



UNIVERSIDAD NACIONAL
AUTÓNOMA DE MÉXICO.

DIVISIÓN DE ESTUDIOS DE POSGRADO
FACULTAD DE INGENIERÍA.

Codificación de imágenes empleando técnicas de selección adaptable.

TESIS
QUE PARA OBTENER EL GRADO DE DOCTOR EN INGENIERÍA
PRESENTA

Rodrigo Montúfar Chaveznava.



Director de Tesis: Dr. Francisco J. García Ugalde.
Departamento de Ingeniería Eléctrica.
División de Estudios de Posgrado.
Facultad de Ingeniería, UNAM.

México D.F.

Junio 1999

TESIS CON
FOLIO DE ORIGEN

273171



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

“Evitemos el orgullo intelectual que genere el saber, y no pongamos a la ciencia al servicio de la vanidad individual, ya que su principal objetivo es anunciar la verdad”.

Agradecimientos.

Son muchas las personas a las que debo expresar mis agradecimientos. Sin duda alguna, sin cada una de ellas este trabajo jamás se hubiera realizado, ya que a ellas debo todas las cosas que he vivido durante el desarrollo de este trabajo.

En primer lugar quiero agradecer a Gerardo Coello Coutiño, Jefe de la Unidad de Cómputo del Instituto de Fisiología Celular todo el apoyo, la confianza y la amistad que me ha brindado desde siempre, al igual que a Ana María Escalante G. y Juan M. Barbosa Castillo, quienes laboran también en dicha Unidad.

Quiero agradecer al Dr. Francisco García Ugalde, el director de esta Tesis, de igual manera por el apoyo, la confianza y la amistad demostrada en estos años.

Agradezco las ideas, discusiones y ayuda que llegue a tener durante la etapa final del desarrollo de este trabajo por parte de Stéphane Pateux del IRISA en Rennes, Francia, y de Gabriel Cristóbal del Instituto de Óptica "Daza de Valdés" de Madrid, España. También agradezco al equipo del proyecto TEMICS por el buen trato, las facilidades y el apoyo brindados durante mi estancia doctoral en Rennes.

Quiero agradecer a mi padre, mi madre y mi hermano por creer en mi, por no dejar de sonreír y pelear conmigo, y estar dispuestos a caminar conmigo aún en el mismo infierno, así como yo lo estoy. De igual manera quiero expresar mi gratitud a toda mi familia y todos mis amigos, simplemente por su amistad, sinceridad y buenos momentos, las cuales aún no hallo donde guardarlos sino en el corazón.

Las imágenes de microscopía electrónica empleadas en este trabajo fueron cedidas amablemente por los estudiantes del Dr. Francisco Fernández de Miguel del Instituto de Fisiología Celular de la UNAM.

Agradezco el apoyo dado a este trabajo por parte del CONACYT, del Programa de Apoyo a los Estudios de Posgrado de la UNAM y de INTTELMEX.

A mis padres.

Reflexión personal.

Comenzaré por definir de donde vengo, porque jamás se debe uno olvidar de sus orígenes, cualesquiera que éstos sean. No conozco a fondo mi historia familiar, sin embargo, sé por qué estoy aquí: soy la razón de ser de mis padres, quienes a su vez han sido la razón de ser de mis abuelos y así sucesivamente, hasta el inicio de la vida misma. Soy la razón de ser de mis ancestros: su realidad y sus sueños, su posesión y sus anhelos, sus aciertos y sus errores, su verdad y su mentira, sus preocupaciones, su seguridad y su alegría, todas sus contradicciones. Sé que mis abuelos han vivido del campo, los animales y algo más, no así mis padres quienes han vivido del comercio y el transporte. Yo vivo de la educación que ellos me han procurado, su mejor legado. Tengo la dicha de no ser tan tonto, ni tan flojo, de poder tener la respuesta a algunas de las preguntas que me hago, ó me hacen, y saber donde buscarlas cuando no las tengo, aunque en ocasiones no sean del todo correctas. Y no sólo tengo esa dicha, sino muchas más, simplemente porque he sido más afortunado que muchos otros, porque siempre he tenido un techo, vestido, alimento y cariño. Todo lo que se necesita para ser feliz, lo indispensable para cabalgar sin dificultad en esta vida. Nada me ha faltado, y sí creo que aún tengo mucho que dar, y sobretodo, mucho que agradecer, de lo cual no quiero jamás olvidarme.

He aprendido muchas cosas que me han servido y servirán en esta vida, pero sé que aún me falta mucho por aprender. Sé que la mejor manera de vivir es con sencillez, con simpleza, con humildad, sin complicaciones, siempre bajo la premisa de que las cosas son ó no son, y que lo extraordinario y maravilloso se encuentra en las personas comunes y en las cosas comunes. He aprendido a disfrutar de la soledad y de la compañía. Sé que la riqueza material es efímera, porque mi paso por este mundo lo es. Jamás he visto un muerto que se vaya con las manos y los bolsillos llenos, y sin embargo, he visto muchas tumbas desiertas y olvidadas.

He visto cosas extraordinarias y cosas absurdas. Personas que prefieren darse a odiar diciendo no, cuando nada les cuesta decir sí para darse a querer, y viceversa. He visto como grandes sueños desaparecen cuando ya están al alcance de la mano, porque simplemente uno no es capaz de comprender el poder de un par de palabras tan simples, breves y sencillas: sí y no. Porque se dice una cuando el corazón grita la otra, engañándose

uno mismo. O bien, porque se es egoísta con el conocimiento, con la verdad, y uno se la guarda para sí mismo, en lugar de compartirla, de dar a conocer esa verdad que libera y da alas, y sobre todo, que rompe tantas cadenas y sentimientos reprimidos.

La vida está llena de decisiones, y éstas deben tomarse fríamente, sin caprichos, sin soberbia, sin desconsideración, sin orgullo, sin envidia, porque éstas marcan la manera en la cual se rige la vida de cada uno, la facilidad con la cual los sueños se harán realidad. No importa si la decisión tomada no es la correcta, finalmente esa decisión es de uno mismo y de nadie más. No se debe uno olvidar que nada nos separará de nuestro destino, de nuestra historia personal.

Veo con tristeza como en nuestra vida diaria, nos esforzamos en envenenarnos con problemas e ideas inexistentes, en complicarnos la vida. Creemos falsamente en dioses falsos, dioses que nosotros mismos creamos, y así, cortamos nosotros mismos nuestras alas, y nos llenamos de pretextos para hacer o dejar de hacer las cosas, nos rendimos ante la primera adversidad, y damos la bienvenida a todo lo que fácil llega. Creemos que todo merecemos, cuando lo que merecemos es sólo aquello por lo que trabajamos. Hablamos de injusticia, cuando es injusto hablar de ella. Nos ponemos un escalón sobre los demás y los juzgamos y señalamos, en lugar de ponernos uno debajo para comprenderlos y admirarlos, nos olvidamos que debemos estar en su calzado para comprenderlos, que para entender el sufrimiento de los demás debemos sufrir como ellos. Vivimos todo el tiempo enojados, y en este enojo nos asfixiamos, nos ahogamos y lentamente vamos muriendo por dentro. Falsamente nos sentimos acosados, observados, engañados; nos sentimos las víctimas en la comedia de nuestra vida.

Sé que si el día de hoy me fue mal, no me debo quejar, sino debo creer que el siguiente será mejor, y más que un acto de fe, es el no desconocer mi posibilidad de hacer el cambio. Por otro lado, he aprendido a encontrarle cierta fascinación al cansancio, al sufrimiento y al dolor, porque es por ellos que siento que estoy realmente vivo.

He aprendido a observar a la gente, sus expresiones, sus gestos, sus manos, sus ojos y su mirada. De esta manera puedo comprender un poco lo que llevan dentro y lo que dejan fuera. Aprendo a respetar y a no encontrar diferencias entre el rostro sucio y el rostro limpio; la mano llena y la mano vacía; el saludo tímido y el saludo fuerte y seguro; el que toca la puerta y el que la abre.

Tengo defectos, algunos los conozco y trato de corregirlos; otros no los conozco y nada puedo hacer; y otros más, aún no los tengo. Y así me acepto, porque aceptándome como soy, puedo aceptar a los demás sin cuestionarlos; porque así sé también que lo perfecto es un absurdo, existe y no existe, porque lo más perfecto que se puede lograr está en la medida de perfección del hombre mismo: su imperfección, la cual ha sido definida por el hombre mismo. Con esta idea, hago todo lo posible por escuchar y aceptar las críticas que se vierten en mi persona, evito cerrarme a estos comentarios, porque sé que sólo así puedo mejorar, evito esa necedad de tener siempre la razón y trato de reconocer y aceptar de buena gana cuando me equivoco, se que no siempre tengo la razón.

Hasta este momento de mi vida, quizás no he tenido todo cuanto he querido, pero sé lo que aún puedo obtener y lo que jamás obtendré; y no por eso me resigno, no por eso me rindo, no por eso voy a dar la espalda; y si no puedo cumplirme algún deseo, tengo el consuelo de haber hecho el mejor intento y a la manera que yo he creído. No tendré los bolsillos llenos de monedas, pero no me importa, ni me interesa, porque siempre he tenido al menos una moneda, y sé como multiplicarla, día a día, aunque sea por uno.

Sé del poder de una sonrisa, de una frase amable, de un beso dulce, de una tierna mirada. Quizás no cambien al mundo, pero si lo hacen mejor. Como lo hace un momento de paz y silencio, una rosa en la mano correcta, un beso en la mejilla adecuada, una noche de sueño ininterrumpido ó hasta un instante de locura deseada.

Y pensar que es tan fácil ser feliz, mucho más fácil que tratar de serlo, y para lograrlo es importante saber compartir, es importante no ponerle precio a los actos, a los sentimientos, a los pensamientos y a las buenas intenciones. Es importante hacer a un lado la fría indiferencia y ese sentimiento de molestia y hastío que tanto pesa y abrirle paso a la modestia y humildad. Y si he de llorar que sea de alegría, y si he de sufrir es porque así vale aquello por lo que sufro, y si he de escribir con sangre mi historia es porque jamás será borrada. Y si he de compartir mi vida, no dudaré en entregarla toda, porque las columnas que soporten mi palabra, mis creencias, mi religión y mi filosofía no han de ser movidas, ni destruidas, por nada, ni nadie, ya que día con día estaré al cuidado de ellas, y porque día con día el esfuerzo de ser mejor no disminuirá ni un ápice.

Con todo lo que tengo y he vivido sólo puedo agradecer a la vida esta oportunidad de vivirla con todo lo que ella implica y jamás renunciaré a ella. Sé que el saco que he de llenar jamás lo llenaré porque debo compartirlo, pero jamás estará vacío, de eso estoy seguro, así puedo seguir compartiendo.

He querido incluir esta reflexión personal simplemente como un capricho. Pero en el fondo, quisiera no fuera sólo eso. Siempre que leo estas líneas, me siento como un hipócrita, dando reglas y preceptos morales que siento no aplico conmigo, siento que no predico con el ejemplo, lo intento, pero en el fondo siento que fallo, siento que el verdadero coraje que se necesita aún se encuentra dormido en mí. Aún me cuesta trabajo distinguir un sí de un no, la verdad de la mentira. A veces le doy una moneda al menos afortunado, y a veces lo ignoro cuando se postra frente a mí. A veces soy considerado e indulgente, y a veces no lo soy. A veces sonrío a quien me cae mal, y a veces lo esquivo y evito ver. A veces busco satisfacer mis dudas, y a veces me quedo con ellas. A veces siento que el corazón me empuja a enfrentar al miedo, que me da fuerza y coraje, que me llena de decisión; y a veces siento que me abandona y soy el ser más pequeño y más cobarde del universo. A veces digo lo que pienso, y a veces miento o callo. A veces lastimo, y a veces conforto. No me defino, y por eso, no me fío de mí, aunque debo reconocer la naturaleza contradictoria del hombre, en consecuencia, mi propia naturaleza; además, todas estas contradicciones son producto de la pasión, de un alma intranquila que se revela y se somete, que juzga y es juzgada, que vive y muere.

Hoy trato de desnudar mi alma, de hacerle sentir frío y calor a la vez. Hoy me pongo frente al espejo y me miro directo a los ojos. Hoy saco todo lo bueno y todo lo malo que tengo y lo enfrento. Hoy quiero dormir profundamente, para así, al despertar, ver que todas esas contradicciones que tengo no existen más, se han marchado lejos, y no volverán. Quiero ser o no ser, quiero creer o no creer, quiero definirme. Simplemente, trato de morir y volver a nacer.

Resumen.

Actualmente las técnicas de aproximación adaptable se han vuelto muy populares, y debido a esta popularidad se han desarrollado una gran cantidad de éstas, entre las cuales se puede mencionar a tres de las más recientes que son: "Matching Pursuit", "Basis Pursuit" y "High Resolution Pursuit".

La técnica de "High Resolution Pursuit" ha sido utilizada para la extracción de características y el reconocimiento de objetos. "High Resolution Pursuit" es una versión diferente al algoritmo "Matching Pursuit", ya que principalmente resalta las correspondencias locales sobre las globales durante el proceso iterativo que realiza. Además, cabe mencionar que tanto "High Resolution Pursuit" como "Matching Pursuit" poseen el mismo orden de complejidad computacional, y por esta razón superan la baja complejidad computacional presente en "Basis Pursuit".

La concepción de los modelos de codificación empleados para la realización de este trabajo es producto de una muy completa investigación de las técnicas de selección adaptable, de los efectos que produce la cuantización al introducirse en estas técnicas, de los métodos tradicionales de compresión y de las medidas de calidad más comunes que se emplean para la evaluación de resultados en el análisis de imágenes. Además, las técnicas de selección adaptable involucran algunos factores dignos de un estudio profundo, como lo son la selección del diccionario que se empleará y el criterio para detener el proceso de selección.

En primera instancia, los modelos propuestos en este trabajo expanden la imagen en estudio a lo largo de un diccionario redundante de funciones, este diccionario ha sido seleccionado previamente con base en los estudios realizados al respecto. A partir de esta expansión, se seleccionan mediante una de las técnicas de selección adaptable como "Matching Pursuit" y "High Resolution Pursuit" los coeficientes correspondientes a las estructuras más importantes de la imagen, empleando el criterio de similitud característico de cada técnica. Se hace uso de diccionarios redundantes ya que estos diccionarios son extremadamente flexibles, sobre todo cuando se trata de obtener las correspondencias de las estructuras más importantes de una señal.

A los coeficientes que se obtienen se les aplica un proceso de cuantización lineal justo al momento que son seleccionados, con la finalidad de evitar la propagación, en las iteraciones siguientes durante el proceso, del tipo de errores que se producirían si la cuantización se realizara sobre todos los coeficientes una vez obtenidos. Con los coeficientes seleccionados se tiene una "óptima" descomposición de la imagen, o bien, una nueva representación con un número reducido de elementos. Esta nueva representación, en cierta manera, corresponde a una imagen codificada con tasa de compresión alta, la cual se busca obtener y se ha planteado como objetivo de este trabajo.

Los átomos seleccionados implícitamente, son representados por un conjunto de parámetros dependiendo del modelo empleado. A estos parámetros se les ordena y aplica una técnica de compresión tradicional para obtener una compresión efectiva de la información.

Empleando un algoritmo de reconstrucción se recupera la imagen original con una alta calidad visual sujeta a evaluación. Esta evaluación se realiza empleando la escala de calidad PQS ("Picture Quality Scale") y los criterios de calidad objetiva más tradicionales.

Se puede observar que, como consecuencia de la investigación que se realiza a lo largo de este trabajo, surge un par de modelos de codificación de imágenes. Estos modelos de codificación son novedosos, lo más completos posible, producen una alta tasa de compresión, y resultados de muy buena calidad visual.

Abstract.

At present, adaptive approximation techniques have become very popular; in consequence, many of them have been developed recently, such as Matching Pursuit, Basis Pursuit and the novel High Resolution Pursuit.

High Resolution Pursuit has been employed for feature extraction and object recognition. High Resolution Pursuit is a Matching Pursuit different version, because it emphasizes the local correspondences over the global correspondences while processing data. Furthermore, as High Resolution Pursuit has the same computational complexity order as Matching Pursuit, they overcome the excessive Basis Pursuit computational complexity.

The conception of the methods used to perform this work are the result of an extensive research related to adaptive approximation techniques, quantization effects in these techniques, traditional compression methods and the most common image quality criteria. Besides, adaptive approximation techniques involve some factors, which need a deep study, such as the dictionary to employ and the algorithm stop criteria.

At first, the proposed models in this work expand an image over an overcomplete dictionary, such dictionary has been selected according the results presented and published in previous works. From such expansion, by adaptive approximation techniques such as Matching Pursuit and High Resolution Pursuit, the coefficients corresponding to the most important image structures are selected. This selection is done by a similarity measure. Overcomplete dictionaries are employed because they are extremely flexible, principally when tracking the best correspondences to the most important signal structures.

Selected coefficients are quantized just when they are chosen in order to avoid bigger error propagation during the process that when quantization is done at the end of the process. The set of selected coefficients is an "optimal" image decomposition, or a new representation with a reduced quantity of elements. This new representation corresponds to a coded image with a low bit rate, which is the goal of this work.

Selected atoms are represented by the parameters set, such parameters depend of the model employed. A sorting process and a traditional compression technique is applied to this set in order to get an effective compression.

By a simple reconstruction algorithm we recover the original image with an expected visual high quality subject to evaluation. At the end, we perform such evaluation by PQS (Picture Quality Scale) and the most common criteria of objective quality.

We can note, as consequence of all the research done along this work, a set of image coding models is presented, such models are integral as possible, and they produce very good visual quality results, with a high compression rate.

Contenido.

1. Introducción	1
1.1. Motivación	3
1.2. Contribución	4
Referencias	5
2. Aproximación adaptable	7
2.1. Diccionarios	7
2.2. Descomposición atómica de tiempo-frecuencia	8
2.3. Técnicas de aproximación adaptable	10
2.3.1. "Matching Pursuit"	10
2.3.2. "Basis Pursuit"	12
2.3.3. "High Resolution Pursuit"	13
Referencias	14
3. Cuantización	15
3.1. Cuantización escalar	15
3.2. Cuantizador regular y rango del cuantizador	16
3.3. Objeto y desempeño de un cuantizador	16
3.4. Cuantización de alta resolución	18
3.5. Cuantización en selecciones adaptables	18
Referencias	20
4. Calidad de la imagen	21
4.1. Definición de la calidad de la imagen	21
4.2. Medición de la calidad de la imagen	22
4.3. Criterios de calidad objetiva	23
4.4. La escala de calidad PQS ("Picture Quality Scale")	24

4.4.1. Factores de distorsión	27
4.4.2. Representación de la componente principal de las medidas de distorsión	27
Referencias	28
5. Modelos de codificación	31
5.1. Antecedentes	32
5.1.1. Esquema de descomposición piramidal	32
5.1.2. Modelo de codificación contornos-textura	34
5.1.3. Otros modelos de codificación	36
5.2. Modelo de codificación por regiones	37
5.2.1. El diccionario empleado	37
5.2.2. La búsqueda de los átomos	40
5.2.3. Búsqueda rápida del producto interno	41
5.2.4. Etapa de cuantización	43
5.2.5. Codificación de los parámetros	44
5.3. Modelo de codificación piramidal	45
5.3.1. Selección adaptable de la base	46
5.3.1.1. La mejor base y la concavidad de Schur	46
5.3.2. Búsqueda exhaustiva	49
5.3.3. Codificación de los parámetros	51
Referencias	51
6. Codificación de imágenes	55
6.1. Experimentación, análisis y resultados	56
6.1.1. Lenna	60
6.1.2. Bárbara	69
6.1.3. Cameraman	76
6.2. Una aplicación real con imágenes de microscopía electrónica	84
6.2.1. Fluorescencia	84
6.2.2. Laminar	92
Referencias	101
7. Conclusiones y perspectivas	103
7.1. Conclusiones	103
7.2. Perspectivas	108
Referencias	109

Lista de figuras.

Figura 5.1. Espacio de Fourier a 4 y 8 escalas	38
Figura 5.2. Diccionario de funciones de Gabor empleado	39
Figura 5.3. Detalle de un átomo de Gabor	40
Figura 5.4. Procedimiento de búsqueda del producto interno	41
Figura 5.5. Típica transformada ondeleta 2-D	45
Figura 6.1. Lenna	61
Figura 6.2. Imágenes de Lenna reconstruidas con MPP	62
Figura 6.3. Imágenes de Lenna reconstruidas con MPR	63
Figura 6.4. Imágenes de Lenna reconstruidas con HRPR	64
Figura 6.5. Efectos producidos por los diferentes valores del paso de cuantización en MPP para Lenna	65
Figura 6.6. Efectos producidos por los diferentes valores del paso de cuantización en MPR para Lenna	65
Figura 6.7. Efectos producidos por los diferentes valores del paso de cuantización en HRPR para Lenna	66
Figura 6.8. Efectos producidos cuando el valor del paso de cuantización en los tres modelos de codificación es 1	66
Figura 6.9. Efectos producidos cuando el valor del paso de cuantización en los tres modelos de codificación es 16	67
Figura 6.10. Efectos producidos cuando el valor del paso de cuantización en los tres modelos de codificación es 32	67
Figura 6.11. Efectos producidos cuando el valor del paso de cuantización en los tres modelos de codificación es 64	68
Figura 6.12. Bárbara	69
Figura 6.13. Imágenes de Bárbara reconstruidas con MPP	70
Figura 6.14. Imágenes de Bárbara reconstruidas con MPR	71
Figura 6.15. Imágenes de Bárbara reconstruidas con HRPR	72
Figura 6.16. Efectos producidos por los diferentes valores del paso de cuantización en MPP para Bárbara	73
Figura 6.17. Efectos producidos por los diferentes valores del paso de cuantización en MPR para Bárbara	73

Figura 6.18. Efectos producidos por los diferentes valores del paso de cuantización en HRPR para Bárbara	74
Figura 6.19. Efectos producidos cuando el valor del paso de cuantización en los tres modelos de codificación es 1	74
Figura 6.20. Efectos producidos cuando el valor del paso de cuantización en los tres modelos de codificación es 16	75
Figura 6.21. Efectos producidos cuando el valor del paso de cuantización en los tres modelos de codificación es 32	75
Figura 6.22. Efectos producidos cuando el valor del paso de cuantización en los tres modelos de codificación es 64	76
Figura 6.23. Cameraman	77
Figura 6.24. Imágenes de Cameraman reconstruidas con MPP	78
Figura 6.25. Imágenes de Cameraman reconstruidas con MPR	79
Figura 6.26. Imágenes de Cameraman reconstruidas con HRPR	80
Figura 6.27. Efectos producidos por los diferentes valores del paso de cuantización en MPP para Cameraman	81
Figura 6.28. Efectos producidos por los diferentes valores del paso de cuantización en MPR para Cameraman	81
Figura 6.29. Efectos producidos por los diferentes valores del paso de cuantización en HRPR para Cameraman	82
Figura 6.30. Efectos producidos cuando el valor del paso de cuantización en los tres modelos de codificación es 1	82
Figura 6.31. Efectos producidos cuando el valor del paso de cuantización en los tres modelos de codificación es 16	83
Figura 6.32. Efectos producidos cuando el valor del paso de cuantización en los tres modelos de codificación es 32	83
Figura 6.33. Efectos producidos cuando el valor del paso de cuantización en los tres modelos de codificación es 64	84
Figura 6.34. Fluorescencia	85
Figura 6.35. Imágenes de fluorescencia reconstruidas con MPP	86
Figura 6.36. Imágenes de fluorescencia reconstruidas con MPR	87
Figura 6.37. Imágenes de fluorescencia reconstruidas con HRPR	88
Figura 6.38. Efectos producidos por los diferentes valores del paso de cuantización en MPP para la imagen de fluorescencia	89
Figura 6.39. Efectos producidos por los diferentes valores del paso de cuantización en MPR para la imagen de fluorescencia	89
Figura 6.40. Efectos producidos por los diferentes valores del paso de cuantización en HRPR para la imagen de fluorescencia	90
Figura 6.41. Efectos producidos cuando el valor del paso de cuantización en los tres modelos de codificación es 1	90
Figura 6.42. Efectos producidos cuando el valor del paso de cuantización en los tres modelos de codificación es 16	91
Figura 6.43. Efectos producidos cuando el valor del paso de cuantización en los tres modelos de codificación es 32	91

Figura 6.44. Efectos producidos cuando el valor del paso de cuantización en los tres modelos de codificación es 64	92
Figura 6.45. Laminar	93
Figura 6.46. Imágenes del cuerpo laminar reconstruidas con MPP	94
Figura 6.47. Imágenes del cuerpo laminar reconstruidas con MPR	95
Figura 6.48. Imágenes del cuerpo laminar reconstruidas con HRPR	96
Figura 6.49. Efectos producidos por los diferentes valores del paso de cuantización en MPP para la imagen laminar	97
Figura 6.50. Efectos producidos por los diferentes valores del paso de cuantización en MPR para la imagen laminar	97
Figura 6.51. Efectos producidos por los diferentes valores del paso de cuantización en HRPR para la imagen laminar	98
Figura 6.52. Efectos producidos cuando el valor del paso de cuantización en los tres modelos de codificación es 1	98
Figura 6.53. Efectos producidos cuando el valor del paso de cuantización en los tres modelos de codificación es 16	99
Figura 6.54. Efectos producidos cuando el valor del paso de cuantización en los tres modelos de codificación es 32	99
Figura 6.55. Efectos producidos cuando el valor del paso de cuantización en los tres modelos de codificación es 64	100

Lista de tablas.

Tabla 5.1. Tamaños de cada átomo asociados a la escala	40
Tabla 5.2. Parámetros que definen un átomo	41
Tabla 5.3. Secuencia de almacenamiento de los parámetros de los átomos	44
Tabla 5.4. Parámetros que definen un átomo	51
Tabla 6.1. Coeficientes de los filtros QMF y DQMF de las ondeletas "Spline" empleados con el modelo piramidal	60
Tabla 6.2. Conjunto de valores obtenidos para Lenna	61
Tabla 6.3. Medidas de error, relación, calidad y compresión para Lenna	62
Tabla 6.4. Conjunto de valores obtenidos para Bárbara	70
Tabla 6.5. Medidas de error, relación, calidad y compresión para Bárbara	71
Tabla 6.6. Conjunto de valores obtenidos para Cameraman	77
Tabla 6.7. Medidas de error, relación, calidad y compresión para Cameraman	78
Tabla 6.8. Conjunto de valores obtenidos para la de imagen de fluorescencia	85
Tabla 6.9. Medidas de error, relación, calidad y compresión para la imagen de fluorescencia	86
Tabla 6.10. Conjunto de valores obtenidos para la imagen laminar	94
Tabla 6.11. Medidas de error, relación, calidad y compresión para la imagen laminar	95
Tabla 7.1. Resultados presentados por modelos de codificación con selección adaptable	107
Tabla 7.2. Tasa de compresión (bpp) para imágenes reconstruidas	107
Tabla 7.3. Valores PSNR y SNR para Lenna a 0.44 bpp y 0.50 bpp	107

Introducción.

Se comenzará diciendo que el término *compresión* se puede referir también al término de *codificación con tasa de bits baja*, ó simplemente al de *codificación*. Por esta razón, se puede decir que la representación de una imagen en términos distintos a aquellos con los que normalmente se representa se considera una imagen codificada. Ahora bien, si la imagen codificada, con relación a la imagen original, consiste de una menor cantidad de términos ó coeficientes diferentes de cero, ó de una mayor cantidad de términos repetidos - en breve, una entropía menor -, ésta se podría comprimir hasta cierto punto empleando un algoritmo de compresión, obteniéndose entonces una imagen comprimida.

Un problema fundamental en la codificación de imágenes consiste en obtener altas tasas de compresión conservando al mismo tiempo una alta calidad subjetiva en las imágenes reconstruidas u obteniendo una pérdida perceptiva mínima.

Por otro lado, tenemos que una señal se puede descomponer en una suma ponderada de elementos extraídos de un diccionario redundante. Estos diccionarios redundantes se han empleado últimamente en diversas aplicaciones por ser extremadamente flexibles, sobre todo, cuando se trata de obtener las correspondencias de las estructuras más importantes de una señal.

Inicialmente S. Mallat propuso un algoritmo de selección adaptable llamado "Matching Pursuit" [1], posteriormente S. Jaggi propuso otro algoritmo llamado "High Resolution Pursuit" [2], estos algoritmos descomponen una señal en una expansión lineal de formas de onda que pertenecen a un diccionario redundante de funciones. A continuación, seleccionan algunas de las formas de onda de manera "óptima" gracias al criterio de similitud que utilizan, esta selección la llevan a cabo de tal manera que las formas de onda seleccionadas correspondan lo mejor posible a las estructuras de la señal en estudio. Cabe señalar que estos algoritmos, en teoría, emplean un número de operaciones mínimo durante su desarrollo, en consecuencia, el tiempo de procesamiento que utilizan es mínimo también.

"High Resolution Pursuit" fue creado y ha sido utilizado para la extracción de características y el reconocimiento de objetos con muy buenos resultados [2, 3]. Mientras "Matching Pursuit" ha sido estudiado de manera más exhaustiva en diversas áreas del procesamiento de señales. Las referencias a dichos trabajos se presentarán en su oportunidad a lo largo del escrito.

La concepción de los modelos propuestos para la realización de este trabajo es producto de una completa investigación de las técnicas de selección adaptable, de los efectos que produce la cuantización al introducirse en estas técnicas, de los métodos tradicionales de compresión y de las medidas de calidad más comunes que se emplean para la evaluación de resultados en el análisis de imágenes. Además, las técnicas de aproximación adaptable involucran algunos factores dignos de un estudio profundo como lo son la selección del diccionario que se empleará y el criterio para detener el proceso de selección.

En primera instancia, los modelos propuestos expanden la imagen en estudio a lo largo de un diccionario redundante de funciones, este diccionario ha sido seleccionado previamente con base en los estudios realizados y los resultados publicados relacionados con la selección de bases óptimas. A partir de esta expansión, se seleccionan mediante una de las técnicas de selección adaptable como "Matching Pursuit" y "High Resolution Pursuit" los coeficientes correspondientes a las estructuras más importantes de la imagen empleando un criterio de similitud. Se hace uso de diccionarios redundantes ya que estos diccionarios son extremadamente flexibles, sobre todo cuando se trata de obtener las correspondencias de las estructuras más importantes de una señal.

A los coeficientes que se obtienen se les aplica un proceso de cuantización lineal justo al momento que son seleccionados, con la finalidad de evitar la propagación, en las iteraciones siguientes durante el proceso, del tipo de errores que se producirían si la cuantización se realizara sobre todos los coeficientes una vez obtenidos. Con los coeficientes seleccionados se tiene una "óptima" descomposición de la imagen, o bien, una nueva representación con un número reducido de elementos. Esta nueva representación, en cierta manera, corresponde a una imagen codificada con tasa de compresión alta, la cual se busca obtener y se ha planteado como objetivo de este trabajo.

Los átomos seleccionados implícitamente se representan por un conjunto de parámetros dependiendo del modelo empleado. A estos parámetros se les aplica una técnica de compresión tradicional para una obtener una compresión efectiva de la información.

Empleando un algoritmo de reconstrucción se recupera la imagen original con una alta calidad visual sujeta a evaluación. Esta evaluación se realiza empleando la escala de calidad PQS ("Picture Quality Scale") y los criterios de calidad objetiva más tradicionales como el error cuadrático medio, la relación señal a ruido y la relación señal a ruido pico. PQS es una escala de calidad que se basa en las medidas cuantitativas de algunos factores de distorsión [4].

1.1 Motivación.

Las técnicas de aproximación adaptable se han vuelto muy populares, empleándose principalmente para obtener representaciones parsimoniosas de una diversidad de clases de señales. Debido a su popularidad, recientemente se han desarrollado varias de estas técnicas, entre las cuales se puede mencionar a "Matching Pursuit" [1], "Basis Pursuit" [5] y "High Resolution Pursuit" [2].

"Matching Pursuit" es la técnica que más se ha estudiado y utilizado en diversas áreas, principalmente en codificación de señales [6, 7, 8, 9, 10]. Por otro lado, se tiene que las propiedades de "Basis Pursuit" y "High Resolution Pursuit" han sido apenas estudiadas [11, 12, 13, 14, 15], ya sea por el alto costo computacional que se llega a presentar en "Basis Pursuit", ó bien, por tratarse de una técnica novedosa como lo es "High Resolution Pursuit". Con esta observación, surge la inquietud de emplear la técnica más novedosa para la codificación de imágenes junto con una técnica ya tradicional, pero aún no muy explotada en codificación de imágenes, y de esta manera comparar los resultados que se obtengan y las diferencias que puedan existir entre estas técnicas.

La selección adecuada de una base de descomposición, o de la mejor base, ha sido un tema de estudio e interés en los últimos años. Por esta razón, en uno de los modelos desarrollados, se propone el empleo de una función de costo, con la finalidad de seleccionar la mejor base de descomposición de la imagen en estudio, y a partir de la cual se generará el diccionario sobre el que actuará la técnica de selección adaptable.

Por otra parte, se han estudiado los efectos que produce la cuantización en "Matching Pursuit" [16, 17, 18], y considerando los resultados obtenidos, surge la idea de introducir también una etapa de cuantización en "High Resolution Pursuit", con la idea de mejorar los resultados que se puedan obtener, tanto con relación a la tasa de compresión, como a la calidad de la imagen reconstruida.

Finalmente, siempre es necesario realizar una evaluación de los resultados que se obtienen. En este caso se trabaja con imágenes fijas, las cuales no son muy fáciles de evaluar, sobre todo pensando en términos de calidad y en la diversidad de aplicaciones que pueden tener. Para la evaluación de los resultados se empleará una escala de calidad llamada PQS ("Picture Quality Scale") [19], la cual cumple con los requisitos que establece el Comité Consultivo Internacional de la Radio (CCIR) con respecto a la calidad de una imagen en TV [20].

Durante el proceso de codificación de una imagen, se llegan a presentar errores aleatorios y errores estructurados. Generalmente, los errores estructurados prevalecen en la codificación de imágenes y degradan las características locales y la calidad perceptual en mayor medida que los errores aleatorios, por esta razón, las medidas de error más comunes como el error cuadrático medio y la relación señal a ruido pico no son del todo adecuadas para evaluar imágenes comprimidas [4]. Considerando estos razonamientos, se

propone el empleo de la medida PQS, la cual contempla los errores estructurales más comunes.

De esta manera se puede decir que los resultados obtenidos son evaluados con relación a una escala de calidad de carácter comercial, y a la vez, con una métrica que se basa en las medidas cuantitativas de diferentes factores de distorsión, los cuales contribuyen colectivamente a determinar la calidad de la imagen percibida. Además, los resultados obtenidos se evaluarán con los criterios de calidad objetiva más tradicionales con fines comparativos.

1.2 Contribución.

La contribución de este trabajo cae dentro del área de codificación de imágenes. Se ha consultado una gran cantidad de bibliografía, a partir de la cual han surgido las ideas que van formando la técnica de codificación que aquí se presenta. A continuación se listan las diferentes áreas que se han estudiado:

1. Técnicas de codificación de imágenes, principalmente aquellas que involucran el uso ondeletas ("wavelets"), ya sean transformadas ó paquetes de éstas.
2. Técnicas de selección adaptable, principalmente "Matching Pursuit" y "High Resolution Pursuit", al igual con sus aplicaciones a voz, imagen y vídeo.
3. Técnicas de selección adaptable de la mejor base de descomposición.
4. Técnicas de cuantización y sus efectos en las técnicas de selección adaptable.
5. Técnicas de evaluación de calidad de imágenes.

En toda la bibliografía consultada no se encontró una técnica o un modelo de codificación como alguno de los que se proponen en este trabajo, en realidad, solamente se tiene referencia del trabajo desarrollado en la codificación de imágenes empleando "Matching Pursuit" [6, 7, 9], sin introducir una etapa de cuantización, ni una selección adaptable de la base como se presenta en uno de los modelos propuestos, ni realizar una evaluación adecuada de los resultados obtenidos. Solamente se tienen referencias de los problemas que se pueden presentar al emplear tanto "Matching Pursuit" como "Basis Pursuit" bajo ciertas condiciones, y de las bondades de "High Resolution Pursuit" [2, 3]. Cabe mencionar que esta última técnica no se ha estudiado más profundamente, y su uso para codificación de imágenes no ha pasado de ser más que una mera propuesta.

En este trabajo se presentan dos técnicas de codificación de imágenes que llegan a ser novedosas, y son producto del trabajo de investigación realizado. Este trabajo es la aplicación práctica de todo el marco teórico desarrollado en las diversas áreas listadas anteriormente, lo cual viene a ser la finalidad del trabajo de todo investigador.

