergodicidad mencionado en la cuarta parte de este capítulo, sin embargo, la longitud

fue aumentada hasta que el histograma de la secuencia no presentaba diferencias muy notables con la función de distribución Laplaciana (este es un criterio subjetivo pero tiene por objeto aumentar la validez de los resultados). En la gráfica 5.5.7 se muestra la comparación de desempeño de los cuantificadores clasificado y el ordinario del vecino más cercano. Como se puede ver, en este tipo de fuente el desempeño de los dos cuantificadores clasificados es menor que el de VQ. Este hecho se hace más notable conforme la tasa del código se hace más pequeña. Esto se debe a que en esta parte, la tasa dedicada a los índices de la clasificación es una parte muy significativa de la tasa total del código, con lo cual se acentúa el hecho de que el CVQ es subóptimo en términos de las condiciones del centroide y del vecino más cercano. Hay que notar que, a diferencia de lo que ocurre con los coeficientes de la 2D-VDCT, en este caso las curvas de los dos cuantificadores clasificados son muy parecidas.



Gráfica 5.5.7. CVQ vs VQ en fuente Laplaciana independiente idénticamente distribuida, 320000 muestras.

Generar una distribución normal es un trabajo especialmente difícil. Si se utiliza el método de la transformación inversa; es imposible obtener una expresión de forma cerrada para la función $F^{-1}(x)$. Aquí se usará una propiedad especial de la distribución Normal para generar las muestras del proceso Gaussiano independiente idénticamente distribuido (*iidG*), utilizado para comparar el desempeño del cuantificador clasificado contra el desempeño del cuantificador del vecino más

cercano ordinario. Box y Muller han demostrado que si V y W son variables aleatorias independientes, uniformes en el intervalo [0,1] y si se hace:

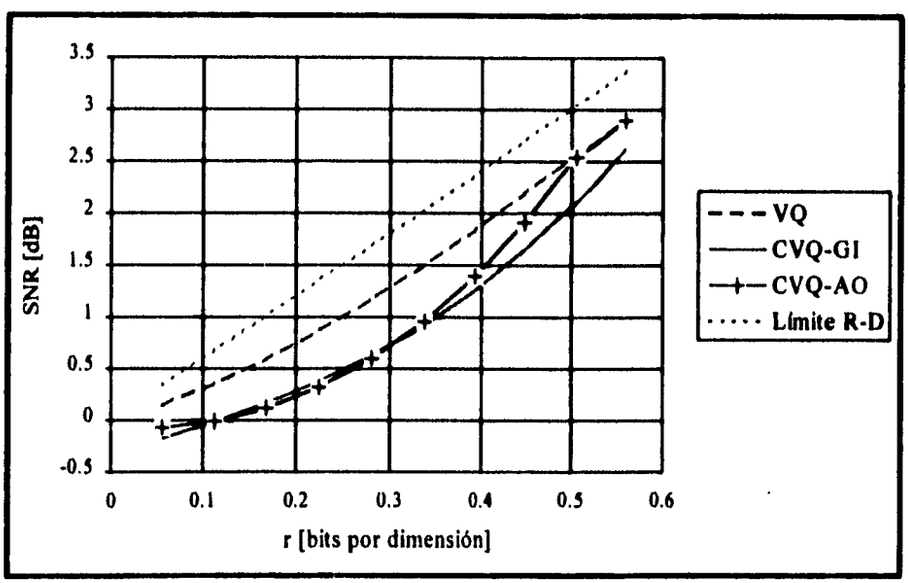
$$X = \cos(2\pi V)\sqrt{-2\ln(W)}$$
$$Y = \text{sen}(2\pi V)\sqrt{-2\ln(W)}$$

entonces X e Y serán variables aleatorias independientes con media cero y variancia uno [VQ60]. Entonces, las propiedades de linealidad de la distribución normal pueden ser utilizadas para transformar X e Y a distribuciones del tipo $N(\mu,\sigma^2)$ por medio de

$$X = \sigma X + \mu$$
$$Y = \sigma Y + \mu$$

y así generar el proceso aleatorio con las características deseadas.

En la gráfica 5.5.8 se pueden ver las curvas de relación señal a ruido en función de la tasa normalizada del código para la fuente *iidG*. De nuevo, el desempeño de los CVQ es menor que el de VQ para casi todo el intervalo de comparación. Sin embargo, aquí hay que notar que a diferencia de lo que pasa en los coeficientes de la 2D-VDCT, el desempeño del clasificador basado en los atributos óptimos da mejores resultados, que el clasificador basado en las geometrías implícitas.



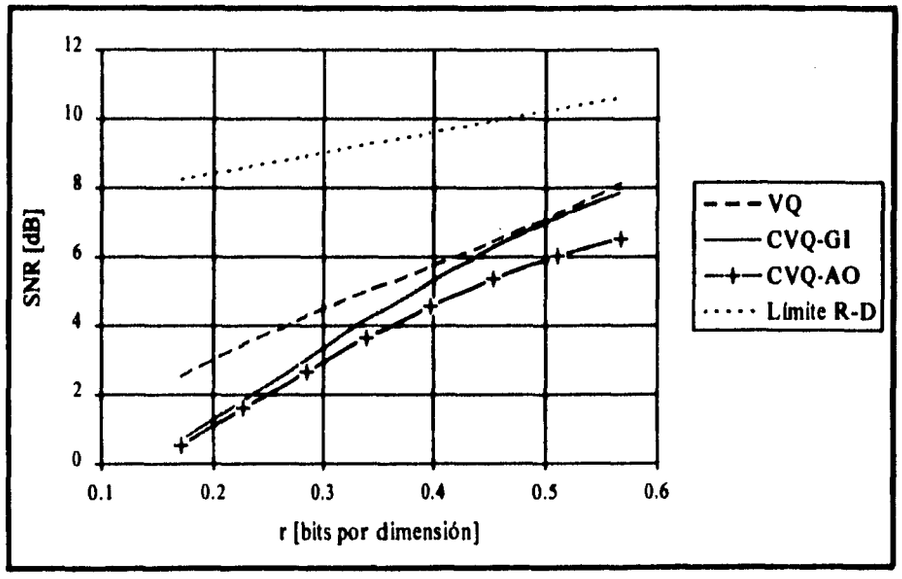
Gráfica 5.5.8. CVQ vs VQ en fuente Gaussiana independiente idénticamente distribuida, 320000 muestras.

Para continuar con la comparación del desempeño, era necesario probar los dos cuantificadores en una fuente con memoria. El proceso Gauss-Markov es uno de los ejemplos más simples de fuentes con memoria. Este proceso está definido como

$$X_{n+1} = aY_n + W_n$$

en donde a es conocida como el coeficiente de regresión y tiene magnitud menor que uno y $\{W_n\}$ es una fuente Gaussiana de media cero y varianza unitaria. Para generar las muestras de este proceso, se utilizo la ecuación anterior mas el generador utilizado en el proceso Gaussiano descrito anteriormente.

En la gráfica 5.5.9 se muestra la curva de comparación entre el desempeño del cuantificador clasificado y el del vecino más cercano sobre una secuencia de entrenamiento para el proceso Gauss-Markov de primer orden, con un coeficiente de autorregresión a=0.9. Al igual que en las fuentes Gaussiana y Laplaciana independientes idénticamente distribuidas, el CVQ tiene un peor desempeño que el de VQ, sin embargo, cuando la tasa del código alcanza los 0.5 bits por dimensión, el desempeño de CVQ es muy similar al de VQ.



Gráfica 5.5.9. CVQ vs VQ en fuente Gauss-Markov primer orden, a=0.9, 320000 muestras.

Resumen

A través de todas las pruebas de comparación realizadas, el criterio de diseño del clasificador que proporciona resultados más consistentes es el de las geometrías implícitas. Además, éste tiene la ventaja adicional de que el tiempo de cálculo requerido para encontrar el conjunto de umbrales es mucho más pequeño (al rededor de 20 veces menos), que el que necesita el algoritmo basado en el criterio de los atributos óptimos.

En base a los resultados obtenidos sobre el conjunto de fuentes utilizadas para hacer las comparaciones, se puede deducir que el cuantificador clasificado es adecuado para codificación de fuentes con cierto tipo de memoria, esto es, en muestras correlacionadas en formas particulares. En los procesos aleatorios en los que el cuantificador clasificado resulta peor que el cuantificador ordinario del vecino más cercano, se tiene una matriz de correlación concentrada principalmente en la diagonal principal. En cambio, en los coeficientes de la 2D-VDCT, para fuentes en las que el cuantificador clasificado es mejor que el cuantificador ordinario del vecino más cercano, se tiene que las matrices de correlación presentan algunos elementos fuera de la diagonal principal con valores grandes.

Otra observación importante, es que si bien el cuantificador clasificado produce mejores resultados que el cuantificador ordinario del vecino más cercano, en la codificación de los coeficientes de la 2D-VDCT, esto es válido únicamente a partir de un cierto valor de la tasa del código. Este hecho limita la aplicación del cuantificador clasificado a aplicaciones en las que no se requieran tasas muy bajas del código. Sin embargo, si se toma en cuenta que los nuevos desarrollos de codificación de transformada (contexto en el que se utilizará al cuantificador clasificado que se propone aquí), indican que este esquema es más apropiado para compresión a tasas medias, ya que esquemas como la codificación subbanda producen mejores resultados a tasas muy bajas (en [VQ5] se muestran resultados que demuestran esto). Entonces la limitación del cuantificador clasificado a tasas bajas, no es un problema que impida su aplicación en el esquema de codificación por transformada.

Finalmente, cuando se comparan los cuantificadores clasificado y el ordinario del vecino más cercano, se observa que aunque el cuantificador clasificado tiene una mayor cantidad de palabras de código, su tiempo de diseño es menor (al rededor del 50%) al del vecino más cercano ordinario. Esto se debe principalmente a que el proceso de diseño de este último cuantificador, requiere de casi el triple de iteraciones para converger.

5.6 Resultados con el esquema de compresión de imágenes por medio del cuantificador vectorial clasificado y la transformada coseno discreta vectorial

En el capítulo 5 se presentó un esquema de codificación de imágenes por medio de la 2D-VDCT. Este esquema utiliza un cuantificador para cada uno de los coeficientes. La cantidad de la tasa total del código asignada a cada uno de los coeficientes, se calcula en base al determinante de la matriz de covarianza del coeficiente. Así, los coeficientes con un rango dinámico grande tendrán una tasa de código mayor a los coeficientes con un rango dinámico pequeño. El hecho de que, entre más correlacionadas entre si estén las componentes del coeficiente, más cercano a cero será el determinante de su matriz de covarianza, y por lo tanto mayor la tasa asignada al coeficiente, proporciona un medio de reservar más tasa a los coeficientes menos susceptibles de ser cuantificados eficientemente, independientemente de su rango dinámico. Esto viene a complementar la justificación de la asignación de tasa, en base al determinante de la matriz de covarianza. El sistema de compresión, utiliza un conjunto de plantillas para eliminar, en un bloque transformado, los coeficientes que no son relevantes para la reconstrucción de la imagen. Esta eliminación de ciertos coeficientes es conocida como codificación zonal.

Para determinar que coeficientes podrán ser preservados en la codificación zonal, y cuales podrán ser eliminados, se utilizó el criterio de concentración de la energía mencionado en el capítulo cuatro. El cálculo del umbral de ese método, se hizo en base a la suposición de que con el sistema de compresión, se busca codificar las imágenes con una relación señal a ruido de alrededor de 33dB, con lo que el nivel de energía de los coeficientes que se van a descartar debe ser menor al 0.0005% de la energía total del bloque. Esta suposición da un valor γ_o de 0.9993. Experimentalmente se encontró que un valor de γ_E=0.9997 requiere que la calidad de cuantificación de los coeficientes que la codificación zonal preserva, esté sólo 0.7 dB por encima del valor de SNR deseado para el bloque. En la asignación de bits basada en el determinante de la matriz de covarianza, este valor de γ_E permite que el incremento en la SNR de codificación de los coeficientes preservados, sea cubierta fácilmente por un ligero ajuste al momento de hacer los redondeos.

Para poder interpretar los resultados de asignación de bits y de codificación zonal, es necesario conocer el significado de cada uno de los coeficientes de la transformación. En la gráfica 5.6.1 se muestran los índices correspondientes al bloque de coeficientes de la 2D-VDCT de 8x8 vectores. El coeficientes C0 es el que corresponde al nivel de *DC*, mientras que el C63 es el que corresponde a las componentes de frecuencias más altas de la señal de entrada. Generalmente el coeficiente de *DC* concentra la mayor cantidad de la energía total del bloque. Los

coeficientes de la primera columna son los que corresponden a los bordes verticales puros, y los que están en el primer renglón, corresponden a los bordes horizontales puros. Sin embargo, la presencia simultánea de los coeficientes correspondientes a los bordes horizontales y verticales puros, puede estar relacionada con bordes diagonales en el dominio de la imagen. Los coeficientes en las diagonales están relacionados con bordes u objetos en el dominio de la imagen con diferentes grados de inclinación [VQ61]. Es esta relación entre la ubicación del coeficiente en el bloque transformado y las características del bloque en el dominio de la imagen, la que justifica la codificación zonal por medio de un conjunto de plantillas.

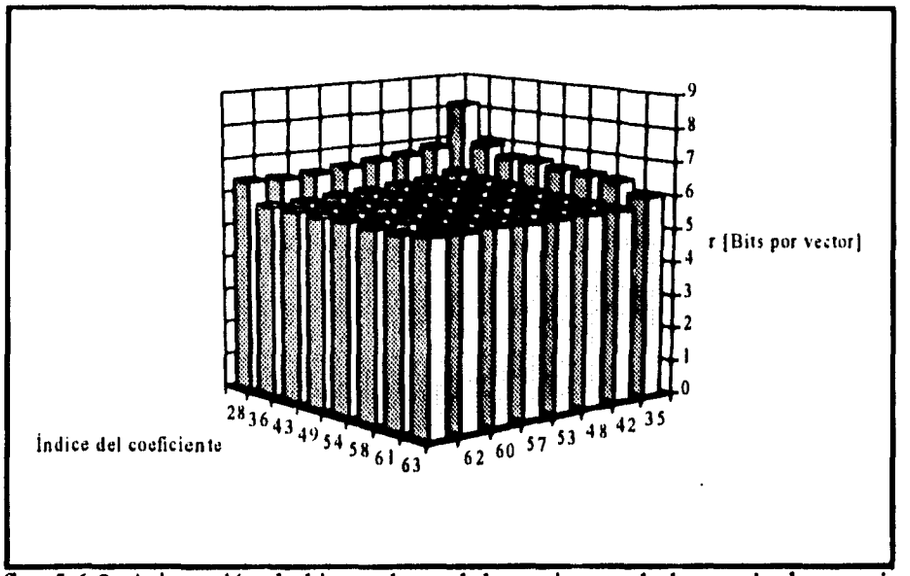
Como se puede ver en la gráfica 5.3.5 los coeficientes con mayor rango dinámico son aquellos que pertenecen a la primera columna y al primer renglón, y son estos los que, bajo el criterio de asignación de bits en base a la matriz de covarianza, deben recibir una mayor cantidad de información.

En la gráfica 5.6.2 se muestra el resultado de hacer la asignación de bits a cada uno de los coeficientes, para una tasa total promedio de r=6 bits por vector. Al comparar está gráfica con la de distribución de la energía (gráfica 5.3.5) se puede comprobar que el método, en efecto, está asignando mas información a los coeficientes con un rango dinámico mayor, esto es: los coeficientes correspondientes a los bordes horizontales y verticales puros, y el coeficiente de *DC*.

El hecho de que en la asignación de la tasa, los coeficientes de frecuencias más altas reciban cantidades mínimas de bits, causa que sean cuantificados burdamente. Esto es muy importante, ya que cuando alguna imagen tenga concentrada su energía en los coeficientes de las frecuencias más altas, el esquema de codificación zonal con asignación fija de bits no podrá realizar una muy buena codificación .

C0	C2	C5	C9	C14	C20	C27	C35
C1	C4	C8	C13	C19	C26	C34	C42
C3	C7	C12	C18	C25	C33	C41	C48
C6	C11	C17	C24	C32	C40	C47	C53
C10	C16	C23	C31	C39	C46	C52	C57
C15	C22	C30	C38	C45	C51	C56	C60
C21	C29	C37	C44	C50	C55	C59	C62
C28	C36	C43	C49	C54	C58	C61	C63

Figura 5.6.1. Índices de los coeficientes de la 2D-VDCT (8x8).



Gráfica 5.6.2. Asignación de bits en base al determinante de la matriz de covarianza de los coeficientes de la secuencia de entrenamiento en el dominio de la 2D-VDCT (8x8,n=16).

Debido a los redondeos hechos en la asignación de bits, la tasa promedio real es ligeramente diferente a la tasa que se tenía prevista originalmente. Con la asignación de bits que se muestra en la gráfica 5.6.2, la tasa promedio real es r=5.47 bits por vector, que es equivalente a 0.3418 bits por pixel.

De los resultados presentados en la parte 4 de este capítulo, se determinó que el cuantificador clasificado basado en el criterio de las geometrías implícitas se desempeña mejor que el que se basa en el criterio de los atributos óptimos, en términos de la distorsión en función de la tasa del código. En consecuencia, el cuantificador clasificado basado en el criterio de las geometrías implícitas es el que se utilizó para construir este esquema de compresión, y el que finalmente se compara contra el cuantificador ordinario del vecino más cercano.

Las secuencias de entrenamiento utilizadas para generar los libros de códigos de los cuantificadores clasificados del compresor de imágenes, fueron extraídas de un conjunto de datos, el cual es el resultado del procesamiento, por medio de la 2D-VDCT (8x8,n=16), de 75 imágenes de la vida real, 8 sintéticas y 4 biomédicas, las que en su mayoría fueron obtenidas de la base de datos de la *University of Southern California* (USC). En base a los criterios de estacionariedad y ergodicidad expuestos en la tercer parte de este capítulo, las longitudes de la secuencia para el diseño de los cuantificadores clasificados son: 4096 vectores para el modo 0, 2048 vectores para los modos 1 y 2, y 1024 vectores para CVQ modo 3.

Como secuencias de prueba se utilizaron los vectores resultantes de procesar, por medio de la 2D-VDCT, 4 imágenes no incluidas en las secuencias de entrenamiento. Estas imágenes son:

Tabla 5.6.1 Imágenes en la secuencia de prueba (8bpp)

Imagen	Dimensiones	Vectores/coef.
Girl.raw	512x512	256
Lady.raw	256x256	64
Lena.raw	512x512	256
Peppers.raw	512x512	256

Para poder medir el desempeño de un sistema de compresión de imágenes, se debe tener una medida que indique la distorsión que se le añade a la imagen al pasar por dicho sistema. Las medidas de desempeño más frecuentemente utilizadas en la literatura de codificación de imágenes son: la comparación subjetiva, es decir, un observador humano compara la imagen procesada con la imagen original, y la relación señal a ruido pico, definida de la siguiente manera [VQ52, VQ63]:

$$PSNR = 10\log_{10}\left(\frac{N_{max}^2}{\sum_{i=1}^{Col}\sum_{j=1}^{Ren}\left(x_{ij}-\hat{x}_{ij}\right)^2}\right)$$

donde *Col* y *Ren* es el número de columnas y renglones de la imagen, N_{max} es el valor máximo que pueden tomar los pixels (por ejemplo N_{max} =256 para imágenes a 8bpp), x_{ij} y $\hat{x}_{ij}$ son los pixels de la imagen original y de la imagen procesada por medio del compresor, respectivamente.

Para realizar la comparación subjetiva de la calidad de codificación del sistema, en las figuras 5.6.2 a 5.6.9 se presentan las imágenes de la secuencia de prueba originales y sus versiones procesadas por el sistema de compresión. La discusión sobre la comparación subjetiva está basada en cuatro tipos de perturbaciones causadas por la codificación de transformada. Estas perturbaciones son:

- *Difuminación.* Causada por la cuantificación burda o supresión de los coeficientes correspondientes a las frecuencias altas.
- *Granularidad y ruido.* Causados por la cuantificación burda de los coeficientes correspondientes a las frecuencias medias.

- *Artefactos*. Causados por la cuantificación burda de los coeficientes correspondientes a las frecuencias bajas y a los bordes verticales y horizontales puros.
- *Blocking*. (discontinuidad artificial en los bordes de bloques transformados independientemente). Relacionada con el error de cuantificación del coeficiente de *DC*, y con el tamaño de la transformación (en este caso 8x8 vectores, lo que es equivalente a un bloque de 32x32 pixels).

Debido a que como se verá más adelante, la naturaleza de las perturbaciones introducidas por los dos métodos de cuantificación que se comparan dentro de este sistema de compresión, es la misma. La discusión de la comparación subjetiva no se centra en confrontar los resultados obtenidos con uno u otro método. En lugar de eso, la discusión estará enfocada principalmente a resaltar el tipo de perturbaciones que introduce el sistema de compresión por codificación de transformada, y relacionarlas con las funciones que se realizan dentro de él.

La comparación cuantitativa, en términos de la relación señal a ruido pico, se presenta en la tabla 5.6.2, en la cual se muestran los resultados de codificar los coeficientes por medio del cuantificador vectorial del vecino más cercano y por medio del cuantificador clasificado propuesto en este trabajo. Los valores de tasa del código que se presentan en la tabla son los correspondientes a la tasa promedio sin supresión de coeficientes. Esto es, la tasa de asignación de bits para el diseño de los códigos. Posteriormente se presenta la tabla 5.6.3 con los valores reales de las tasas de codificación. En la comparación cuantitativa, es claro que el desempeño del sistema basado en el cuantificador clasificado, es superior al del que se basa en el cuantificador ordinario del vecino más cercano.

Tabla 5.6.2. PSNR de la imágenes codificadas por medio de los esquemas 2D-VDCT+VQ y 2D-VDCT+CVQ.