A pesar de que actualmente existen numerosas técnicas de codificación de imágenes y día a día surgen más, lo cual es muy natural, debido principalmente a la creciente demanda de ahorro y uso adecuado de recursos en varios sectores, con este trabajo, se

sigue contribuyendo a la búsqueda de respuestas y presentación de alternativas en el área de la codificación de imágenes.

La introducción de los modelos de codificación propuestos en este trabajo, los cuales nunca antes habían sido introducidos, amplía el campo de conocimiento y estudio de esta área. Además que todas las consideraciones previamente descritas y contempladas en estos modelos tampoco han sido presentadas en algún modelo de codificación conocido. Se está haciendo uso de una pequeñísima parte del conocimiento generado a lo largo de la existencia del hombre para proporcionar respuestas y alternativas a uno de los múltiples problemas e inquietudes que se llegan a presentar en uno de los juegos que juega el hombre. Esto no es más que un grano de arena en un inmenso desierto, pero que en algún momento puede caer dentro del calzado de cualquiera para llamar su atención, aunque sea brevemente, y jamás debe descartarse esa posibilidad.

Referencias.

- [1] S. Mallat y Z. Zhang, Matching Pursuits with Time-Frequency Dictionaries, *IEEE Transactions on Signal Processing*, Vol. 41, No. 12, Dic. 1993.
- [2] S. Jaggi, W. C. Carl, S. Mallat y A. S. Willsky, High Resolution Pursuit for Feature Extraction, *Reporte Técnico, MIT*, Noviembre 1995.
- [3] S. Jaggi, Multiscale Geometric Feature Extraction and Object Recognition, *Tesis Doctoral, LIDS, MIT*, Febrero 1997.
- [4] M. Miyahara, K. Kotani y V. R. Algazi. Objective picture quality scale (PQS) for image coding. *IEEE Transactions on Communications*, 1998.
- [5] S. Chen y D. Donoho, Basis Pursuit, *Reporte Técnico, Departamento of Estadística, Universidad de Stanford*, Mayo 1995.
- [6] F. Bergeaud y S. Mallat, Processing images and sounds with matching pursuits, *SPIE*, Vol. 2569, pp. 2-13, Septiembre 1995.
- [7] F. Bergeaud, F. y S. Mallat, Matching pursuit of images, *SPIE*, Vol. 2491, pp. 2-13, Abril 1995.
- [8] R. Rezaifar y H. Jafarkhani, Wavelet based speech coding using orthogonal matching pursuit, *Reporte Técnico, DEE and ISR, University of Maryland*.
- [9] A. Serir, B. Sansal, A. Seghier y A. Serir, Satellite image multi-scale representation by matching pursuit, *SPIE*, Vol. 2646, pp. 146-152, Diciembre 1995.
- [10] M. Vetterli y T. Kalker, Matching pursuit for compression and application to motion compensated video coding. *Reporte Técnico, Department of EECS, UC Bekerly, CA*.
- [11] S. Chen y D. Donoho, Atomic decomposition by basis pursuit. *Reporte Técnico, Statistics Department, Stanford University*, Mayo 1995.
- [12] S. Chen y D. Donoho, Examples of Basis Pursuit, *SPIE*, Vol. 2569, pp. 564-574.
- [13] R. Gribonval, E. Bacry, S. Mallat, P. Depalle y X. Rodet, Analysis of sound signals with high resolution matching pursuit, *Proceedings on TFTS'96*, pp. 125-128, 1996.
- [14] R. Montúfar-Chaveznava y F. García-Ugalde, Quantized High Resolution Pursuit, *Proceedings on TFTS'98*, 1998.

- [15] R. Montúfar-Chaveznava y F. García-Ugalde, Image analysis by Basis Pursuit, *SPIE*, Vol. 3164, pp. 642-649, 1997.
- [16] V. K. Goyal, Quantized overcomplete expansions: analysis, synthesis and algorithms, *Memorandum No. UCB/ERL M95/97*, Julio 1995.
- [17] V. K. Goyal y M. Vetterli, Consistency in quantized Matching Pursuit, *Proceedings on ICASSP-96*, May 7-10, Atlanta GA, 1996.
- [18] V. K. Goyal y M. Vetterli, Dependent coding in quantized Matching Pursuit, *SPIE*, Vol. 3024, pp. 2-12.
- [19] M. Miyahara, K. Kotani y V. R. Algazi. Objective picture quality scale (PQS) for image coding. *IEEE Transactions on Communications*, 1996.
- [20] CCIR, Method for the Subjective Assessment of the quality of television pictures, *Rec. 500-I*, 1978.

Aproximación adaptable.

Recientemente las técnicas de aproximación adaptable se han vuelto muy populares dentro la comunidad científica encargada del estudio de señales, principalmente debido a su empleo en la búsqueda de representaciones de grandes clases de funciones.

En una aproximación adaptable, el objetivo principal consiste en encontrar la representación de una función f como una suma ponderada de elementos a partir de un diccionario redundante. Es decir, se busca que f se represente de la siguiente manera:

$$f = \sum_{\gamma \in \Gamma} \alpha_{\gamma} g_{\gamma} \quad (2.1)$$

Donde el conjunto $\{g_{\gamma} | \gamma \in \Gamma\}$ corresponde a un diccionario que se extiende a lo largo del espacio de todas las posibles funciones, y que además es redundante.

Varias de las posibles representaciones de f se pueden encontrar en un diccionario redundante. Por esta razón, se han sugerido muchos métodos para encontrar la "óptima" representación de la función (2.1). Entre estos métodos se puede mencionar al método de "Frames" [1], la Mejor Base Ortogonal [2], "Matching Pursuit" [3], "Basis Pursuit" [4], y "High Resolution Pursuit" [5], el cual es el más reciente.

2.1 Diccionarios.

Un acontecimiento importante que se presenta en el uso de las técnicas de aproximación adaptable es la selección del diccionario a utilizar. Las técnicas de aproximación adaptable se emplean para determinar la "óptima" descomposición de f en la forma (2.1) a partir de un diccionario redundante. Recientemente, los diccionarios redundantes se han empleado exhaustivamente ya que tienen la virtud de poder acomodar fácilmente a aquellos elementos que poseen un amplio rango de características tiempo-frecuencia y corresponden a las estructuras más importantes de la función en estudio. El problema de la descomposición puede ser más simple o trivial si se emplean diccionarios ortogonales.

El principal problema con los diccionarios ortogonales es que éstos no son lo suficientemente flexibles como para incluir los diferentes tipos de estructuras que se presentan en una función¹. Por ejemplo, suponiendo que f es la suma de una senoidal y un pulso de Dirac, dos de los posibles diccionarios ortogonales que se podrían emplear son los diccionarios de Fourier y de Dirac. Una descomposición de f , ya sea sobre el diccionario de Fourier o sobre el diccionario de Dirac podría no soportar una descripción mínima de la función ya que ninguno de estos diccionarios contiene todas las estructuras de la función global [6].

A partir de la analogía empleada en [3], estos diccionarios son similares a un lenguaje con un vocabulario muy limitado. Sin embargo, se pueden emplear oraciones completas para reemplazar a las palabras no disponibles. El diccionario que mejor se adapte a la descomposición de f debe ser un diccionario redundante el cual contenga tanto a la base de Fourier como a la base de Dirac. En resumen, podemos decir que el diccionario a emplear se debe seleccionar de modo que contenga a los elementos que correspondan a las estructuras más importantes de la función en estudio y, generalmente, los diccionarios redundantes cumplen con este requisito.

2.2 Descomposición atómica en tiempo-frecuencia.

La descomposición de señales en familias de funciones que se encuentran bien localizadas tanto en el tiempo como en la frecuencia han tenido varias aplicaciones en el procesamiento de señales y en el análisis armónico. Dichas funciones son llamadas *átomos de tiempo-frecuencia*. Dependiendo de la selección de los átomos de tiempo-frecuencia, la descomposición puede tener diferentes propiedades. La transformada de Fourier de tiempo corto y la transformada ondeleta ("wavelet") son ejemplos de una aplicación de la descomposición de señales en el dominio tiempo-frecuencia. Para extraer la información de señales complejas, es necesario adaptar la descomposición tiempo-frecuencia a ciertas estructuras en particular.

A continuación se discuten los requisitos necesarios para efectuar la adaptabilidad.

Una familia general de átomos de tiempo-frecuencia puede ser generada por medio del escalamiento, la traslación y la modulación de la ventana de una simple función, similar a la transformada de Fourier de tiempo corto, esto es, $g(t) \in L^2(\mathbb{R})$. Suponiendo que $g(t)$ es real, continuamente diferenciable y cuyo orden es $O(1/t^2 + 1)$. Además suponiendo

¹ Las funciones fundamentales de las técnicas de transformación para codificación pueden juzgarse por su localización simultánea en el espacio y frecuencia. Esta localización no puede ser mejor que una cantidad universal fija, como lo muestra el principio de desigualdad o incertidumbre de Heisenberg.

Una función diferente de cero con soporte compacto no puede tener también una transformada de Fourier de soporte compacto. La transformada de Fourier de una función con soporte compacto es completamente analítica, y una función entera la cual se desvanece sobre un conjunto abierto debe ser cero siempre. Existen varios mejoramientos a este resultado, dificultando el análisis tiempo-frecuencia [7].

que $|g| = 1$, que la integral de $g(t) \neq 0$ y que $g(0) \neq 0$. Para cualquier escala $s > 0$, frecuencia de modulación ξ y traslación u , se denota $\gamma = (s, u, \xi)$ y se define

$$g_\gamma(t) = s^{-1/2} g(t - u/s) e^{\xi t} \quad (2.2)$$

El índice γ es un elemento del conjunto $\Gamma = \mathbf{R}^+ \times \mathbf{R}^2$. El factor $s^{-1/2}$ normaliza en 1 la norma de $g_\gamma(t)$. Si $g(t)$ es par, lo cual generalmente ocurre, $g_\gamma(t)$ está centrada en la abscisa u . La mayor parte de su energía se encuentra concentrada en la vecindad de u , cuyo tamaño es proporcional a s . Sea $G(\omega)$ la transformada de Fourier de $g(t)$. La ecuación (2.2) quedaría así:

$$G_\gamma(\omega) = s^{1/2} G(s(\omega - \xi)) e^{-i(\omega - \xi)u} \quad (2.3)$$

Debido a que $|G(\omega)|$ es una función par, $|G_\gamma(\omega)|$ se encuentra centrada en la frecuencia $\omega = \xi$. Su energía está concentrada en la vecindad de ξ , cuyo tamaño es proporcional a s^{-1} .

La familia $D = (g_\gamma(t))_{\gamma \in \Gamma}$ es extremadamente redundante y sus propiedades han sido estudiadas por Torresani [8]. Para representar de manera eficiente cualquier función $f(t)$, se debe seleccionar a un subconjunto apropiado de átomos $(g_{\gamma_n}(t))_{n \in \mathbf{N}}$, con $\gamma_n = (s_n, u_n, \xi_n)$, de modo que $f(t)$ se pueda escribir como

$$f(t) = \sum_{\gamma \in \Gamma} a_n g_{\gamma_n}(t) \quad n = -\infty \dots +\infty \quad (2.4)$$

Dependiendo de la elección de los átomos $g_{\gamma_n}(t)$, la expansión de los coeficientes a_n proporciona información explícita sobre algunas de las propiedades de $f(t)$. Las ventanas empleadas en la transformada de Fourier de tiempo corto y en la transformada ondeleta corresponden a diferentes familias de átomos de tiempo-frecuencia, las cuales son bases de $L^2(\mathbf{R})$.

En una ventana de la transformada de Fourier, todos los átomos g_{γ_n} tienen una constante de escala $s_n = s_0$, y de esta manera son localizados principalmente a lo largo de un intervalo cuyo tamaño es proporcional a s_0 . Si las estructuras de la señal principal se localizan sobre una escala de tiempo de orden s_0 , los coeficientes de expansión a_n proporcionarán la información relacionada con su localización y su contenido de frecuencia. Sin embargo, una ventana de la transformada de Fourier no se adapta lo suficientemente bien como para describir aquellas estructuras que son mucho más pequeñas o mucho mayores que s_0 . Para analizar las componentes de diferentes tamaños, es necesario emplear átomos de tiempo-frecuencia a diferentes escalas.

Contrariamente a la ventana de la transformada de Fourier, la transformada ondeleta descompone la señal en átomos de tiempo-frecuencia a diferentes escalas, llamados ondeletas. Una familia de ondeletas $(g_{\gamma_n}(t))_{n \in \mathbf{N}}$ se construye relacionando el parámetro de frecuencia ξ_n con el de escala s_n mediante $\xi_n = \xi_0/s_n$, donde ξ_0 es una

constante. La familia resultante está compuesta de dilataciones y traslaciones de una simple función, multiplicada por el complejo del parámetro de la fase. Los coeficientes de expansión a_n de funciones sobre las familias de ondeletas caracterizan el comportamiento de la escala de las estructuras de las señales. Esto es importante para el análisis de fractales y comportamientos singulares. Sin embargo, los coeficientes de expansión en un marco ondeleta no proporcionan una estimación precisa del contenido de la frecuencia de las formas de onda cuya transformada de Fourier está bien localizada, especialmente a altas frecuencias. Esto se debe a la restricción que existe en el parámetro de la frecuencia ξ_n , el cual permanece inversamente proporcional a la escala s_n^2 .

Para señales $f(t)$ que contengan escalamiento y estructuras altamente oscilantes, no se puede definir a priori las restricciones apropiadas para los parámetros de escala y modulación de los átomos de tiempo-frecuencia $g_\gamma(t)$ empleados en la expansión (2.4). Es necesario seleccionar adaptablemente los elementos del diccionario $D = (g_\gamma(t))_{\gamma \in \Gamma}$, dependiendo de las propiedades locales de $f(t)$.

2.3 Técnicas de aproximación adaptable.

La tarea principal detrás de una descomposición adaptable tiempo-frecuencia consiste en encontrar aquellos procedimientos que puedan expandir funciones sobre un conjunto de formas de onda, y seleccionar algunas de estas formas de onda bajo cierto criterio a lo largo de un diccionario grande y redundante. "High Resolution Pursuit" y "Matching Pursuit" son algoritmos generales que realizan esta tarea, ya que llevan a cabo dicha descomposición adaptable y utilizan un criterio de similitud para la selección de las formas de onda.

"High Resolution Pursuit" tiene por objeto combinar la rapidez computacional de "Matching Pursuit" [3] y la superresolución de "Basis Pursuit" [4]. El algoritmo de "High Resolution Pursuit" es estructuralmente similar a "Matching Pursuit", y por lo tanto, posee la misma complejidad computacional de éste. A diferencia de "Matching Pursuit", "High Resolution Pursuit" emplea una medida de similitud que resalta las correspondencias locales sobre las correspondencias globales, con lo cual es posible obtener una superresolución similar a la que exhibe "Basis Pursuit".

2.3.1 "Matching Pursuit".

Se puede definir a un diccionario D como una familia de átomos o vectores $(g_\gamma)_{\gamma \in \Gamma}$, los cuales pertenecen al espacio de Hilbert H , y poseen una norma unitaria $\|g_\gamma\| = 1$. Sea V

² A Heisenberg se le acredita la observación hecha en mecánica cuántica, donde no es posible especificar simultáneamente tanto la localización, como el momento de una forma de onda con arbitraria certeza. El producto de las incertidumbres de la localización y el momento debe exceder a la cuantización. Sin necesidad de desarrollar toda la mecánica cuántica, se puede visualizar la prueba de este hecho y relacionarla con la presente discusión de átomos de tiempo-frecuencia.

el espacio lineal cerrado del diccionario de vectores. Las expansiones lineales y finitas de los vectores en D son densas en el espacio V . Se puede decir que el diccionario está completo sí y sólo sí $V = H$. Para un diccionario de átomos de tiempo-frecuencia, $H = L^2(\mathbf{R})$, y cada uno de los vectores g_γ es un átomo definido por la ecuación (2.2). Si las expansiones lineales de los átomos de tiempo-frecuencia son densas en $L^2(\mathbf{R})$, entonces este diccionario está completo [8].

Sea $f \in H$. Se desea calcular una expansión lineal de f sobre un conjunto de vectores seleccionados de D , de modo que se obtenga la mejor correspondencia de su estructura interna. Esto se realiza empleando una aproximación sucesiva de f con proyecciones ortogonales sobre elementos de D . Sea $g_{\gamma_0} \in D$. El vector f se puede descomponer en

$$f = \langle f, g_{\gamma_0} \rangle g_{\gamma_0} + Rf, \quad (2.5)$$

donde Rf es el vector residual después de aproximar f en dirección de g_{γ_0} . Claramente g_{γ_0} es ortogonal a Rf , por lo tanto

$$|f|^2 = |\langle f, g_{\gamma_0} \rangle|^2 + |Rf|^2. \quad (2.6)$$

Para minimizar $|Rf|$, se debe elegir $g_{\gamma_0} \in D$ tal que $|\langle f, g_{\gamma_0} \rangle|$ sea máximo. En algunos casos, esto se puede lograr con tan solo encontrar un vector g_{γ_0} que sea el mejor en el sentido que cumpla con la condición

$$|\langle f, g_{\gamma_0} \rangle| \geq \alpha \sup_{\gamma \in \Gamma} |\langle f, g \rangle|, \quad (2.7)$$

donde α es un factor de optimización que satisface $0 < \alpha \leq 1$.

“Matching Pursuit” es un algoritmo iterativo que subdescompone el residuo Rf proyectándolo sobre un vector de D , de modo que este vector corresponda lo mejor posible al residuo Rf ; este proceso se lleva a cabo de la misma manera como anteriormente se hizo para f . “Matching Pursuit” efectúa nuevamente este procedimiento con el siguiente residuo que se obtiene y así sucesivamente, hasta cumplir con alguna condición establecida, tal como el número de átomos a obtener o el valor de la relación $|R^n f| / |f|$.

“Matching Pursuit” se lleva a cabo de la siguiente manera:

$$\text{Sea } R^0 f = f.$$

Suponiendo que se ha calculado el residuo de orden n : $R^n f$, para $n \geq 0$. A continuación se escoge un elemento $g_\gamma \in D$ que corresponda lo mejor posible al residuo $R^n f$, bajo la siguiente condición:

$$|C(R^n f, g_{\gamma_n})| = \sup_{\gamma \in \Gamma} |C(R^n f, g_\gamma)| \quad (2.8)$$

donde $C(R^n f, g_\gamma)$ es una función de correlación que mide la similitud entre f y g_γ .

Entonces, el residuo $R^n f$ se subdescompone en:

$$R^n f = C(R^n f, g_\gamma) g_\gamma + R^{n+1} f \quad (2.9)$$

el cual define al residuo de orden $n + 1$. En "Matching Pursuit", S. Mallat y Z. Zhang [3] introdujeron inicialmente la siguiente función de correlación:

$$C(f, g_\gamma) = \langle f, g_\gamma \rangle.$$

Se prueba entonces que el error $|R^n f|$ decae hacia cero. De modo que realizando iterativamente (2.9) se obtiene:

$$f = \sum C(R^n f, g_\gamma) g_\gamma \quad n = 0, 1, 2, \dots, \infty \quad (2.10)$$

La estructura del algoritmo de "Matching Pursuit" permite que las operaciones se efectúen rápidamente hasta cumplir con la condición que se ha establecido.

2.3.2 "Basis Pursuit".

"Basis Pursuit" es una de las aproximaciones más recientes, desarrolladas para el procesamiento de señales. Esta aproximación encuentra la representación de una señal dentro un diccionario redundante empleando un proceso de optimización convexo.

La esencia de este método consiste en obtener la descomposición de una señal f , bajo la condición de minimizar la norma l^1 de los coeficientes presentes en la representación, esto lo logra optimizando la siguiente declaración

$$\min \|\alpha\|_1 \text{ sujeta a } \Phi\alpha = f.$$

Debido a la no diferenciabilidad de la norma l^1 , este principio de optimización conduce a descomposiciones que pueden tener algunas propiedades muy diferentes de las del Método de "Frames" [9]; en particular, las descomposiciones pueden ser más reducidas.

Como este método consiste en una optimización global, puede resolver la descomposición de manera estable, lo cual no lo puede hacer "Matching Pursuit". Ya que "Basis Pursuit" siempre proporciona una descomposición en una base óptima, y no necesariamente en una base ortogonal, parece ser un mejor método que "Best Orthogonal Basis", principalmente resolviendo estructuras no ortogonales.

2.3.3 "High Resolution Pursuit"

"Matching Pursuit" es un algoritmo económico en el sentido de que en cada paso optimiza la energía que adquiere. Esto comúnmente lo lleva a escoger aquellas características que se acoplan globalmente a la estructura de la señal, sin embargo, no son las que mejor se adaptan a las estructuras locales.

Para evitar este problema, S. Chen y D. Donoho [4] introdujeron "Basis Pursuit", el cual efectúa una completa optimización minimizando $\sum_{\gamma \in \Gamma} |\alpha_\gamma|$ sobre todas las posibles descomposiciones $f = \sum_{\gamma \in \Gamma} \alpha_\gamma g_\gamma$.

Sin embargo, este algoritmo está casado con algunos problemas de gran escala de programación lineal, y por esta razón emplea una gran cantidad de tiempo para el cálculo.

"High Resolution Pursuit" es una versión diferente a "Matching Pursuit". Emplea una función de correlación distinta, la cual permite resaltar las correspondencias locales sobre las globales presentando la rapidez que muestra la estructura del algoritmo "Matching Pursuit".

Para cada átomo de tiempo-frecuencia g_γ se introduce un conjunto de índices subatómicos I_γ . I_γ corresponde a pequeños átomos g_{γ_i} , donde $\gamma_i \in I_\gamma$ con un soporte de tiempo incluido en el soporte de g_γ y modulado a la misma frecuencia.

Suponiendo que el átomo g_γ es elegido en un seguimiento, entonces

$$Rf = f - C(f, g_\gamma) g_\gamma$$

es el residuo de la señal f producido en este seguimiento.

Para toda $\gamma_i \in I_\gamma$, $\langle Rf, g_{\gamma_i} \rangle$, representa la cantidad de energía de Rf localizada en el soporte tiempo-frecuencia de g_{γ_i} . Esta cantidad debe ser menor a la energía de la señal $\langle f, g_{\gamma_i} \rangle$ en la misma posición. Además, el decremento correspondiente $\langle C(f, g_\gamma) g_\gamma, g_{\gamma_i} \rangle$ de la energía de la señal no puede ser mayor a la energía de la señal misma.

Esto se formaliza de la siguiente manera:

$$|\langle Rf, g_{\gamma_i} \rangle| \leq |\langle f, g_{\gamma_i} \rangle| \quad (2.11)$$

$$|\langle C(f, g_\gamma) g_\gamma, g_{\gamma_i} \rangle| \leq |\langle f, g_{\gamma_i} \rangle| \quad (2.12)$$

A partir de estas relaciones se deriva una nueva función de correlación $C(f, g_\gamma)$, la cual maximiza la cantidad de la energía de la señal que el seguimiento puede adquirir cuando selecciona al átomo g_γ :

$$C(f, g_\gamma) = \varepsilon \min_{\gamma_i \in I_\gamma} |\langle f, g_{\gamma_i} \rangle| / |\langle g_\gamma, g_{\gamma_i} \rangle|$$

donde ε se evalúa de la siguiente manera:

- Si $\langle f, g_{\gamma_i} \rangle$ tiene el mismo signo, para toda $\gamma_i \in I_\gamma$, entonces ε es el signo común.
- De otra manera $\varepsilon = 0$.

En "Matching Pursuit" el producto interno, empleado como función de correlación entre el átomo y la señal, se olvida si la señal contiene energía sobre todo el soporte tiempo-frecuencia del átomo. Por el contrario, la nueva función de correlación evita crear energía en las posiciones de tiempo donde no la hay.

Debido a la nueva función de correlación, dos átomos seleccionados por la descomposición poseen un soporte de tiempo más pequeño que en la descomposición usual con "Matching Pursuit", en consecuencia, debido a las desigualdades de Heisenberg, poseen además un soporte con frecuencia mayor. De este modo, "High Resolution Pursuit" efectúa una mayor descomposición tiempo-frecuencia que "Matching Pursuit", sin embargo su resolución en frecuencia se decrementa.

Referencias.

- [1] I. Daubechies, Time-frequency localization operators: a geometric phase space approach, *IEEE Transactions on Information Theory*, Vol. 34, No. 4, pp. 605-612, 1988.
- [2] R. R. Coifman y M. V. Wickerhauser, Entropy-based algorithms for best-basis selection, *IEEE Transactions on Information Theory*, Vol. 38, pp. 713-718, 1992.
- [3] S. Mallat y Z. Zhang, Matching Pursuits with Time-Frequency Dictionaries, *IEEE Transactions on Signal Processing*, Vol. 41, No. 12, Dic. 1993.
- [4] S. Chen y D. Donoho, Basis Pursuit, *Reporte Técnico, Departamento of Estadística, Universidad de Stanford*, Mayo 1995.
- [5] S. Jaggi, W. C. Carl, S. Mallat y A. S. Willsky, High Resolution Pursuit for Feature Extraction, *Reporte Técnico, MIT*, Noviembre 1995.
- [6] S. Jaggi, Multiscale Geometric Feature Extraction and Object Recognition, *Tesis Doctoral, LIDS, MIT*, Febrero 1997.
- [7] M. V. Wickerhauser, Comparison of picture compression methods: wavelet, wavelet packet, and local cosine transform coding, *Reporte Técnico, Departamento de Matemáticas, Universidad de Washington*.
- [8] B. Torresani, Wavelets associated with representations of the affine Wely-Heisenberg group, *Journal of Mathematical Physics*, Vol. 32, No. 5, pp. 1273-1279, Mayo 1991.
- [9] L. Daubechies, Ten Lectures on wavelets, *SIAM*, 1992.

Cuantización.

En general, la cuantización es el corazón de la conversión analógica a digital. En su forma más simple, un cuantizador observa a un solo número y selecciona el valor más cercano dentro de un conjunto finito predeterminado de valores numéricos permitidos. Ordinariamente, el valor de entrada es analógico, esto es, que puede tomar cualquier valor dentro de un rango continuo de posibles valores; sin embargo, la salida es digital, la cual se especifica únicamente por un entero de un conjunto $\{y_1, y_2, y_3, \dots, y_N\}$, donde N es el tamaño del conjunto de valores de salida.

3.1 Cuantizador escalar.

Podemos definir a un cuantizador escalar de N valores (en una dimensión) Q como la función $Q: \mathfrak{R} \rightarrow C$ donde \mathfrak{R} es el eje real y

$$C \equiv \{y_1, y_2, y_3, \dots, y_N\} \subset \mathfrak{R}$$

es el conjunto de salida de tamaño $|C| = N$. Los valores de salida y_i , en ocasiones están referidos como niveles de salida, puntos de salida ó valores de reproducción.

En todos los casos de interés práctico N es finito, por lo que un número finito de dígitos binarios es suficiente para especificar el valor de salida. Para un cuantizador con un conjunto de salida finito, se asume que los índices de los valores de salida son elegidos de modo que $y_1 < y_2 < \dots < y_N$.

A cada punto N del cuantizador está asociado una porción del eje real \mathfrak{R} , el cual se encuentra dividido en N celdas o átomos R_i , donde $i = 1, 2, \dots, N$. La i -ésima celda está dada por $R_i = \{x \in \mathfrak{R} : Q(x) = y_i\} \equiv Q^{-1}(y_i)$, que equivale a la imagen inversa de y_i bajo Q .

Como un cuantizador Q se puede describir completamente a través de sus niveles de salida $\{y_i; i = 1, 2, \dots, N\}$ y su celda correspondiente $\{R_i; i = 1, 2, \dots, N\}$, algunas veces se puede escribir $Q = \{y_i, R_i; i = 1, 2, \dots, N\}$ para hacer explícita dicha descripción [1].

3.2 Cuantizador regular y rango del cuantizador.

Se dice que un cuantizador es regular si cumple con las siguientes condiciones:

- a) Cada celda R_i es un intervalo de la forma (x_{i-1}, x_i) junto con uno ó ambos puntos extremos, e
- b) $y_i \in (x_{i-1}, x_i)$.

Los valores x_i son en ocasiones llamados puntos límite, puntos de decisión, niveles de decisión o puntos extremos. Un cuantizador regular tiene la propiedad característica de que si dos valores de entrada, sean a y b con $a > b$, son cuantizados con el mismo valor de salida, w , cualquier otra entrada que esté entre a y b será cuantizada también con el mismo valor de salida w . Esta propiedad corresponde a la noción intuitiva de cuantización como una manera de aproximar un número real con una representación de precisión finita. Como una manera arbitraria de asignar puntos terminales al igual que de proporcionar una descripción completa, se asume que para un cuantizador regular

$$R_i = (x_{i-1}, x_i] = \{r : x_{i-1} < r \leq x_i\}; i = 0, 1, \dots, N-1$$

donde la notación $(a, b]$ significa que se incluye al intervalo (a, b) junto con el valor terminal derecho b . Esta definición es la misma para $i = N$ si x_N es finita. Si x_N es infinita, entonces $R_N = (x_{N-1}, x_N)$.

Para un cuantizador regular, el intervalo de puntos finales se ordena de acuerdo a:

$$x_0 < y_1 < x_1 < y_2 < x_2 < \dots < y_N < x_N.$$

Usualmente, $x_0 = -\infty$ y $x_N = +\infty$ y las celdas R_0 y R_N son celdas saturadas.

El rango de un cuantizador, B , se define como la longitud total de las celdas, de modo que para un cuantizador regular no limitado, $B = x_{N-1} - x_1$. Ocasionalmente los valores de entrada pueden estar restringidos dentro de un subconjunto del eje real. En particular, si los valores de entrada caen dentro de una región limitada, entonces x_0 y x_N pueden ser finitos dado que no hay la necesidad de definir la operación del cuantizador para los valores de entrada que no se llegarán a presentar. En este caso, el cuantizador no contiene celdas saturadas y el rango está dado por $B = x_N - x_0$.

3.3 Objeto y desempeño de un cuantizador.

Todo cuantizador puede verse como un efecto combinado de dos operaciones sucesivas: codificación E y decodificación D . La codificación está definida como una operación $E : R \rightarrow I$, donde $I = \{1, 2, \dots, N\}$, y la decodificación está definida como una operación $D : I \rightarrow C$. De modo que si $Q(x) = y_i$ entonces $E(x) = i$ y $D(i) = y_i$. Dadas estas

definiciones tenemos que $Q(x) = D(E(x))$. En ocasiones, en el área de la ingeniería, se asume que la cuantización genera un índice $E(x)$ y un valor de salida, y , en tanto que la decodificación es definida como una "cuantización inversa".

El propósito de la cuantización es proporcionar una descripción precisa y limitada de un valor de entrada previamente desconocido. Debido a que previamente no se conoce el valor de entrada es necesario cuantizar. Así la entrada debe modelarse como una variable aleatoria, teniendo cierto carácter estadístico específico, usualmente definido por su función de densidad de probabilidad. En consecuencia, el error introducido en la cuantización es aleatorio también. Para asegurar de manera conveniente el desempeño de un cuantizador en particular, se necesita un número que indique la degradación de la calidad o la distorsión ocasionada durante su tiempo de utilización, considerando una entrada particular definida estadísticamente. Además, se requiere de algún tipo de medida total de desempeño, la cual usualmente se basa en un promedio estadístico que considere la entrada de la función de densidad de probabilidad junto con las características específicas del cuantizador.

La medida de distorsión más común entre dos números se encuentra en el desarrollo de una predicción lineal óptima: el error cuadrático. Otra medida de interés es el error absoluto.

Para una medida total de desempeño que considere el desempeño total de un cuantizador se puede emplear el valor en el peor de los casos, ó el promedio estadístico, de alguna medida que se adapte a la distorsión. Para una entrada limitada, el error absoluto en el peor de los casos, ó error máximo, es el valor máximo del error cuadrático obtenido a partir de todas las posibles entradas. Cabe señalar que el error máximo depende solamente del soporte de la función de densidad de probabilidad, el cual es el conjunto de valores de entrada para los cuales la función de densidad es diferente de cero. Para una entrada no limitada, con un cuantizador finito, donde el número de niveles N es finito, el error en el peor de los casos es infinito y por lo tanto no es una medida de desempeño significativa.

El promedio estadístico de la distorsión es usualmente una medida de desempeño informativa y significativa [1]. El promedio estadístico del error cuadrático de la medida de distorsión es referido como distorsión promedio. Esta distorsión promedio es comúnmente llamada error cuadrático medio (MSE). Dada una variable aleatoria de entrada, X , y un cuantizador, $Q = \{ y_i, R_i; i = 1, 2, \dots, N \}$, la distorsión promedio esta definida por

$$D = E[(X - Q(X))^2]$$

El promedio de la distorsión del error cuadrático es la medida más común, empleada para valorar el desempeño de una compresión de datos ó de un sistema de cuantización.

El error cuadrático de distorsión es fácil de usar, y en general se basa en una motivación intuitiva, pero bien se sabe que no corresponde adecuadamente a una medida subjetiva de calidad para señales visuales destinadas a la percepción humana.

3.4 Cuantización de alta resolución.

En la mayoría de las aplicaciones de cuantización escalar, el número de niveles N , se elige grande de modo tal que la salida cuantizada sea lo más cercana posible a la entrada original. Se emplea el término alta resolución para referirse al caso de cuantización cuando el promedio de distorsión es mucho menor que la varianza de entrada.

3.5 Cuantización en selecciones adaptables.

Para efectuar una operación de cuantización teniendo un buen desempeño en la tasa de distorsión, no se espera que éste se realice con la cuantización adecuada del conjunto de coeficientes de un marco y la conservación de todos ellos. Más bien, para conseguir una tasa de codificación relativamente baja, se espera que el mejor desempeño de la tasa de distorsión resulte de obtener solamente una pequeña cantidad de coeficientes. Esta observación motiva el uso de algoritmos de selección adaptable para la conservación de unos cuantos productos internos de la operación.

La introducción de la cuantización al momento que se desarrolla el algoritmo de selección adaptable, proporciona un error de menor magnitud en comparación con el error que se obtendría si la cuantización se efectuara una vez concluido el proceso de selección. La cuantización de los coeficientes destruye la ortogonalidad de las proyecciones y sus residuos [2].

Existen algunos problemas de diseño que deben considerarse cuando se desea efectuar una cuantización en una técnica de selección, por ejemplo, se debe seleccionar un diccionario, se debe diseñar un cuantizador escalar, y se debe decidir cuantos productos internos cuantizados se van a conservar. En un principio, se pueden optimizar cada uno de estos problemas para una fuente con una distribución dada, una medida de distorsión y una tasa de distorsión.

En algunas aplicaciones de procesamiento de señales, una señal se descompone en una combinación lineal de elementos de una base y su desempeño depende enormemente de la selección de esta base. Por ejemplo, en la compresión de imágenes, se cree ampliamente que la base DCT ("Discrete Cosine Transform") se desempeña bien en regiones uniformes, sin embargo, lo hace de manera muy pobre en regiones con muchos contornos. Si se encuentran algunas representaciones con respecto a un conjunto saturado (opuesto a la base), tendremos una gran flexibilidad para aproximar adecuadamente a una gran variedad de señales.

En el caso de dimensión finita, una vez que se ha seleccionado el conjunto redundante, el problema de encontrar una representación eficiente se puede abstraer de la siguiente manera. Definimos el vector de una señal como $f \in \mathfrak{R}^n$. Los vectores de expansión se encuentran en las columnas de $A \in \mathfrak{R}^{n \times m}$. Entonces Ax , $x \in \mathfrak{R}^m$ es una representación eficiente de b si $\|Ax - b\|$ es pequeña y x tiene un pequeño número de entradas diferentes de cero. Sin tener restricciones en A , es difícil hallar una x de modo que:

- Para un $\varepsilon \in \mathfrak{R}^+$ dado, $M \in \mathbb{Z}^+$, determinar si existe x con no más de M entradas diferentes de cero tal que $\|Ax - b\| \leq \varepsilon$ [3, 4].
- Para una $M \in \mathbb{Z}^+$ dada, encontrar x , tal que minimice $\|Ax - b\|$ junto con todos los vectores con no más de M entradas diferentes de cero [4].

"Matching Pursuit", así como cualquier otra técnica de selección adaptable, proporciona una aproximación subóptima y económica a este problema. En el caso de una dimensión finita, esto equivale al uso de una heurística económica conocida para encontrar algunas soluciones aproximadas de ecuaciones lineales sin necesidad de ortogonalización [3, 5].