Imagen	PSNR VQ [dB] r_T=0.375bpp	PSNR CVQ [dB] r_T=0.3418bpp	PSNR CVQ [dB] r_T=0.5870bpp
Girl.raw	28.0538	28.9292	33.9896
Lady.raw	26.1632	27.1461	34.8276
Lena.raw	29.1547	31.3803	38.8757
Peppers.raw	28.7946	30.1514	38.7625

Debido a que el uso de las transformaciones vectoriales en sistemas de compresión de imágenes, es bastante reciente, no existen muchos resultados con los que se pueda comparar el desempeño del sistema de compresión propuesto en este trabajo. Weiping Li en [VQ22] presenta resultados de un sistema de compresión por codificación de transformada en base a la 2D-VDCT, (8x8). El criterio de asignación

de bits que se utiliza en ese trabajo es el mismo que se aplica aquí. Con ese esquema W. Li. reporta la codificación de Lena con una SNR=29.1328dB con asignación de bits basada en una tasa de código de 0.37bpp, lo cual concuerda muy cercanamente con el resultado de 2D-VDCT+VQ@0.375bpp presentado en la tabla 5.6.2. El mismo autor presenta en [VQ5] un sistema de compresión por codificación de transformada. En ese sistema la asignación de bits se realiza de forma dinámica. Sin embargo, como no se menciona la forma concreta de codificación de los índices de los vectores y de los códigos, no se puede hacer una comparación rigurosa con los resultados presentados en la tabla 5.6.2. En ese trabajo W. Li reporta una SNR=30.1 dB en la codificación de Lena con una tasa de 0.17bpp.

Como se puede ver en la tabla 5.6.3, la tasa real de codificación en el sistema de compresión 2D-VDCT+CVQ, disminuye significativamente debido a la eliminación de coeficientes que realiza la codificación zonal. Si se toman en cuenta conjuntamente los valores de tasa de la tabla 5.6.3 y los valores de SNR de la tabla 5.6.2, entonces el sistema de compresión 2D-VDCT+CVQ (sin algún tipo de codificación entrópica) presentado aquí, tiene un desempeño cercano al del sistema presentado por W. Li en [VQ5].

Tabla 5.6.3. Tasas reales de las imágenes en el sistema de compresión 2D-VDCT+CVQ con asignación fija de bits y 12 plantillas.

Imagen	r[bpp]@r_T=0.3418	r[bpp]@r_T=0.5870
Girl	0.2495	0.4864
Lady	0.3127	0.5541
Lena	0.2324	0.4632
Peppers	0.2803	0.5173

Con el propósito de ilustrar el grado más fuerte de degradación que sufren los coeficientes de frecuencias altas, se incluye en la secuencia de imágenes de prueba las versiones de una misma imagen a dos diferentes resoluciones, 256x256 pixels, Lady, y 512x512 pixels, Lena. De entrada, en la comparación cuantitativa por medio de la PSNR, el hecho de que los coeficientes de alta frecuencia se vean más degradados que el resto de los coeficientes, o que sean truncados por las plantillas, se ve reflejado en que la codificación de Lady por medio de los dos esquemas bajo comparación, produce relaciones señal a ruido pico más bajas en 3dB (VQ) y 4db (CVQ) que las obtenidas en Lena.

El problema de la codificación burda de los coeficientes correspondientes a las frecuencias más altas, puede ser evitado en el sistema de compresión por codificación zonal, variando la asignación de bits de acuerdo al número de coeficientes preservados por cada elemento del conjunto de plantillas. Sin embargo, por razones de tiempo, este método no será utilizado en el presente trabajo.

Discusión de la calidad subjetiva

En el sentido subjetivo, la degradación de la imagen debida a la codificación burda, o supresión de los coeficientes de frecuencias altas de la 2D-VDCT, puede percibirse por la difuminación de los detalles finos de la imagen. En las figuras 5.6.2 y 5.6.4, al comparar los originales; de Lady y Lena, con las versiones procesadas por el sistema de compresión, se puede percibir que la imagen se aprecia más difuminada a menor resolución.

Para comparar la naturaleza de la distorsión introducida por el esquema de compresión con los dos tipos de cuantificadores, y determinar si es igual o diferente, se incluye en la figura 5.6.2 la versión de la imagen codificada por medio de 2D-VDCT+VQ. Como se puede apreciar, las dos imágenes codificadas presentan el mismo tipo de degradación, esto es, ambas se aprecian ruidosas y difuminadas, y en ambas hay una ligera degradación de *blocking* de 32 pixels asociada con el tamaño de la transformada. Este fenómeno de distorsión se debe principalmente al error de cuantificación de los coeficientes de *DC*, ya que se percibe como una diferencia en el tono promedio. El *blocking* es más notable en la parte de la imagen que esta arriba del sombrero, y en una porción del hombro. A diferencia de la cuantificación vectorial en el dominio de la imagen [VQ52], en la codificación por transformada vectorial la degradación de escalera en los bordes diagonales es poco notable. Sin embargo, la cuantificación burda de algunos coeficientes causa la aparición de artefactos granulados como el que se puede apreciar en el hombro y partes del rostro de la mujer de la figura. Este tipo de artefactos se puede apreciar en los dos esquemas de cuantificación que se presentan aquí. En la figura 5.6.3 se presenta la misma imagen (a resolución de 256x256) pero a una tasa mayor. En este caso se aprecia el mismo tipo de perturbaciones pero a una magnitud ligeramente menor.

En la tabla 5.6.3 se presenta el índice de plantilla para cada uno de los bloques de la imagen Lady en el dominio transformado. Como se puede apreciar, la gran mayoría de bloques fue codificado con la plantilla 12. Esta plantilla no suprime coeficientes, por lo que la tasa de codificación real no es muy pequeña en comparación con la tasa de asignación de bits (ver tabla 5.6.2). Las plantillas que suprimen coeficientes están localizadas en las regiones correspondientes al fondo de la imagen.

En las figura 5.6.4 y 5.6.5, en las cuales se presenta la imagen Lena a resolución de 512x512, se puede apreciar la aparición de algunas franjas horizontales. Este tipo de artefactos está relacionado con el error de cuantificación de los coeficientes del primer renglón de la 2D-VDCT, los cuales corresponden a los bordes horizontales (ver figura 5.6.1). El efecto de *blocking* en Lena puede notarse en las regiones que están alrededor del sombrero y en algunas partes del hombro y rostro. Las granularidades debidas al error de cuantificación de los coeficientes correspondientes a las frecuencias medias puede apreciarse en el fondo, en el sombrero, alrededor de los ojos, y en el borde diagonal del hombro. La difuminación

causada por la cuantificación burda de los coeficientes de las frecuencias altas, se puede apreciar en el pelo, en las plumas, y más notablemente en las fibras del sombrero, las cuales casi no pueden apreciarse en la imagen codificada.

En la tabla 5.6.4 se puede observar que en aproximadamente 44% de los bloques de la imagen Lena se utiliza plantillas con supresión de coeficientes. Los bloques en los que se suprimen coeficientes son los que corresponden a las partes homogéneas de la imagen, tales como el fondo, el espejo y el hombro de la mujer. En una parte significativa, casi el 10% de los bloques, de la imagen, se está aplicando la plantilla 9, la cual suprime el 94% de los coeficientes. Esto causa una reducción significativa de la tasa de codificación, en comparación con la tasa de asignación de bits (ver tabla 5.6.2). Las zonas en las que no se suprimen coeficientes, como por ejemplo en el pelo y la región de los ojos, la imagen tiene alto contenido de frecuencias altas. Esto significa que los resultados obtenidos con el criterio de la energía, utilizado para adaptar la codificación zonal, son coherentes con las características de la imagen.

En las figuras 5.6.6 y 5.6.7 se tiene otro buen ejemplo de la difuminación causada por la cuantificación burda de los coeficientes de alta frecuencia. Este tipo de perturbación puede apreciarse en toda la región del pelo, las pestañas y las cejas. Las granularidades y ruido están presentes principalmente, en la punta de la nariz y alrededor de los ojos. En este caso los efectos de *blocking*, notables en las orillas de las mejillas y en la parte del fondo cercana al pelo, no se deben tanto a la discontinuidad de tono, sino a que pueden apreciarse como discontinuidad en el nivel de ruido de la imagen. Esto quiere decir que aquí no está influyendo el ruido de cuantificación de los coeficientes de *DC*, y el factor más importante está siendo la transformación independiente de cada uno de los bloques de la imagen. La aparición de artefactos se puede apreciar en las partes izquierda e inferior de la imagen, en la cual se ven franjas verticales y horizontales respectivamente. Al comparar el conjunto de índices de codificación zonal de la tabla 5.6.5 con la imagen, se puede ver que las plantillas que suprimen más coeficientes están localizadas en las regiones más homogéneas de la imagen, por lo que también en este caso el criterio de la energía utilizado para la adaptación de la codificación zonal, es adecuado. En el caso de la imagen Girl, el porcentaje de bloques en los cuales la plantilla suprime la mitad de los coeficientes, es de 40%, mientras que el porcentaje de bloques con plantillas con supresión del 94% de los coeficientes, es de 7%. Esto se ve reflejado en que la tasa real de codificación está moderadamente por debajo de la tasa de asignación de bits (ver tabla 5.6.3).

A diferencia de la imágenes Girl y Lena, en Peppers (figuras 5.6.8 y 5.6.9) el *blocking* es notable como discontinuidad de tono (pimiento largo de la izquierda). En esta imagen casi no es notable la difuminación, aún en los bordes. Sin embargo, existe una gran cantidad de artefactos en las cuatro orillas de la imagen, así como *blocking* apreciable en los diferentes niveles de ruido. En esta imagen, la reducción

de la tasa debida a la codificación zonal es pequeña (ver tabla 5.6.3). Esto se debe a que sólo el 30% de los bloques son codificados con plantillas que suprimen la mitad de los coeficientes, y únicamente el 3% de los bloques es codificado con la plantilla que suprime el 94% de los coeficientes.

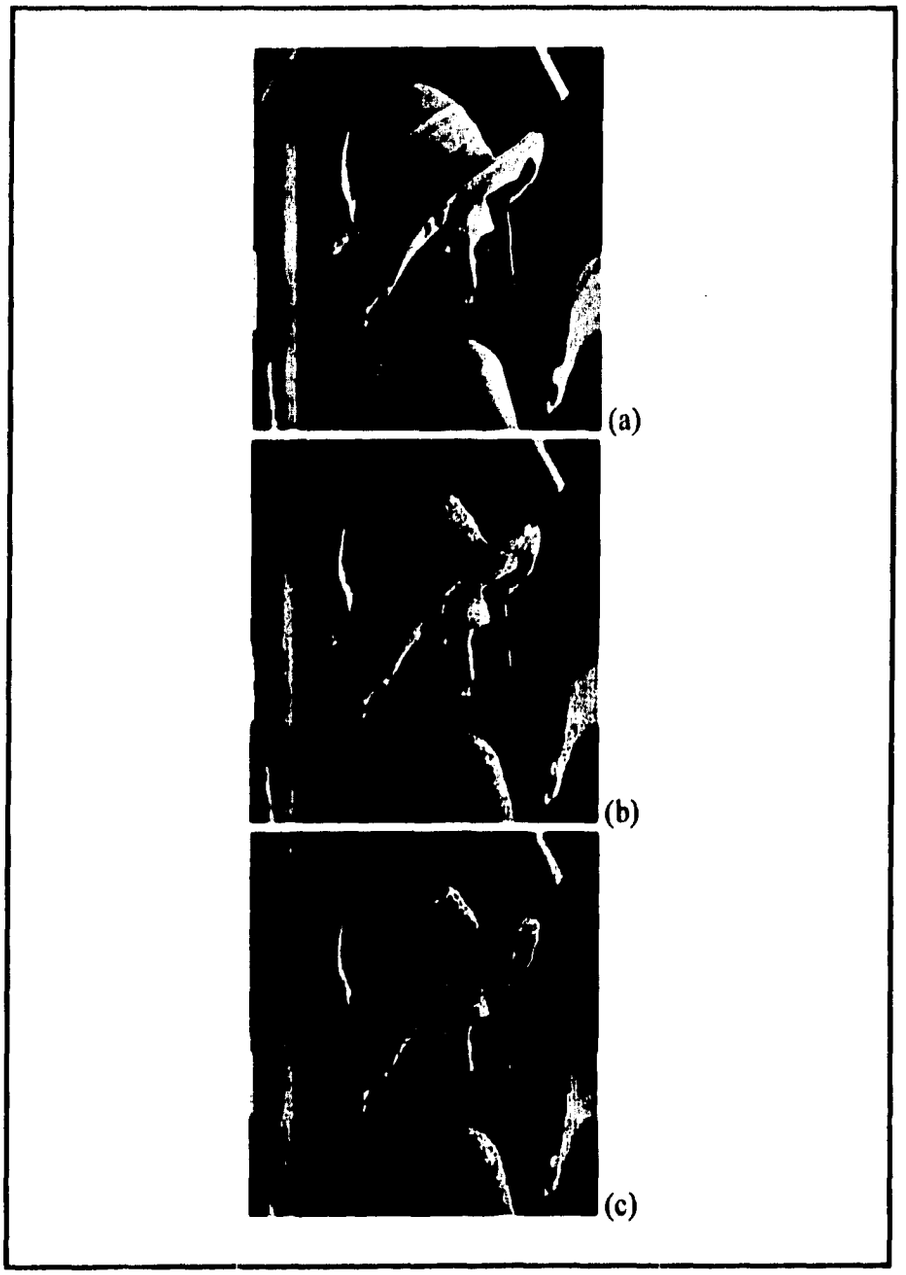


Figura 5.6.2. Original @8bpp (a), 2D-VDCT+CVQ @0.3418bpp (b) y 2D-VDCT+VQ @ 0.375bpp de la imagen Lady, 256x256, pixels de la base de datos de la University of Southern California (USC).

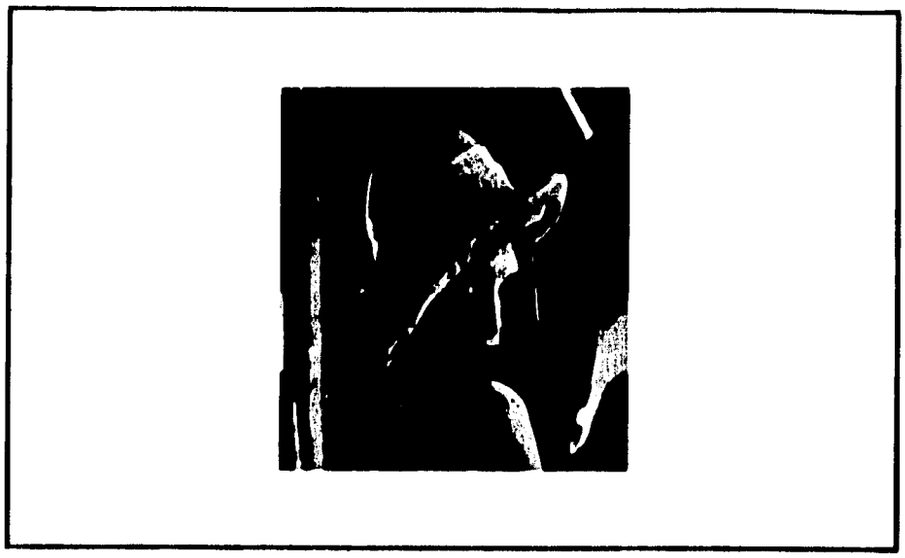


Figura 5.6.3. Versión 2D-VDCT+CVQ @ 0.587bpp de la imagen Lady 256x256 de la base de datos de la University of Southern California (USC).

Tabla 5.6.3. Mapa de plantillas para la imagen Lady en el dominio de la 2D-VDCT (8x8).

5	5	7	12	12	5	12	12
12	12	12	12	12	12	12	12
5	12	12	12	12	12	12	12
5	12	12	12	12	12	12	7
5	12	12	12	12	12	12	12
12	12	12	12	12	12	5	12
12	12	12	12	12	12	12	12
12	12	12	12	10	12	12	12

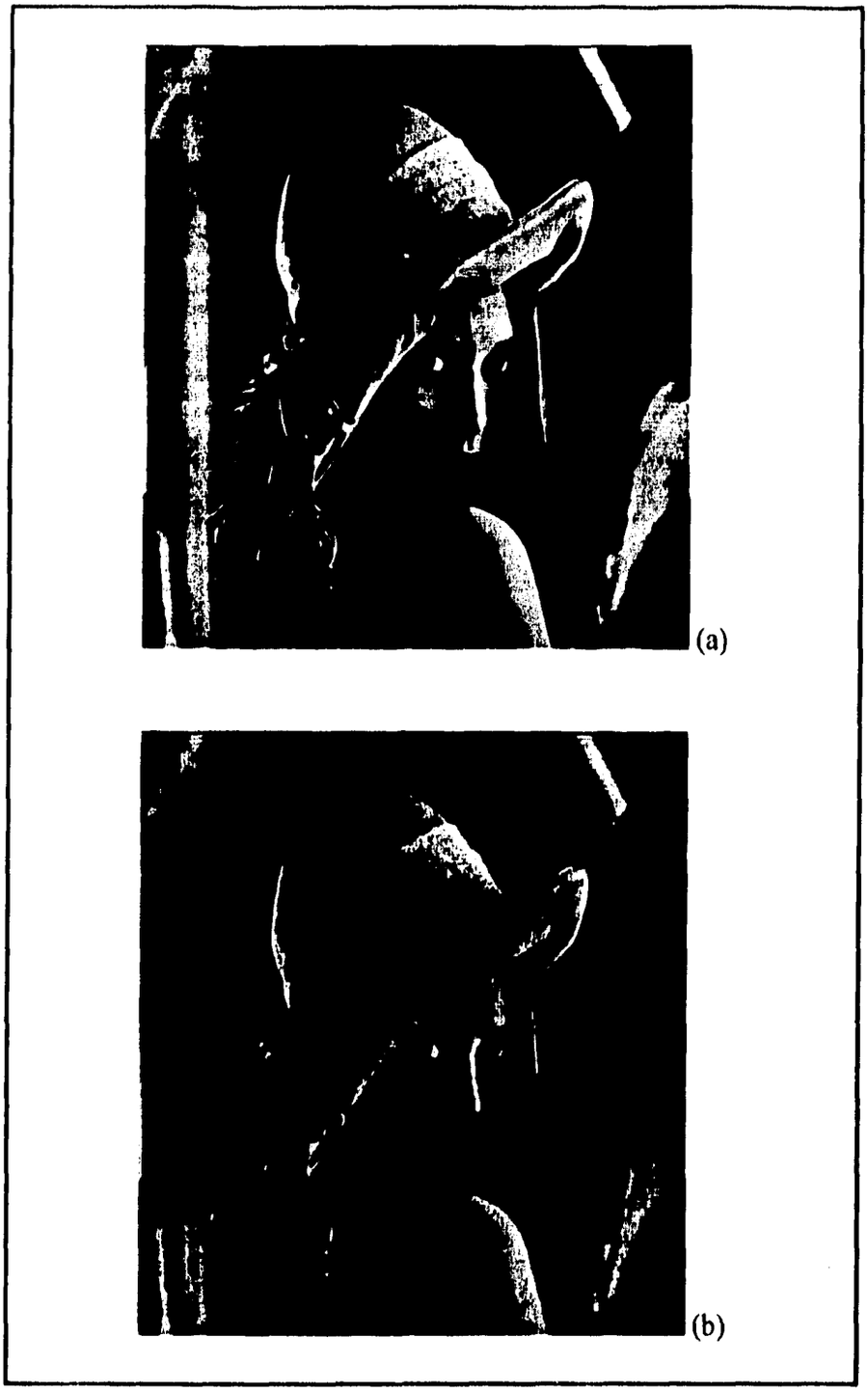


Figura 5.6.4. Original @8bpp (a) y 2D-VDCT+CVQ @0.3418bpp (b) de la imagen Lena, 512x512 pixels, de la base de datos de la University of Southern California (USC).

Figura 5.6.5. Versión 2D-VDCT+CVQ @ 0.587bpp de la imagen Lena 512x512 de la base de datos de la University of Southern California (USC).

Tabla 5.6.4. Mapa de plantillas para la imagen Lena en el dominio de la 2D-VDCT (8x8)

7	1	7	7	7	7	7	7	7	7	7	7	1	12	7	12
10	5	5	9	9	7	7	7	12	10	10	9	9	5	12	12
5	5	5	5	12	7	5	12	12	12	10	5	9	12	12	12
5	5	5	12	12	12	12	12	10	12	12	12	12	12	12	12
5	5	5	5	12	12	12	12	5	5	12	12	12	12	12	11
5	5	5	5	12	12	12	12	12	12	10	12	12	12	12	9
5	5	5	12	12	12	12	12	12	10	12	12	12	12	10	9
5	5	12	12	12	12	12	12	12	7	11	12	12	5	9	10
5	5	12	12	12	12	12	12	12	12	12	12	12	5	10	5
5	5	12	12	12	12	12	5	5	5	5	12	5	9	9	9
5	5	12	12	12	12	12	5	10	10	12	12	5	9	5	9
5	5	12	12	12	12	12	12	10	10	12	12	9	9	5	10
12	1	12	12	12	12	12	12	7	12	12	5	5	10	5	12
12	5	12	12	12	12	12	12	9	9	10	12	9	1	5	5
12	5	12	12	12	12	12	12	9	9	9	12	7	1	12	11
12	1	12	12	12	12	12	10	9	9	9	5	5	11	12	12

Figura 5.6.6. Original @8bpp (a) y 2D-VDCT+CVQ @0.3418bpp (b) de la imagen Girl, 512x512 pixels, de la base de datos de la University of Southern California (USC).

Figura 5.6.7. Versión 2D-VDCT+CVQ @ 0.587bpp de la imagen Lady 256x256 de la base de datos de la University of Southern California (USC).