En las aplicaciones de compresión, los coeficientes deben cuantizarse. Como solamente los valores de los coeficientes cuantizados estarán disponibles para la reconstrucción, el uso de valores cuantizados durante el desarrollo del algoritmo produce globalmente un error de cuantización menor con respecto a si se efectuara la cuantización al final del proceso. Así, un algoritmo de selección con cuantización es un algoritmo en el cual cada coeficiente es cuantizado antes de obtener el siguiente residual.

En el uso de un algoritmo de selección adaptable para compresión, además del diseño del diccionario, existen dos factores que determinan su desempeño: el método de reconstrucción y el método empleado para la codificación con pérdidas bajas de los índices y los coeficientes cuantizados. En [6] se presenta un trabajo referido al problema de reconstrucción para "Matching Pursuit", donde se propone el uso de la consistencia para una mejora en la reconstrucción, y se presentan los resultados obtenidos en señales sintéticas, imágenes y vídeo.

Aún cuando se tenga un método de reconstrucción óptimo, el desempeño de la codificación puede ser pobre. Existen algunas razones por las cuales se presenta un desempeño tan pobre, junto con estas razones está una carencia de la estructura en la fuente que puede explotarse en el diseño del diccionario y en la ineficacia de la codificación con pérdidas bajas.

Sin embargo, si se asume que se emplea una función de cuantización escalar en cada iteración del algoritmo, y que el cuantizador escalar proyecta un intervalo simétrico hacia cero, con base en simples observaciones, se puede notar que un método de codificación alternativo muy simple puede aumentar grandemente la tasa de compresión. Por ejemplo, si se asume que el coeficiente cuantizado $\hat{a}_j = 0$, entonces $\hat{a}_k = 0$ para toda $k > j$ debido a que el residuo no cambia; además, si $\hat{a}_j = 0$, entonces los índices i_j no poseen mucha

información. Más aún, en el caso en que las especificaciones de i_j y \hat{a}_j definan un conjunto no convexo, esta información es difícil de emplear para la reconstrucción [7]. De modo que se proponen algunos ajustes en [6] para el algoritmo de selección adaptable y lo refieren como codificación dependiente.

Un codificador de selección adaptable con cuantización puede ser competitivo con un codificador tradicional de transformación a tasa de bits baja. Este hecho se debe al alto gasto en la codificación de los índices de los elementos del diccionario empleado, o en el caso de un diccionario ortogonal, de los índices de los coeficientes de la transformación empleada. La flexibilidad del marco de trabajo permite experimentar empleando diversos conjuntos redundantes para la representación; dado que el desempeño es competitivo también con un diccionario ortogonal, esta operación no está fuera de contexto.

Tal y como se ha mencionado, en la práctica se espera que al introducir una etapa de cuantización al momento que se efectúa la selección de los coeficientes de los átomos en los modelos propuestos, los valores de dichos coeficientes formen en su totalidad un conjunto reducido que permita obtener, después de efectuar la compresión efectiva, una tasa de compresión alta comparable con la tasa que se pueda obtener al emplear un codificador tradicional. El tamaño del conjunto de valores de los coeficientes cuantizados depende del valor del paso de cuantización que se emplee, por esta razón se propone emplear diferentes valores para el paso de cuantización que permitan obtener y comparar los resultados. Por otro lado, introducir la etapa de cuantización justo en el momento que se selecciona un coeficiente reduce el error global a que si ésta etapa introdujera al final del proceso, de esta manera el desempeño del codificador no se ve afectado enormemente.

Referencias.

- [1] A. Gersho y R. M. Gray. Vector quantization and signal compression, *Kluwer International Series in Engineering and Computer Science*, SECS 159, 1991.
- [2] V. K. Goyal, M. Vetterli y N. T. Thao, Quantization of Overcomplete Expansions, *Reporte Técnico*, DEE-UCB.
- [3] B. K. Natarajan, Sparse approximate solutions to linear systems, *SIAM J. Computing*, Vol. 24, No. 2, pp. 227-234, Abril de 1995.
- [4] G. Davis, Adaptive Nonlinear Approximations, *Tesis de Doctorado*, Univ. de New York, Septiembre 1994.
- [5] G. H. Golub y C. F. Van Loan, *Matrix Computations*, Johns Hopkins Univ. Press, Baltimore, Segunda Edición, 1989.
- [6] V. K. Goyal y M. Vetterli, Dependent coding in quantized matching pursuit, *SPIE*, Vol. 3024, pp. 2-12, 1997.
- [7] V. K. Goyal y M. Vetterli, Consistency in quantized matching pursuit, *Proc. IEEE ICASSP*, pp. 1787-1790, Mayo 1996.

Calidad de la imagen.

Hoy en día, donde la información juega un papel muy importante, las imágenes son utilizadas ampliamente como un medio efectivo de comunicación de la información. Los ejemplos más típicos los tenemos en las imágenes producidas por la TV, los satélites y los sistemas médicos. En estos casos, la calidad de la imagen que percibe el observador es decisiva para determinar la calidad total del sistema. La calidad perceptual de las imágenes producidas por estos sistemas está condicionada en gran medida por las medidas realizadas al momento del diseño. Para obtener la calidad de imagen deseada y tener un ciclo de diseño efectivo en costo, estas decisiones deben basarse en el conocimiento de calidad de la imagen y en los factores que influyen en ella.

4.1 Definición de calidad de la imagen.

Cada persona tiene una noción de calidad de imagen. Esta noción puede depender principalmente de su contexto. Es difícil encontrar una definición general de calidad de imagen que sea aplicable a todos los contextos. Roufs y Bouma [1] definieron calidad perceptual de imagen como "el grado de excelencia de la imagen".

El término *calidad subjetiva de la imagen*, es comúnmente empleado en lugar de *calidad perceptual de la imagen*. Aunque ambos términos se encaminan al mismo concepto, el uso de la palabra "subjetiva" puede algunas veces causar confusión [2]. El término *calidad subjetiva* puede implicar que las impresiones son personales y pueden diferir ampliamente a través de las personas. Sin embargo, se sabe que las personas suelen realizar juicios consistentes de la calidad de una imagen, y que dichos juicios, aunque sean de distintas personas, suelen coincidir de manera general. El término *calidad subjetiva de la imagen* puede significar también que los componentes estéticos de la imagen están involucrados. Para evitar tales confusiones, es preferible emplear el término *calidad perceptual de la imagen* [2].

La calidad perceptual de la imagen expresa la impresión total de un observador, y por esta razón es considerada un atributo global psicológico de la imagen. Además, para

percibir el atributo global de calidad de la imagen, los observadores de imágenes también perciben otros atributos básicos de la imagen como son la intensidad, el brillo, el contraste, el ruido, etc. Estos atributos básicos afectan la impresión total de la calidad de la imagen.

Otro concepto comúnmente empleado en la búsqueda de la calidad de una imagen, similar al atributo global de la calidad perceptual de la imagen, es el imperfecto perceptual (global o total). Percibir un imperfecto global es un concepto que se opone al de calidad perceptual [3, 4]. Lo cual significa entonces, *un grado de degradación percibido en la imagen*. La calidad perceptual de una imagen puede decrecer por varias razones, por ejemplo, debido al ruido en la imagen o a un difuminado. Los procesos físicos, tales como difuminado y ruido, que llevan al decremento en la calidad perceptual de la imagen, y que ocasionan un daño físico a las imágenes, son llamados imperfectos físicos o degradaciones.

Debido a que no se cuenta con una definición estricta de calidad de imagen, se emplean en su lugar definiciones operacionales. Las definiciones operacionales pueden variar dependiendo del contexto. En este momento, es importante efectuar una distinción entre los dos tipos de contextos en los cuales la calidad de la imagen se emplea. Primero, existe una calidad relacionada con la realización de una tarea basada en la imagen. Los ejemplos son: la lectura en una unidad de despliegue de vídeo; la detección de un objetivo dentro de la imagen, tal y como la detección de un tumor en una imagen de tomografía computarizada o de un tanque en una fotografía aérea. La calidad de la imagen ante este tipo de medio ambientes es llamada *calidad orientada al desempeño* [5]. Esto es diferente de la calidad de una imagen ante un medio ambiente de entretenimiento como la TV o el cine. En el caso de la TV o el cine, la calidad se refiere principalmente a la apreciación o el interés, por lo tanto, se emplea el término de *calidad orientada a la apreciación*. Estos dos tipos de calidad pueden influir recíprocamente una en la otra [6].

4.2 Medición de la calidad de la imagen.

Para ser útil, la calidad visual se debe poder medir. Las medidas deben ser confiables y reproducibles. Tradicionalmente, la calidad de la imagen se ha medido empleando métodos psicofísicos, donde las personas (sujetos) participan como observadores y jueces de las imágenes (estímulos). Estos experimentos involucran cuatro componentes principales: los estímulos, las condiciones de visibilidad, la metodología (instrucciones) y los sujetos. Los estímulos generan una sensación en los sujetos. La fuerza de dichas sensaciones se expresa como una respuesta del sujeto de acuerdo a la instrucción [2]. Por ejemplo, la instrucción puede ser, empleando una escala de 1 a 10, juzgar un atributo del estímulo, que pudiera ser la intensidad. La confianza y reproducción de los datos depende de las escenas seleccionadas y de los sujetos, al igual que de las condiciones de visibilidad y de la instrucción. Las recomendaciones al respecto, han sido propuestas por el Comité Consultivo Internacional de la Radio (CCIR) [7].

Para asegurar la validez y la generalidad de los resultados experimentales, se debe tener especial cuidado en los efectos que a continuación se mencionan. Las escenas deben

seleccionarse cuidadosamente, recordando que la variación de los datos entre escenas puede ser considerablemente mayor que entre sujetos [2, 8]. Aunque el problema de la selección de la escena es enormemente insoluble, se le puede dar la vuelta empleando varias escenas de diversos tipos. El número y tipo de sujetos debe elegirse considerando el hecho de que existen diferencias entre los sujetos nativos y los experimentados. El empleo de sujetos familiarizados con cada una de las escenas ó el tipo de distorsiones puede ocasionar resultados equivocados [9]. La metodología depende del tipo de medio ambiente: orientada al desempeño u orientada a la apreciación. Uno de los métodos más rápidos y confiables para expresar la sensación percibida, es el uso de una escala compuesta por categorías definidas por números ó adjetivos. Por ejemplo, para la calidad orientada a la apreciación, el CCIR recomienda el uso de una escala de categoría con adjetivos: Excelente, Bueno, Moderado, Pobre, Malo. La escala con categorías numéricas ha demostrado ser muy confiable para escalar la calidad y sus atributos [10, 11].

Las pruebas realizadas en laboratorio por organismos internacionales tales como el CCIR han demostrado que los resultados obtenidos de valorar la calidad de la imagen son confiables y reproducibles. Un excelente ejemplo lo tenemos en la prueba de escalamiento COST, en la cual se compararon cuatro algoritmos de codificación de vídeo con una escala de categorías, empleando adjetivos en cinco lenguajes Europeos [12]. Los resultados para los cinco países fueron notablemente similares.

4.3 Criterios de calidad objetiva.

La evaluación de la calidad entre una imagen original y una imagen reconstruida después de un proceso de codificación-decodificación es seguramente uno de los problemas más cruciales que se puede encontrar. Se han realizado varias pruebas de carácter subjetivo en las cuales se involucran observadores humanos para evaluar los niveles de calidad en la reconstrucción, sin embargo, estas pruebas suelen ser experimentalmente difíciles y largas, y los resultados dependen de las condiciones de prueba; además, estas pruebas no proporcionan métodos constructivos que permitan mejorar el desempeño de los codificadores y son difíciles de emplear como parte de un diseño. Por esta razón, las medidas objetivas no sólo alivian las dificultades expuestas, sino ayudan a ampliar el campo de la codificación de imágenes, permitiendo la posibilidad de efectuar ajustes sucesivos para mejorar u optimizar la calidad percibida.

A continuación se listan los criterios de calidad objetiva más comunes, los cuales no se adecuan perfectamente a las propiedades del sistema visual del hombre. Una alternativa se encuentra en ponderar estas medidas objetivas por medio de algunas funciones de carácter perceptual relacionadas con el modelado del sistema visual del hombre, tal y como lo hace la PQS ("Picture Quality Scale").

Si $I(i, j)$ e $\hat{I}(i, j)$ denotan respectivamente a la imagen original y reconstruida, entonces se definen los siguientes criterios de calidad objetiva [13]:

Error cuadrático medio (MSE):

$$MSE = \frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N (I(i, j) - \hat{I}(i, j))^2$$

donde M y N son los tamaños de la imagen en el sentido horizontal y vertical.

Relación señal a ruido (SNR):

$$SNR_{db} = 10 \log_{10} \frac{\sum_{i=1}^M \sum_{j=1}^N I^2(i, j)}{\sum_{i=1}^M \sum_{j=1}^N (I(i, j) - \hat{I}(i, j))^2}$$

Relación señal a ruido pico (PSNR):

$$PSNR_{db} = 10 \log_{10} \frac{255^2 MN}{\sum_{i=1}^M \sum_{j=1}^N (I(i, j) - \hat{I}(i, j))^2}$$

4.4 La escala de calidad PQS ("Picture Quality Scale").

En la codificación de imágenes, los artefactos introducidos por los codificadores tradicionales suelen ser incontrolables ó impredecibles, debido a que estos codificadores están diseñados para reducir la redundancia estadística, en lugar de producir una degradación progresiva de la imagen. Esta característica permite tan solo efectuar una compresión muy pequeña para poder conservar una calidad aceptable en la imagen.

La investigación realizada en el área psicofísica de la percepción visual del hombre ha revelado que el sistema visual humano no presenta la misma sensibilidad a los diferentes tipos de distorsión que se presentan en una imagen, afectando directamente la calidad de la imagen percibida. Se ha reportado una nueva metodología para determinar una métrica objetiva para la codificación de imágenes. Esta nueva metodología arroja como resultado una escala de calidad en imágenes ("Picture Quality Scale": PQS) aplicable a la codificación de imágenes acromáticas sobre todo el rango de calidad definida por los resultados subjetivos de opinión media ("Mean Opinion Score": MOS). PQS considera las propiedades de la percepción visual tanto para las características globales, como para los disturbios localizados. Además, PQS se aproxima mucho a MOS, con un coeficiente de correlación mayor a 0.92, comparado con el coeficiente de correlación de 0.57 obtenido para MSE [14].

La escala PQS está basada en las medidas cuantitativas de diversos factores de distorsión. Debido a que estos factores de distorsión están correlacionados, se realiza un análisis en la componente principal para transformarlos en fuentes de error no correlacionadas. Después, estos errores son llevados a un valor de la escala PQS, empleando el modelo obtenido a partir del análisis de regresión lineal con el MOS, el cual es una escala subjetiva de cinco valores de calidad de la imagen en términos de las distorsiones percibidas. Estos valores de error son:

- 5 Imperceptible.
- 4 Perceptible, pero no molesto.
- 3 Ligeramente molesto.
- 2 Molesto.
- 1 Muy molesto.

Considerando las descripciones anteriores, se puede decir que la métrica de PQS está basada en las propiedades perceptuales de la visión humana y en la vasta experiencia ingenieril respecto a la observación de modificaciones en imágenes a causa de la codificación [14, 15]. Las distorsiones ocasionadas por la codificación son identificadas típicamente como borrosidad, sombreado, difuminado, etc. La severidad de estas distorsiones hacia un observador humano difieren de acuerdo a una estructura espacial específica de los artefactos. Por esta razón, se deben combinar diferentes tipos de distorsiones para que reflejen su grado de visibilidad subjetiva [14], tal y como lo hace la escala PQS. La escala PQS ha sido utilizada exitosamente en diversas aplicaciones, tales como el diseño de un sistema electro-óptico para imágenes [16], la optimización de codificadores [17], y la comparación en el desempeño de codificadores [18, 19, 20]. En un trabajo relacionado, Xu *et al.* [21] proponen una métrica de error basada en segmentación empleando un método similar basado en la correlación con una escala de imperfecciones, mientras que Davies *et al.* [22] desarrollaron una métrica similar empleando una red neuronal no lineal. Todos estos métodos consideran las distorsiones desde una perspectiva de "alto nivel" identificando la estructura en los errores. Algunos métodos similares se han aplicado al vídeo [23].

A continuación describimos brevemente al algoritmo PQS, todos los detalles se pueden consultar en [14]. Primero, la señal de la imagen es transformada en una señal proporcional a la percepción visual de luminancia empleando una ley de potencia para entonces obtener los errores ponderados de frecuencia a través del filtrado con una función de sensibilidad de contraste. Las distorsiones percibidas en la imagen son identificadas y los factores de calidad objetiva correspondientes que cuantifican cada degradación de la imagen son calculados como funciones de los errores ponderados de frecuencia. Las distorsiones percibidas nos llevan a obtener medidas numéricas en términos de los factores de distorsión $\{ F_i \}$. En la primera etapa, cada uno de los factores de las imágenes es obtenido realizando cálculos locales sobre los errores ponderados de frecuencia. Enseguida, se obtienen los factores de distorsión F_i , almacenando los factores correspondientes a las imágenes (se debe notar que el factor F_1 no emplea las mismas fuentes de transformación que los otros factores, este se incluye en la escala PQS debido a

que es comúnmente empleada la distorsión ponderada de la función de sensibilidad de contraste).

La escala PQS emplea cinco factores, los primeros dos contabilizan los errores aleatorios, el tercero a los artefactos de bloques, y los dos últimos corresponden a los errores estructurales y los errores de máscara cerca de los contornos. El valor global de la escala PQS está dado por la combinación lineal de las componentes principales, $\{ Z_i \}$, calculadas a partir de los factores de distorsión, $\{ F_i \}$, de modo que

$$PQS = b_0 + \sum_{i=1}^J b_i Z_i$$

donde $\{ b_i \}$ son parámetros por ser determinados.

La escala PQS está diseñada para corresponder lo mejor posible con la evaluación subjetiva de la calidad de una imagen [19]. El valor subjetivo de la calidad es obtenido bajo condiciones de visibilidad controladas, empleando la escala de imperfecciones de ITU-R recomendación 500, que nos da el MOS que va de 1 a 5 [24]. Los coeficientes $\{ b_i \}$ son calculados empleando análisis de regresión múltiple [25] entre los factores de distorsión y los valores subjetivos MOS determinados experimentalmente. La aplicabilidad de los pesos del análisis de regresión múltiple depende del acomodo del conjunto de datos empleados en la regresión. En la versión actual de PQS, el conjunto de datos consiste de cinco diferentes imágenes, las cuales fueron distorsionadas empleando JPEG y los codificadores de sub-banda y ondeleta estándar [14]. El coeficiente de correlación entre las escalas PQS y MOS es $R=0.92$, el cual representa una gran mejora comparada con la correlación $R=0.54$ cuando sólo se emplea el error cuadrático medio de la frecuencia ponderada (F_2). La descripción detallada de la escala PQS y su desempeño en todo el rango de calidad se encuentra en [14].

En la práctica, para evaluar la calidad de las imágenes reconstruidas a partir de los coeficientes obtenidos con los modelos de codificación propuestos en este trabajo, se emplea la versión 1.0 de PQS desarrollada en el Centro para el procesamiento de imágenes y cómputo integrado de la Universidad de California en Davis, E.U. Esta versión ha sido probada exhaustivamente por los miembros de este centro, principalmente en imágenes comúnmente empleadas en el medio del procesamiento de imágenes como Lenna, Bárbara y Cameraman [14]. Estas imágenes forman parte del conjunto de prueba que se empleará en este trabajo. Su uso es muy sencillo, simplemente se proporciona como entrada al programa PQS la imagen original y la imagen reconstruida, y éste realiza todas las operaciones necesarias para obtener cada uno de los valores de los cinco factores de distorsión contemplados en el modelo PQS. Estos factores de distorsión se describen en la siguiente sección, y sus valores son empleados para proporcionar la calificación de calidad PQS. A la salida producida por el programa, simplemente se le aplica una operación de aproximación al entero más cercano para tener una correspondencia con la escala de cinco valores de calidad MOS, esta operación se realiza con la finalidad de mejorar la resolución perceptual de PQS.

4.4.1 Factores de distorsión.

Como se mencionó, la versión actual de la escala PQS contempla cinco factores de distorsión de los cuales, los primeros dos son derivados de errores aleatorios y los tres últimos de errores estructurales. A continuación se describen brevemente estos factores de distorsión. Las fórmulas para el cálculo de las medidas numéricas actuales son detalladas en [14]. Nótese que las medidas de distorsión perceptual efectivas están en función de la resolución del despliegue y la distancia visual.

Factor de distorsión F_1 . Es la diferencia ponderada entre la imagen original y la comprimida. La función de ponderación adoptada corresponde al estándar ponderado del ruido en TV propuesto por el CCIR. Para este factor se supone que la distancia visual es cuatro veces la altura de la imagen.

Factor de distorsión F_2 . Es también la diferencia ponderada entre la imagen original y la comprimida. La función de ponderación se deriva de un modelo del sistema visual del hombre [26]. Adicionalmente, se incluye una función indicadora para contabilizar el umbral perceptual de visibilidad.

Factor de distorsión F_3 . Refleja los disturbios de fin de bloque. El sistema visual del hombre es realmente sensible a los errores lineales y estructurales en las características de las imágenes. En los codificadores de bloques, la imagen de error contiene discontinuidades al final de los bloques, lo cual explica los artefactos de bloques percibidos en la imagen comprimida.

Factor de distorsión F_4 . Contabiliza los errores correlacionados generales. Los errores con una fuerte correlación son más perceptibles que los patrones aleatorios. La fuerte correlación en la imagen de error sugiere una distorsión más aparente que la contabilizada mediante la magnitud de los errores.

Factor de distorsión F_5 . Es una medida de la longitud de los errores que ocurren en la mayoría de los codificadores en la vecindad de las transiciones de alto contraste (contornos). Se pueden llegar a presentar dos efectos psicofísicos en la vecindad de los contornos de alto contraste. Por un lado, la visibilidad del ruido, junto con los errores, decrece; esto está referido como una "máscara visual". Por el otro lado, los desalineamientos y la borrosidad en los contornos se incrementan, lo que es completamente reprobable.

4.4.2 Representación de la componente principal de las medidas de distorsión.

Debido a que los factores de distorsión $\{F_i\}_{1 \leq i \leq 5}$ están correlacionados, se realiza un análisis de la componente principal para decorrelacionar las medidas de distorsión e identificar las fuentes dominantes. Los datos del MOS consisten de 675 evaluaciones

subjetivas: 9 observadores, 5 imágenes, 3 codificadores y 5 niveles de calidad para cada codificador. Para cada una de las 75 imágenes codificadas, se tiene que el MOS es promediado obteniéndose 75 evaluaciones MOS. Además se obtiene una matriz de covarianza C_F de $\{F_i\}$ a partir del conjunto de imágenes codificadas en estudio. Un análisis con valores y vectores característicos de los factores de distorsión contra las evaluaciones MOS, indica que los tres valores característicos más grandes contabilizan el 98% del error de la energía total. En consecuencia, los tres vectores característicos correspondientes transforman $\{F_i\}$ en una representación de la componente principal, $\{Z_i\}_{1 \leq i \leq 3}$, y entonces tenemos que

$$PQS = b_0 + \sum_{i=1}^3 b_i Z_i$$

donde $\{b_i\}_{0 \leq i \leq 3}$ son los coeficientes de la regresión parcial obtenidos por una regresión lineal múltiple de $\{Z_i\}$ contra el resultado MOS [14].

Referencias.

- [1] J. A. J. Roufs y H. Bouma. Towards linking perception research and image quality, *Proceedings of the Society for Info. Display*, SID 21, pp. 247-270, 1980.
- [2] J. A. J. Roufs. Perceptual image quality: concept and measurement, *Philips Journal of Research*, 47, 35-62, 1992.
- [3] H. de Ridder. Minkowski-metrics as a combination rule for digital-image-coding impairments, in B. E. Rogowitz, (Ed.), *SPIE Human Vision, Visual Processing and Digital Display III, Proc. SPIE*, Vol. 1666, pp. 16-26, 1992.
- [4] M. Nijenhuis. Sampling and Interpolating of static images: a perceptual view, *Tesis de Doctorado, Eindhoven University of Technology*, 1993.
- [5] B. R. Hunt y G. F. Sera. Power-law stimulus-response models for measures of image quality in nonperformance environments, *IEEE Trans. Syst. Man Cybern.*, Vol. SMC-8, pp. 781-791, 1978.
- [6] J. A. J. Roufs y M. C. Boschman. Visual comfort and performance, J.A.J. Roufs (Ed.) *Vision and visual dysfunction*, Vol. 15, J.R. Cronly-Dillon (Ed.), *The Man-machine interphase*, MacMillan Press, London. pp 24-40, 1991.
- [7] CCIR. Method for the subjective assessment of the quality of television pictures, *Recommendation 500-3*. In: *Recommendations and reports of the CCIR. International Telecommunication Union, Geneva*, 1986.
- [8] J. A. J. Roufs, V. Koselka y A. van Tongeren. Global brightness contrast and the effect on perceptual image quality, in Rogowitz and Allebach, (Ed.), *SPIE, Human vision, visual processing and digital display V*, SPIE, Vol. 2179, pp. 80-89, 1994.
- [9] J. Westerink. Influences of subject expertise in quality assessment of digitally coded images, *Society for Info. Display, SID International Symposium, May 15-19, Baltimore, Digest of Technical papers*, No. 20, pp. 124-127, 1989.

- [10] J. A. J. Roufs, H. de Ridder y J. Westerink, Perceptual image quality metrics, *Applied Vision Topical Meeting San Francisco, Applied Vision Technical Digest Series, Washington: Optical Society of America*, 1990.
- [11] H. de Ridder y G. M. M. Majoor. Numerical category scaling: an efficient method for assessing digital image coding impairments, *SPIE Human Vision and electronic imaging: models, methods and applications, Proc. SPIE*, Vol. 1249, pp. 65-77, 1990.
- [12] J. W. Allnatt, N. Gleiss, F. Kretz, A. Sciarappa y E. van der Zee. Definition and the validation of methods of subjective assessment of visual telephone picture quality CSE2T, *Reporte Técnico XI*, pp. 59-65, 1983.
- [13] C. Labit y G. Tziritas, Motion analysis for image sequence coding, Colección: *Advances in image communication*, No. 3, Ed. Elsevier, 1994.
- [14] M. Miyahara, K. Kotani y V. R. Algazi. Objective picture quality scale (PQS) for image coding. *IEEE Transactions on Communications*, 1998.
- [15] M. Miyahara, Quality assessments for visual service. *IEEE Communications Magazine*, pp. 51-60, Octubre 1998.
- [16] J. L. Olives, B. Lamiscarre y M. Gzalet, Optimization of electro-optical imaging system with an image quality measure. *Proceedings of SPIE, Very High Resolution and Quality Imaging*, pp. 158-167, 1997.
- [17] V. R. Algazi, G. E. Ford, M. Mow y A. Najmi, Design of subband coders for high quality based on perceptual criteria, *Proceedings of SPIE, Application of Digital Image Processing XVI*, Vol. 2028, pp. 40-49, 1993.
- [18] J. Lu, V. R. Algazi y R. R. Estes Jr., A comparative study of wavelet image coders, *Optical Engineering*, Vol. 35, No. 9, pp. 2605-2619, Septiembre 1996.
- [19] V. R. Algazi, N. Avadhanam, R. R. Estes Jr., Quality Measurement and Use of Pre-processing in Image Compression, *Elsevier Preprint*, 25 de Junio de 1998.
- [20] V. R. Algazi, R. R. Estes Jr., Comparative Performance of Wavelet and JPEG Coders at High Quality, *Reporte Técnico CIPIC*, Universidad de California en Davis.
- [21] W. Xu y G. Hauske, Picture quality evaluation based on error segmentation, *Proceedings of SPIE, Visual Communication and Image Processing*, Vol. 2308, pp. 3:1454-1465, 1994.
- [22] I. R. L. Davies, D. Rose y R. J. Smith, Automated image quality assessments, *Proceedings of SPIE, Human Vision, Visual Processing and Digital Display IV*, Vol. 1913, pp. 27-36, 1993.
- [23] A. A. Webster, C. T. Jones, M. H. Pinson, S. D. Voran *et al.*, An objective video quality assessment system based on human perception, *Proceedings of SPIE, Human Vision, Visual Processing and Digital Display IV*, Vol. 1913, pp. 15-26, 1993.
- [24] ITU-R Rec. BT. 500-8: Methodology for the subjective assessment of the quality of television pictures, 1998.
- [25] M. G. Kendall, Multivariate Analysis, *Charles Griffin*, 1975.
- [26] Y. Horita y M. Miyahara. Image coding and quality estimation in uniform perceptual space. *Reporte Técnico IE87-115, IEICE*, Enero 1987.

Modelos de codificación.

En este capítulo, con base en todo el trabajo de investigación y las pruebas preliminares realizadas, se describirán los modelos de codificación propuestos. Antes de pasar al detalle de los modelos que se proponen, se mencionarán algunos de los modelos de codificación de imágenes que históricamente preceden a los que se presentan en este trabajo, y los cuales vienen a ser su punto de partida.

En esencia, como antecedente, se presentarán aquellos modelos de codificación que emplean la descomposición tiempo-frecuencia de alguna u otra manera para realizar su labor codificadora. Cabe hacer notar que desde hace tiempo la descomposición tiempo-frecuencia ha comenzado a jugar un papel importante en todos aquellos trabajos relacionados con el procesamiento de imágenes, en consecuencia, estos trabajos se han vuelto muy populares; he aquí la importancia de estudiar con más detenimiento a los modelos de codificación, que implícitamente efectúan este tipo de descomposición.

Los trabajos que se presentan como antecedente han surgido a partir de la investigación que han realizado, por una parte, Burt y Adelson, y por otra parte, S. Mallat y su equipo de colaboradores. Este último grupo ha estudiado arduamente en los últimos años a la transformada ondeleta, sus propiedades y características.

Concretamente en este trabajo se presentan dos propuestas, debido a que los métodos de selección adaptable son muy flexibles y dan la libertad de poder elegir la manera de emplearlos, por esta razón son muy utilizados hoy en día en diversas aplicaciones, tales como la codificación de imágenes residuales de vídeo [1, 2, 3, 4, 5, 6, 7], el análisis de señales de audio [8, 9, 10], la eliminación de ruido [11, 12], el reconocimiento de formas y la detección de contornos [13, 14, 15].

En cada uno de los modelos propuestos, se selecciona en primer lugar el diccionario a utilizar, éste debe adaptarse lo mejor posible a las estructuras de la imagen para obtener los mejores resultados, a pesar de que cada imagen tenga la característica de ser única. A continuación se expande la imagen sobre el diccionario y se inicia la selección de los átomos, introduciendo una etapa intermedia de cuantización en el algoritmo. En esta etapa

de cuantización se debe emplear un paso de cuantización óptimo que permita obtener buenos resultados con respecto a la tasa de compresión y la calidad de la imagen; finalmente se desarrolla la estrategia que realice una compresión efectiva de la información obtenida, para un almacenamiento o transmisión final.

5.1 Antecedentes.

Las técnicas de codificación se clasifican en dos categorías: técnicas de codificación de la primera generación y técnicas de codificación de la segunda generación. Las técnicas más importantes de la primera generación utilizan bases ortonormales para representar una imagen, mientras las técnicas de la segunda generación explotan las características de la imagen y la psicofísica de la percepción visual del hombre.

En esta sección, se menciona y se hace referencia a dos de las técnicas más representativas para efectuar una codificación compacta en imágenes, una perteneciente a la primera generación y otra a la segunda generación. Cabe señalar que estas técnicas son algunas de la gran cantidad de técnicas que existen actualmente, y que han dado hincapié para el desarrollo de este trabajo.

Cabe hacer notar que actualmente, existen muchos codificadores de imágenes, de naturaleza muy diversa, y que llegan a producir excelentes resultados, sin embargo la gran mayoría llegan a presentarse como elementos aislados o individuales, los cuales podrían considerarse como una parte de un codificador realmente completo. Un codificador completo debe integrar, además de alguna técnica de codificación, procesos de cuantización, selección, compresión efectiva y evaluación. De ahí la importancia del trabajo que aquí se presenta, en el cual, además de introducir un par de modelos de codificación novedosos, se considera e involucra los aspectos antes mencionados para que estos modelos sean lo más completos posible. Esta integridad hace que sean muy flexibles y puedan utilizarse indiscriminadamente sin importar la naturaleza de las imágenes.

5.1.1 Esquema de descomposición piramidal.

El esquema de descomposición piramidal Laplaciano, introducido por Burt y Adelson [16] corresponde a una técnica de codificación de imágenes de la primera generación, este algoritmo se basa en la transformada ondeleta rápida.

Se tiene que una imagen es una función de soporte compacto, valuada en el eje real en $L^2(\mathbb{R}^2)$. En la práctica, esta imagen está cuantificada de manera discreta y puede representarse como una matriz de píxeles. Por otro lado, en general, una imagen puede tener características complejas en diferentes escalas, y por lo tanto, su representación en una resolución adecuada requiere de una matriz de gran tamaño. Esto implica un costo excesivo en términos de cálculo, memoria y tiempo de proceso cuando se desea representar adecuadamente, almacenar o transmitir dicha imagen.

Una característica común, encontrada en las imágenes, es la alta correlación que existe entre los valores de una vecindad de píxeles, la cual suele ser el resultado de una alta redundancia de información contenida en la imagen. La representación de imágenes como una matriz de píxeles discretos es, en consecuencia, frecuentemente deficiente. La tarea principal de un esquema de codificación es entonces, la eliminación de la redundancia, lo cual lleva a una compresión significativa de la información. Realmente, esto es uno de los objetivos de los diferentes esquemas de descomposición y reconstrucción de imágenes diseñados para llevar a cabo la compresión de imágenes. Típicamente, estos esquemas caen dentro de dos categorías, llamadas esquemas predictivos causal y no causal.

En los esquemas de codificación predictiva causal, los píxeles son codificados como renglones secuenciales. Antes de codificar cada valor del píxel, este se "predice" a partir de los valores de los píxeles previamente codificados. Este valor predicho representa la información redundante, y por lo tanto, es sustraído del valor del píxel actual, y solo se codifica la diferencia (error de predicción). Estos esquemas de codificación son referidos como *causales* debido a que utilizan solamente los píxeles previamente codificados para la predicción. El valor de un píxel es decodificado calculando su valor predicho a partir de los valores de los píxeles previamente decodificados y agregándole el error de la predicción codificada.

Los esquemas predictivos no causales (basados en la vecindad) son más naturales y dan como resultado una predicción más exacta, produciendo una compresión más alta. Sin embargo, estos esquemas son difíciles de realizar ya que no permiten una codificación secuencial, pero involucran el uso de transformadas en imágenes o la solución de sistemas de ecuaciones simultáneas muy grandes, lo cual produce una sobrecarga computacional muy elevada.

El esquema presentado por Burt y Adelson [16] es un híbrido de los esquemas causal y no causal, debido a que goza de un cálculo fácil en el esquema causal, mientras conserva las ventajas y naturaleza de los esquemas no causales. Más aún, inherentemente posee la propiedad de proporcionar multiresolución y, en consecuencia, se adapta fácilmente a las técnicas de análisis de imágenes basadas en ondeletas.

La idea básica de este esquema es el siguiente. Sea $c^0 = (c^0(i, j))_{R \times C}$ la imagen original representada como una matriz. Si se aplica un filtro paso-bajas a c^0 , entonces, como la imagen resultante no contiene las componentes espaciales de alta frecuencia de c^0 , ésta se puede submuestrear adecuadamente sin pérdida de información, para obtener de esta manera, una versión reducida c^1 de c^0 . Interpolando los valores entre los puntos de muestreo, la imagen reducida c^1 se puede expandir en una imagen c^0 , la cual tiene el mismo tamaño que c^0 , pero es una versión difusa de ésta.

El valor predicho del píxel (i, j) está dado por $c^0(i, j)$. El error predicho es obtenido sustrayendo el valor predicho del valor actual del píxel:

$$d^0 = c^0 - c^0$$

En lugar de codificar c^0 , las imágenes d^0 y c^1 son codificadas, esto debido, por un lado, a que d^0 está enormemente decorrelacionada, y en consecuencia, puede representarse con una menor cantidad de bits que c^0 ; y por otro lado, a que como c^1 se obtiene a partir de c^0 empleando un filtraje paso-bajas, ésta puede codificarse con una tasa de muestreo menor. La reducción en la tasa de muestreo es proporcional a la reducción del ancho de banda. Este proceso se realiza iterativamente con el fin de poder obtener una mayor compresión de datos. En el siguiente paso, c^1 es filtrada con un filtro paso-bajas para obtener c^2 , la cual se expande para obtener la imagen predicha $c^{\sim 1}$ de c^1 . La imagen de error d^1 es nuevamente obtenida sustrayendo $c^{\sim 1}$ de c^1 . Este proceso produce una secuencia de imágenes de error d^0, d^1, \dots, d^n , una cada vez más pequeña que la anterior. Estas imágenes se acomodan en un esquema piramidal. Los diferentes niveles de esta estructura de datos piramidal corresponden a distintos rangos de frecuencia espacial de la imagen original. Cada uno de estos niveles se puede obtener efectuando la convolución de la imagen original con la diferencia de dos funciones con peso distinto, y como esta operación es similar a la aplicación de operadores Laplacianos en técnicas de mejoramiento de imágenes, el esquema Burt-Adelson es llamado también esquema de pirámide Laplaciana.