Tabla 5.6.5. Mapa de plantillas para la imagen Girl en el dominio de la 2D-VDCT (8x8)

7	1	7	7	7	7	7	7	7	7	7	7	1	12	7	12
10	5	5	9	9	7	7	7	12	10	10	9	9	5	12	12
5	5	5	5	12	7	5	12	12	12	10	5	9	12	12	12
5	5	5	12	12	12	12	12	10	12	12	12	12	12	12	12
5	5	5	5	12	12	12	12	5	5	12	12	12	12	12	11
5	5	5	5	12	12	12	12	12	12	10	12	12	12	12	9
5	5	5	12	12	12	12	12	12	10	12	12	12	12	10	9
5	5	12	12	12	12	12	12	12	7	11	12	12	5	9	10
5	5	12	12	12	12	12	12	12	12	12	12	12	5	10	5
5	5	12	12	12	12	12	5	5	5	5	12	5	9	9	9
5	5	12	12	12	12	12	5	10	10	12	12	5	9	5	9
5	5	12	12	12	12	12	12	10	10	12	12	9	9	5	10
12	1	12	12	12	12	12	12	7	12	12	5	5	10	5	12
12	5	12	12	12	12	12	12	9	9	10	12	9	1	5	5
12	5	12	12	12	12	12	12	9	9	9	12	7	1	12	11
12	1	12	12	12	12	12	10	9	9	9	5	5	11	12	12

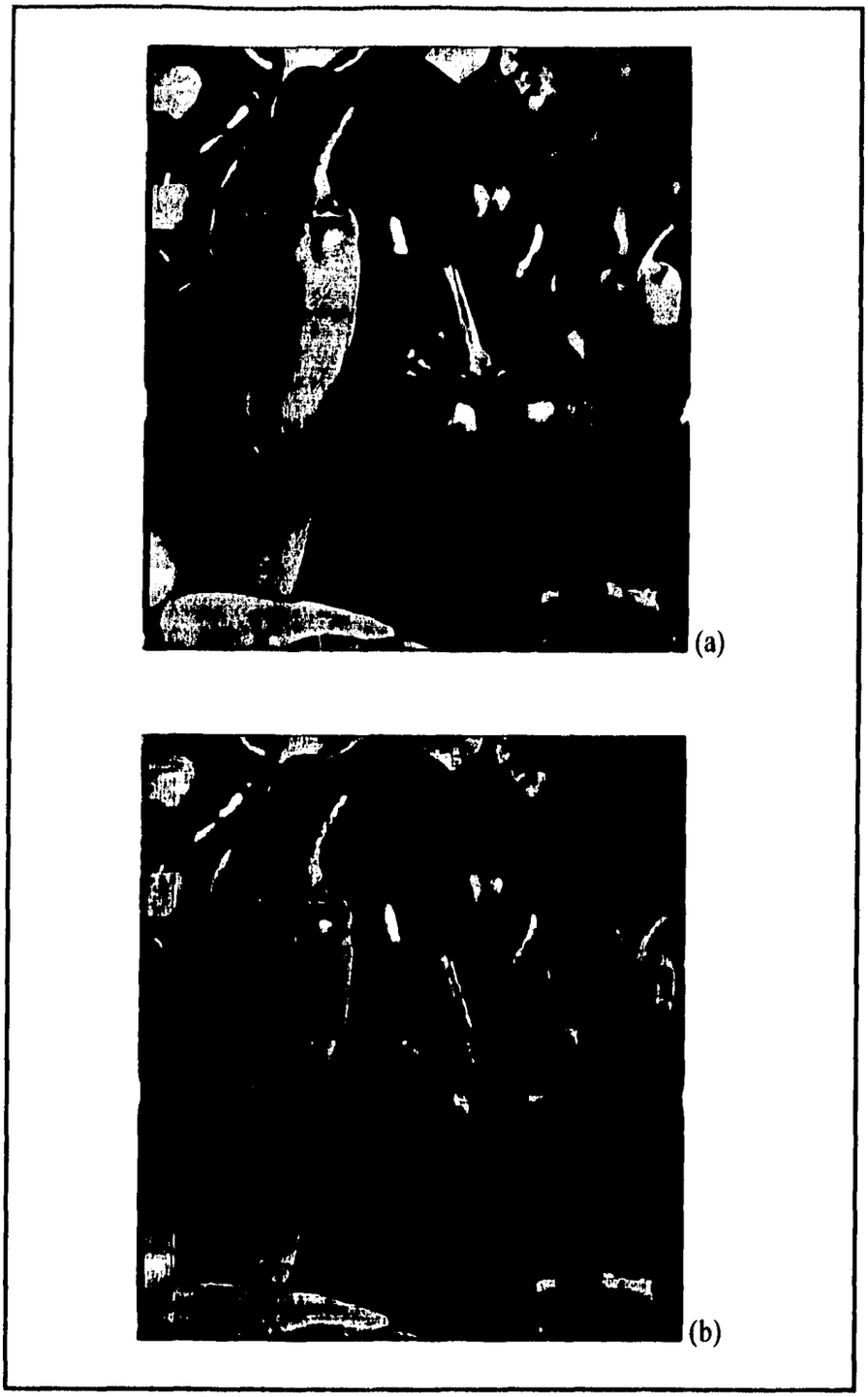


Figura 5.6.8. Original @8bpp (a) y 2D-VDCT+CVQ @0.3418bpp (b) de la imagen Peppers, 512x512 pixels, de la base de datos de la University of Southern California (USC).

Figura 5.6.9. Versión 2D-VDCT+CVQ @ 0.587bpp de la imagen Peppers 512x512 de la base de datos de la University of Southern California (USC).

Tabla 5.6.6. Mapa de plantillas para la imagen Peppers en el dominio de la 2D-VDCT (8x8)

7	1	7	7	7	7	7	7	7	7	7	7	1		7	12
10	5	5	9	9	7	7	7	12	10	10	9	9	5	12	12
5	5	5	5	12	7	5	12	12	12	10	5	9	12	12	12
5	5	5	12	12	12	12	12	10	12	12	12	12	12	12	12
5	5	5	5	12	12	12	12	5	5	12	12	12	12	12	11
5	5	5	5	12	12	12	12	12	12	10	12	12	12	12	9
5	5	5	12	12	12	12	12	12	10	12	12	12	12	10	9
5	5	12	12	12	12	12	12	12	7	11	12	12	5	9	10
5	5	12	12	12	12	12	12	12	12	12	12	12	5	10	5
5	5	12	12	12	12	12	5	5	5	5	12	5	9	9	9
5	5	12	12	12	12	12	5	10	10	12	12	5	9	5	9
5	5	12	12	12	12	12	12	10	10	12	12	9	9	5	10
12	1	12	12	12	12	12	12	7	12	12	5	5	10	5	12
12	5	12	12	12	12	12	12	9	9	10	12	9	1	5	5
12	5	12	12	12	12	12	12	9	9	9	12	7	1	12	11
12	1	12	12	12	12	12	10	9	9	9	5	5	11	12	12

Para completar la comparación subjetiva, se presentan las imágenes de error obtenidas de restar pixel a pixel las imágenes originales con las imágenes procesadas por el sistema de compresión. En estas imágenes de error, se pueden apreciar más claramente los cuatro tipos de perturbaciones debidas a la codificación de transformada.

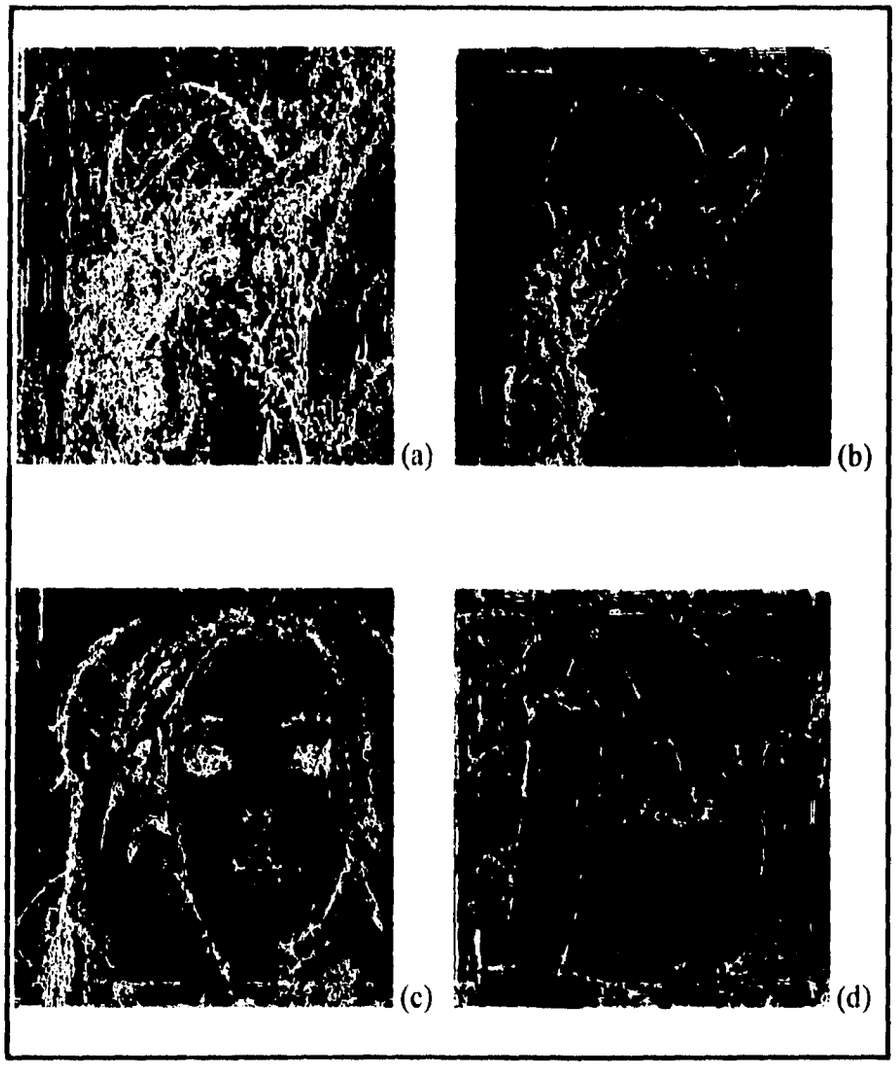


Figura 5.6.10. Señal de error entre los originales y las versiones codificadas, a 0.3418bpp, por medio del esquema 2D-VDCT-CVQ, de las imágenes. Lady (a), Lena (b), Girl (c) y Peppers (d), de la base de datos de la University of Southern California (USC).

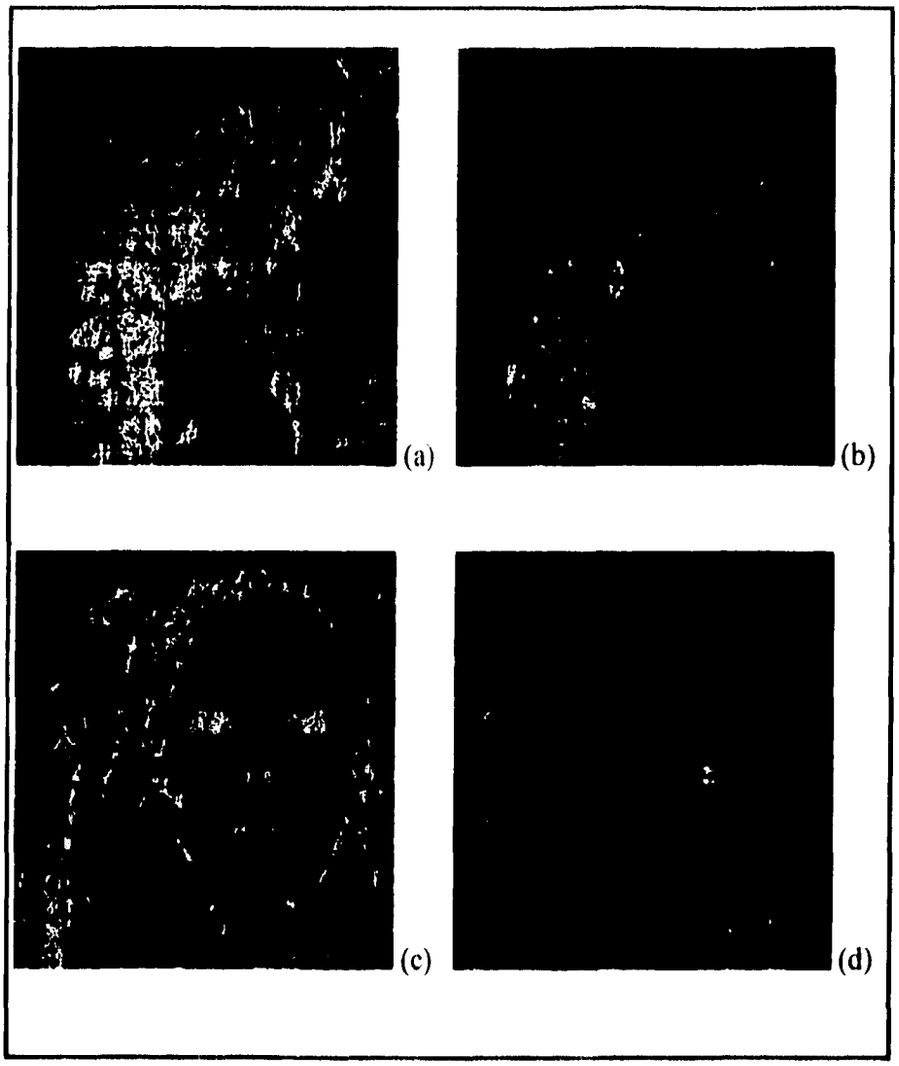


Figura 5.6.11. Señal de error entre los originales y las versiones codificadas, a 0.5870bpp, por medio del esquema 2D-VDCT+CVQ, de las imágenes. Lady (a), Lena (b), Girl (c) y Peppers (d), de la base de datos de la University of Southern California (USC).

5.7 Herramientas de desarrollo y comparación

El desarrollo de métodos de codificación de fuente, tales como la cuantificación vectorial son procesos estrechamente ligados a la simulación por computadora, debido principalmente a que se trata de métodos numéricos. Por otro lado, muchas de las ideas que pueden llevar al planteamiento de un nuevo método, la mayoría de las veces son motivadas por los resultados de experimentos planteados de una forma intuitiva. Además, como no existen modelos probabilísticos que puedan simular plenamente el comportamiento de las señales de la vida real, tales como voz e imágenes, el diseño de sistemas de procesamiento para este tipo de señales generalmente se basa en métodos estadísticos basados en simulación por computadora.

Para realizar las pruebas necesarias durante las diferentes partes de la etapa experimental, fue necesario construir una serie de herramientas de *software*. A continuación se presenta una lista de dichos programas y algunas de sus funciones principales.

- Caracterización y procesamiento de imágenes. *VQ013*.
  - Recorte de imágenes
  - Amplificación, exploración de la imagen por medio de un cursor, 3 ajustes de tono, búsqueda de máximos y mínimos.
  - Diferencia entre imágenes, cálculo de PSNR y MSE.
  - Decimación, elongación, suma y producto por un escalar.

- Cálculo de la 2D-VDCT y la 2D-BDCT. *VQ016*
  - 2D-VDCT y 2D-BDCT directas e inversas.
  - Amplificación, exploración de la imagen por medio de un cursor, 3 ajustes de tono, búsqueda de máximos y mínimos.
  - Diferencia entre imágenes, cálculo de PSNR y MSE.
  - Cálculo de los valores muestra de: BIT, ICF, varianza, media y norma.
  - Clasificación de coeficientes, etiquetación por umbrales de cuantificación vectorial clasificada.
  - Etiquetación de bloques para codificación zonal.

- Caracterización de secuencias de entrenamiento para fuentes aleatorias n-dimensionales. *VQ017 y VQ019*
  - Cálculo de los valores muestra de: BIT, ICF, varianza, media y norma.
  - Estimación de BIT e ICF.
  - Concentración de energía para una clasificación por medio de la norma.

- Curva de acumulación de la energía.
- Eliminación y restauración de la media muestra y de la norma muestra.
- Histogramas de distribución de las componentes y de la norma de los vectores.
- Generación de vectores aleatorios con distribuciones Gaussiana, Laplaciana y Gauss-Markov de primer orden.
- Almacenamiento, recorte y unión de secuencias.
- Cuantificación escalar de media y norma muestra por medio de: Lloyd-Max y SQ uniforme de 8 16 y 32 bits.
- Cálculo, en base a una secuencia de entrenamiento, de la matriz de correlación de las componentes del vector.

- Diseño de cuantificadores vectoriales: VQ, CVQ_{AO} y CVQ_{GI}.*VQ020*
  - Gráficas de SNR por vector para codificación por medio de VQ y CVQ.
  - Codificación por medio de VQ y CVQ y cálculo de la tasas de codificación.
  - Etiquetación de vectores por medio del clasificador basado en la norma.
  - Codificación de grupos de secuencias de coeficientes de la 2D-VDCT y cálculo de la tasa de codificación
  - Diseño del clasificador en base a los atributos óptimos y a las geometrías implícitas.
  - Diseño de libros de códigos de VQ y CVQ para diferentes tasas de código.
  - Cálculo de la longitud promedio de Huffman para el clasificador y los libros de códigos de VQ y CVQ.
  - Almacenamiento, unión y recorte de secuencias de entrenamiento.

Los algoritmos de diseño de cuantificadores vectoriales, utilizados durante la parte experimental de este trabajo de tesis, requieren de una cantidad grande de cálculos. Es por eso que su construcción en lenguajes tipo intérprete tales como *MatLab* es poco práctica. En busca de acelerar el tiempo de cálculo de esos procesos, y construir herramientas de simulación *amigables*, se eligió programar en *Pascal* para *Windows*.

Conclusiones

A partir de 1990 empiezan a aparecer estudios que buscan determinar la forma de la etapa de procesamiento de la señal que optimiza el desempeño de la etapa de cuantificación vectorial en sistemas de compresión. En estos trabajos se han propuesto esquemas de procesamiento tales como: transformaciones vectoriales, bancos de filtros vectoriales, etc. Los esquemas de procesamiento basados en esos bloques de procesamiento de la señal que se han propuesto hasta ahora, utilizan cuantificación vectorial ordinaria del vecino más cercano.

En este trabajo se utilizó una etapa de procesamiento vectorial, específicamente la transformada coseno discreta vectorial, 2D-VDCT, como punto de inicio a lo que puede ser un estudio más profundo sobre nuevas formas de aplicación de la VQ en sistemas de compresión con procesamiento orientado a vectores.

En primer lugar, para justificar el uso de una transformación *vectorial* en lugar de una transformación *escalar* en un sistema de compresión, se compararon dos esquemas de procesamiento de la señal, la 2D-VDCT y la 2D-BDCT. Esta comparación se realizó en función de los atributos que hacen que una etapa de procesamiento de la señal optimice el desempeño de la etapa de cuantificación vectorial en un sistema de compresión. Estos atributos son: el nivel de la correlación entre vectores y el nivel de la correlación entre las componentes de los vectores, en el dominio transformado.

Se comprobó que la etapa de procesamiento de la señal orientada a vectores, la 2D-VDCT, realiza mejor la tarea de preservación de la correlación intravector que la etapa orientada a escalares, la 2D-BDC. Además, el nivel de descorrelación entre los vectores logrado en ambas transformaciones es similar. También se encontró que las matrices de covarianza de los coeficientes de la 2D-VDCT tienen componentes fuera de la diagonal principal con valores grandes. Esto significa que existen algunas componentes de los vectores que son muy dependientes entre sí. En cambio, en el

caso de señales procesadas por medio de la 2D-BDCT, las matrices de covarianza de las componentes de los vectores se encuentran concentradas fundamentalmente en la diagonal principal. Otra característica importante de los vectores en el dominio de la 2D-VDCT, es que el grado de correlación intervector que la transformación no puede eliminar, se presenta en la forma de patrones de distribución de la energía bien definidos. Esto es, dependiendo de las características de la señal original, la energía de la señal en el dominio transformado se distribuye en un grupo de coeficientes arreglados en una forma geométrica bien definida.

Hubiera sido deseable realizar la comparación de los dos sistemas de procesamiento de la señal, 2D-VDCT y 2D-BDCT, por medio de los resultados *tasa-distorsión* de VQ y CVQ, sobre secuencias de entrenamiento obtenidas al procesar un conjunto de imágenes. Sin embargo, el gran tamaño de las secuencias de entrenamiento requerido para justificar la estacionariedad y ergodicidad hacen que sea necesario tener mucho espacio de almacenamiento, además de que el tiempo de cálculo necesario para hacer todas las pruebas no hubieran permitido terminar este trabajo dentro de los límites de tiempo establecidos.

Una vez que se tuvieron claras las características de los vectores en el dominio de la 2D-VDCT se determinó que una forma de explotar estas características en particular podría ser la utilización de un cuantificador vectorial que explotara la dependencia estadística entre estos vectores, esto es, la utilización de cuantificación vectorial con memoria.

Como el resultado de observaciones experimentales, se determinó que el uso de un conjunto de cuantificadores clasificados en los cuales el modo de operación de cada uno de ellos se seleccionara en base a la distribución de la energía en todo el conjunto de vectores, permite introducir memoria a la etapa de cuantificación de una manera sencilla, con la ventaja adicional de que la información lateral requerida por este esquema de cuantificación vectorial con memoria, se puede introducir fácilmente en la información requerida por un bloque de codificación zonal, en un sistema de compresión por transformada.

La filosofía de diseño que se utilizó para los cuantificadores vectoriales es la que se basa en una secuencia de entrenamiento. Los métodos de diseño de cuantificadores vectoriales basados en secuencias de entrenamiento se justifican únicamente cuando la fuente que genera la secuencia es estacionaria (o al menos asintóticamente estacionaria en la media) y cumple con la propiedad ergódica, esto es que los valores esperados puedan ser aproximados por promedios muestra de término largo. Durante el diseño de un cuantificador vectorial, la forma de verificar que estas dos propiedades se cumplen consiste en lo siguiente: se fija una longitud de la secuencia de entrenamiento, se diseña el cuantificador y se prueba en una secuencia (generada por la misma fuente que generó la secuencia de entrenamiento) que no está incluida en la secuencia de entrenamiento. Si el desempeño de prueba es cercano al desempeño de diseño entonces la longitud de la secuencia de

entrenamiento es adecuada. En caso contrario, hay que repetir el proceso de diseño con una secuencia de entrenamiento más larga.

El método descrito arriba fue utilizado en este trabajo para justificar el tamaño de las secuencias de entrenamiento utilizadas durante el proceso de construcción de los cuantificadores vectoriales bajo comparación: el clasificado (con memoria) y el ordinario del vecino más cercano (sin memoria).

El algoritmo que se utilizó para el diseño de esos cuantificadores fue el LBG (debido a Linde, Buzo y Gray), al cual se le añadieron el criterio de la distorsión parcial y los métodos de búsqueda rápida de la esfera y del anillo. Para validar el desempeño del programa de computadora que realiza dicho algoritmo, se generaron libros de códigos de distintas tasas sobre una secuencia de entrenamiento de 16 imágenes. Los libros generados se usaron para codificar la imagen *Lena*, la cual no estaba incluida en la secuencia de entrenamiento, y los resultados que se obtuvieron son muy parecidos a los que se presentan en [VQ52].

Como se mencionó en el capítulo cuatro, en este trabajo se propone la aplicación de cuantificación vectorial con memoria, a través de un conjunto de cuantificadores clasificados, en un sistema de compresión que utiliza una etapa de procesamiento de la señal orientado a vectores. Para la construcción del clasificador de los cuantificadores vectoriales, se formularon dos métodos: el de los atributos óptimos y el de las geometrías implícitas. Estos dos criterios de diseño fueron comparados en términos del tiempo de cálculo requerido para construir el clasificador y del desempeño *tasa-distorsión* que el cuantificador vectorial puede alcanzar.

A través de todas las pruebas de comparación realizadas, el método de diseño del clasificador que proporciona mejores resultados es el que se basa en las geometrías implícitas. Este método tiene la ventaja de que el tiempo de cálculo requerido para encontrar el conjunto de umbrales es mucho más pequeño (alrededor de 20 veces menos) que el que necesita el algoritmo basado en el criterio de los atributos óptimos. Además, el mejor desempeño *tasa-distorsión* del cuantificador vectorial clasificado, se obtiene cuando el clasificador se diseño en base al criterio de las geometrías implícitas.