5.1.2 Modelo de codificación contornos-textura.

La representación multiescala de los contornos de una imagen, en combinación con una técnica de codificación de la textura, que puede basarse en el algoritmo de la ondeleta rápida piramidal, corresponde a una técnica de codificación de imágenes de la segunda generación.

Como se apuntó anteriormente, desde el punto de vista de la transmisión y almacenamiento de imágenes, es altamente deseable poder codificar la imagen con unos cuantos bits por pixel (bpp) en la medida de lo posible, para de esta manera economizar, el tiempo requerido para transmitir o recibir la imagen, al igual que la cantidad de espacio (memoria) necesaria para almacenar la imagen. Esto significa que alguna cantidad de la información o de los detalles, tiene que descartarse de tal modo que no degrade la calidad de la imagen de manera significativa, mientras se consigue un costo efectivo en la compresión de la imagen.

Mallat y Zhong [17] desarrollaron un algoritmo de codificación compacta basado en la caracterización aproximada de una imagen empleando la información de sus contornos multiescala. Mientras este algoritmo obtiene factores de compresión tan altos como 30, ignora totalmente la información de la textura, lo cual suele ser desventajoso para algunas aplicaciones, ya que la textura lleva consigo información importante y suplementaria de la imagen.

Froment y Mallat [18] desarrollaron una técnica de codificación de imágenes de doble etapa en la cual la imagen se codifica primero empleando solamente la información de los contornos multiescala, y posteriormente se codifica la diferencia entre la imagen original y la imagen de contornos codificada, esta diferencia contiene la información de la

textura, y la codificación se realiza empleando un algoritmo piramidal con ondeletas ortogonales. La codificación basada en contornos es más precisa que la codificación basada en textura [19].

En algunas de las aplicaciones de codificación de imágenes es necesario que los datos brutos sean interpretados por medio de patrones visuales significativos, lo cual impone un tratamiento espacial de los datos generados por la fuente. La codificación contornos-textura realiza un tratamiento espacial de dichos datos, el cual consiste en separar la información original en dos componentes: *contornos* y *textura*.

Dado que la información de los contornos es una información geoméricamente muy estructurada, el sistema de visión humano se torna muy sensible a las degradaciones provocadas por la codificación de esta información, por lo tanto es necesario codificar ésta con mayor precisión. Por otro lado, como la información de la textura es menos estructurada que la información de los contornos, ésta se puede codificar introduciendo un mayor grado de distorsión, sin provocar una fuerte deformación en la imagen decodificada.

La propuesta de Froment y Mallat es la siguiente:

En primer lugar, para codificar los contornos de una imagen éstos se deben de detectar, para lo cual, se descompone la imagen a procesar a lo largo de una estructura multiresolución empleando la transformada ondeleta, de esta manera se obtiene un conjunto de imágenes a diferentes resoluciones, tantas como se determinen. Enseguida se detectan los contornos correspondientes a cada una de estas imágenes resultantes. Los contornos se forman por puntos, cada uno de los puntos se constituye de dos parámetros: ángulo y módulo del gradiente; esto implica que por cada imagen de contornos obtenida se tienen dos mapas de información. El ángulo y módulo del gradiente se obtienen con la ayuda de la transformada ondeleta. Una vez que se tienen los mapas de contornos, se elige uno en especial a criterio personal (aquel que se asemeje más a la imagen original o aquel que tenga una menor cantidad de ruido), sobre el cual se llevará a cabo la codificación, y se desechan los mapas restantes.

La codificación de este mapa consiste en proponer dos valores de umbral, uno correspondiente a la longitud y otro a la media del módulo del gradiente de los contornos. Con base en estos valores de umbral se desechan los contornos cuya longitud y media del módulo sean menores a los valores de umbral propuestos para obtener un mapa de contornos codificado.

Al mapa de contornos codificado, se le aplica un algoritmo de reconstrucción para obtener una aproximación de la imagen original. La información de esta primera aproximación se resta de la imagen original, permitiendo calcular el error, el cual representa la información de textura.

La textura se analiza por medio de una base ortogonal ondeleta, sobre la cual se descompone su información a manera de una estructura piramidal. Sobre la estructura

piramidal se efectúa también una operación de umbral, la cual consiste en eliminar regiones cuadradas, cuya área sea proporcional a la escala de descomposición, y cuya energía sea menor al valor de umbral propuesto.

Esta nueva estructura contiene los datos codificados correspondientes a la textura.

Para recuperar la imagen original, simplemente se aplica la transformada inversa a los datos de la textura codificada y se suman a la aproximación obtenida con la información de los contornos codificados.

Combinando los parámetros de las codificaciones de los contornos y de la textura se obtienen diferentes tasas de compresión, las cuales dependiendo de la aplicación se evalúan cualitativa y cuantitativamente.

Con esta técnica se logran factores de compresión de 100 para imágenes de 512×512 pixeles. Detalles más precisos, y un estudio más completo de esta técnica se pueden encontrar en [18, 20].

5.1.3 Otros modelos de codificación.

Como se mencionó, actualmente existe una gran cantidad de modelos de codificación de imágenes, clasificados en dos generaciones. Sin embargo, todos estos modelos tienen el mismo objetivo, el cual ya ha sido mencionado ininidad de veces: una representación compacta de la imagen con una alta calidad visual en la reconstrucción.

Algunos modelos emplean filtros y bancos filtros para realizar su labor codificadora [21, 22], los cuales deben diseñarse cuidadosamente; otros modelos efectúan diferentes operaciones sobre árboles jerárquicos, los cuales son el resultado de la descomposición de la imagen [23, 24, 25, 26]; otro modelo, por ejemplo, se encuentra en una aplicación práctica, donde el FBI, para la codificación de su banco de imágenes de huellas digitales, emplea paquetes de ondeletas, junto con técnicas de selección de la base de descomposición [27]; también algunos modelos hacen uso de métodos que involucran un tratamiento especial en el espacio y la frecuencia [28, 29, 30].

Aunque las tendencias actuales son el uso de medidas de distorsión para regular la acción codificadora [31, 32] y de modelos de percepción humana [33], principalmente para obtener buena calidad en los resultados; el empleo de funciones de transformación tradicionales como la transformada de Gabor [34, 35, 36], la transformada ondeleta [37, 38, 39] con sus derivaciones [40, 41], y la transformada coseno [42, 43, 44], ha sido y sigue siendo la técnica más utilizada hoy en día, simplemente debido a su facilidad y sus ventajosas propiedades; tan es así, que uno de los formatos comerciales de imágenes más común, como JPEG [45], utiliza la transformada discreta coseno con buenos resultados.

5.2 Modelo de codificación por regiones.

Este modelo se basa en la idea presentada en los trabajos de Ralph Neff *et al.* relacionados con codificación de vídeo [1, 2, 3, 4, 5, 6, 7], donde se emplea "Matching Pursuit" para la codificación de las imágenes residuales del vídeo, y donde el primer paso consiste en la selección adecuada de un diccionario base. Esta selección se realiza empleando una técnica de entrenamiento, donde a partir de un conjunto de imágenes residuales se eligen los elementos del diccionario que mejor se adapten a las estructuras de las imágenes.

Con el diccionario seleccionado, se emplea una estrategia llamada "búsqueda de la energía" para seleccionar una ventana sobre la cual se efectuará el algoritmo de selección. Esta búsqueda se realiza iterativamente, y en cada paso se obtiene un elemento del diccionario que mejor se adapte a alguna de las estructuras presentes en la ventana seleccionada.

5.2.1 El diccionario empleado.

El diccionario empleado con este método consiste de una colección redundante de funciones separables de Gabor 2-D. Este tipo de diccionarios es muy utilizado en el análisis de imágenes, fundamentalmente porque las funciones que lo componen suelen adaptarse muy bien a las características principales de las imágenes. En experimentos realizados en el área de fisiología y visión, se ha observado que la respuesta en la corteza visual de animales se aproxima a una función de Gabor al momento de aplicarle un estímulo [46, 47]. Como se mencionó, con una técnica de entrenamiento se obtuvo al conjunto de funciones de Gabor que forman al diccionario utilizado.

El conjunto empleado en este modelo se define en términos de una ventana Gaussiana:

$$g(t) = \sqrt{2} e^{-\pi t^2} \quad (5.1)$$

Se puede entonces definir a las funciones discretas de Gabor de 1-D como un conjunto de ventanas Gaussianas escaladas:

$$g(i) = K_s g\left(\frac{i - \frac{N}{2} + 1}{s}\right) \quad i \in \{0, 1, \dots, N - 1\}; \quad (5.2)$$

Donde s es el parámetro de escala positiva. La constante K_s se elige de modo tal que la secuencia resultante tenga norma unitaria. Si se considera que B es el conjunto de todas las escalas, podemos definir las funciones de Gabor separables de la siguiente manera:

$$G_{a,b}(i,j) = g_a(i)g_b(j) \quad i,j \in \{0, 1, \dots, N-1\} \quad \text{y } a,b \in B \quad (5.3)$$

En la práctica, se selecciona un conjunto finito de funciones base y se permite la existencia de todos los productos separables de estas funciones en 1-D dentro del diccionario 2-D.

Además, en este trabajo se selecciona al conjunto de escalas diádicas $s = 2^j$, donde $j \in \{0, 1, \dots, 7\}$, por tres razones fundamentales:

1. Considerando el empleo de "High Resolution Pursuit" con este modelo, y dado que este algoritmo depende directamente del factor de escala, se sugiere el manejo de escalas diádicas por razones de cómputo binario.
2. Los estudios realizados con funciones de Gabor muestran que basta con emplear tan sólo cuatro escalas o canales para obtener buenos resultados [46, 48, 49]. De modo que con ocho escalas se abarca adecuadamente el espacio de Fourier. Esto se puede apreciar en la Figura 5.1.
3. Finalmente, las pruebas preliminares realizadas con los algoritmos de selección adaptable presentaron buenos resultados empleando tan solo este rango de escalas.

Cabe mencionar que cada función 1-D posee un tamaño asociado N , el cual permite el uso de grandes funciones. Además, este tamaño asociado permite tener un mejor almacenamiento e incrementa la velocidad de la búsqueda. El tamaño para cada elemento de la base 1-D se determina imponiendo un umbral de intensidad en la ventana Gaussiana $g(t)$. Los mayores valores de escala, generalmente se trasladan a tamaños más grandes. En la Tabla 5.1 se muestran los tamaños asociados a cada escala.

Cabe hacer notar que en el contexto de los seguimientos adaptables, el diccionario que se muestra en la Figura 5.2, representa una base redundante para la imagen en estudio. Este conjunto consiste del posicionamiento de todas y cada una las formas del diccionario sobre cada uno de los píxeles de la imagen. Se puede apreciar el detalle de una de estas formas en la Figura 5.3.

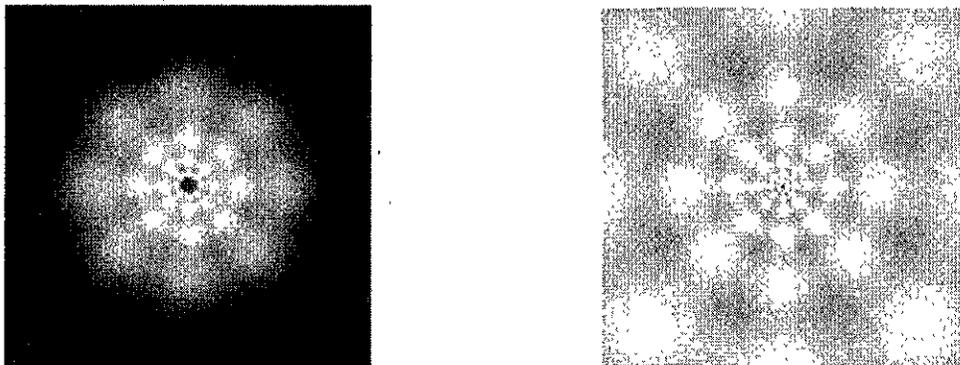


Figura 5.1 Espacio de Fourier a 4 y 8 escalas.

Ahora bien, para demostrar que se trata de una base redundante, basta considerar al subconjunto que contiene sólo al elemento de 1×1 pixel en la esquina superior izquierda. Como este elemento se puede colocar en cualquier posición de la imagen, él por sí sólo forma la base estándar de la imagen. Empleando las 64 formas de la Figura 5.2, tenemos un conjunto de bases mucho más grande. Para entender el tamaño del diccionario, se considera una imagen de 256×256 pixeles, la cual contiene 65,536 posiciones, y si en cada posición se puede colocar una de los 64 formas del diccionario, se tiene un total de 4,194,304 elementos base. A todo esto, se puede observar la existencia de un intercambio, el cual dice que el costo de representar cada elemento se incrementa con el tamaño del conjunto base. Cabe mencionar que los experimentos realizados al respecto muestran que este intercambio es extremadamente favorable en tasas de bits bajas [3].

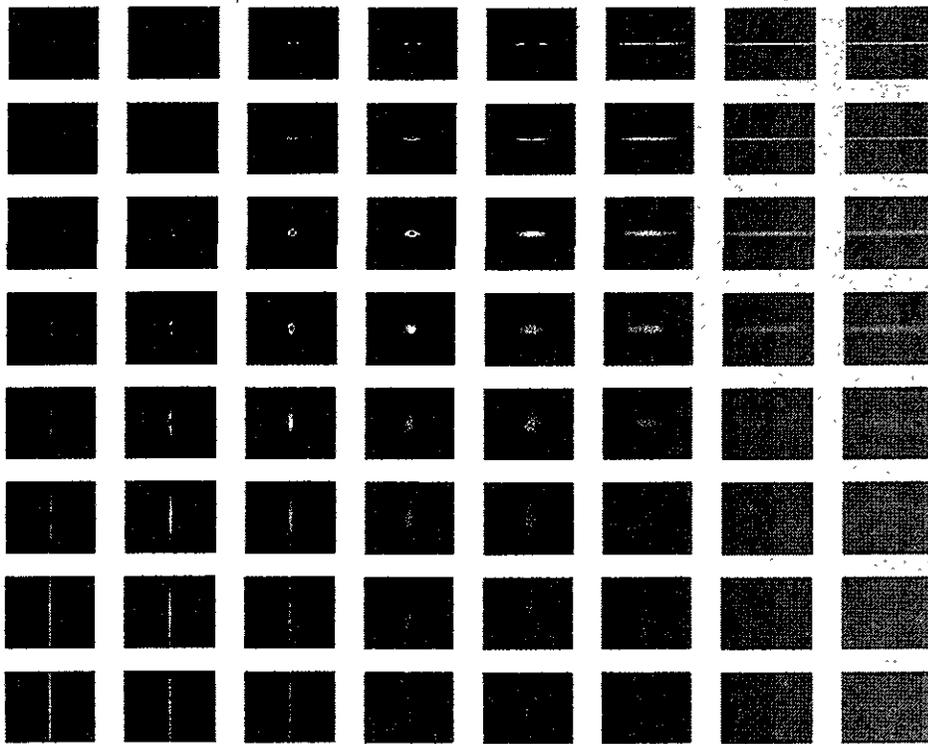


Figura 5.2 Diccionario de funciones de Gabor empleado.

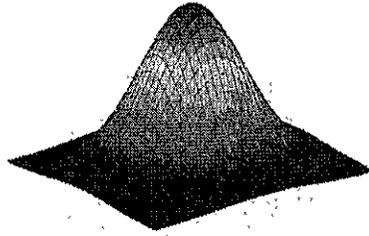


Figura 5.3 Detalle de un átomo de Gabor.

	Escala s	Tamaño N
1	1	1
2	2	3
3	4	5
4	8	9
5	16	17
6	32	33
7	64	65
8	128	129

Tabla 5.1 Tamaños de cada átomo asociados a la escala.

5.2.2 La búsqueda de los átomos.

En 5.2.1 se definieron las estructuras del diccionario 2-D que se utilizarán para descomponer la imagen con las técnicas de selección adaptable. La aplicación directa de estas técnicas requiere de examinar cada una de las estructuras del diccionario 2-D en todas las posibles localidades de la imagen y calcular todos los valores de la medida de similitud. Para reducir esta búsqueda a un nivel más manejable, se considera la técnica empleada por Ralph Neff *et al.* [1, 2, 3, 5] en la codificación de imágenes residuales.

El procedimiento de la búsqueda se puede observar en la Figura 5.4. La imagen se divide primero en bloques o regiones, a continuación se calcula la norma de cada una de estas regiones como se ilustra en la Figura 5.4(a). A este procedimiento se le llama "búsqueda de la energía". El centro del bloque con mayor valor de energía, visto en la Figura 5.4(b), es adoptado como un estimado inicial para la búsqueda de los valores de la medida de similitud. Enseguida, el diccionario se emplea exhaustivamente en una ventana $S \times S$ alrededor del estimado inicial, como se ve en la Figura 5.4(c). En la práctica, se emplean bloques de 8×8 para la búsqueda de la energía [7], y la ventana de búsqueda utilizada tiene un valor de $S = 24$ para "Matching Pursuit" y de $S = 16$ para "High

Resolution Pursuit". Los valores de S fueron determinados con base a las pruebas realizadas durante el desarrollo del trabajo, enfatizando la velocidad de búsqueda.

La búsqueda exhaustiva se puede ver de la siguiente manera. Cada estructura de tamaño $N_1 \times N_2$ del diccionario se centra en cada localidad de la ventana de búsqueda, y se calcula la medida de similitud entre la estructura y la región correspondiente de la imagen con tamaño $N_1 \times N_2$. Por fortuna, parte de esta búsqueda se puede efectuar rápidamente empleado el cálculo de productos internos separables.

El valor de la mayor medida de similitud, junto con la estructura del diccionario correspondiente y la posición en la imagen forman un conjunto de cinco parámetros, los cuales se muestran en la Tabla 5.2. Se puede decir que estos cinco parámetros definen un átomo, es decir, una estructura codificada dentro de la imagen.

a, b	Estructura de elemento del diccionario con mejor similitud
x, y	Posición del elemento con mejor similitud en la imagen.
P	Valor de la mayor medida de similitud; proyección del dato de la imagen en (x, y) dentro de $G_{a,b}(i, j)$.

Tabla 5.2 Parámetros que definen un átomo.

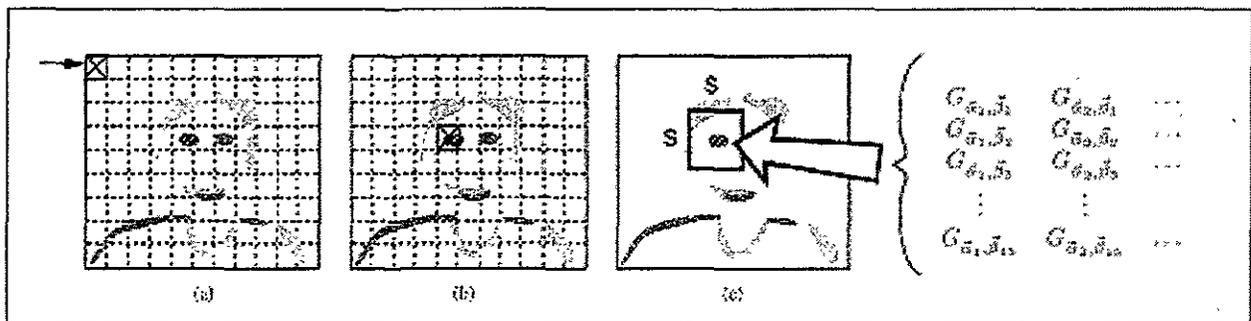


Figura 5.4 Procedimiento de búsqueda del producto interno.

(a) Búsqueda de la energía. (b) Ventana con mayor energía.

(c) Búsqueda exhaustiva del producto interno en la ventana de $S \times S$.

5.2.3 Búsqueda rápida del producto interno.

La naturaleza separable de las estructuras del diccionario se puede emplear para reducir la búsqueda del producto interno de manera significativa. Recordemos que en cada etapa en la cual el algoritmo encuentra un átomo se requiere calcular el producto interno entre cada forma del diccionario 2-D y la región de la imagen correspondiente en cada

localidad de la ventana de búsqueda de tamaño $S \times S$; y en el caso de "High Resolution Pursuit", además, se debe calcular el producto interno entre cada forma de onda y cada uno de los elementos de la subfamilia asociada.

Considerando solamente el caso de "Matching Pursuit", para calcular el número de operaciones necesarias para encontrar un átomo, es necesario introducir primero una notación que describa las funciones base y sus tamaños asociados. Suponiendo que h y v son índices escalares en la Tabla 5.1 que representan los componentes 1-D de una función base separable 2-D. Específicamente, h y v son los índices que corresponden a a y b en la ecuación (5.3). El conjunto del diccionario puede escribirse de manera compacta como:

$$G_{h,v}(i,j) = g_h(i)g_v(j) \quad i \in \{0, \dots, L_h - 1\}, j \in \{0, \dots, L_v - 1\} \text{ y } h, v \in \{0, \dots, B - 1\} \quad (5.4)$$

donde B es el número de elementos base 1-D empleados para generar al conjunto 2-D, y L_h y L_v son los tamaños asociados a las bases de la componente horizontal y vertical, respectivamente.

Suponiendo que se ignora la separabilidad. En este caso, el producto interno entre un elemento de base dado de 2-D y la región de la imagen correspondiente se puede escribir como

$$p = \sum_{i=0}^{L_h-1} \sum_{j=0}^{L_v-1} G_{h,v}(i,j)I(i,j) \quad (5.5)$$

El cálculo de p requiere de $L_h L_v$ operaciones acumuladas de multiplicación. Para encontrar un sólo átomo se requiere que el producto interno se calcule para cada una de las combinaciones de h y v en cada localidad de la ventana de búsqueda $S \times S$. El número total de operaciones sería entonces

$$T_{no-sep} = M^2 \sum_{h=0}^{B-1} \sum_{v=0}^{B-1} L_h L_v \quad (5.6)$$

Empleando los parámetros de la Tabla 5.1, con la ventana de búsqueda con $S = 24$, se tiene un total de 39,538,944 operaciones de multiplicación acumuladas.

Asumiendo un diccionario separable, el producto interno se puede escribir como

$$p = \sum_{i=0}^{L_h-1} \sum_{j=0}^{L_v-1} g_h(i)g_v(j)I(i,j) = \sum_{i=0}^{L_h-1} g_h(i) \sum_{j=0}^{L_v-1} g_v(j)I(i,j) \quad (5.7)$$

El cálculo de un solo producto interno 2-D es equivalente a tomar L_h productos internos verticales 1-D, cada uno de longitud L_v , y siguiendo con un simple producto interno horizontal de longitud L_h . La búsqueda del átomo requiere calcular exhaustivamente el producto interno de cada posición empleando todas las combinaciones

de h y v . Es muy natural el calcular previamente los productos internos verticales 1-D necesarios con un g_v particular, e ir de manera cíclica a través de los productos internos horizontales 1-D empleando todas las g_h posibles. Más aún, los resultados del prefiltrado vertical 1-D son aplicables también a los productos internos en las localidades adjuntas en la ventana de búsqueda. Esto motiva al uso de una "matriz de prefiltrado" grande para almacenar todos los resultados del filtrado vertical 1-D de una g_v en particular. El número total de operaciones para calcular un solo átomo será:

$$T_{sep} = \sum_{v=0}^{B-1} (L_v M (M + L_{max}) + M^2 \sum_{h=0}^{B-1} L_h) \quad (5.8)$$

Nótese que L_{max} es el tamaño de la función de base 1-D más grande. Empleando los valores de la Tabla 5.1, se requieren 2,169,360 operaciones de multiplicación acumuladas para encontrar cada átomo. Esto produce un factor de mejoramiento en la velocidad de alrededor de 18 sobre el caso general no separable.

Para el caso de "High Resolution Pursuit", a los valores obtenidos hay que multiplicarlos por un factor F , donde F representa el tamaño de la subfamilia asociada a cada átomo. Para el diccionario empleado en este trabajo, el valor de F sería en promedio cinco con un desplazamiento de $K = 2$ en la escala, y siete para $K = 3$.

5.2.4 Etapa de cuantización.

Como solamente los valores de los coeficientes cuantizados estarán disponibles para la reconstrucción, el uso de valores cuantizados durante el desarrollo del algoritmo produce globalmente un error de cuantización menor con respecto a sí se efectuara la cuantización al final del proceso. Así, un algoritmo de selección adaptable con cuantización es un algoritmo en el cual cada coeficiente es cuantizado antes de obtener el siguiente residual.

Recordando la manera en la cual estos algoritmos en cada iteración actualizan el residual se tiene

$$R^{n+1}f = R^n f - C(R^n f, g_{\gamma^n}) g_{\gamma^n} \quad (5.9)$$

Considerando que $a = C(R^n f, g_{\gamma^n})$, y definiendo la función de cuantización $Q(a) = \hat{a}$. Se tiene ahora que

$$R^{n+1}f = R^n f - \hat{a} g_{\gamma^n} \quad (5.9)$$

Explícitamente,

$$Q(a) = \text{Round}(a/val)*val \quad (5.10)$$

Donde *val* es el valor del paso de cuantización y *Round* es la función que aproxima al entero más próximo el resultado de la operación efectuada.

En este trabajo se emplea $val \in \{ 1, 16, 32, 64 \}$, para el análisis de los resultados de la codificación.

5.2.5 Codificación de los parámetros.

Los átomos, explícitamente los cinco parámetros de cada uno de ellos, obtenidos durante el proceso de selección son agrupados y ordenados en primer lugar de acuerdo a las estructuras que los constituyen (*a, b*) y a continuación, de acuerdo a su posición (*x, y*).

Para cada par de estructuras (*a, b*) se conserva sólo la primera posición (*x₀, y₀*) de ésta en la imagen y las siguientes posiciones son trasladadas a un parámetro relativo *d* que representa el desplazamiento de dicha estructura sobre la imagen. Los valores de los coeficientes se conservan tal cual debido a que éstos varían de acuerdo al paso de cuantización que se utilice.

Por ejemplo, si se obtienen 6,000 coeficientes en el proceso, se tendrían 30,000 parámetros a codificar. Ahora bien, considerando que se tienen 64 estructuras, representadas por los 64 pares (*a, b*), se tendrán también 64 pares (*x₀, y₀*) de posiciones iniciales, 5,936 desplazamientos *d* y los 6,000 coeficientes *p* de los átomos, lo cual da un total de 12,188 parámetros a transmitir o almacenar, es decir, solamente el 40% del total de los parámetros obtenidos en primera instancia.

Trasladando los parámetros de las estructuras y de posición obtenidos a bits, con este arreglo se reduce el número de bits necesarios para la representación de la imagen de 122,000 a sólo 12,188 bits, es decir, a un 10% del total inicial, conservándose sin cambio los coeficientes de los átomos.

En la Tabla 5.3 se muestra la secuencia que se obtiene para almacenar los coeficientes.

(a_0, b_0)	(x_0, y_0)	p_0	d_1	p_1	d_2	p_2	...	d_{m-1}	p_{m-1}
(a_1, b_1)	(x_1, y_1)	p_m	d_m	p_{m+1}	d_{m+1}	p_{m+2}	...	d_{n-1}	p_{n-1}
...									
(a_7, b_7)	(x_7, y_7)	p_s	d_s	p_{s+1}	d_{s+1}	p_{s+2}	...	d_{t-1}	p_{t-1}

Tabla 5.3 Secuencia de almacenamiento de los parámetros de los átomos.

La secuencia resultante se codifica empleando el algoritmo universal para compresión secuencial de datos [50].

5.3 Modelo de codificación piramidal.

En este modelo, se expande la imagen empleando la clásica transformada ondeleta 2-D [51], la cual se puede observar en la Figura 5.5. Con base en los estudios, las pruebas efectuadas y la experiencia personal, se opta por el uso de ondeletas "Spline", explícitamente, ondeletas biortogonales "Spline" de reconstrucción perfecta [52]. Principalmente porque estas funciones son una derivación de las ondeletas de Gabor, que en cuestiones de visión, por sus características, son las que mejores resultados producen [47].

Existen diferentes tipos de funciones "Spline", por lo cual, antes de iniciar el algoritmo se obtiene una medida de costo, definida por el método adaptable de selección de bases [52]. Esta medida de costo se utiliza para seleccionar aquella función que mejor se adapte a la imagen.

A diferencia del método por regiones, en este método se efectúa una búsqueda exhaustiva a lo largo de toda la descomposición. Haciendo una similitud con el paquete de ondeletas, esta transformada se puede considerar un paquete de ondeletas reducido, el cual contiene los elementos suficientes para efectuar una reconstrucción perfecta [52], por lo que no es necesario el uso de un diccionario de mayor amplitud.

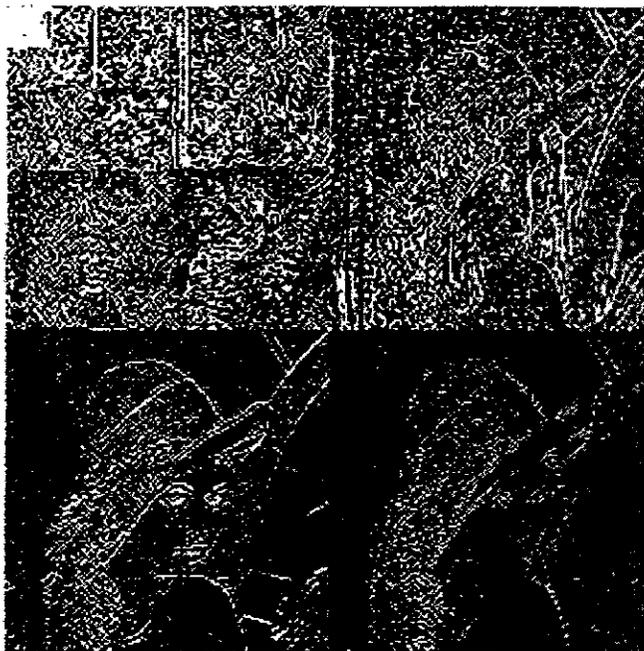


Figura 5.5 Típica transformada ondeleta 2-D.

5.3.1 Selección adaptable de la base.

Para optimizar las aproximaciones no lineales de las bases, se puede elegir de manera adaptable la base dependiendo de las características estadísticas de la señal. Esto se efectúa minimizando una función de costos cóncava [52]. Por ejemplo, las bases de paquetes de ondeletas y de cosenos locales son familias grandes de bases ortogonales que incluyen diferentes tipos de átomos locales de tiempo-frecuencia. La mejor base paquete de ondeletas o la mejor base coseno local descompone la señal en átomos de tiempo-frecuencia que mejor se adaptan a las estructuras de tiempo-frecuencia de la señal.

5.3.1.1 La mejor base y la concavidad de Schur.

Considerando que un diccionario D es la unión de bases ortonormales en el espacio de una señal de dimensión finita N :

$$D = \bigcup_{\lambda \in \Lambda} B^\lambda.$$

Cada base ortonormal es una familia de N vectores.

$$B^\lambda = \{g_m^\lambda\}_{1 \leq m \leq N}.$$

Los paquetes de ondeletas y los árboles de cosenos locales son ejemplos de diccionarios donde las bases comparten algunos vectores en común.

Si se desea optimizar una aproximación no lineal de f eligiendo la mejor base en D . Se dice que I_M^λ es el índice del conjunto de M vectores de B^λ que maximiza $\langle f, g_m^\lambda \rangle$. La mejor aproximación no lineal de f en B^λ es

$$f_M^\lambda = \sum_{m \in I_M^\lambda} \langle f, g_m^\lambda \rangle g_m^\lambda.$$

La aproximación del error es

$$\varepsilon^\lambda[M] = \sum_{m \in I_M^\lambda} \left| \langle f, g_m^\lambda \rangle \right|^2 = \|f\|^2 - \sum_{m \in I_M^\lambda} \left| \langle f, g_m^\lambda \rangle \right|^2. \quad (5.11)$$

Definición: Se dice que $B^\alpha = \{g_m^\alpha\}_{1 \leq m \leq N}$ es mejor base que $B^\gamma = \{g_m^\gamma\}_{1 \leq m \leq N}$ al aproximar a f sí para toda $M \geq 1$

$$\varepsilon^\alpha[M] \leq \varepsilon^\gamma[M]. \quad (5.12)$$

Esta comparación de bases es una relación de orden parcial entre las bases de D . Ni B^α , ni B^γ es mejor si existen M_0 y M_1 , tal que

$$\varepsilon^\alpha[M_0] < \varepsilon^\gamma[M_0] \quad \text{y} \quad \varepsilon^\alpha[M_1] < \varepsilon^\gamma[M_1] \quad (5.13)$$

Insertando (5.11) se puede probar que la mejor condición de base (5.12) equivale a:

$$\forall M \geq 1, \quad \sum_{m \in I_M^\alpha} \left| \langle f, g_m^\alpha \rangle \right|^2 \geq \sum_{m \in I_M^\gamma} \left| \langle f, g_m^\gamma \rangle \right|^2 \quad (5.14)$$

El siguiente teorema se deriva del criterio basado en las funciones cóncavas de Schur.

Teorema: Una base B^α es mejor que una base B^γ para aproximarse a f sí y sólo sí para todas las funciones cóncavas $\Phi(u)$

$$\sum_{m=1}^N \Phi \left(\frac{\left| \langle f, g_m^\alpha \rangle \right|^2}{\|f\|^2} \right) \leq \sum_{m=1}^N \Phi \left(\frac{\left| \langle f, g_m^\gamma \rangle \right|^2}{\|f\|^2} \right) \quad (5.15)$$

La prueba de este teorema se basa en el resultado clásico de la teoría de la mayorización [53].

En la práctica, dos bases se comparan empleando una simple función cóncava $\Phi(u)$. El costo de aproximar f en una base B^λ está definido por la suma cóncava de Schur.

$$C(f, B^\lambda) = \sum_{m=1}^N \Phi \left(\frac{\left| \langle f, g_m^\lambda \rangle \right|^2}{\|f\|^2} \right).$$

El teorema anterior demuestra que si B^α es una mejor base que B^γ para aproximar a f , entonces

$$C(f, B^\alpha) \leq C(f, B^\gamma) \quad (5.16)$$

Esta condición es necesaria pero no suficiente para garantizar que B^α es mejor que B^γ , dado que se ha probado con una función cóncava simple. Coifman y Wickerhauser [54] encontraron una *mejor base* B^α en D minimizando el costo de f .

$$C(f, B^\alpha) = \min_{\lambda \in \Lambda} C(f, B^\lambda).$$

No existe una mejor base en D que aproxime a f . Sin embargo, existen varias bases en D que son equivalentes en el sentido de (5.12). En este caso, la elección de la mejor base depende de la función cóncava Φ en particular.

A continuación se presentan algunos ejemplos de funciones de costo cóncavas, en la práctica se emplea la función de costo de entropía, la cual se eligió de manera arbitraria.

Función de costo de entropía: La función de costo de entropía $\Phi(x) = -x \log x$ es cóncava para $x \geq 0$. Su costo correspondiente es llamado entropía de la distribución de la energía.

$$C(f, B) = -\sum_{m=1}^N \frac{\langle f, g_m \rangle^2}{\|f\|^2} \log_e \frac{\langle f, g_m \rangle^2}{\|f\|^2} \quad (5.17)$$

Donde

$$0 \leq C(f, B) \leq \log_e N \quad [21]. \quad (5.18)$$

Cabe hacer notar que esta entropía, a priori, no está relacionada con el número de bits requeridos para codificar los productos internos $\langle f, g_m \rangle$. El Teorema de Shanon muestra que la cota inferior del número de bits para codificar individualmente cada $\langle f, g_m \rangle$ es la entropía de la distribución de probabilidad de los valores tomados por $\langle f, g_m \rangle$. Esta distribución de probabilidad puede ser muy diferente a la distribución de las energías normalizadas $\frac{\langle f, g_m \rangle^2}{\|f\|^2}$. Por ejemplo, si $\langle f, g_m \rangle = A$ para $0 \leq m < N$, entonces $\frac{\langle f, g_m \rangle^2}{\|f\|^2} = \frac{1}{N}$ y el costo $C(f, B) = \log_e N$ es máximo. Mientras que la distribución de probabilidad del producto interno es una Dirac discreta localizada en A y su entropía es mínima e igual a cero.