Originalmente el cuantificador clasificado se pensó como una forma de añadir memoria a la cuantificación vectorial de los coeficientes de la 2D-VDCT. Esto se hace de la siguiente forma. En base a la distribución de la energía en los coeficientes del bloque transformado, una plantilla de supresión de coeficientes, indica que coeficientes se preservan y en que modo de operación deben ser cuantificados. Las características de cada plantilla se obtienen al observar las estadísticas de un conjunto de bloques transformados, y son fijas. De esta forma toda la información de clasificación de cada uno de los coeficientes es conocida tanto en el codificador como en el decodificador, por lo que la única información secundaria que es necesario enviar con cada bloque codificado, es el índice de la plantilla. Como se

tiene un conjunto pequeño de plantillas, la tasa requerida para identificar la plantilla correspondiente a cada bloque es despreciable, en comparación con la tasa requerida por el conjunto de índices de cada uno de los coeficientes. Bajo estas consideraciones, la comparación de CVQ y VQ, en base a la curva *tasa-distorsión*, se debería hacer sin incluir, en CVQ, la tasa requerida por el clasificador. Sin embargo, los resultados experimentales demuestran que, al incluir en la tasa de cada uno de los coeficientes la tasa requerida por la clasificación, el desempeño de CVQ sigue siendo mejor que el de VQ (a partir de un cierto valor de la tasa del código). Teóricamente, no existe un cuantificador con memoria que se desempeñe mejor que el cuantificador del vecino más cercano ordinario. Esto llevó a que inicialmente la incongruencia de los resultados experimentales con la teoría se explicara de la siguiente forma: al permitir que los diferentes modos de operación del cuantificador clasificado tengan distintas longitudes de código, se está introduciendo cierta forma de codificación entrópica en este cuantificador.

Por lo tanto, para tener una comparación real entre los dos esquemas de cuantificación vectorial, también se debe realizar una comparación *tasa-distorsión* utilizando codificación entrópica de los índices de los vectores de código.

Después de realizar la comparación de esta forma se observó que, si bien el desempeño de CVQ permanecía siendo mejor que el de VQ, la brecha entre ambos disminuyó notablemente.

Otra observación importante es que si bien el cuantificador clasificado produce mejores resultados que el cuantificador ordinario del vecino más cercano, en la codificación de los coeficientes de la 2D-VDCT, esto es válido únicamente en un intervalo de la tasa del código. Este hecho limita la aplicación del cuantificador clasificado a aplicaciones en las que no se requieran tasas muy bajas del código. Sin embargo, si se toma en cuenta que los nuevos desarrollos de codificación de transformada (contexto en el que se utilizará al cuantificador clasificado que se propone aquí) indican que este esquema es más apropiado para compresión a tasas medias, ya que esquemas como la codificación subbanda producen mejores resultados a tasas muy bajas (en [VQ5] se muestran datos que demuestran esto) entonces la limitación del cuantificador clasificado a tasas bajas no es un problema que impida su aplicación en el esquema de codificación por transformada.

Finalmente, cuando se comparan los cuantificadores clasificado y el ordinario del vecino más cercano, se observa que, aunque el cuantificador clasificado tiene una mayor cantidad de palabras de código, su tiempo de diseño es menor (alrededor del 50%) al del vecino más cercano ordinario. Esto se debe principalmente a que el proceso de diseño de este último cuantificador, requiere de casi el triple de iteraciones para converger.

Lógicamente, el conjunto de cuantificadores vectoriales clasificados, al utilizarse como un medio de realizar cuantificación vectorial con memoria en los coeficientes de la 2D-VDCT, no es adecuado para la codificación de fuentes sin

memoria. Para comprobar este hecho, se aplicó un cuantificador clasificado a secuencias de entrenamiento de procesos Gaussiano y Laplaciano independientes idénticamente distribuidos y de un proceso Gauss-Markov de primer orden. Como era de esperarse, el desempeño del cuantificador clasificado fué menos bueno que el del cuantificador ordinario del vecino más cercano. Este resultado es una consecuencia de que, al utilizar un sólo cuantificador clasificado la información de los índices de clasificación tiene que ser incluida con el índice de cuantificación de cada uno de los vectores. Esto contrasta con el caso de la codificación de los coeficientes de la 2D-VDCT, en el que los índices de clasificación de un conjunto de vectores son agrupados juntos en sólo un índice de plantilla. En resumen, el uso de un sólo CVQ no permite introducir memoria a la cuantificación vectorial.

Para el contexto de la codificación de imágenes, en este trabajo se presentó un esquema de compresión por transformada basado en la 2D-VDCT, un conjunto de cuantificadores clasificados y un bloque de codificación zonal. Este es un esquema de compresión muy simple con asignación fija de bits(incluso no se incluyó codificación entrópica en las palabras del código de los cuantificadores), pero los resultados *tasa-distorsión* obtenidos, lo sitúan cerca de esquemas bastante más complicados, como el sistema de compresión con asignación dinámica de información [VQ5].

El sistema presenta los problemas típicos de los esquemas de compresión por transformada. Esto es, las imágenes presentan problemas de difuminación y *blocking* cuando son codificadas a tasas bajas. Otro de los inconvenientes es que el espacio de almacenamiento requerido por los CVQ para construir un bloque de cuantificación con asignación dinámica de bits, es bastante más grande que el requerido por los VQ.

En el esquema de codificación que se propuso en este trabajo, la clasificación es hecha en base a plantillas preestablecidas. Una refinación que posteriormente se puede hacer a este esquema, es utilizar los índices de clasificación de la plantilla como una predicción del índice real de clasificación. El índice real se obtendría aplicando el clasificador al vector. La diferencia entre el índice predicho y el real se transmitiría como una señal de error.

Bibliografía

VQ1. Robert M. Gray, "Vector Quantization", IEEE ASSP Magazine, pp. 4-29, April 1984.

VQ2. Yoseph Linde, Andrez Buzo and Robert M. Gray, "An Algorithm for Vector Quantizer Design", IEEE Transactions on Communications, Vol. COM-28, pp. 84-95, January 1980.

VQ3. William H. Equitz, "A New Vector Quantizer Clustering Algorithm", IEEE Transactions on Acoustics, Speech and Signal Processing, Vol. 37, pp. 1568-1575, Ocober 1989.

VQ4. C. M. Huang, Q. Bi, G. S. Stiles and R. W. Harris, "Fast Full Search Equivalent Encoding Algorithms for Image Compression Using Vector Quantization", IEEE Transactions on Image Processing, Vol. 1, No. 3, pp. 413-416.

VQ5. Weiping Li and Ya-Qin Zhang, "Vector-Based Signal Processing and Quantization for Image and Video Compression", Proceedings of the IEEE, Vol. 83, No. 2, pp317-335, February 1995.

VQ6. Morris Goldberg and Huifang Sun, "Image Sequence Coding Using Vector Quantization", IEEE Transactions on Communications, Vol. COM-34, pp. 703-710, July 1986.

VQ7. Hsia-hui Shen and Richard L. Baker, "A Finite State/Frame Difference Interpolative Vector Quantizer for Low Rate Image Sequence Coding", IEEE

International Conference on Acoustics, Speech and Signal Processing, Vol. 2, pp. 118-1191, April 1988.

VQ8. Christine I. Poldilchuk, Nikil S. Jayant and Nariman Farvadin, "Three-Dimensional Subband Codign of Video", IEEE Transactions on Image Processing, Vol. 4, No. 2, pp.125-139, February 1995.

VQ9. Kiyoharu Aizawa and Thomas S. Huang, "Model-Based Image Coding: Advanced Video Coding Techniques for Very Low Bit-Rate Applications", Proceedings of the IEEE, Vol. 83, No. 2, pp.259-271, February 1995.

VQ10. Taner Ozcelik, James, C. Brailean and Aggelos K. Katsaggelos, "Image and Video Compression Algorithms Based on Tecovery Techniques Using Mean Field Annealing", Proceedings of the IEEE, Vol. 83, No. 2, pp. 304-316, February 1995.

VQ11. Sakae Okubo, "Reference Model Methology-A Tool for the Collaborative Creation of Video Coding Standards", Proceedings of the IEEE, Vol. 83 No. 2, pp. 139-150, February 1995.

VQ12. Kiran Challapali *et al*, "The Grand Alliance System for US HDTV", Proceedings of the IEEE, Vol. 83, No. 2, pp. 158-173, February 1995.

VQ13. Robert D. Dony and Simon Haykin, "Neural Network Approaches to Image Compression", Proceedings of the IEEE, Vol. 83, No. 2, pp. 288-303, February 1995.

VQ14. Leonardo Chiariglione, "The Development of an Integrated Audiovisual Coding Standard: MPG", Proceedings of the IEEE, Vol. 83, No. 2, pp.151-163, February 1995.

VQ15. Jianping Pan and Thomas R. Fischer, "Two-Stage Vector Quantization-Lattice Vector Quantization", IEEE Transactions on Information Theory, Vol. 41, No. 1, pp.155-163, January 1995.

VQ16. Dae Gwon Jeong and Jerry D. Gybson, "Image Coding with Uniform and Piecewise-Uniform Vector Quantizers", IEEE Transactions on Image Processing, Vol. 4, No. 2, pp. 140-146, February 1995.

VQ17. Weiping Li and Ya-Quin Zhang, "Scanning the Issue. Special Issue on Advances in Image and Video Compression", Proceedings of the IEEE, Vol. 83, No. 2, pp.135-138, February 1995.

VQ18. Jonatan N. Bradley and Christopher MBrislawn, "Image Compression by Vector Quantization of Multiresolution Decompositions", Proceedings of the CNLS, 11th Annual Conference on Experimental Mathematics: Computational Issues in Nonlinear Scinece. May 1991.

VQ19. Michel Barlaud *et al*, "Pyramidal Lattice Vector Quantization for Multiscale Image Coding", 13S Lboratory, URA 1376 CNRS, University of Nice- Sophia Antipolis, April 1992.

VQ20. Olivier Egger, Wei Li and Murat Kunt, "High Compression Image Coding Using an Adptive Morphological Subband Decomposition", Proceedings of the IEEE, Vol. 83, No. 2, pp. 272-287, February 1995.

VQ21. Mohammad Gharabi-Alkhansari and Thomas S. Huang, "A Fractal-Based Image Block-Coding Algorithm", IEEE International Conference on Acoustics, Speech and Signal Processing, Vol. 5, pp. V345-V348, 1995.

VQ22. Weiping Li and Ya-Qin Zhang, "New Insights and Results on Transform Domaing VQ of Images", IEEE International Conference on Acoustics, Speech and Signal Processing, Vol. 5, pp. V609-V612, 1994.

VQ23. Brian DeCleene and Henrik Sorensen, "Multiresolution Vector Transform Codign for Video Compression", IEEE International Conference on Acoustics, Speech and Signal Processing, Vol. 5, pp. V413-V416, 1994.

VQ24. K. Metin Uz, M. Zaphiro and Martin Czigler, "Optimal Bit Allocation in the Presence of Quantizer Feedback", IEEE International Conference on Acoustics, Speech and Signal Processing, Vol. 5, pp. V385-V388, 1994.

VQ25. Skjalg Lepsoy, Geir E. Oien and Tor A. Ramstad, "Attractor Image Compression with a Fast Non-Iterative Decoding Algoritm", IEEE International Conference on Acoustics, Speech and Signal Processing, Vol. 5, pp. V337-V340, 1994.

VQ26. Lester Thomas and Farzin Deravi, "Pruning of the Transform Space In Block-Based Fractal Image Compression", IEEE International Conference on Acoustics, Speech and Signal Processing, Vol. 5, pp. V341-V344, 1994.

VQ27. Yair Shoham and Allen Gersho, "Efficient Bit Allocation For an Arbitrary Set of Quantizers", IEEE Transactions on Acoustics, Speech and Signal Processing, Vol. 36, No. 9, pp.1445-1453, September 1988.

VQ28. Weiping Li, "On Vector Transformation", IEEE Transactions on Signal Processing, Vol. 41, No. 11, pp.3114-3126, November 1990.

VQ29. M. Antonini, M. Barlaud, P. Mathieu and I. Daubechies, "Image Coding using Vector Quantization in the Wavelet Transform Domain", IEEE Transactions on Image Processing, Vol. 1, pp. 5-220, April 1992.

VQ30. Thomas R. Fischer, "Geometric Source Coding and Vector Quantization", IEEE Transactions on Information Theory, Vol. 35, No. 1, pp137-145, January 1989.

VQ31. Stephen Wond, Loren Zaremba, David Gooden, and H. K. Huang, "Radiologic Image Compression-A Review", Proceedings of the IEEE, Vol., 83, No. 2, pp194-219, February 1995.

VQ32. J. P. Laresp and C. Labit, "Vector Quantization in Transformed Image Coding", IRISA/ Centre INRIA de Rennes, 35042 RENNES Cedex, France.

VQ33. Federico Kuhlmann, James A. Bucklew, "Piecewise Uniform Vector Quantizers", IEEE Transactions on Information Theory, Vol. 34, No. 5, pp1259-1263, September 1988.

VQ34. Randall C. Reininger and Jerry D. Gibson, "Distributions of the Two-dimensional DCT Coefficientes for Images", IEEE Transactions on Communications, Vol. 31, No. 6, pp835-839, June 1983.

VQ35. Yariv Ephraim, Hanoch Lev-Ari and Robert M. Gray, "Asymptotic Minimum Discrimination Information Measure for Asymptotically Wakly Stationary Process", IEEE Transactions on Information Theory, Vol. 34, No. 5, pp1033-1040, September 1988.

VQ36. Daniel F. Fuhrmann and Michael I. Miller, "On the Existence of Positive-Definite Maximum-Likelihood Estimates of Structured Covariance Matrices, IEEE Transactions on Information Theory, Vol 34, No. 4, pp722-729, July 1988.

VQ37. Dimitri Kazakos, "New Results on Robust Quantization", IEEE Transactions on Communications, Vol. COM-31, No. 8, pp965-973, August 1983.

VQ38. Layne T. Watson, Robert M. Haralick and Oscar A Zuñiga, "Constrained Transform Coding and Surface Fitting", IEEE Transactions on Communications, Vol. COM-31, No. 5, pp. 717-726, May 1983.

VQ39. Jean-Pierre Adoul and Michel Barth, "Neares Neighbor Algorithm for Spherical Codes from the Leech Lattice", IEEE Transactions on Information Theory, Vol. 34, No. 5, pp. 1188-1197, September 1988.

VQ40. Mikhail Shnaider and Andrew P. Paplinski, "Wavelet Transform in Image Coding", Faculty of Computing and Information Technology, Departament of Robotics and Digital Technology, Technical Report 94-11, October 1994.

VQ41. Dae Gwon Jeong and Jerry D. Gibson, "Uniform and Piecewise Uniform Lattice Vector Quantization for Memoryless Gaussian and Laplacian Sources", IEEE Transactions on Information Theory, Vol. 39, No. 3, pp.786-804, May 1993.

VQ42. John C. Kieffer, Teresa M. Jahns and Viktor A. Obuljen, "New Results on Optimal Entropy-Constrained Quantization", IEEE Transactions on Information Theory, Vol. 34, No. 5, pp. 1250-1258, September 1988.

VQ43. Peter F. Swaszek and John B. Thomas, "Multidimensional Spherical Coordinates Quantization", IEEE Transactions on Information Theory, Vol. IT-29, No. 4, pp570-576, July 1983.

VQ44. James A. Buckew and Neal C. Gallagher Jr., "Quantization Schemes for Bivariate Gaussian Random Variables", IEEE Transactions on Information Theory, Vol. IT-25, No. 4, pp. 537-543.

VQ45. Willian A. Pearlman, "Polar Quantization of a Complex Gaussian Random Variable", IEEE Transactions on Communications, Vol. COM-27, No. 6, pp. 892-899, June 1979.

VQ46. Stephen G. Wilson, "Magnitude/Phase Quantization of Independent Gaussian Variates", IEEE Transactions on Communications, Vol. COM-28, No. 11, pp. 1924-1929, November 1980.

VQ47. Peter F. Swaszek and Tsu W. Ku, "Asymptotic Performance of Unrestricted Polar Quantizers", IEEE Transactions on Information Theory, Vol. IT-32, No. 2, pp.330-333, March 1986.

VQ48. David A. Huffman, "A Method for the Constuction of Minimum-Redundancy Codes", Proceedings of the IRE, pp. 1098-1101, September 1952.

VQ49. Robert M. Gray, *Source Coding Theory*. Kluwer Academic Publishers, 1990.

VQ50. A. Murat Tekalp, *Digital Video Processing*. Prentice Hall, 1995.

VQ51. M. Barlaud, *Wavelets in Image Communication*. Elsevier, Amsterdam, 1994.

VQ52. Allen Gersho and Robert M. Gray, *Vector Quantization and Signal Compression*. Kluwer Academic Publishers, 1992.

VQ53. Michael J. Sabin and Robert M. Gray, "Product Code Vector Quantizers for Waveform and Voice Coding", IEEE Transactions on Acoustics, Speech and Signal Processing, Vol. ASSP-32, pp. 474-488, June 1984.

VQ54. Kevin T. Malone and Thomas R. Fischer, "Contour-Gain Vector Quantization", IEEE Transactions on Acoustics, Speech and Signal Processing, Vol. 36, No.6, pp. 862-870, June 1988.

VQ55. Thomas R. Fischer, "A Pyramid Vector Quatizer", IEEE Transactions on Information Theory, Vol IT-32, No. 4, pp. 568-583, July 1986.

VQ56. David J. Sakrison, "A Geometric Treatment of the Source Encoding of a Gaussian Random Variable", IEEE Transactions on Information Theory, Vol. IT-14, No. 3, pp. 481-486, May 1968.

VQ57. Vladimir Cuperman, "Joint Bit Allocation and Dimensions Optimization for Vector Transform Quantization", IEEE Transactions on Information Theory, Vol 39, No. 1, pp. 302-305, January 1993.

VQ58. Weiping Li, "Vector Transform and Image Coding", IEEE Transactions on Circuits and Systems for Video Technology, Vol. 1, pp. 297-307, December 1991.

VQ59. W. Ding, "Optimal Vector Transform for Vector Quantization", IEEE Signal Processing Letters, December 1993.

VQ60. Norman S. Matloff, *Probability Modeling and Computer Simulation*, PWS Kent Publishing Company, Boston Massachusetts, 1988.

VQ61. K. R. Rao and P. Yip, *Discrete Cosine transform Algorithms, Advantages, Applications*, Academic Press, Inc., San Diego, CA., 1990.

VQ62. Huey-Chen Tseng and Thomas R. Fischer, "Transform and Hybrid Transform/DPCM Coding of Images Using Pyramid Vector Quantization", IEEE Transactions on Communications, Vol., 35, No. 1, pp. 79-86, January 1987.

VQ63. Ahmet M. Eskicioglu and Paul S. Fisher, "Image Quality Measures and Their Performance", IEEE Transactions on Communication, Vol. 43, No. 12, pp. 2959-2965, December 1995.

VQ64. Yamada, "Asymptotic Performance of Block Quantizers", IEEE Transactions on Information Theory, Vol. IT-26, No. 1, pp. 6-14, January 1980.

VQ65. Philip A. chou, Tom Lookabaugh and Robert M. Gray, "Entropy-Constrained Vector Quantization", IEEE Transactions on Acoustics, Speech and Signal Processing, Vol. 37, No. 1, pp. 31-42, January 1989.

VQ66. Peter Noll and Rainer Zelinski, "Bounds on Quantizer Performance in the Low Bit Region", IEEE Transactions on Communications, Vol. COM-26, No. 2, pp. 300-304, February1978.

VQ67. Thomas R. Fischer and Roy M. Dicharry, "Vector Quantizer Design for Memoryless Gaussian, Gamma, and Laplacian Sources", IEEE Transactions on Communications, Vol. COM-32, No. 9, pp. 1065-1069, September 1984.

VQ68. Allen Gersho, "Asymptotically Optimal Block Quantization", IEEE Transactions on Information Theory, Vol. IT-25, No. 4, July 1979.

VQ69. Wolfgang Mauersberger, "Experimental Results on the Performance of Mismatched Quantizers", IEEE Transactions on Information Theory, Vol. IT-25, No. 4, pp.381-386, July 1979.

VQ70. R. A. King and N. M. Nasrabadi, "Image Coding Using Vector Quantization in the Transform Domain", Patern Recognition Letters, Vol. 1, pp. 323-329, July 1983.

VQ71. C. E. Shannon, "A Mathematical Theory of Communication", Bell Systems Technical Journal, Vol. 27, pp. 379-423, 623-656, May 1968.

Resumen

A partir de 1990 empiezan a aparecer estudios sobre etapas de procesamiento de la señal orientadas a la optimización de la cuantificación vectorial en sistemas de compresión. En este trabajo se utiliza una etapa de procesamiento vectorial, específicamente la transformada coseno discreta vectorial, 2D-VDCT, como punto de inicio a lo que puede ser un estudio más profundo sobre nuevas formas de aplicación de la VQ en sistemas de compresión con procesamiento orientado a vectores. También se presenta un estudio comparativo entre dos etapas de procesamiento de la señal, una escalar y otra vectorial, en base a los atributos que permiten que una señal sea codificada eficientemente por medio de cuantificación vectorial. Observaciones hechas sobre imágenes en el dominio de la transformada 2D-VDCT revelaron que existe una dependencia estadística entre los coeficientes, debida a las características de la imagen original. Estas observaciones indican que existe una cierta cantidad de correlación en el dominio de la 2D-VDCT que puede ser explotada por la cuantificación interdependiente de los coeficientes, esto es, mediante el uso de cuantificadores vectoriales con memoria. En este trabajo se presenta una forma de introducir memoria en la cuantificación vectorial de los coeficientes de la 2D-VDCT por medio de un cuantificador vectorial clasificado. Para la parte de clasificación, se presentan dos métodos de diseñó del clasificador basados en la norma de los vectores. Para realizar el diseño de los cuantificadores vectoriales, tanto clasificado como ordinario, fue necesaria la justificación de la estacionariedad y ergodicidad en base a las longitudes de las secuencias de entrenamiento. Una vez que se diseñaron estos cuantificadores, se procedió a la comparación del desempeño en base a curvas *tasa-distorsión*, obtenidas sobre secuencias de entrenamiento de los distintos coeficientes. Finalmente, los dos esquemas de cuantificación se compararon dentro de un esquema de compresión de imágenes.