Función de costo I^p : Para $p < 2$, la función $\Phi(x) = x^{p/2}$ es cóncava para $x \geq 0$. El costo resultante es

$$C(f, B) = \sum_{m=1}^N \frac{\langle f, g_m \rangle^p}{\|f\|^p} \quad (5.19)$$

Donde

$$1 \leq C(f, B) \leq N^{1-\frac{p}{2}} \quad [52]$$

Este costo mide la norma I^p de los coeficientes de f en B :

$$C^{\frac{1}{p}}(f, B) = \frac{\|f\|_{B,p}}{\|f\|}.$$

La cual se deriva de

$$\varepsilon[M] \leq \frac{\|f\|_{B,p}^2}{\frac{2}{p} - 1} M^{1-\frac{2}{p}}$$

donde el error de aproximación $\epsilon[M]$ está acotado por

$$\epsilon[M] \leq \frac{\|f\|^2 C^{\frac{2}{p}}(f, B)}{\frac{2}{p} - 1} \frac{1}{M^{\frac{2}{p} - 1}}$$

La minimización del costo I^p puede interpretarse como una reducción en el factor del decremento C tal que

$$\epsilon[M] \leq \frac{C}{M^{\frac{2}{p} - 1}}.$$

5.3.2 Búsqueda exhaustiva.

Una vez seleccionada la base a emplear, se realiza la expansión de la imagen empleando la transformada ondeleta 2-D con dicha base.

A continuación se inicia el algoritmo de selección adaptable de manera exhaustiva, es decir, se realiza el seguimiento ó la búsqueda sobre cada uno de los coeficientes obtenidos de la expansión, los cuales para una imagen de 256×256 pixeles son 65,536 dispuestos en cuatro escalas.

En cada paso, una vez seleccionado el coeficiente con mejor adaptación, se efectúa el proceso de cuantización de dicho coeficiente, ya que éste se empleará para actualizar la imagen residual y de esta manera se reducirá el error total en la reconstrucción a causa de dicha cuantización.

Al igual que en el modelo anterior, se emplea la función de cuantización (5.10), con el conjunto de valores del paso de cuantización $val \in \{ 1, 16, 32, 64 \}$.

Cuando se emplea "Matching Pursuit", la búsqueda exhaustiva no presenta complejidad computacional alguna ya que ésta se realiza tan solo sobre los 65,536 coeficientes obtenidos de la descomposición cuando estamos hablando de una imagen de 256×256 pixeles. Sin embargo, cuando se emplea "High Resolution Pursuit", esta búsqueda se efectúa, para la misma imagen, sobre casi cuatro millones de coeficientes considerando un factor de desplazamiento $K = 2$, a lo cual se debe agregar el cálculo de los productos internos entre las subfamilias de cada coeficiente en cada iteración, teniéndose en promedio veinte millones de coeficientes, los cuales incrementan en gran medida el tiempo de búsqueda de cada átomo.

Si el cálculo de productos internos entre familias se efectuara en una etapa inicial para reducir el número de cálculos en cada iteración, tal y como se realiza en el modelo por

regiones, su almacenamiento requeriría de una gran cantidad de memoria lo cual también decrementa la velocidad de búsqueda.

Con estas consideraciones, y con base en las siguientes observaciones, con este modelo solamente se realizó la codificación de imágenes empleando "Matching Pursuit":

1. En las pruebas preliminares, bajo las mismas condiciones, el cálculo de un átomo con el modelo por regiones necesitaba de al menos doscientas veces menos tiempo que con el modelo piramidal. Por ejemplo, el obtener seis mil átomos con el modelo por regiones requiere de ocho horas, mientras en ese mismo tiempo el modelo piramidal apenas obtiene trescientos átomos.
2. Una computadora con 512 Mbytes de memoria, dedicada solamente a la tarea de obtener los productos internos de las subfamilias y retenerlos en memoria, no fue suficiente para realizar dicha tarea. Una opción alternativa es su almacenamiento en disco duro, sin embargo, esta alternativa no se presenta como una buena opción ya que el acceso a disco duro incrementa también el tiempo de cálculo.
3. Las imágenes en general, poseen características muy diversas, y se puede observar que principalmente los contornos que la forman, no así las texturas, suelen parecer más estructuras globales que locales, por lo cual, la esperanza de obtener mejores resultados, tanto en la tasa de compresión como en la calidad de la imagen reconstruida no puede ser significativa.

En resumen, el uso de este modelo con "High Resolution Pursuit" resulta no operacional, ó poco práctico.

Sin embargo, a pesar de estos inconvenientes para su uso con "High Resolution Pursuit", este modelo tiene un par de ventajas sobre el modelo de regiones.

En primer lugar, necesita solamente del valor de la mayor medida de similitud o del coeficiente del átomo encontrado, junto con su posición en la imagen expandida para efectuar la reconstrucción. Estos valores forman un conjunto de tan solo tres parámetros, los cuales se muestran en la Tabla 5.4, a diferencia del conjunto de cinco parámetros que se obtienen cuando se emplea el modelo por regiones. Esto implica una mejora en la tasa de compresión.

Se puede decir que estos tres parámetros definen un átomo, es decir, una estructura codificada dentro de la imagen, donde la posición del coeficiente en la imagen expandida proporciona también la información de la escala y de la frecuencia del átomo.

En segundo lugar, como se observará cuando se analicen los resultados obtenidos con cada uno de los modelos propuestos, el modelo piramidal proporciona los mejores resultados tanto en la escala de calidad PQS como en los criterios objetivos en las imágenes reconstruidas.

x, y	Posición del elemento con mejor similitud en la imagen.
p	Valor de la mayor medida de similitud; proyección del dato de la imagen en (x, y) dentro de $G_{a,b}(i, j)$.

Tabla 5.4 Parámetros que definen un átomo.

5.3.3 Codificación de los parámetros.

En esta última etapa, para realizar una codificación efectiva, los coeficientes p de los átomos obtenidos se colocan de acuerdo a sus parámetros de posición (x, y) sobre una matriz de ceros del tamaño de la imagen, ya que las posiciones no se repiten.

A esta matriz simplemente se le aplica un codificador de entropía como es el tradicional codificador de Huffman [55].

Referencias.

- [1] R. Neff, A. Zakhor y M. Vetterli. Very low bit rate video coding using matching pursuits, *Proceedings of SPIE VCIP*, Vol. 2308, No. 1, pp. 47-60, Septiembre 1994.
- [2] R. Neff y A. Zakhor. Very low bit rate video coding based on matching pursuits, *IEEE Trans. on CSVT*, Vol. 7, No. 5, pp. 158-171, Octubre 1997.
- [3] R. Neff y A. Zakhor. Matching pursuit video coding at very low bit rates, *IEEE Data Compression Conference, Snowbird, Utha*, pp. 411-420, Marzo 1995.
- [4] M. Vetterli y Ton Kalker. Matching Pursuit for compression and application to motion compensated video coding, *Proceedings of ICIP*, pp. 725-729, Noviembre 1994.
- [5] O. Al-Shaykh, E. Miloslavsky, T. Nomura, R. Neff y A. Zakhor. Video compression using matching pursuits, *Reporte Técnico*, Departamento de Ingeniería Eléctrica y Ciencias de la Computación, Universidad de California en Berkeley.
- [6] M. R. Banham y J. C. Brailean. A selective update approach to matching pursuit video coding, *IEEE Trans. on circuits and systems for video technology*, Vol. 7, No. 1, pp. 119-129, Febrero 1997.
- [7] S. S. Cheung y A. Zakhor. Matching pursuit experimental video codec, *Reporte Técnico*, Lab. de procesamiento de imágenes y vídeo, Universidad de California en Berkeley.
- [8] R. Gribonval, E. Bacry, S. Mallat, P. Depalle y X. Rodet. Analysis of sound signals with high resolution matching pursuit, *Proceedings of TFTS'96*, pp. 125-128, Junio 1996.
- [9] R. Gribonval, E. Bacry, S. Mallat, P. Depalle y X. Rodet. Sound signals decomposition using a high resolution matching pursuit, *Proceedings of ICMC'96*, pp. 293-296, Agosto 1996.
- [10] R. Gribonval. Algorithme de matching pursuit applique à l'analyse temps-frequence des signaux sonores, *Reporte DEA*, 1995.

- [11] S. Chen y D. Donoho. Examples of basis pursuit, *Proceedings of SPIE*, Vol. 2569, pp. 564-574.
- [12] S. Chen. Basis pursuit, *Tesis Doctoral*, Departamento de estadística, Universidad de Stanford, 1995.
- [13] S. Jaggi, W. C. Carl, S. Mallat y A. S. Willsky, High Resolution Pursuit for Feature Extraction, *Reporte Técnico, MIT*, Noviembre 1995.
- [14] S. Jaggi, Multiscale geometric feature extraction and object recognition, *Tesis Doctoral*, LIDS-TH-2384, Febrero 1997.
- [15] R. Montúfar-Chaveznava y F. García-Ugalde, Image analysis by Basis Pursuit, *SPIE*, Vol. 3164, pp. 642-649, 1997.
- [16] P. J. Burt y E. H. Adelson. The Laplacian pyramid as a compact image code, *IEEE Trans. in Comm.*, VOL. 31, No. 4, pp. 532-540, Abril 1983.
- [17] S. Mallat y S. Zhong. Charactersitics of signals from multiscale images. *Reporte Técnico*, NYU Ciencias de la Computación, 1991.
- [18] J. Froment y S. Mallat. Second generation compact image coding with wavelets. *Wavelet analysis and its applications*, Academic Press, 2, 1992.
- [19] L. Prasad y S. S. Iyengar. *Wavelet analysis with applications to image processing*, CRC Press, 1997.
- [20] R. Montúfar. Codificación de imágenes por medio de un modelo contornos-textura. *Tesis de Maestría*. DEPFI, UNAM, 1995.
- [21] J. E. Odegard, R. A. Gopinath y C. S. Burrus, Design of linear phase modulated filter banks for subband image compression, *Proc. of ICIP, Austin, Texas*, 1994.
- [22] T. Strutz y E. Müller, Wavelet filter design for image compression, *Proceedings of TFTS'96*, pp. 273-276, 1996.
- [23] J. M. Shapiro, Embedded image coding using zerotrees of wavelet coefficients, *IEEE Trans. on signal processing*, Vol. 41, No. 12, pp. 3445-3462, Diciembre 1993.
- [24] A. Said y W. A. Pearlman, A new, fast, and efficient image codec based on set partitioning in hierarchical trees, *IEEE Trans. on circuits and systems for video technology*, Vol. 6, No. 3, pp. 243-250, Junio 1996.
- [25] H. R. Rabiee, R. L. Kashiap y S. R. Safavian, Multiresolution segmentation-based image coding with hierarchical data structures, *Ref. IEEE 0-7803-3192-3/96*, pp. 1870-1873, 1996.
- [26] E. H. Adelson y E. P. Simonelli, Subband image coding with three-tap pyramids, *Picture coding Symposium, Cambridge MA*, 1990.
- [27] C. M. Brislawn, J. N. Bradley, R. J. Onyshczak y T. Hoper, The FBI compression standard for digitized fingerprint images, *Proc. of SPIE*, Vol. 2847, pp. 344-355, 1996.
- [28] Z. Xiong, K. Ramchandran y M. T. Orchard, Space-frequency methods in image compression, *Proc. of SPIE*, Vol. 2847, pp. 332-343, 1996.
- [29] Z. Xiong, K. Ramchandran y M. T. Orchard, Space-frequency quantization for wavelet image coding, *Proc. of SPIE*, Vol. 2847, pp. 388-399, 1996.
- [30] J. R. Smith y S. Chang, Frequency and spatially adaptive wavelet packets, *Proc. of IEEE ICASP, Detroit, Mi.*, 1995.
- [31] M. Gharavi-Alkhansari, A model for entropy coding in matching pursuit, *Ref. IEEE 0-8186-8821-1/98*, pp. 778-782, 1998.

- [32] A. Ortega y K. Ramchandran, Rate-distortion methods for image and video compression, *IEEE Signal Processing Magazine*, pp. 23-50, Noviembre 1998.
- [33] N. Jayant, J. Johnston y R. Safranek, Signal compression based on models of human perception, *Proceedings of the IEEE*, Vol. 81, No. 10, pp. 1385-1422, Octubre 1993.
- [34] H. Wang y H. Yan, Efficient implementation of Gabor transforms for image compression, *Electronic letters*, Vol. 28, No. 9, 23 de Abril de 1992.
- [35] H. Wang y H. Yan, Adaptive Gabor discrete cosine transforms for image compression, *Electronic letters*, Vol. 28, No. 18, 27 de Agosto de 1992.
- [36] J. G. Daugman, Complete discrete 2-D Gabor transforms by neural networks for image analysis and compression, *IEEE Trans. on acoustics, speech and signal processing*, Vol. 36, No. 7, pp. 1169-1179, Julio 1988.
- [37] A. R. Calderbank, I. Daubechies, W. Sweldens y B. Yeo, Lossless image compression using integer to integer wavelet transforms, *Reporte Técnico*, AT&T Labs Research, 1996.
- [38] J. E. Odegard y C. S. Burrus, Smooth biorthogonal wavelets for applications in image compression, *Proc. of DSP Workshop*, Loen, Noruega, 1996.
- [39] R. Devore, B. Jawerth y B. J. Lucier, Image compression through wavelet transform coding, *IEEE Trans. on Information Theory*, Vol. 38, pp. 719-746, Marzo 1992.
- [40] F. G. Meyer y R. R. Coifman, Directional image compression with brushlets, *IEEE Proc. of TFTS '96*, pp. 189-192, 1996.
- [41] Q. Liu, A. K. Chan y C. K. Chui, Hybrid technique using spline-wavelet packets and vector quantization for high-rate image compression, *Proc. of SPIE*, Vol. 2034, pp. 194-204, 1993.
- [42] G. Aharoni y A. Averbuch, Local cosine transform, *Proc. of SPIE*, Vol 2034, pp. 205-217, 1993.
- [43] K. R. Rao y P. Yip, *Discrete cosine transform*, Academic Press, Nueva York, 1990.
- [44] H. S. Malvar. The lapped transforms for efficient transform/subband coding, *IEEE Trans. on Acoustic, speech and signal processing*, Vol. 38, No. 6, pp. 969-978, Junio 1990.
- [45] G. K. Wallace, The JPEG still picture compression standard, *Comm. of the ACM*, Vol. 34, No. 4, pp. 30-45, Abril 1991.
- [46] F. Bergeaud y S. Mallat, Processing images and sound with matching pursuits, *SPIE*, Vol. 2569, pp. 2-13, Septiembre 1995.
- [47] F. Bergeaud y S. Mallat, Matching pursuit of images, *IEEE Int. Symp. on TFTS Analysis*, pp. 330-333, 1994.
- [48] O. Nestares, R. Navarro, J. Portilla y A. Taberero, Efficient spatial-domain implementation of a multiscale image representation based on Gabor functions, *Comunicado personal (Enviado a Journal of Electronic Imaging, manuscrito revisado)*, Febrero 1997.
- [49] J. G. Daugman, Spatial visual channels in the Fourier plane, *Visual Res.*, No. 24, pp. 891-910, 1994.
- [50] J. Ziv, A universal algorithm for sequential data compression, *IEEE Trans. on Information Theory*, Vol. IT-23, No. 3, Mayo 1977.
- [51] S. Mallat, A theory for multiresolution signal decomposition: The wavelet representation, *IEEE Trans. On PAMI*, Vol. 11, No. 7, pp. 674-693, Julio 1989.
- [52] S. Mallat, *A wavelet tour of signal processing*, Academic Press, 1998.

- [53] A. W. Marshall e I. Olkin, *Inequalities: Theory of Majorization and its applications*, Academic Press, Boston, 1979.
- [54] R. R. Coifman y M. V. Wickerhauser. Entropy-based algorithms for best basis selection, *IEEE Trans. on Information Theory*, Vol. 38, No. 2, pp. 713-718, Marzo 1992.
- [55] D. A. Huffman, A method for the construction of minimum-redundancy codes, *Proc. IRE*, Vol. 40, pp. 1098-1101, 1952.

Codificación de imágenes.

En este capítulo se presentan los resultados obtenidos con los dos modelos de codificación propuestos. En la primera parte se presentan los resultados obtenidos al emplear los modelos de codificación sobre tres imágenes comunes, muy utilizadas por su alta calidad y sus características en pruebas de procesamiento de señales como son Lenna, Bárbara y Cameraman. En la segunda parte se presentan los resultados obtenidos al emplear los modelos con dos imágenes de microscopía electrónica, empleadas para reconstrucción tridimensional y comportamiento celular.

Además, se analizan los resultados obtenidos con cada una de las imágenes utilizadas. Los principales factores a analizar con cada uno de los modelos de codificación propuestos son los criterios de calidad en la reconstrucción, junto con la tasa de compresión, además de los efectos producidos por los diferentes valores empleados en el paso de cuantización.

El análisis se realiza para cada una de las imágenes de manera independiente, ya que éstas poseen características muy diversas y como consecuencia de estas características los resultados obtenidos varían. Por ejemplo, Bárbara posee una mayor cantidad de textura en su pantalón, su tocado y el mantel, mientras que en Lenna, la textura más notoria está en su cabello; por otro lado, la imagen de fluorescencia es la imagen del conjunto que tiene mucho menos energía que cualquier otra.

Tan diversas son las características entre cada una de las imágenes que esta diversidad se ve reflejada al momento de obtener los átomos, por ejemplo, el caso más notorio se encuentra en las imágenes con más estructuras y texturas, tales como Bárbara y el cuerpo laminar, ya que cuando se emplea un paso de cuantización de 64 con estas imágenes, el modelo obtiene los 6,000 átomos establecidos como criterio de paro en el algoritmo, mientras que para las otras imágenes, y más claramente con la imagen de fluorescencia, el algoritmo se detiene mucho antes de encontrar los 6,000 átomos, haciendo válido el segundo criterio de paro, debido a que llega a encontrarse un coeficiente de valor cero.

6.1. Experimentación, análisis y resultados.

Lenna, Bárbara y Cameraman son tres imágenes empleadas exhaustivamente como imágenes de prueba en varios laboratorios y en casi todas las tareas de procesamiento de imágenes que encontramos en la literatura. Las características de estas imágenes, utilizadas para probar los modelos propuestos, son las siguientes:

1. Tamaño: 256×256 píxeles.
2. Niveles de gris: 256.
3. Tamaño de cada pixel: 8 bits.
4. Tamaño total de la imagen: 524,288 bits.

Se emplean los siguientes modelos:

1. "Matching Pursuit" por regiones (MPR).
2. "Matching Pursuit" piramidal (MPP).
3. "High Resolution Pursuit" por regiones (HRPR).

Cada modelo emplea para la etapa intermedia de cuantización el conjunto de valores $val \in \{ 1, 16, 32, 64 \}$. La razón por la cual se ha elegido este conjunto de valores es debido a que al ser una potencia de dos, en términos globales, entre ellos se reducen los niveles de cuantización de manera efectiva, y además, la representación binaria de estos valores, en términos de eficiencia y ahorro, es utilizada en su total dimensión.

Los algoritmos detienen el proceso al cumplirse una de las siguientes condiciones:

1. Encontrar 6,000 átomos.
2. El valor del coeficiente encontrado es cero, dado que los siguientes valores encontrados serían cero también.

Encontrar 6,000 átomos es una decisión intuitiva, ya que se considera que con esa cantidad de elementos, además de obtener una reconstrucción de buena calidad, se pueden observar los efectos de cuantización. Cabe mencionar que en [1] se dice que con 2,500 ó 3,000 átomos se obtiene una reconstrucción con muy buena calidad visual.

El diccionario empleado en los modelos por regiones (MPR y HRPR) es el diccionario de funciones Gabor definido en 5.2.1, el cual se aprecia en la Figura 5.2.

El diccionario empleado en el modelo piramidal (MPP) es seleccionado mediante la función de costo de entropía presentada en 5.3.1.1. La selección se realiza sobre un conjunto de trece pares de filtros espejo en cuadratura (QMF y DQMF) cuyos coeficientes podemos observar en la Tabla 6.1. Los coeficientes de estos pares de filtros han sido extraídos de [2]. De esta manera, para cada imagen, de acuerdo al resultado de la función de costo obtenida con cada uno de los pares de filtros, se selecciona adaptablemente aquel cuyo costo sea el menor.

Finalmente, para cada una de las imágenes de prueba se presenta el análisis de los resultados obtenidos. Se puede observar que entre cada una de las imágenes existen algunas diferencias relacionadas con sus características particulares, tales como la textura, los contornos, la energía, las regiones, etcétera. De esta manera, a partir del análisis realizado se podrá elegir adecuadamente el modelo de codificación de acuerdo con los resultados obtenidos ó con el comportamiento observado en el decremento de la energía en la imagen durante el proceso.

#	Filtro QMF	Filtro DQMF
1	0 0.707106781186547524400844362104849 0.707106781186547524400844362104849	0 0.707106781186547524400844362104849 0.707106781186547524400844362104849
2	0 -0.0883883476483184405501055452631061 0.0883883476483184405501055452631061 0.707106781186547524400844362104849 0.707106781186547524400844362104849 0.0883883476483184405501055452631061 -0.0883883476483184405501055452631061	0 0.707106781186547524400844362104849 0.707106781186547524400844362104849
3	0 0.0165728151840597076031447897368324 -0.0165728151840597076031447897368324 -0.121533978016437855756395124736771 0.121533978016437855756395124736771 0.707106781186547524400844362104849 0.707106781186547524400844362104849 0.121533978016437855756395124736771 -0.121533978016437855756395124736771 -0.0165728151840597076031447897368324 0.0165728151840597076031447897368324	0 0.707106781186547524400844362104849 0.707106781186547524400844362104849
4	-0.176776695296636881100211090526212 0.353553390593273762200422181052425 1.06066017177982128660126654315727 0.353553390593273762200422181052425 -0.176776695296636881100211090526212	0.353553390593273762200422181052425 0.707106781186547524400844362104849 0.353553390593273762200422181052425
5	0.0331456303681194152062895794736648 -0.0662912607362388304125791589473296 -0.176776695296636881100211090526212 0.419844651329512592613001339999754 0.994368911043582456188687384209944 0.419844651329512592613001339999754 -0.176776695296636881100211090526212 -0.0662912607362388304125791589473296 0.0331456303681194152062895794736648	0.353553390593273762200422181052425 0.707106781186547524400844362104849 0.353553390593273762200422181052425

6	-0.00690533966002487816797699572368017 0.0138106793200497563359539914473603 0.0469563096881691715422435709210251 -0.107723298696388099420441133289411 -0.169871355636612002932234094802532 0.447466009969612105284909322894475 0.966747552403482943516779401315223 0.447466009969612105284909322894475 -0.169871355636612002932234094802532 -0.107723298696388099420441133289411 0.0469563096881691715422435709210251 0.0138106793200497563359539914473603 -0.00690533966002487816797699572368017	0.353553390593273762200422181052425 0.707106781186547524400844362104849 0.353553390593273762200422181052425
7	0.00151054305063044209924496781455504 -0.00302108610126088419848993562911007 -0.0129475118625466465649568669819003 0.0289161098263541773284036695929107 0.0529984818906909399392234421792453 -0.134913073607736057206850553951401 -0.163829183434090234535254223544312 0.462571440475916526277359001040025 0.951642121897178522524329723169673 0.462571440475916526277359001040025 -0.163829183434090234535254223544312 -0.134913073607736057206850553951401 0.0529984818906909399392234421792453 0.0289161098263541773284036695929107 -0.0129475118625466465649568669819003 -0.00302108610126088419848993562911007 0.00151054305063044209924496781455504	0.353553390593273762200422181052425 0.707106781186547524400844362104849 0.353553390593273762200422181052425
8	0 -0.353553390593273762200422181052425 1.06066017177982128660126654315727 1.06066017177982128660126654315727 -0.353553390593273762200422181052425	0 0.176776695296636881100211090526212 0.530330085889910643300633271578637 0.530330085889910643300633271578637 0.176776695296636881100211090526212
9	0 0.0662912607362388304125791589473296 -0.198873782208716491237737476841989 -0.154679608384557270962684704210436 0.994368911043582456188687384209944 0.994368911043582456188687384209944 -0.154679608384557270962684704210436 -0.198873782208716491237737476841989 0.0662912607362388304125791589473296	0 0.176776695296636881100211090526212 0.530330085889910643300633271578637 0.530330085889910643300633271578637 0.176776695296636881100211090526212

10	0 -0.0138106793200497563359539914473603 0.041432037960149269007861974342081 0.0524805814161890740766251674999693 -0.26792717880896527291750743407879 -0.0718155324642587329469607555262737 0.966747552403482943516779401315223 0.966747552403482943516779401315223 -0.0718155324642587329469607555262737 -0.26792717880896527291750743407879 0.0524805814161890740766251674999693 0.041432037960149269007861974342081 -0.0138106793200497563359539914473603	0 0.176776695296636881100211090526212 0.530330085889910643300633271578637 0.530330085889910643300633271578637 0.176776695296636881100211090526212
11	0 -0.0645388826287615974906170013996162 -0.0406894176092047544311284575361263 0.418092273222035359691039182452041 0.788485616406371246825474372225903 0.418092273222035359691039182452041 -0.0406894176092047544311284575361263 -0.0645388826287615974906170013996162 0	0.0378284555072640369336415779310193 -0.0238494650195568430594885438634899 -0.110624404418437179741064759139149 0.377402855612830605259910724915914 0.852698679008893810015690724521109 0.377402855612830605259910724915914 -0.110624404418437179741064759139149 -0.0238494650195568430594885438634899 0.0378284555072640369336415779310193
12	0 -6.79744372783698944660235516549766e-4 0.0020392331183510968339807065496493 0.00506031921961198103247064217875937 -0.0206189126411055346546938106686762 -0.0141127879301758447558029850102713 0.0991347824942321571990197448580834 0.0123001362694193142367090236328053 -0.320191968360778569551383320462395 0.00205002271156988570611817060546755 0.942125700678206737299086425937976 0.942125700678206737299086425937976 0.00205002271156988570611817060546755 -0.320191968360778569551383320462395 0.0123001362694193142367090236328053 0.0991347824942321571990197448580834 -0.0141127879301758447558029850102713 -0.0206189126411055346546938106686762 0.00506031921961198103247064217875937 0.0020392331183510968339807065496493 -6.79744372783698944660235516549766e-4	0 0.176776695296636881100211090526212 0.530330085889910643300633271578637 0.530330085889910643300633271578637 0.176776695296636881100211090526212

13	0	0
	0.00302108610126088419848993562911007	0.176776695296636881100211090526212
	-0.00906325830378265259546980688733022	0.530330085889910643300633271578637
	-0.0168317654213106405344439270764704	0.530330085889910643300633271578637
	0.0746639850740189951912512662622918	0.176776695296636881100211090526212
	0.0313329787073628846871956180961988	
	-0.301159125922834999100896725999001	
	-0.0264992409453454699696117210896226	
	0.951642121897178522524329723169673	
	0.951642121897178522524329723169673	
	-0.0264992409453454699696117210896226	
	-0.301159125922834999100896725999001	
	0.0313329787073628846871956180961988	
	0.0746639850740189951912512662622918	
	-0.0168317654213106405344439270764704	
	-0.00906325830378265259546980688733022	
	0.00302108610126088419848993562911007	

Tabla 6.1 Coeficientes de los filtros QMF y DQMF de las ondeletas "Spline" empleados con el modelo piramidal.

6.1.1 Lenna.

En la Figura 6.1 se aprecia la imagen original de Lenna, mientras que en las Figuras 6.2, 6.3 y 6.4 se aprecian las reconstrucciones obtenidas empleando MPP, MPR y HRPR respectivamente. En cada conjunto de reconstrucciones obtenidas a partir de cada uno de los modelos de codificación, se pueden apreciar las diferencias existentes al emplear cada uno de los valores del conjunto de pasos de cuantización.

En la Tabla 6.2 se muestra el conjunto de valores obtenidos durante los procesos de codificación, entre los cuales tenemos la energía de la imagen original, el número de átomos obtenidos con cada modelo de acuerdo al criterio de paro empleado y la energía residual al final de cada proceso. Se observa que se tienen estos valores con cada modelo para cada paso de cuantización. En esta tabla y las siguientes, *val* corresponde al valor del paso de cuantización empleado con el modelo correspondiente.

En la Tabla 6.3 se presentan los resultados obtenidos para los criterios de calidad empleados, estos valores corresponden al error cuadrático medio (MSE), la relación señal a ruido (SNR), la relación señal a ruido pico (PSNR), la escala de calidad para imágenes (PQS) y la tasa de compresión en bits por pixel (bpp); cada valor se obtuvo para cada una de las imágenes reconstruidas de Lenna.



Figura 6.1. Lenna.

Imagen original Energía = 26,134			
MPP, <i>val</i> = 1 No. átomos = 6,000 Residual = 232.47	MPP, <i>val</i> = 16 No. Átomos = 6,000 Residual = 238.17	MPP, <i>val</i> = 32 No. Átomos = 6,000 Residual = 249.10	MPP, <i>val</i> = 64 No. átomos = 4,415 Residual = 362.26
MPR, <i>val</i> = 1 No. átomos = 6,000 Residual = 182.92	MPR, <i>val</i> = 16 No. Átomos = 6,000 Residual = 210.23	MPR, <i>val</i> = 32 No. Átomos = 6,000 Residual = 251.68	MPR, <i>val</i> = 64 No. átomos = 5,120 Residual = 467.57
HRPR, <i>val</i> = 1 No. átomos = 6,000 Residual = 224.46	HRPR, <i>val</i> = 16 No. Átomos = 6,000 Residual = 255.81	HRPR, <i>val</i> = 32 No. Átomos = 6,000 Residual = 291.14	HRPR, <i>val</i> = 64 No. átomos = 3,390 Residual = 521.24

Tabla 6.2. Conjunto de valores obtenidos para Lenna.

	Cuantización 1			Cuantización 16			Cuantización 32			Cuantización 64		
MSE	28.240	21.913	36.128	29.753	30.470	45.414	32.255	42.946	59.445	60.886	112.16	142.46
SNR	26.441	27.542	25.371	26.214	26.111	24.377	25.863	24.602	23.208	23.104	20.451	19.412
PSNR	33.622	34.724	32.552	33.400	33.292	31.559	33.045	31.802	30.390	30.286	27.623	26.594
POS	5	5	4	5	4	4	4	3	3	3	1	1
bpp	1.0000	1.0380	1.0188	0.6555	0.7600	0.7000	0.6110	0.7000	0.6383	0.3760	0.3375	0.3150
	MPP	MPR	HRPR	MPP	MPR	HRPR	MPP	MPR	HRPR	MPP	MPR	HRPR

Tabla 6.3. Medidas de error, relación, calidad y compresión para Lenna.



Figura 6.2. Imágenes de Lenna reconstruidas con MPP.



Figura 6.3. Imágenes de Lenna reconstruidas con MPR.

En las gráficas de las Figuras 6.5, 6.6 y 6.7 se puede observar la manera en la cual se comporta la energía de la imagen con relación al valor del paso de cuantización en cada uno de los modelos a medida que se van extrayendo los átomos de la imagen.

En las gráficas de las Figuras 6.8, 6.9, 6.10 y 6.11 se observa la manera en la cual se comporta la energía residual en cada uno de los modelos cuando operan bajo las mismas condiciones de cuantización, este decremento se puede observar también con relación al número de átomos que se van obteniendo.

Para el análisis de los resultados, como primer paso se definirán las características principales de Lenna de la manera más sencilla. Estas características son:

Energía de la imagen: 26,000.

Textura: 10 %

Número perceptual de regiones visibles: 20.

Estas características son subjetivas, y son calculadas de la siguiente manera: la energía de la imagen corresponde a su norma aproximada al mil más cercano; el porcentaje de la textura corresponde la parte proporcional que ésta ocupa en la imagen, la cual se obtiene, después de cuadrricular uniformemente la imagen, contabilizando el número de regiones donde predomina la textura; y el número perceptual de regiones visibles se obtiene contabilizando las regiones resultantes después de aplicarle a la imagen original un filtro de media con un programa comercial. Estos cálculos se realizan de la misma manera para las otras imágenes en estudio.



Figura 6.4. Imágenes de Lenna reconstruidas con HRPR.

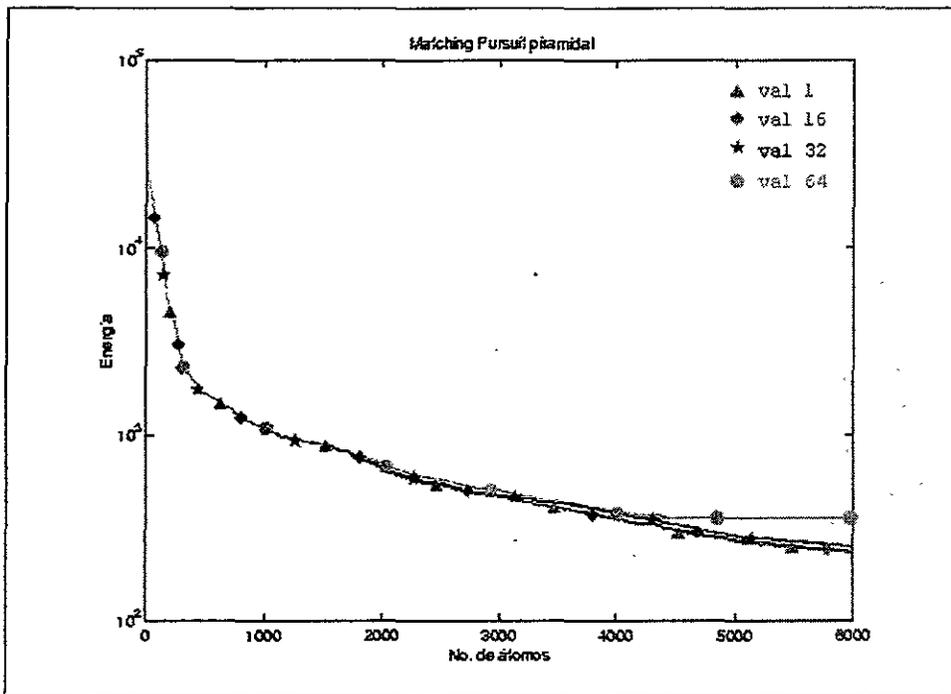


Figura 6.5. Efectos producidos por los diferentes valores del paso de cuantización en MPP para Lenna.

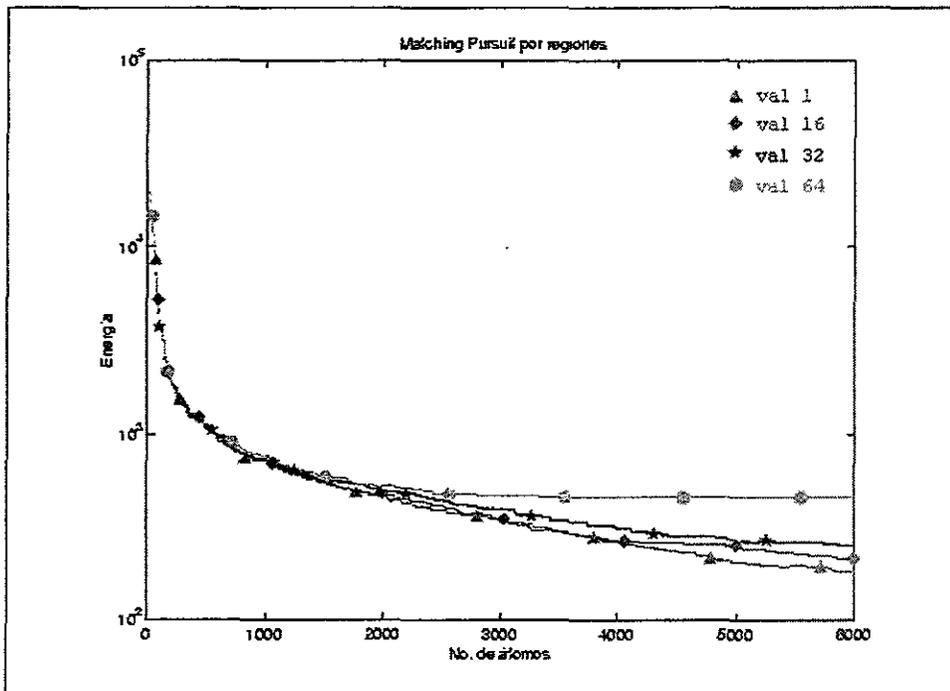


Figura 6.6. Efectos producidos por los diferentes valores del paso de cuantización en MPR para Lenna.

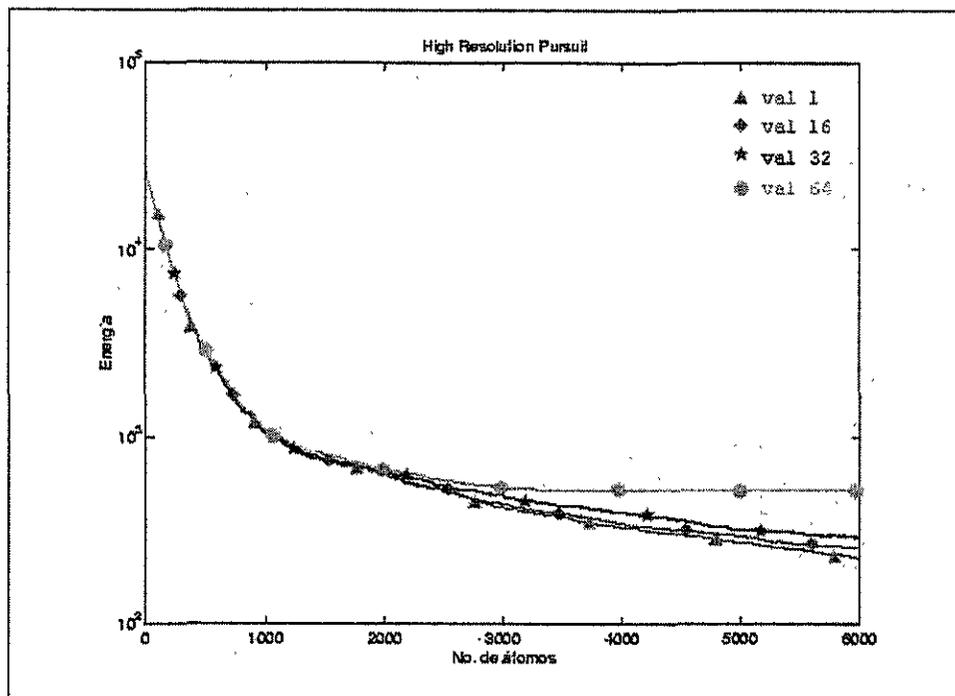


Figura 6.7. Efectos producidos por los diferentes valores del paso de cuantización en HRPR para Lenna.

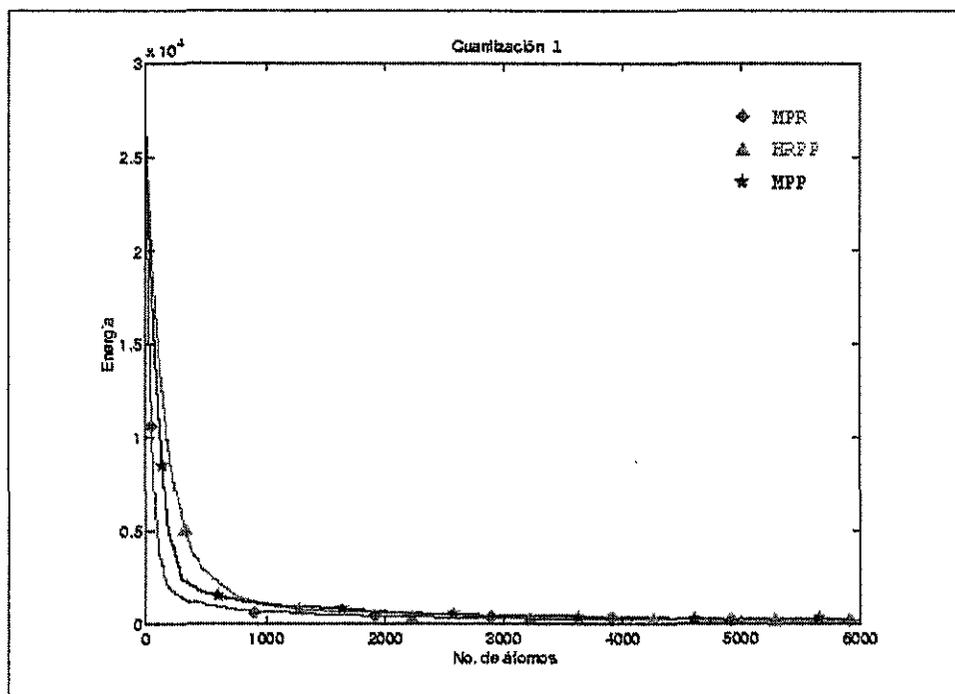


Figura 6.8. Efectos producidos cuando el valor del paso de cuantización en los tres modelos de codificación es 1.

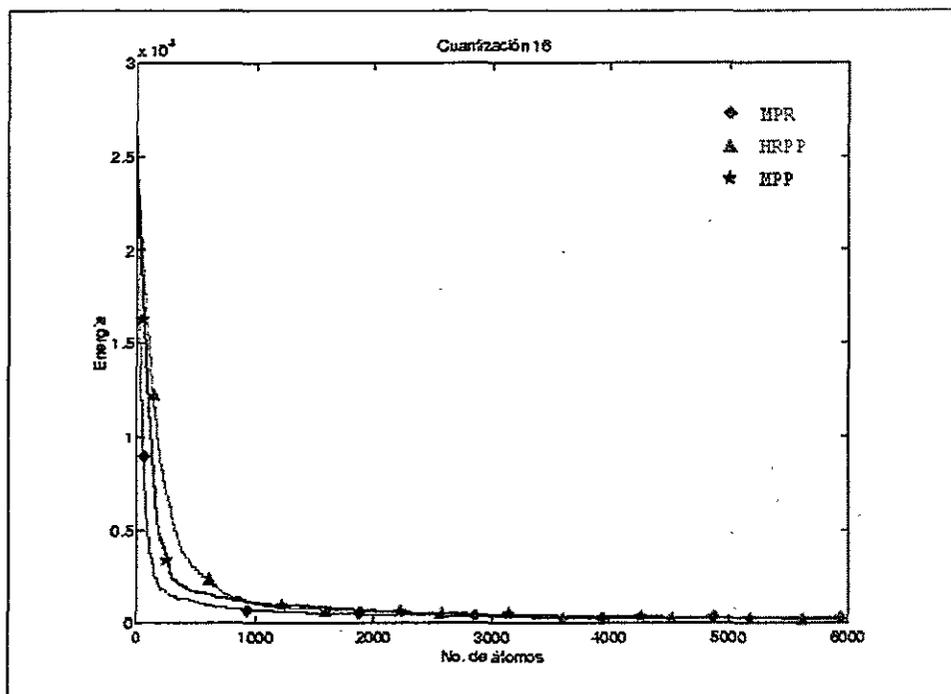


Figura 6.9. Efectos producidos cuando el valor del paso de cuantización en los tres modelos de codificación es 16.

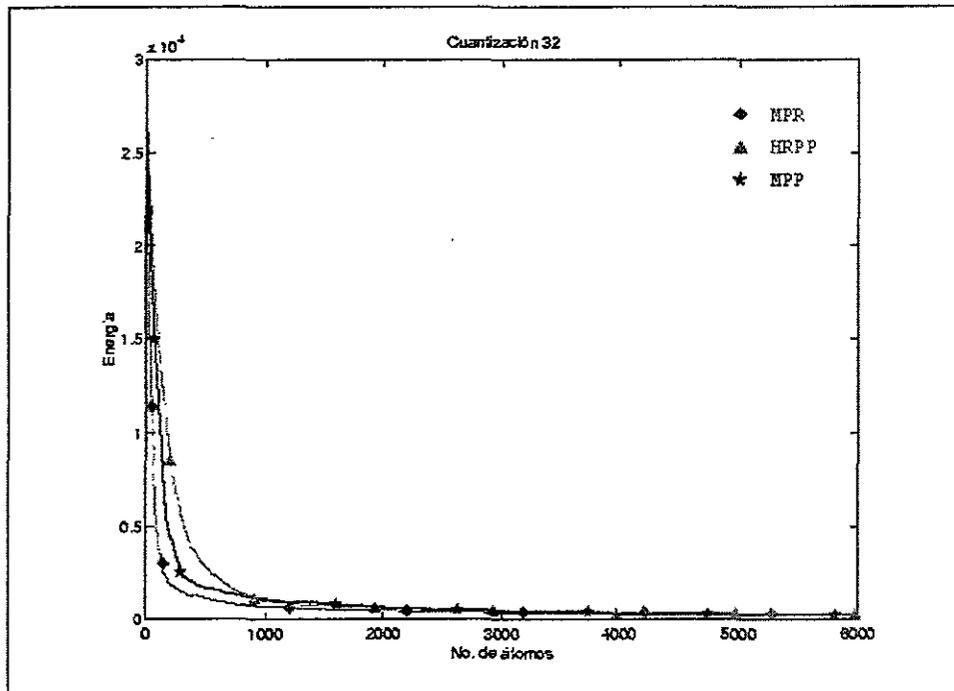


Figura 6.10. Efectos producidos cuando el valor del paso de cuantización en los tres modelos de codificación es 32.

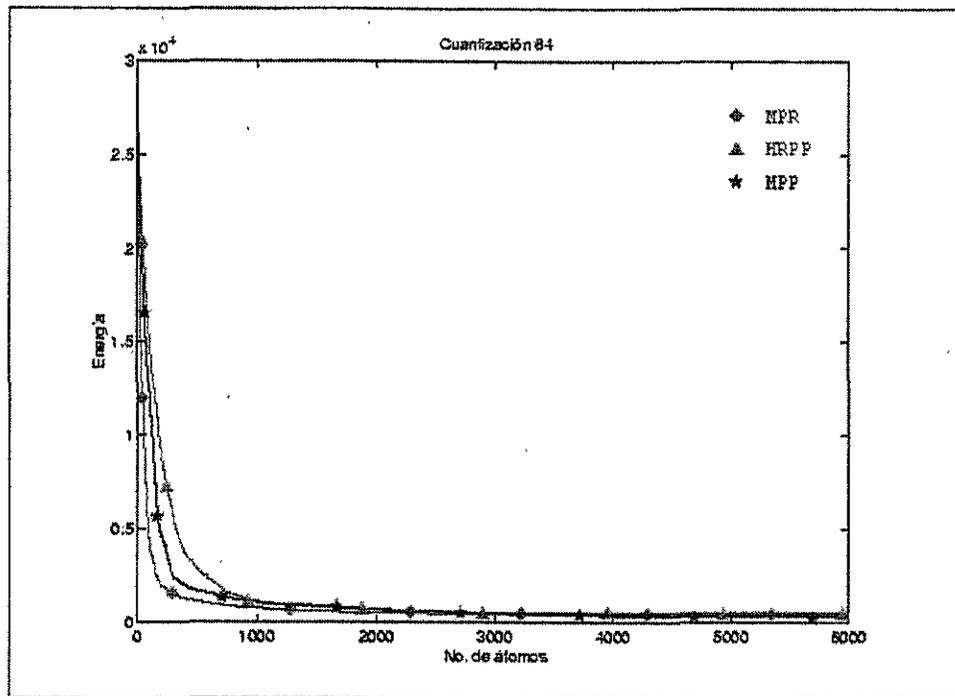


Figura 6.11. Efectos producidos cuando el valor del paso de cuantización en los tres modelos de codificación es 64.

Ahora bien, de acuerdo a los criterios de calidad obtenidos con todos los modelos y pasos de cuantización, se elige aquel modelo que tenga una mayor incidencia en los mejores resultados, esto se observa en la Tabla 6.3.

Mayor incidencia en mejores resultados: MPP.

A continuación, en cada modelo de codificación se observa el momento en el cual el decremento en la energía de la imagen deja de ser casi el mismo para los valores del paso de cuantización, en este punto es donde los efectos de la cuantización comienzan a ser notorios, visto de otra manera, es el punto en el cual la diferencia entre los valores de los coeficientes es menor al paso de cuantización. Esto corresponde a las Figuras 6.5, 6.6 y 6.7 respectivamente.

MPP: 1,000 átomos.

MPR: 250 átomos.

HRPR: 500 átomos.

Finalmente, se analizan los resultados de acuerdo a la rapidez con la cual decrece el valor de la energía de la imagen, también para cada paso de cuantización. Esta observación corresponde a las Figuras 6.8, 6.9, 6.10 y 6.11, y nuevamente se selecciona aquel modelo que presenta un decremento mayor de energía con un menor número de átomos.

Mayor incidencia en mejores resultados: MPR.

6.1.2 Bárbara.



Figura 6.12. Bárbara.

En la Figura 6.12 se presenta la imagen original de Bárbara, y en las Figuras 6.13, 6.14 y 6.15 se muestran las reconstrucciones realizadas con MPP, MPR y HRPR respectivamente, de la misma manera como se efectuaron con Lenna.

En la Tabla 6.4 se muestra el valor de la energía de la imagen original, además del número total de átomos obtenidos durante el proceso de acuerdo al criterio de paro utilizado, junto con el valor de la energía residual para cada uno de los modelos utilizados.

En la Tabla 6.5 se presentan las medidas del error cuadrático medio (MSE), la relación señal a ruido (SNR), la relación señal a ruido pico (PSNR), la escala de calidad para imágenes (PQS) y la tasa de compresión obtenida (bpp) para cada una de las imágenes reconstruidas de Bárbara.

En las Figuras 6.16, 6.17 y 6.18 se observa el comportamiento de la energía de la imagen con relación al valor del paso de cuantización en cada uno de los modelos de codificación de acuerdo al número de átomos que se obtienen.

Por último, las Figuras 6.19, 6.20, 6.21 y 6.22 presentan la manera en la cual se comporta la energía residual en cada uno de los modelos de codificación cuando estos operan bajo las mismas condiciones de cuantización, nuevamente, esta medida se realiza con relación al número de átomos que los seguimientos obtienen.

Imagen original Energía = 29,723			
MPP, <i>val</i> = 1 No. átomos = 6,000 Residual = 540.30	MPP, <i>val</i> = 16 No. átomos = 6,000 Residual = 539.82	MPP, <i>val</i> = 32 No. Átomos = 6,000 Residual = 566.05	MPP, <i>val</i> = 64 No. átomos = 6,000 Residual = 556.82
MPR, <i>val</i> = 1 No. átomos = 6,000 Residual = 525.44	MPR, <i>val</i> = 16 No. átomos = 6,000 Residual = 526.96	MPR, <i>val</i> = 32 No. Átomos = 6,000 Residual = 563.24	MPR, <i>val</i> = 64 No. átomos = 6,000 Residual = 590.94
HRPR, <i>val</i> = 1 No. átomos = 6,000 Residual = 525.79	HRPR, <i>val</i> = 16 No. átomos = 6,000 Residual = 527.61	HRPR, <i>val</i> = 32 No. Átomos = 6,000 Residual = 570.99	HRPR, <i>val</i> = 64 No. átomos = 6,000 Residual = 602.01

Tabla 6.4. Conjunto de valores obtenidos para Bárbara.



Figura 6.13. Imágenes de Bárbara reconstruidas con MPP.

	Cuantización 1			Cuantización 16			Cuantización 32			Cuantización 64		
MSE	154.47	192.39	192.11	156.02	202.58	202.83	163.16	230.70	233.01	173.71	252.57	254.25
SNR	26.242	19.090	19.096	26.199	18.866	18.861	26.004	18.302	18.258	25.733	17.908	17.879
PSNR	20.043	25.289	25.295	20.000	25.065	25.059	19.806	24.500	24.457	19.534	24.107	24.078
PQS	5	2	2	4	2	2	4	2	1	3	1	1
bpp	0.9520	0.9600	0.9600	0.6560	0.6900	0.6880	0.5880	0.6300	0.6285	0.3760	0.5550	0.5600
	MPP	MPR	HRPR	MPP	MPR	HRPR	MPP	MPR	HRPR	MPP	MPR	HRPR

Tabla 6.5. Medidas de error, relación, calidad y compresión para Bárbara.



Figura 6.14. Imágenes de Bárbara reconstruidas con MPR.



Figura 6.15. Imágenes de Bárbara reconstruidas con HRPR.

Las características de Bárbara son:

Energía de la imagen: 30,000.

Textura: 40 %

Número perceptual de regiones visibles: 40.

A partir de la Tabla 6.5 se tiene:

Mayor incidencia en mejores resultados: MPP.

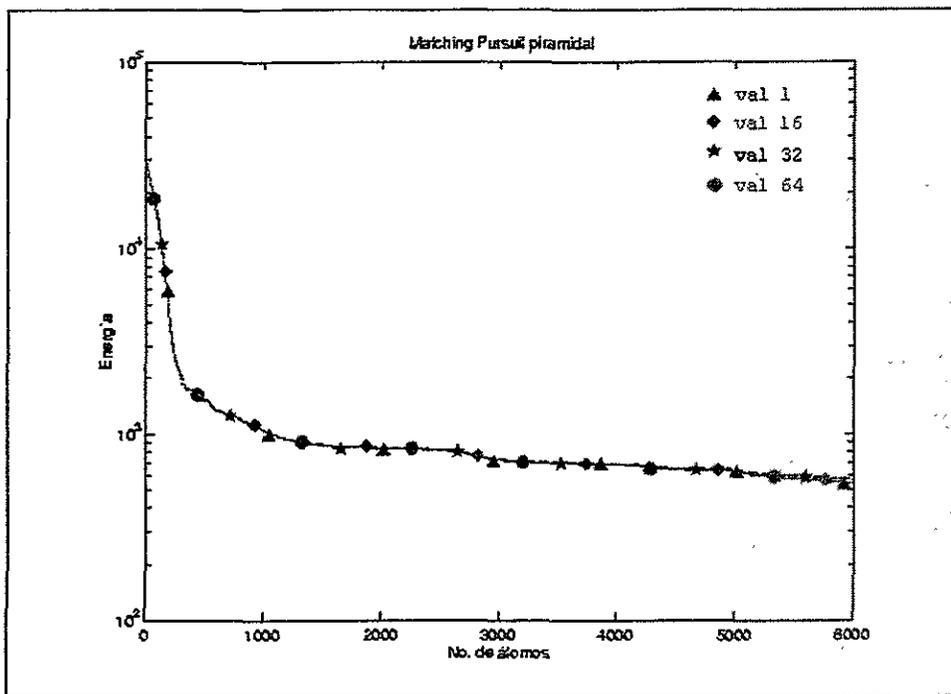


Figura 6.16. Efectos producidos por los diferentes valores del paso de cuantización en MPP para Bárbara.

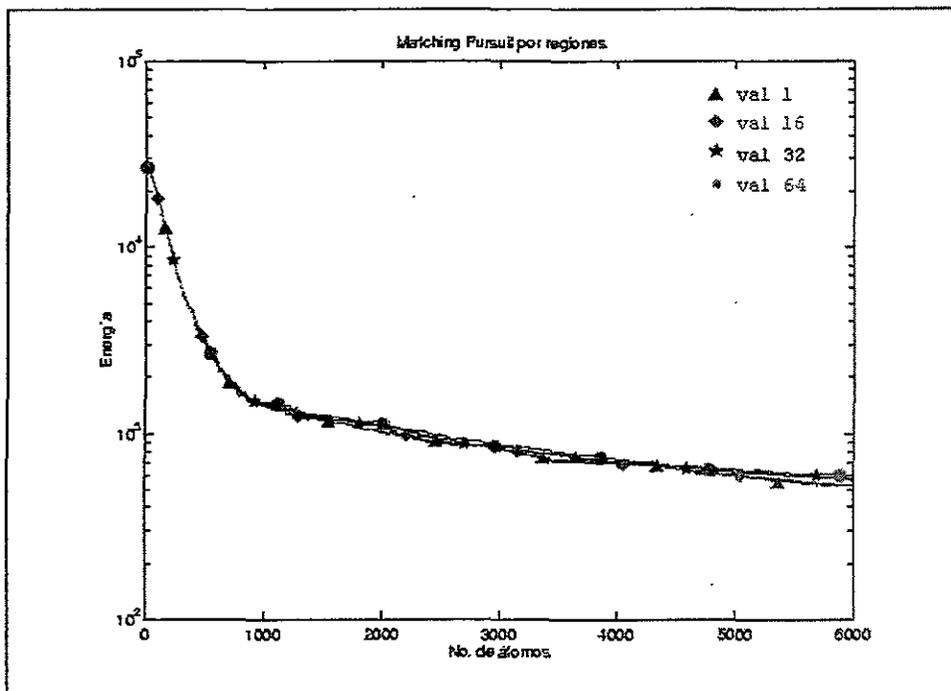


Figura 6.17. Efectos producidos por los diferentes valores del paso de cuantización en MPR para Bárbara.

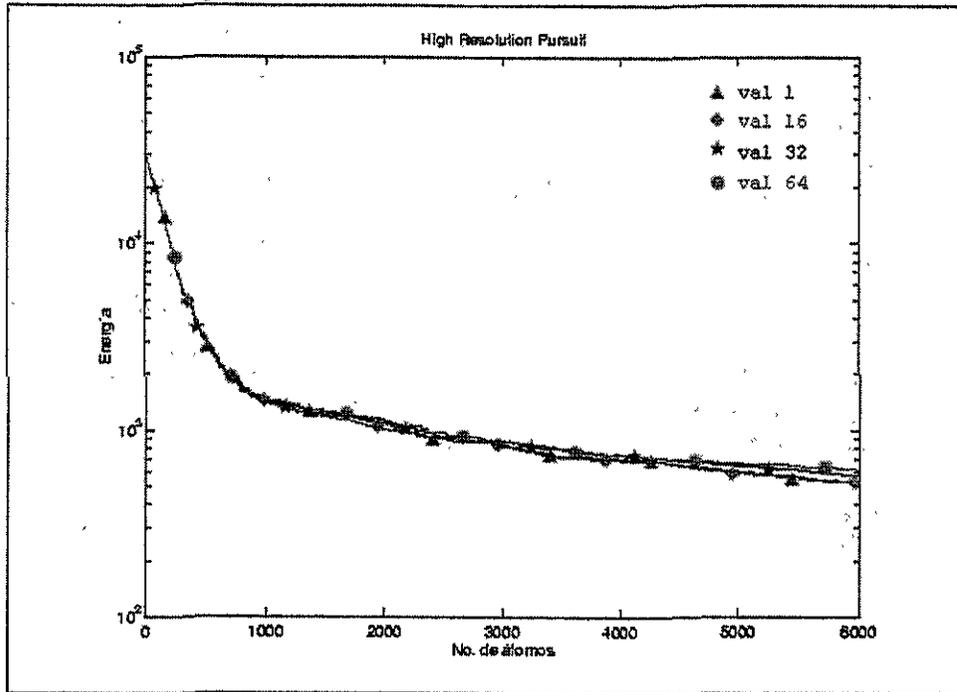


Figura 6.18. Efectos producidos por los diferentes valores del paso de cuantización en HRPR por regiones para Bárbara.

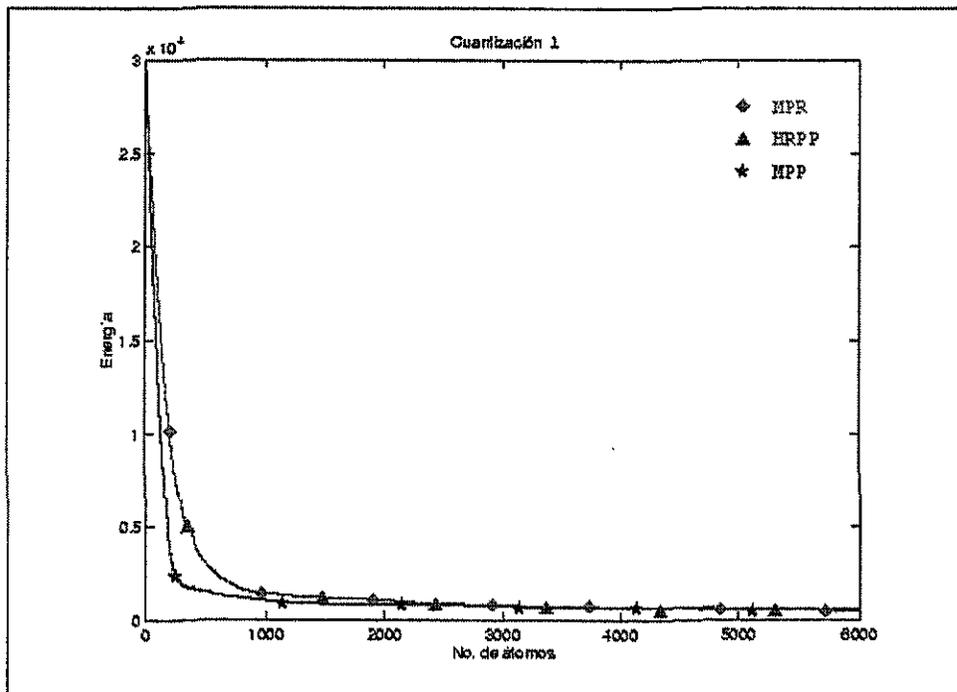


Figura 6.19. Efectos producidos cuando el valor del paso de cuantización en los tres modelos de codificación es 1.

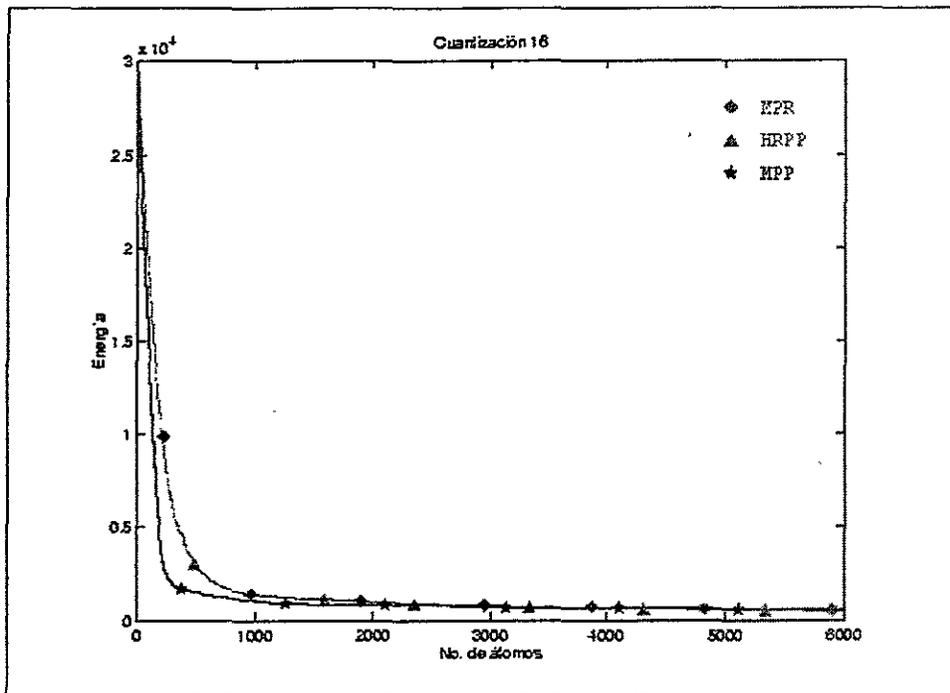


Figura 6.20. Efectos producidos cuando el valor del paso de cuantización en los tres modelos de codificación es 16.

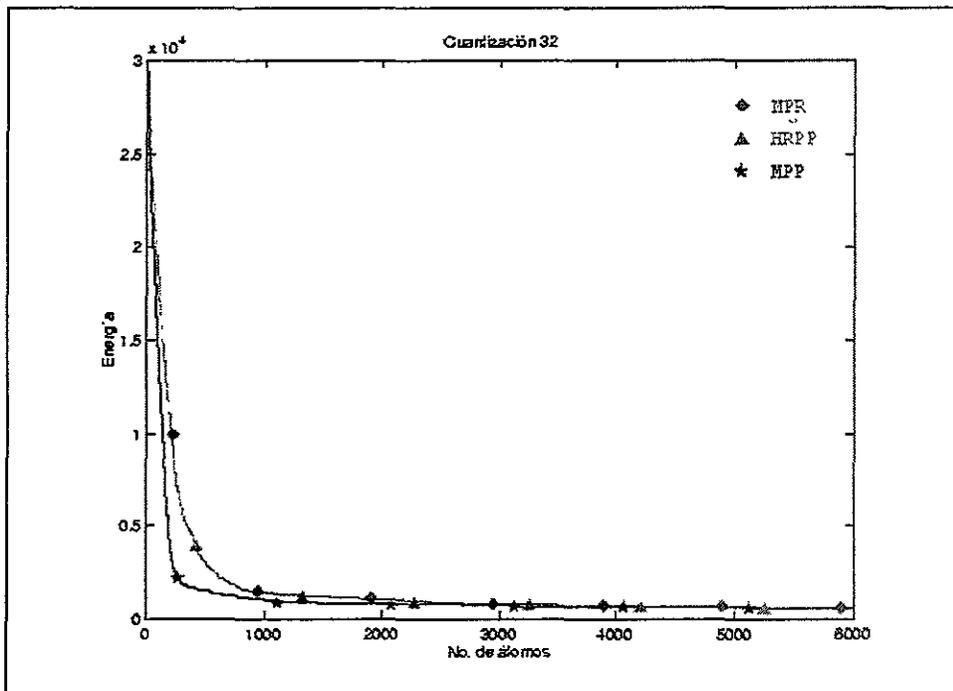


Figura 6.21. Efectos producidos cuando el valor del paso de cuantización en los tres modelos de codificación es 32.

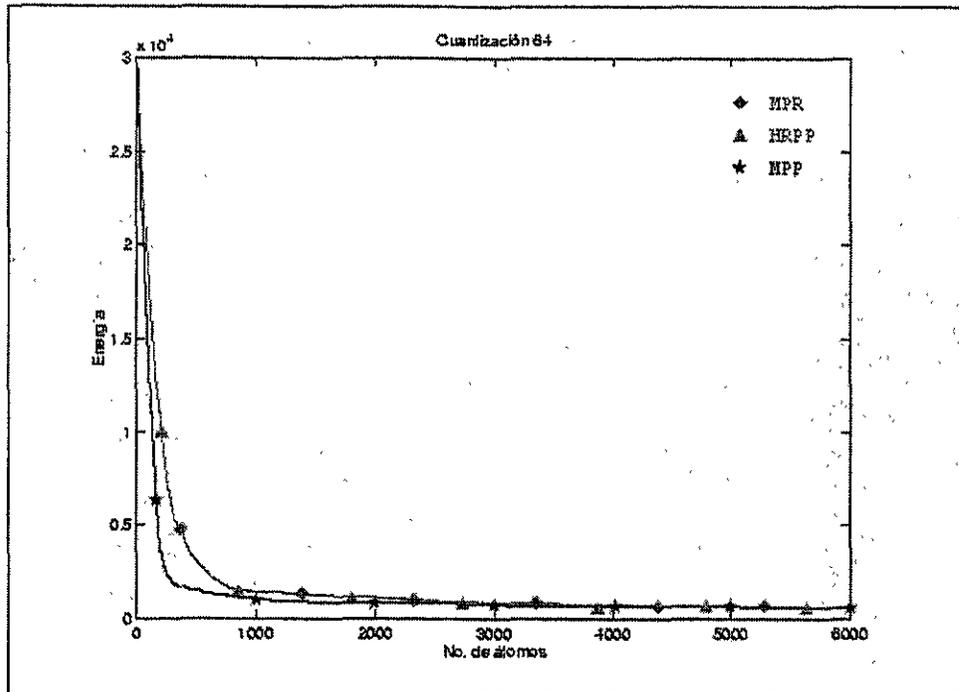


Figura 6.22. Efectos producidos cuando el valor del paso de cuantización en los tres modelos de codificación es 64.

A partir de las Figuras 6.16, 6.17 y 6.18, se tiene:

- MPP: +6,000 átomos.
- MPR: +6,000 átomos.
- HRPR: +6,000 átomos.

Finalmente, de las Figuras 6.19, 6.20, 6.21 y 6.22, se tiene:

Mayor incidencia en mejores resultados: MPP.

6.1.3 Cameraman.

En la Figura 6.23 tenemos la imagen original Cameraman, mientras que en las Figuras 6.24, 6.25 y 6.26 tenemos el conjunto de imágenes reconstruidas con MPP, MPR y HRPR respectivamente.

En la Tabla 6.6 se tiene el valor de la energía de Cameraman, además se tiene el número de átomos obtenidos durante el proceso de acuerdo al criterio de paro empleado y la energía residual obtenida al final de cada uno de los procesos realizados.



Figura 6.23. Cameraman.

En la Tabla 6.7 se tienen medidas del error cuadrático medio (MSE), la relación señal a ruido (SNR), la relación señal a ruido pico (PSNR), la escala de calidad para imágenes (PQS) y la tasa de compresión (bpp) para cada imagen reconstruida.

Para cada valor de cuantización, la manera en la cual la energía de la imagen decrece mientras se van obteniendo los átomos, se observa en las gráficas de las Figuras 6.27, 6.28 y 6.29. Mientras en las gráficas de las Figuras 6.30, 6.31, 6.32 y 6.33 se observa ese decremento pero entre los modelos de codificación, los cuales tiene el mismo paso de cuantización.

Imagen original Energía = 32,036			
MPP, <i>val</i> = 1 No. átomos = 6,000 Residual = 293.90	MPP, <i>val</i> = 16 No. Átomos = 6,000 Residual = 292.57	MPP, <i>val</i> = 32 No. átomos = 6,000 Residual = 301.70	MPP, <i>val</i> = 64 No. átomos = 4,887 Residual = 381.92
MPR, <i>val</i> = 1 No. átomos = 6,000 Residual = 292.45	MPR, <i>val</i> = 16 No. Átomos = 6,000 Residual = 307.53	MPR, <i>val</i> = 32 No. átomos = 6,000 Residual = 343.67	MPR, <i>val</i> = 64 No. átomos = 4,988 Residual = 529.08
HRPR, <i>val</i> = 1 No. átomos = 6,000 Residual = 292.72	HRPR, <i>val</i> = 16 No. Átomos = 6,000 Residual = 306.54	HRPR, <i>val</i> = 32 No. átomos = 6,000 Residual = 352.35	HRPR, <i>val</i> = 64 No. átomos = 4,247 Residual = 563.98

Tabla 6.6. Conjunto de valores obtenidos para Cameraman.

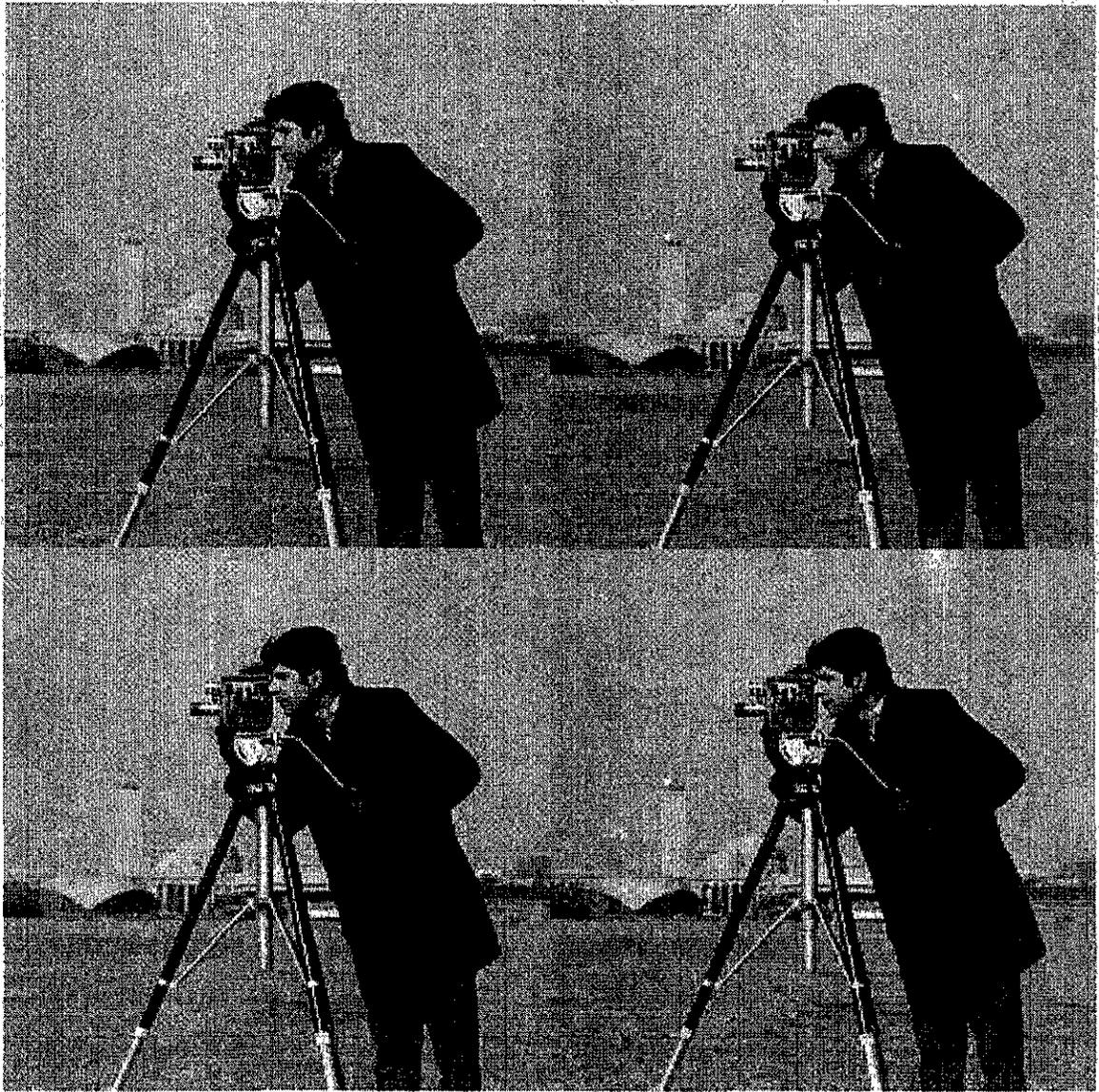


Figura 6.24. Imágenes de Cameraman reconstruidas con MPP.

	Cuantización 1			Cuantización 16			Cuantización 32			Cuantización 64		
MSE	39.349	58.895	58.906	40.977	68.172	67.977	45.372	83.080	83.886	83.806	152.19	163.22
SNR	26.599	24.848	24.847	26.423	24.212	24.225	25.980	23.353	23.312	23.316	20.725	20.421
PSNR	32.185	30.430	30.429	32.005	29.795	29.807	31.563	28.936	28.894	28.898	26.307	26.003
PQS	4	4	4	4	3	3	4	2	3	3	1	1
bpp	0.8800	0.9660	0.9660	0.5840	0.6900	0.6904	0.5120	0.6225	0.6225	0.3600	0.4575	0.4000
	MPP	MPR	HRPR	MPP	MPR	HRPR	MPP	MPR	HRPR	MPP	MPR	HRPR

Tabla 6.7. Medidas de error, relación, calidad y compresión para Cameraman.

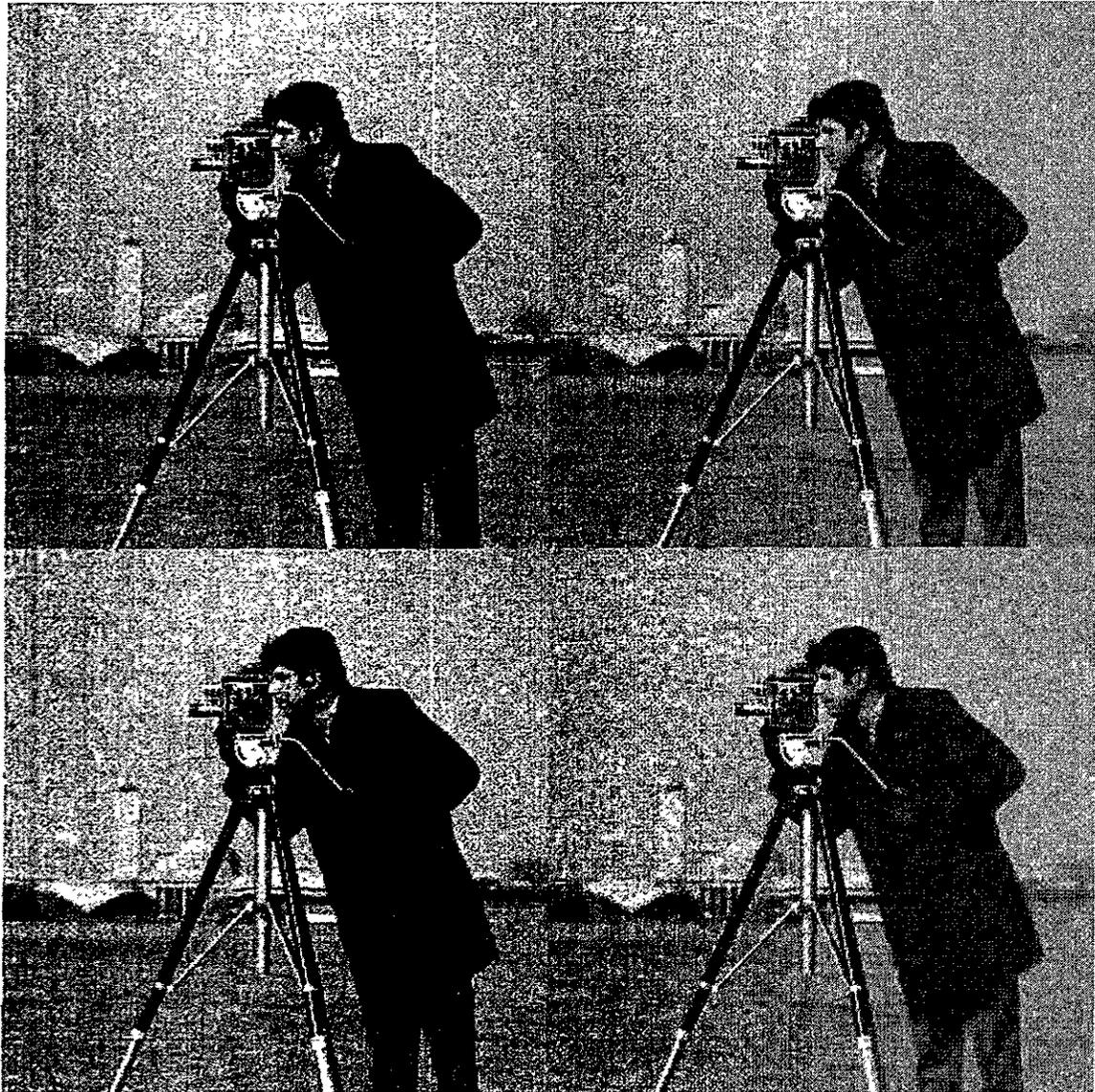


Figura 6.25. Imágenes de Cameraman reconstruidas con MPR.

Las características de Cameraman son:

Energía de la imagen: 32,000.

Textura: 0 %

Número perceptual de regiones visibles: 20.

De la Tabla 6.7. se tiene:

Mayor incidencia en mejores resultados: MPP.

**ESTA TESIS NO SALE
DE LA BIBLIOTECA**

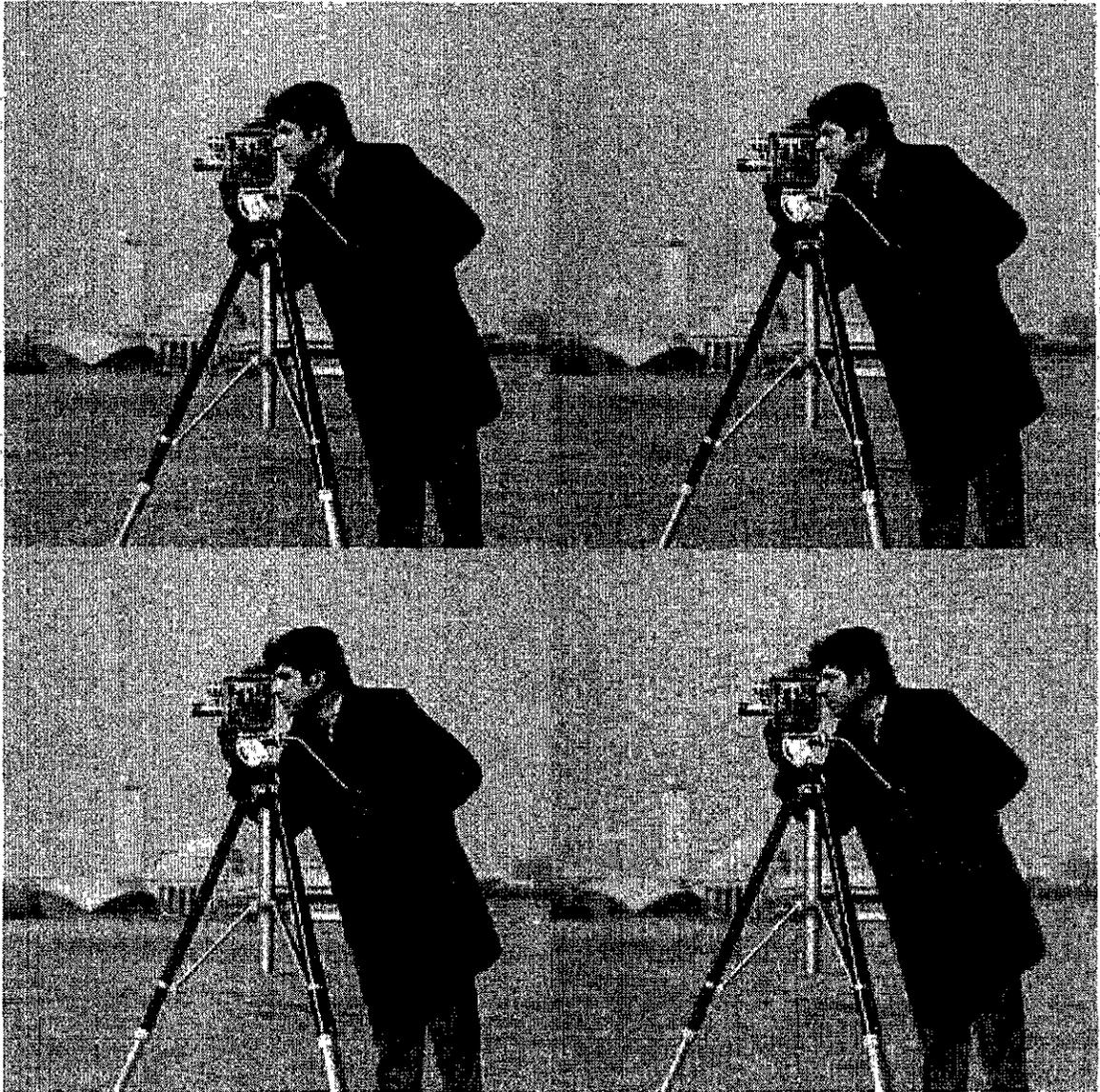


Figura 6.26. Imágenes Cameraman reconstruidas con HRPR.

De las gráficas en las Figuras 6.27, 6.28 y 6.29 se tiene:

MPP: 1,000 átomos.

MPR: 250 átomos.

HRPR: 500 átomos.

Finalmente, de las gráficas en las Figuras 6.30, 6.31, 6.32 y 6.33 se observa:

Mayor incidencia en mejores resultados: MPP.

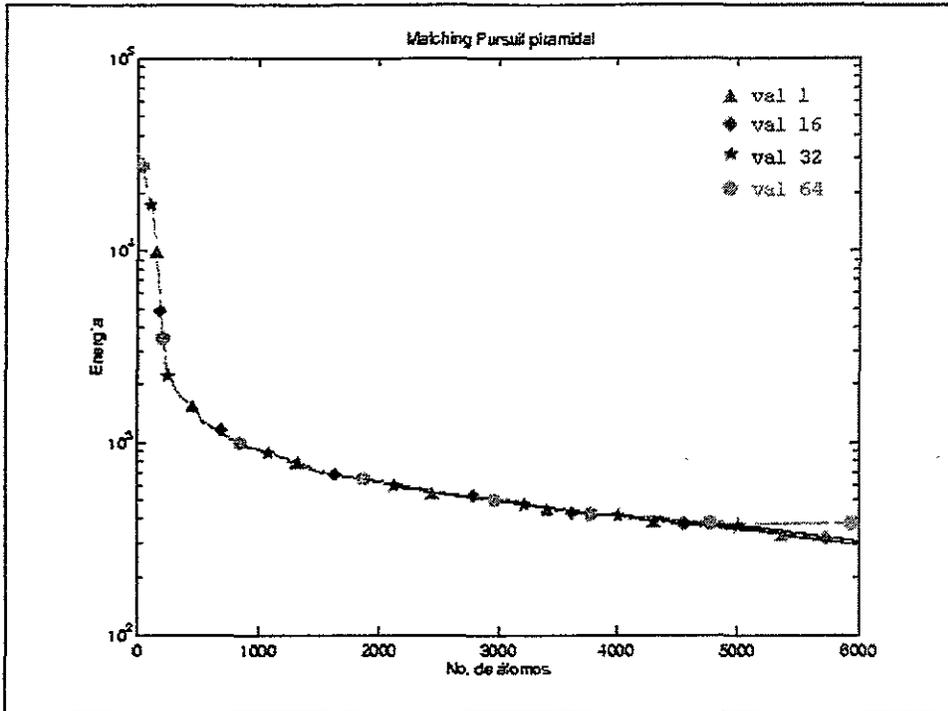


Figura 6.27. Efectos producidos por los diferentes valores del paso de cuantización en MPP para Cameraman.

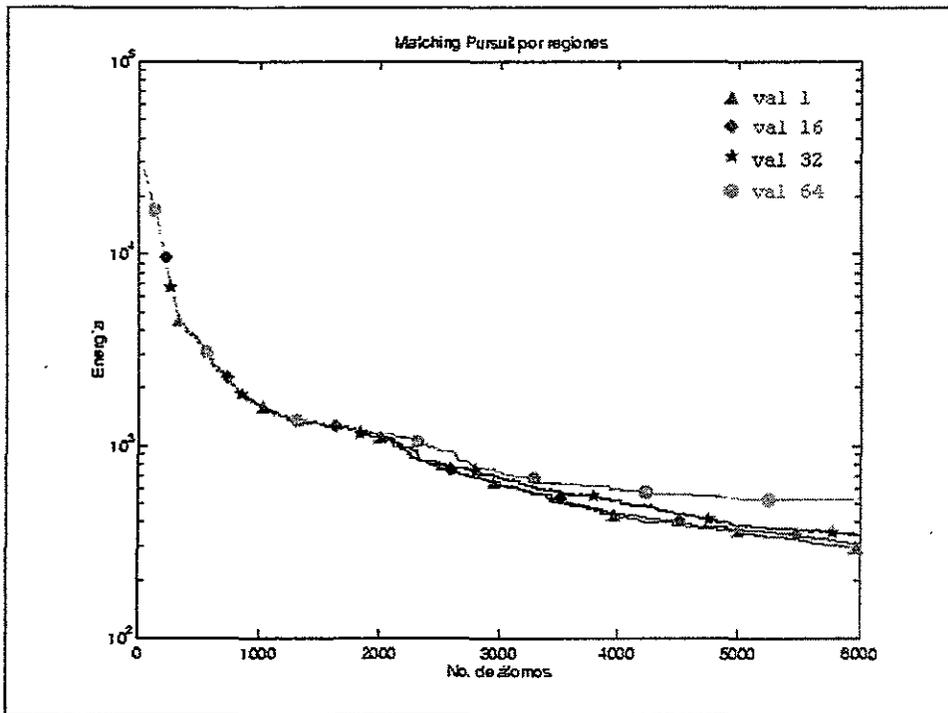


Figura 6.28. Efectos producidos por los diferentes valores del paso de cuantización en MPR para Cameraman.

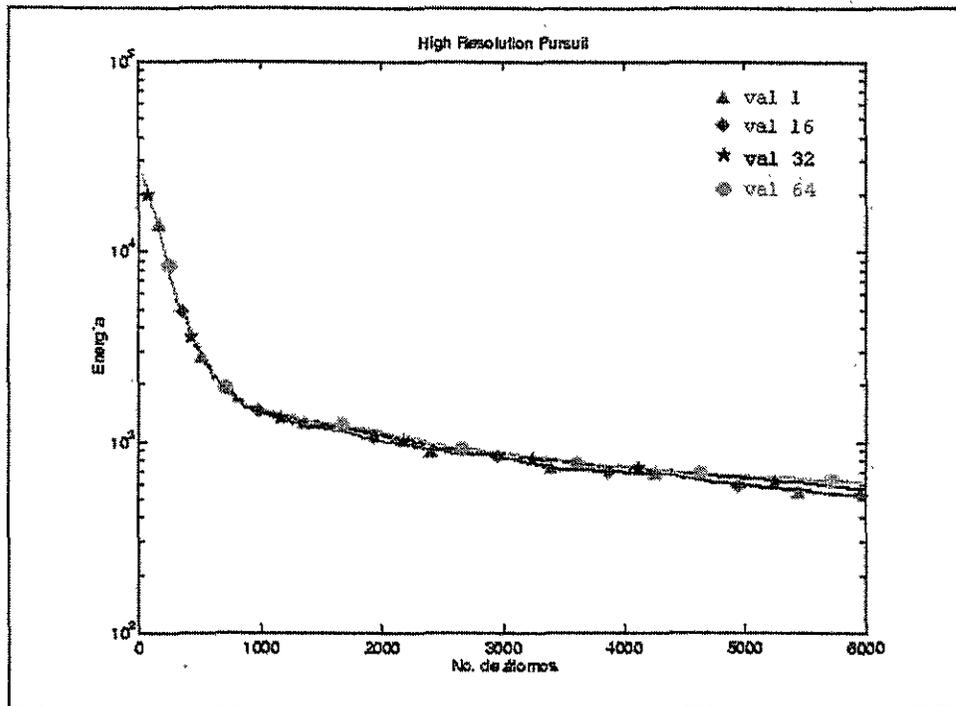


Figura 6.29. Efectos producidos por los diferentes valores del paso de cuantización en HRPR para Cameraman.

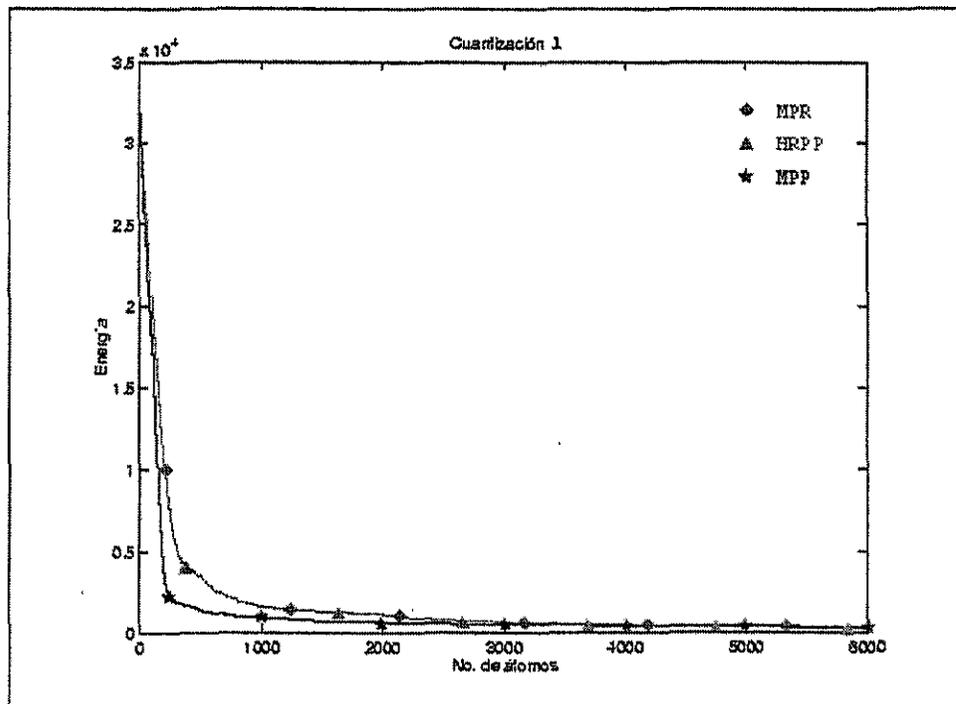


Figura 6.30. Efectos producidos cuando el valor del paso de cuantización en los tres modelos de codificación es 1.

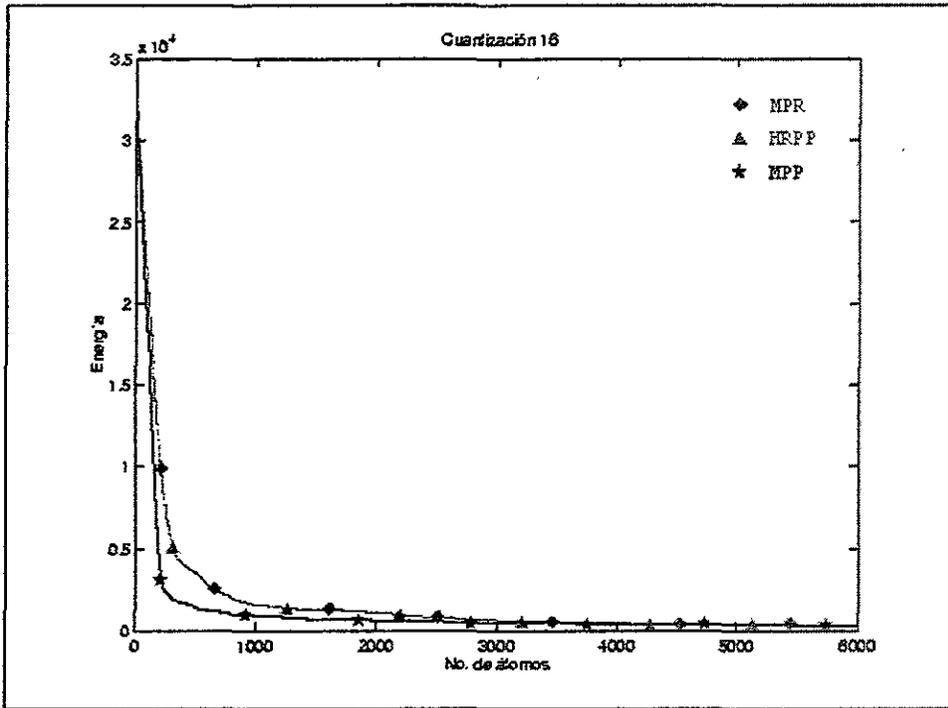


Figura 6.31. Efectos producidos cuando el valor del paso de cuantización en los tres modelos de codificación es 16.

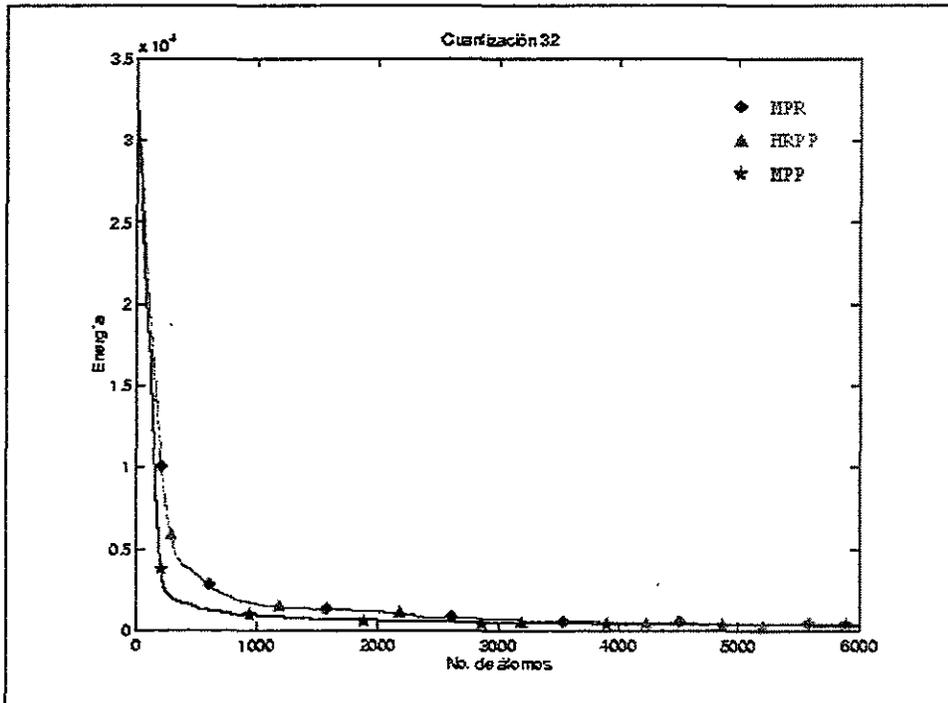


Figura 6.32. Efectos producidos cuando el valor del paso de cuantización en los tres modelos de codificación es 32.

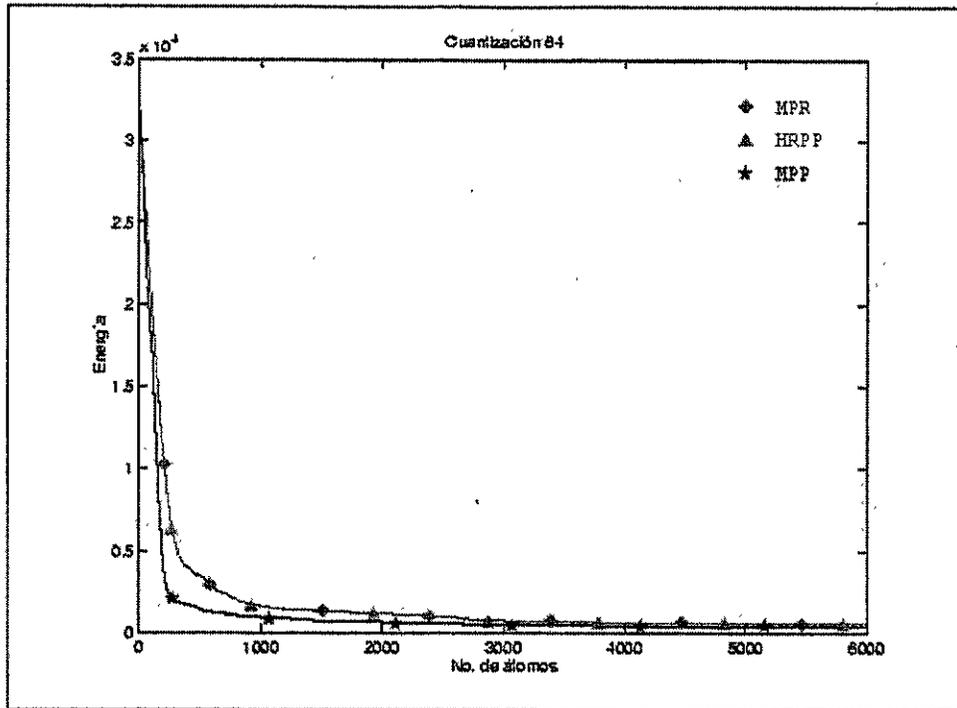


Figura 6.33. Efectos producidos cuando el valor del paso de cuantización en los tres modelos de codificación es 64.

6.2. Una aplicación real con imágenes de microscopía electrónica.

En esta parte se presentan los resultados obtenidos al emplear los modelos de codificación con un par de imágenes de microscopía electrónica. Estas imágenes forman parte del trabajo que se realiza en el laboratorio del Dr. Francisco Fernández de Miguel del Instituto de Fisiología Celular de la UNAM.

Estas imágenes poseen las mismas características que las imágenes de prueba empleadas en 6.1 en cuanto a tamaño y composición. De igual manera estas imágenes son procesadas con los mismos tres modelos de codificación y sus respectivos diccionarios, el mismo conjunto de valores del paso de cuantización y las mismas condiciones de paro.

6.2.1 Fluorescencia.

En esta sección se emplea una imagen de fluorescencia, la cual posee estructuras muy simples con relación a las imágenes Lenna, Bárbara y Cameraman. Esta imagen ha sido aumentada 20,000 veces con ayuda de un microscopio electrónico, y pertenece a una serie de imágenes donde se estudia el crecimiento del axón de una neurona. En la Figura 6.34 tenemos dicha imagen. En las Figuras 6.35, 6.36 y 6.37 se presentan todas las reconstrucciones realizadas con MPP, MPR y HRPP respectivamente.

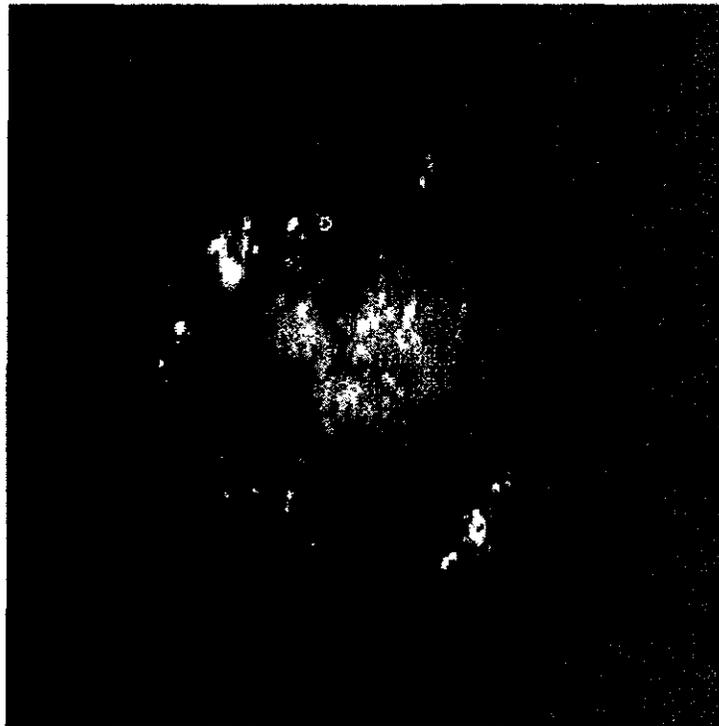


Figura 6.34. Fluorescencia.

En la Tabla 6.8 se tiene el valor la energía de la imagen de fluorescencia, junto con el número de átomos obtenidos durante el proceso, siguiendo el criterio de paro empleado y se proporciona también el valor de la energía residual al final del proceso en cada caso.

Imagen original Energía = 7,669			
MPP, <i>val</i> = 1 No. átomos = 6,000 Residual = 48.55	MPP, <i>val</i> = 16 No. átomos = 6,000 Residual = 76.11	MPP, <i>val</i> = 32 No. átomos = 4,671 Residual = 298.51	MPP, <i>val</i> = 64 No. átomos = 2,258 Residual = 311.44
MPR, <i>val</i> = 1 No. átomos = 6,000 Residual = 31.56	MPR, <i>val</i> = 16 No. átomos = 4,961 Residual = 81.73	MPR, <i>val</i> = 32 No. átomos = 3,560 Residual = 150.85	MPR, <i>val</i> = 64 No. átomos = 1,593 Residual = 289.89
HRPR, <i>val</i> = 1 No. átomos = 6,000 Residual = 31.56	HRPR, <i>val</i> = 16 No. átomos = 4,961 Residual = 81.73	HRPR, <i>val</i> = 32 No. átomos = 3,560 Residual = 150.85	HRPR, <i>val</i> = 64 No. átomos = 1,235 Residual = 289.86

Tabla 6.8. Conjunto de valores obtenidos para la imagen de fluorescencia.

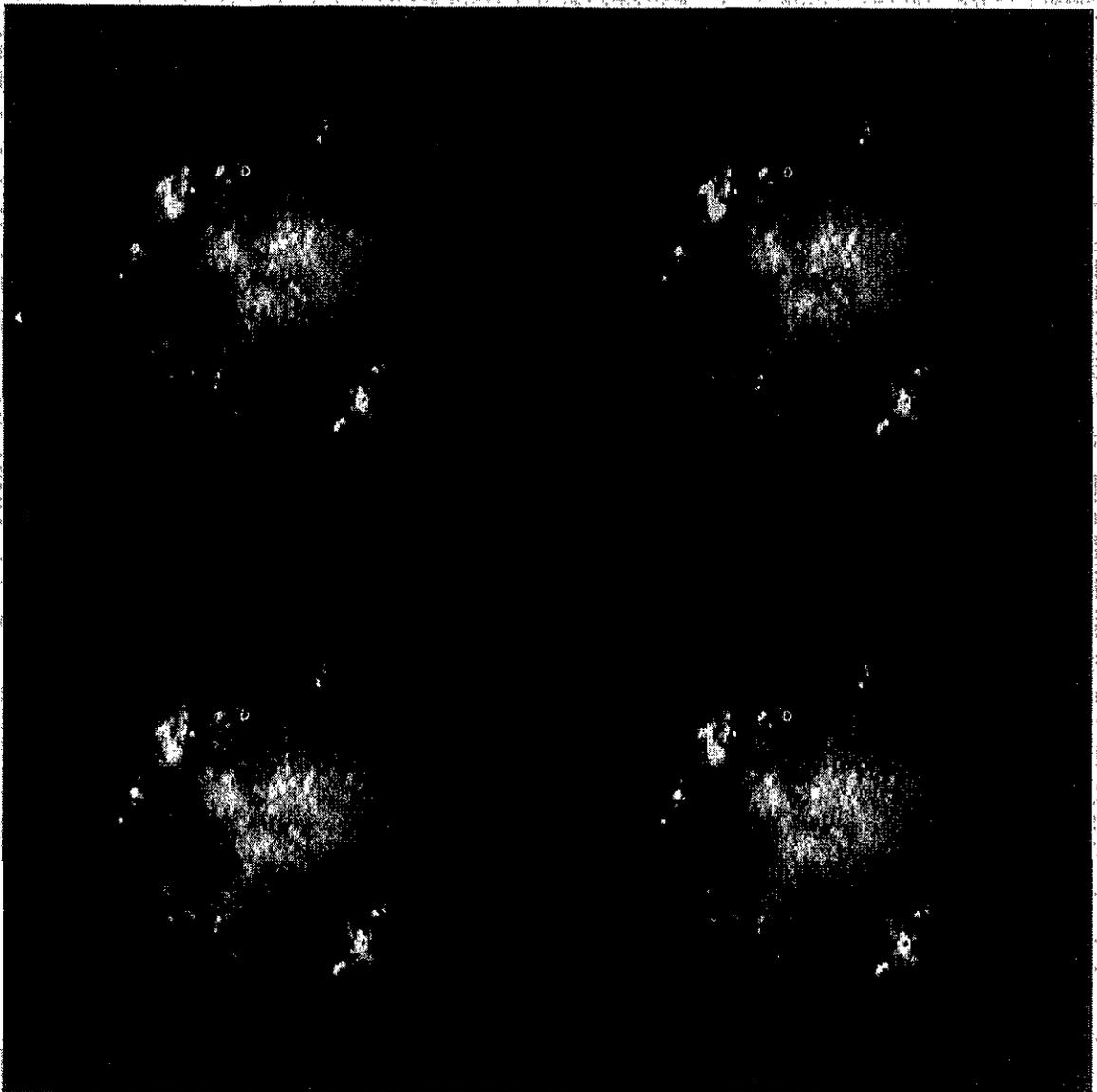


Figura 6.35. Imágenes de fluorescencia reconstruidas con MPP.

La Tabla 6.9 presenta los valores obtenidos para las medidas de criterio de calidad y compresión empleadas para las imágenes reconstruidas.

	Cuantización 1			Cuantización 16			Cuantización 32			Cuantización 64		
MSE	0.8361	0.3835	0.3835	2.2910	2.6498	2.6498	14.885	9.4010	9.4010	35.988	34.269	34.271
SNR	32.184	35.568	35.568	27.806	27.174	27.174	19.679	21.675	21.675	15.845	16.057	16.057
PSNR	48.908	52.293	52.292	44.531	43.899	43.899	36.403	38.399	38.399	32.569	32.782	32.782
PQS	5	5	5	5	5	5	5	5	5	4	4	4
bpp	0.8800	0.9600	0.9600	0.5840	0.5700	0.5690	0.4000	0.3675	0.3675	0.1680	0.1500	0.1125
	MPP	MPR	HRPR	MPP	MPR	HRPR	MPP	MPR	HRPR	MPP	MPR	HRPR

Tabla 6.9. Medidas de error, relación, calidad y compresión para la imagen de fluorescencia.

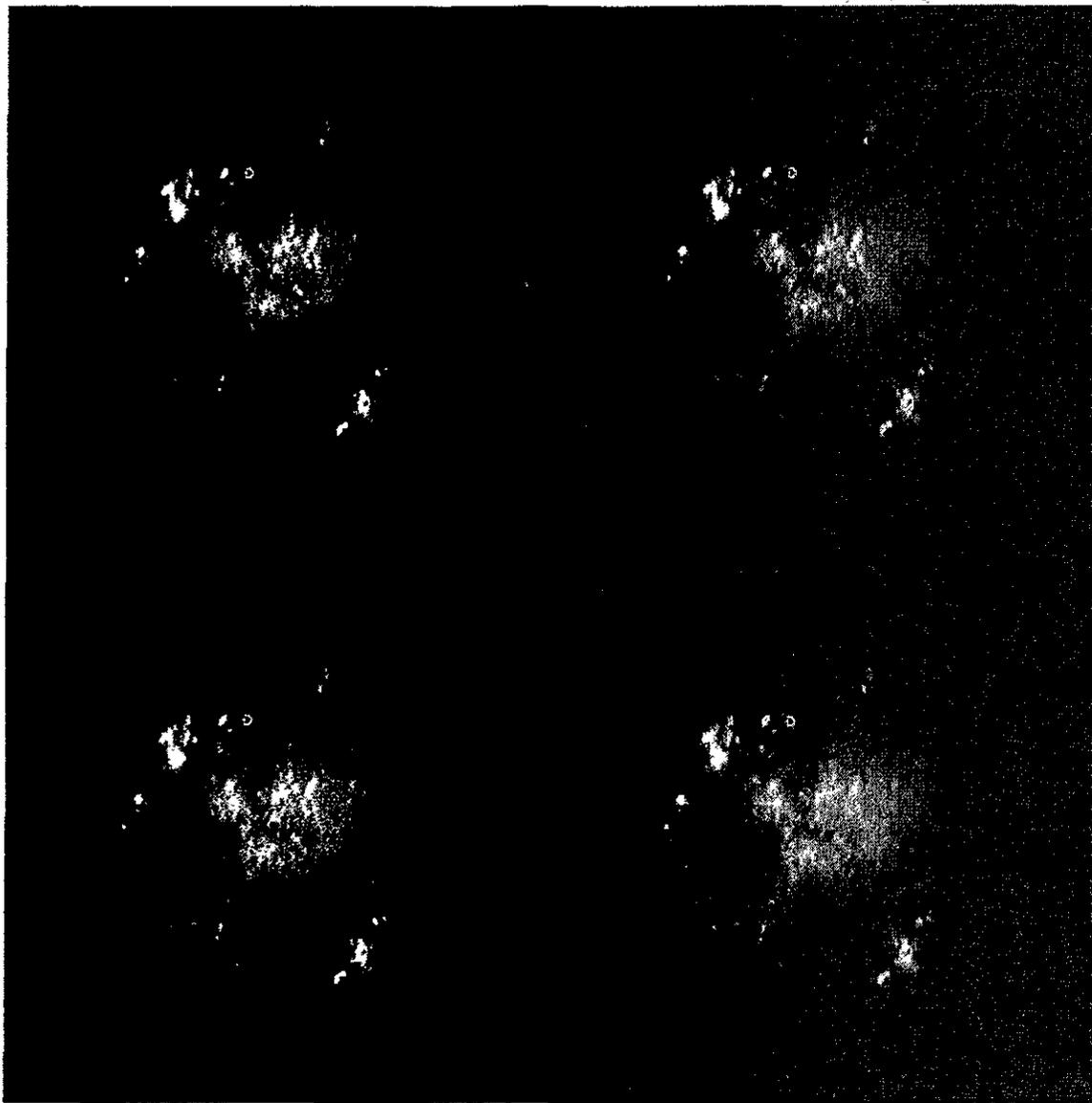


Figura 6.36. Imágenes de fluorescencia reconstruidas con MPR.

Nuevamente, como en 6.1, se observa en las gráficas de las Figuras 6.38, 6.39 y 6.40 el comportamiento de la energía residual de la imagen conforme se van obteniendo los átomos a lo largo del proceso con cada uno de los modelos propuestos y cada valor del paso de cuantización.

De la misma manera, se observa en las gráficas de las Figura 6.41, 6.42, 6.43 y 6.44 el comportamiento de dicha energía residual en cada uno de los modelos, pero ahora cuando operan con las mismas condiciones de cuantización, esta observación se efectúa tomando como referencia el número de átomos que se van extrayendo de la imagen mientras se efectúa el algoritmo.

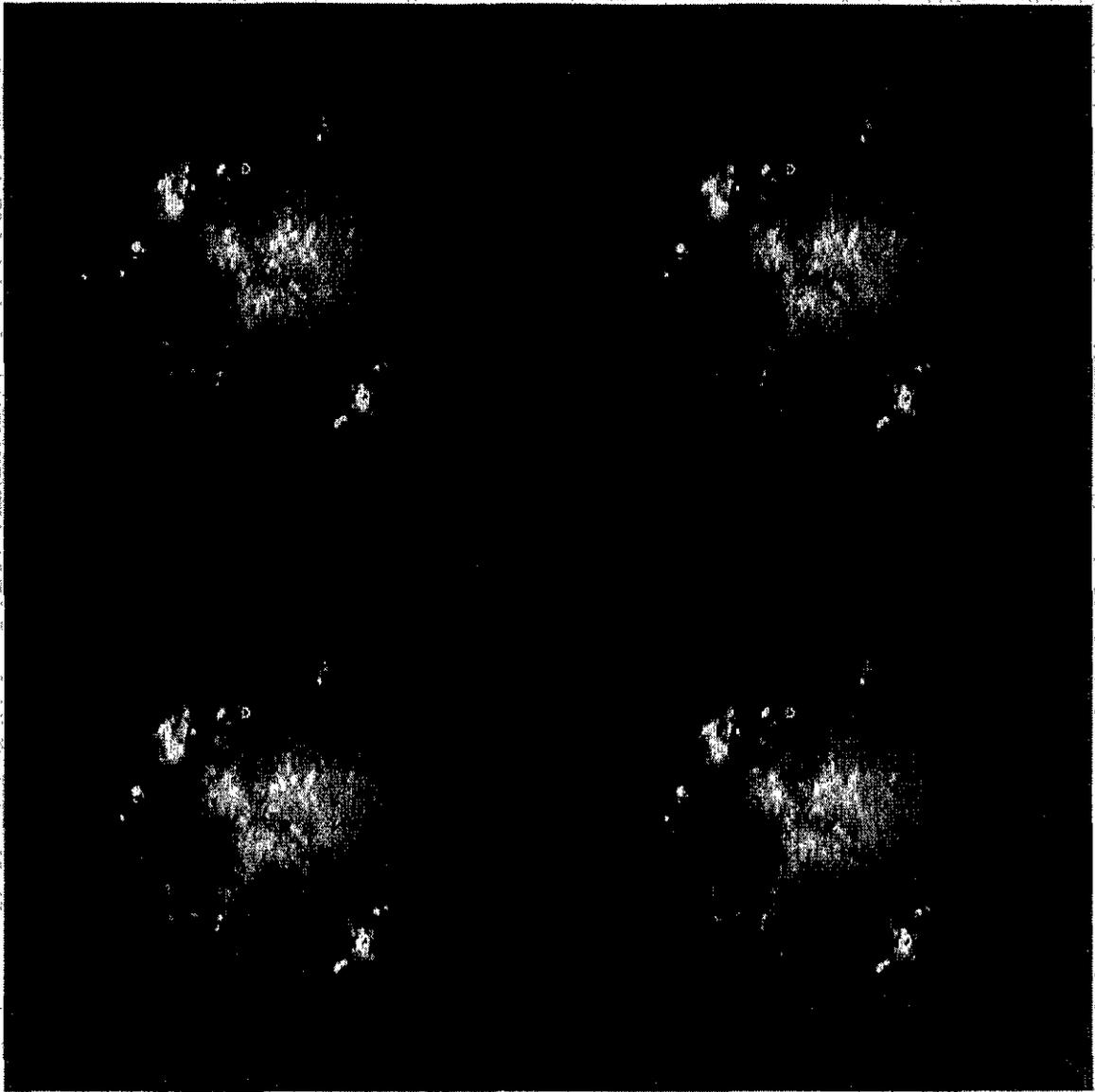


Figura 6.37. Imágenes de fluorescencia reconstruidas con HRPR.

Las características de la imagen de fluorescencia son:

Energía de la imagen: 8,000.

Textura: 10 %

Número perceptual de regiones visibles: 5.

Se observa en la Tabla 6.9:

Mayor incidencia en mejores resultados: MPR y HRPR.

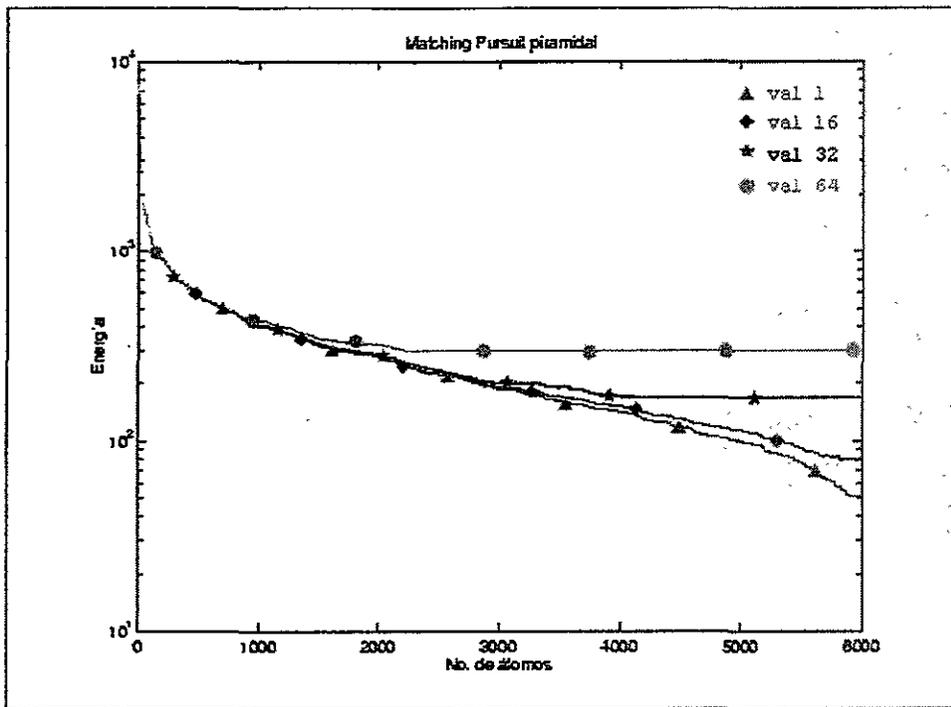


Figura 6.38. Efectos producidos por los diferentes valores del paso de cuantización en MPP para la imagen de fluorescencia.

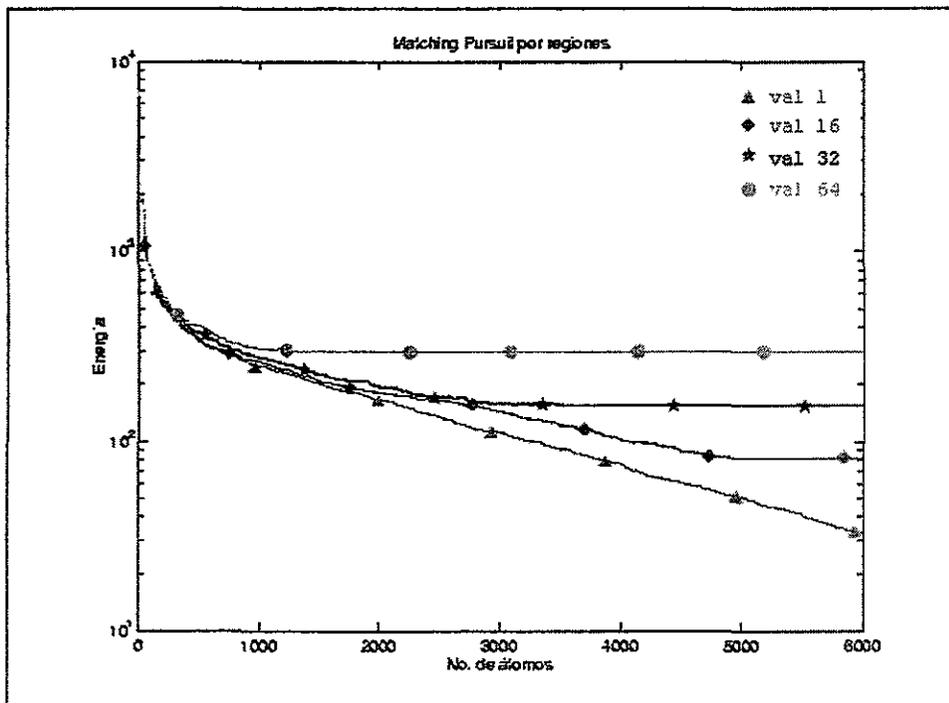


Figura 6.39. Efectos producidos por los diferentes valores del paso de cuantización en MPR para la imagen de fluorescencia.

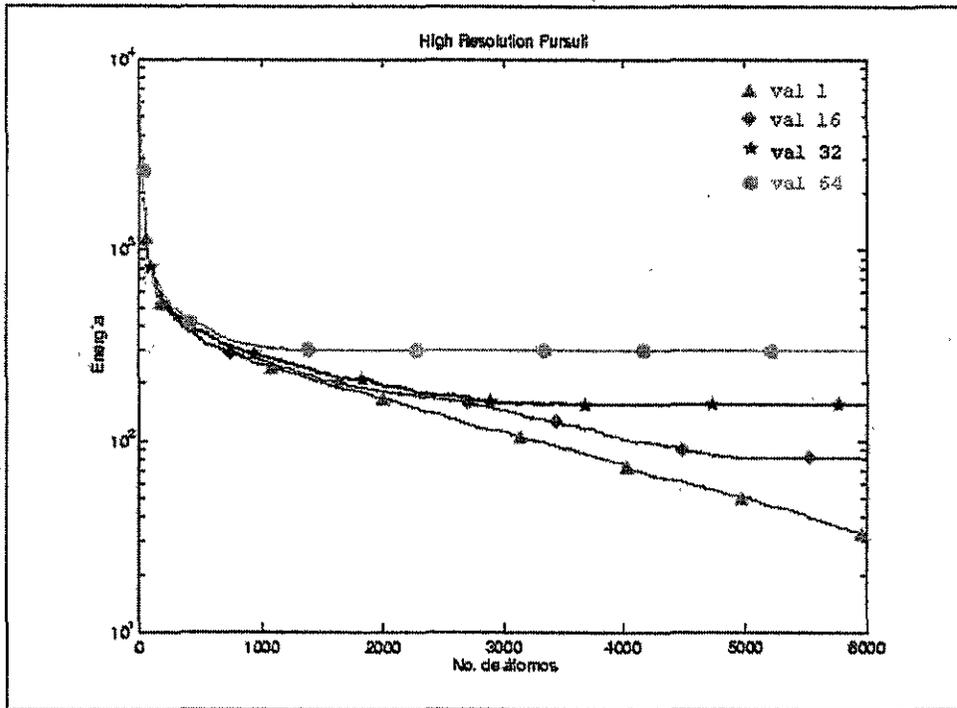


Figura 6.40. Efectos producidos por los diferentes valores del paso de cuantización en HRPR para la imagen de fluorescencia.

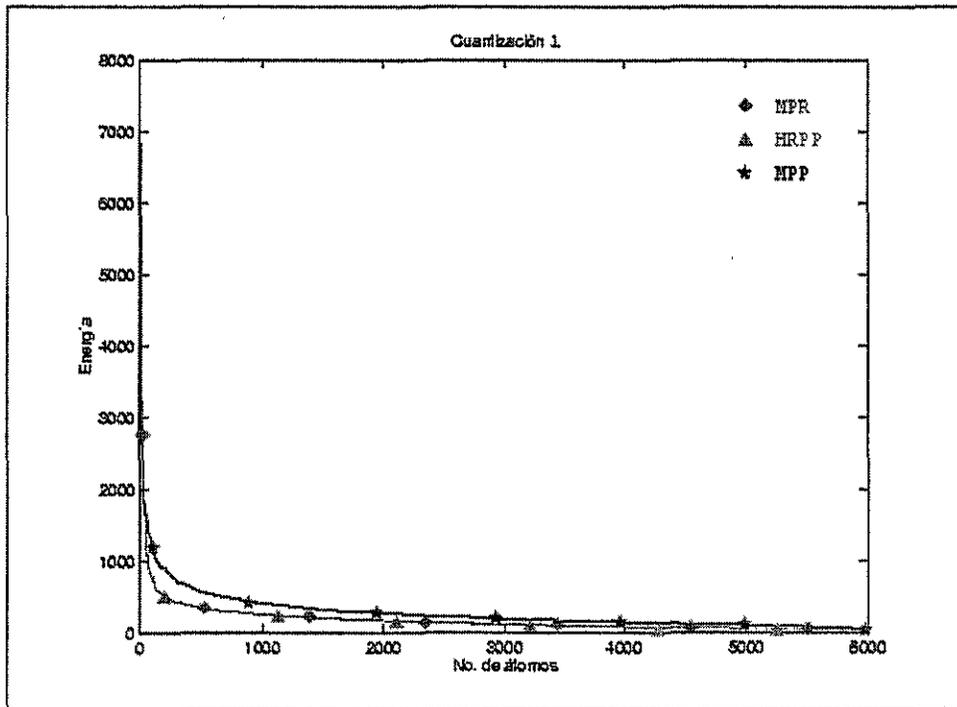


Figura 6.41. Efectos producidos cuando el valor del paso de cuantización en los tres modelos de codificación es 1.

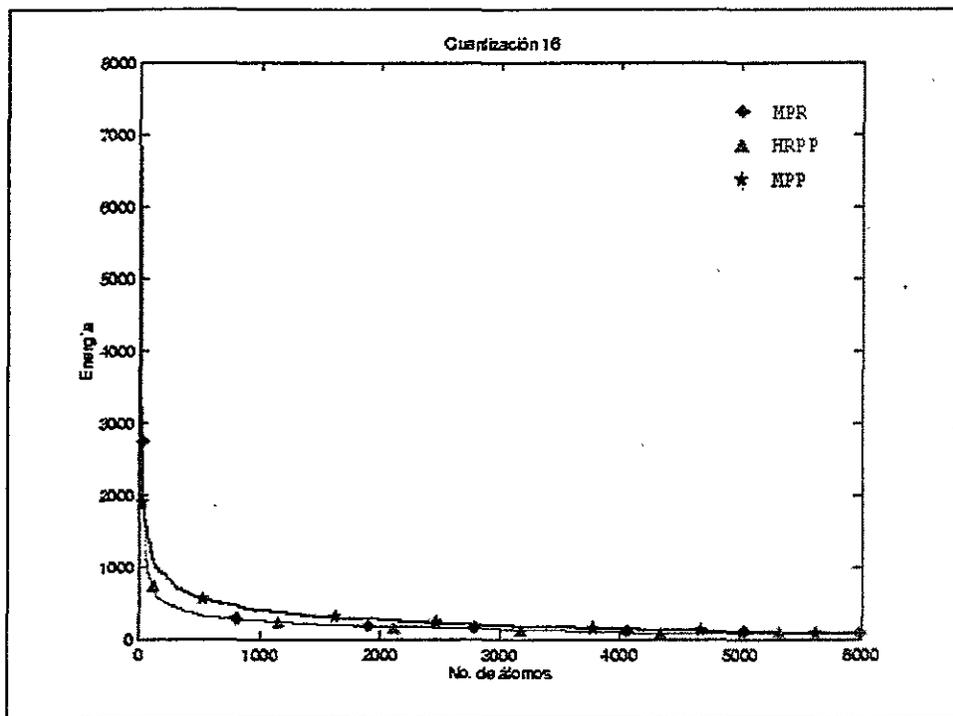


Figura 6.42. Efectos producidos cuando el valor del paso de cuantización en los tres modelos de codificación es 16.

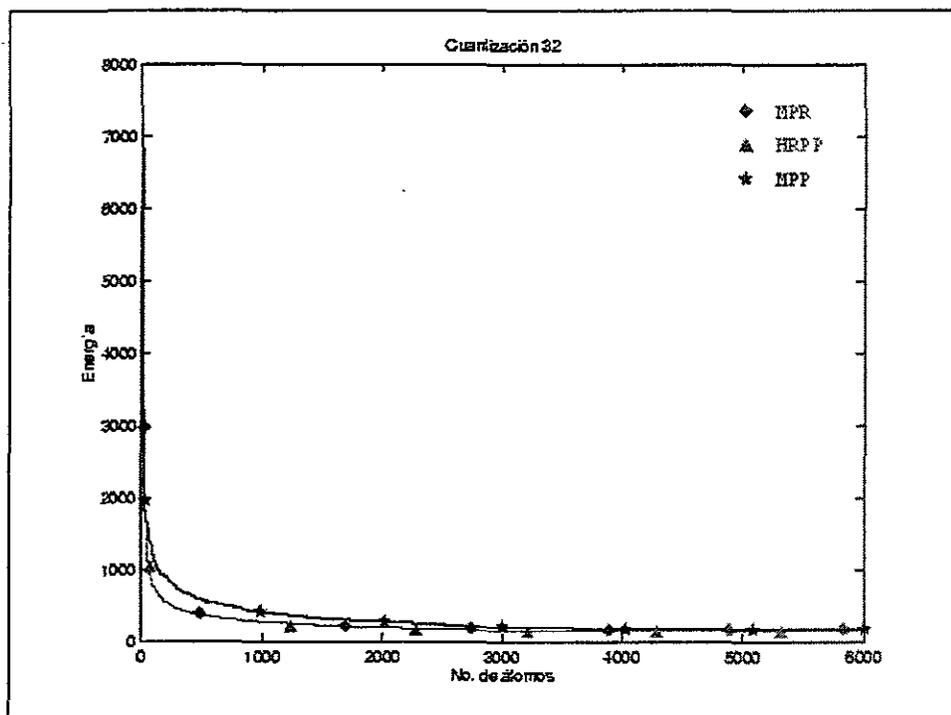


Figura 6.43. Efectos producidos cuando el valor del paso de cuantización en los tres modelos de codificación es 32.

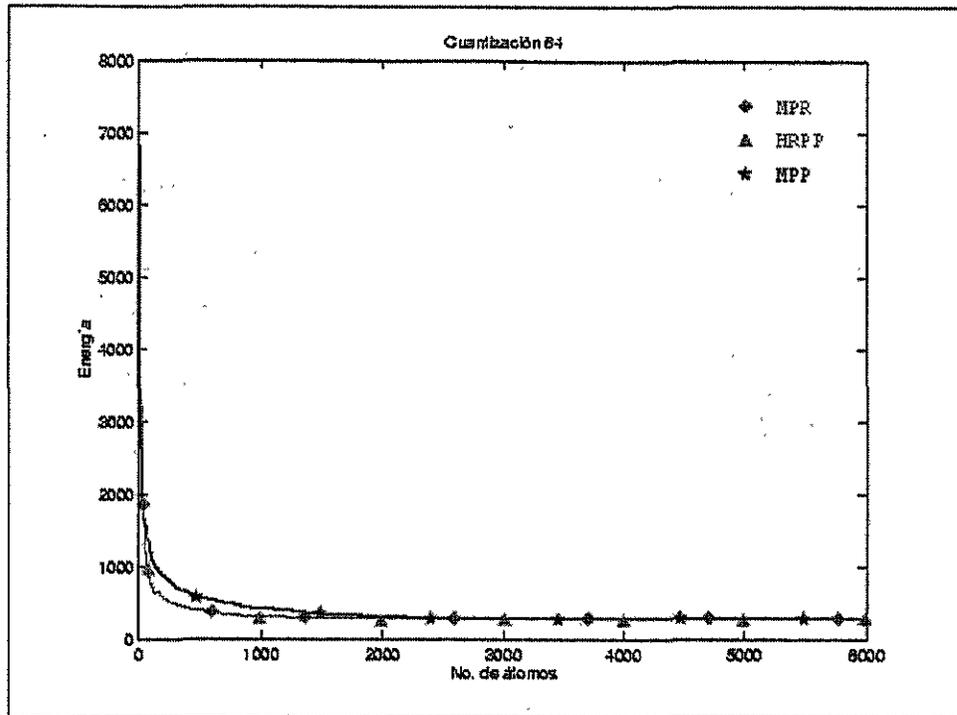


Figura 6.44. Efectos producidos cuando el valor del paso de cuantización en los tres modelos de codificación es 64.

De las gráficas en las Figuras 6.38, 6.39 y 6.40 respectivamente se tiene:

- MPP: 300 átomos.
- MPR: 200 átomos.
- HRPR: 150 átomos.

Por último, de las Figuras 6.41, 6.42, 6.43 y 6.44.

Mayor incidencia en mejores resultados: MPR y HRPR.

6.2.2. Laminar.

La imagen de la Figura 6.45 corresponde a la imagen de un cuerpo laminar y las reconstrucciones realizadas se muestran en las Figuras 6.46, 6.47 y 6.48, para MPP, MRP y HRPR respectivamente.

El cuerpo laminar dentro de la imagen corresponde a la mancha gris más grande y notoria. Además, en esta imagen se aprecian las vesículas sinápticas que corresponden a las regiones circulares más oscuras y las mitocondrias, de las cuales su contorno es la característica más notoria para observar su presencia. Junto con estos elementos, que son

los más importantes en la imagen, se encuentran los canales de alimentación los cuales son dos pequeños hilos que nacen en las esquinas superiores del cuerpo laminar.

Esta imagen, aumentada 20,000 veces, también se obtuvo con ayuda de un microscopio electrónico, y pertenece a una serie de imágenes a partir de las cuales se realiza una reconstrucción tridimensional para observar la manera en la cual el cuerpo laminar sintetiza sustancias y alimenta a las vesículas sinápticas.

En la Tabla 6.10 se presenta la energía de la imagen del cuerpo laminar, al igual que el número de átomos obtenidos durante el proceso y la energía residual una vez que ha concluido cada uno en su oportunidad y con sus parámetros correspondientes.

La Tabla 6.11 contiene todos los valores obtenidos para la medida del error cuadrático medio (MSE), la relación señal a ruido (SNR), la relación señal a ruido pico (PSNR), la escala de calidad para imágenes (PQS) y la tasa de compresión (bpp) para cada una de las imágenes reconstruidas del cuerpo laminar.

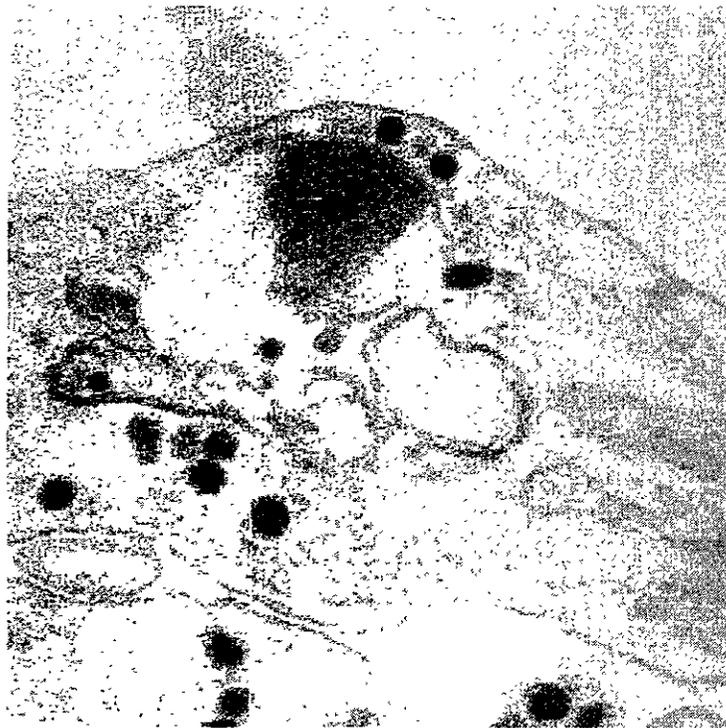


Figura 6.45. Laminar.

Imagen original Energía = 53,176			
MPP, <i>val</i> = 1 No. átomos = 6,000 Residual = 447.47	MPP, <i>val</i> = 16 No. átomos = 6,000 Residual = 451.22	MPP, <i>val</i> = 32 No. átomos = 6,000 Residual = 457.25	MPP, <i>val</i> = 64 No. átomos = 6,000 Residual = 482.14
MPR, <i>val</i> = 1 No. átomos = 6,000 Residual = 459.54	MPR, <i>val</i> = 16 No. átomos = 6,000 Residual = 463.81	MPR, <i>val</i> = 32 No. átomos = 6,000 Residual = 504.97	MPR, <i>val</i> = 64 No. átomos = 6,000 Residual = 562.11
HRPR, <i>val</i> = 1 No. átomos = 6,000 Residual = 458.11	HRPR, <i>val</i> = 16 No. átomos = 6,000 Residual = 471.86	HRPR, <i>val</i> = 32 No. átomos = 6,000 Residual = 494.83	HRPR, <i>val</i> = 64 No. átomos = 6,000 Residual = 554.269

Tabla 6.10. Conjunto de valores obtenidos para la imagen laminar.

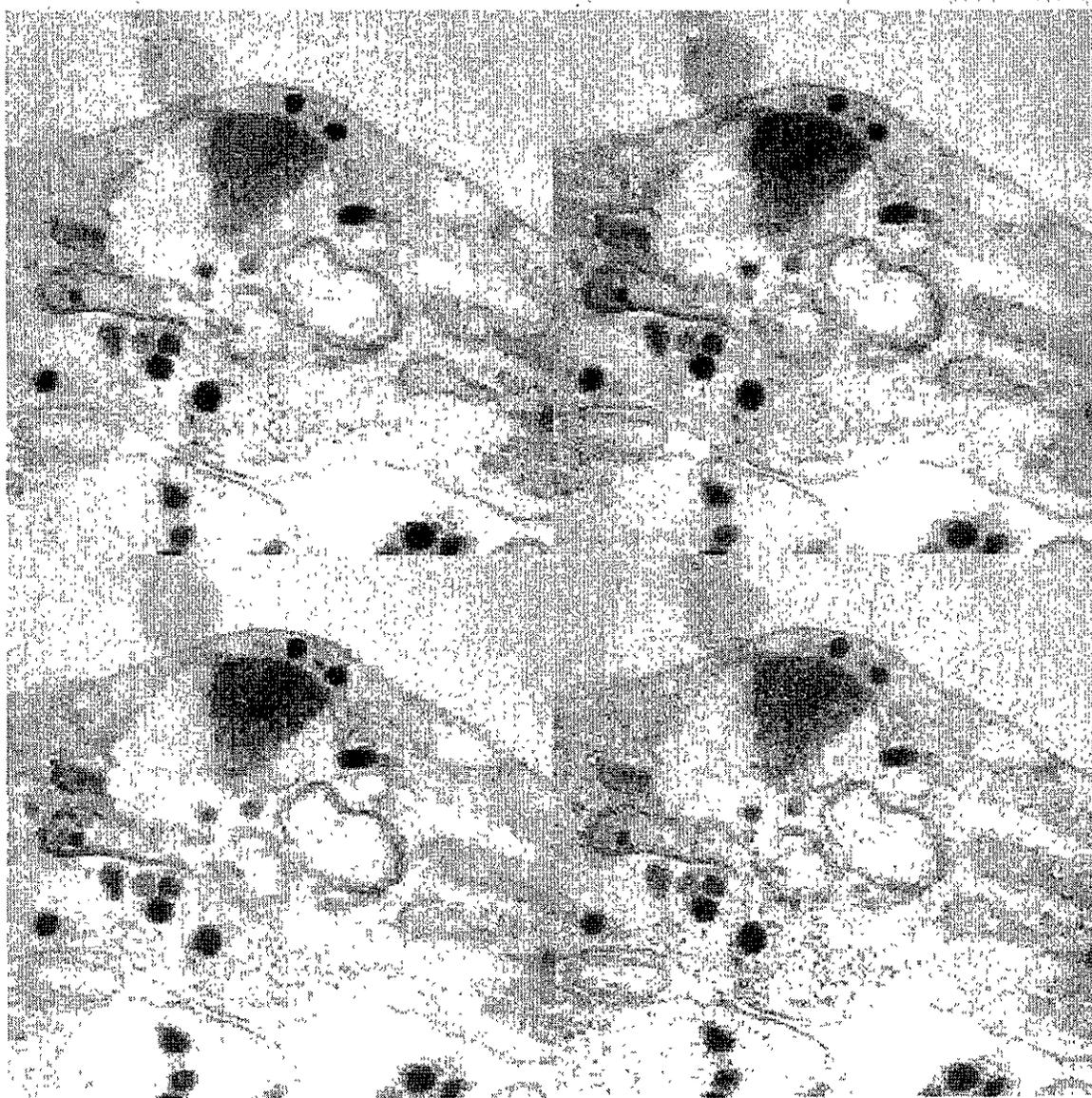


Figura 6.46. Imágenes del cuerpo laminar reconstruidas con MPP.

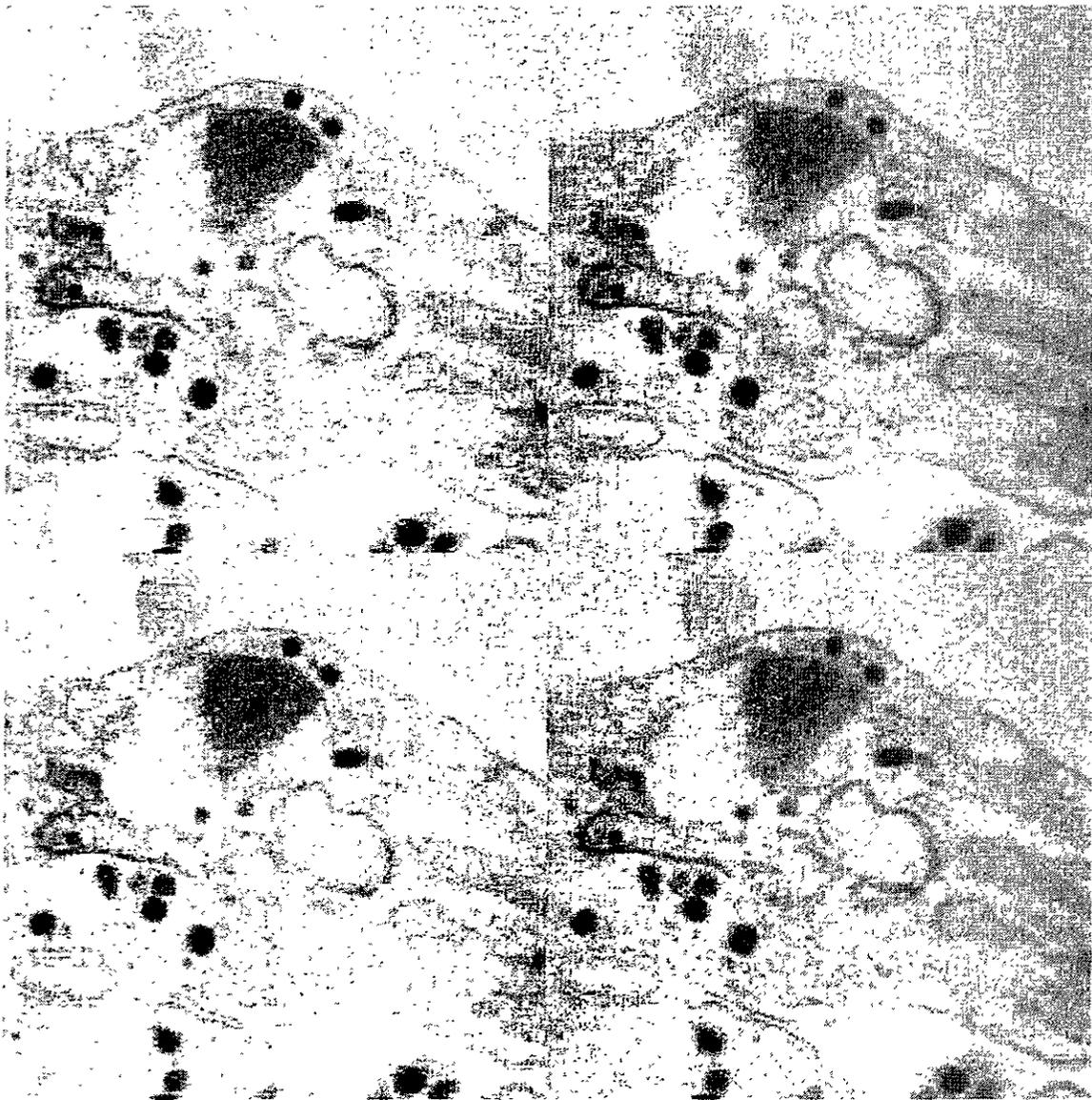


Figura 6.47. Imágenes del cuerpo laminar reconstruidas con MPR.

	Cuantización 1			Cuantización 16			Cuantización 32			Cuantización 64		
MSE	182.73	201.88	201.90	185.02	212.47	212.70	189.45	235.28	235.90	221.45	282.05	280.92
SNR	23.838	23.406	23.405	23.784	23.183	23.179	23.682	22.741	22.729	23.004	21.953	21.971
PSNR	25.513	25.080	25.079	25.459	24.858	24.853	25.356	24.415	24.404	24.678	23.628	23.645
PQS	4	5	5	4	4	4	4	4	4	3	3	3
bpp	0.9520	1.0350	1.0342	0.6560	0.7575	0.7463	0.5840	0.6900	0.6888	0.3760	0.6225	0.6211
	MPP	MPR	HRPR	MPP	MPR	HRPR	MPP	MPR	HRPR	MPP	MPR	HRPR

Tabla 6.11. Medidas de error, relación, calidad y compresión para la imagen laminar.

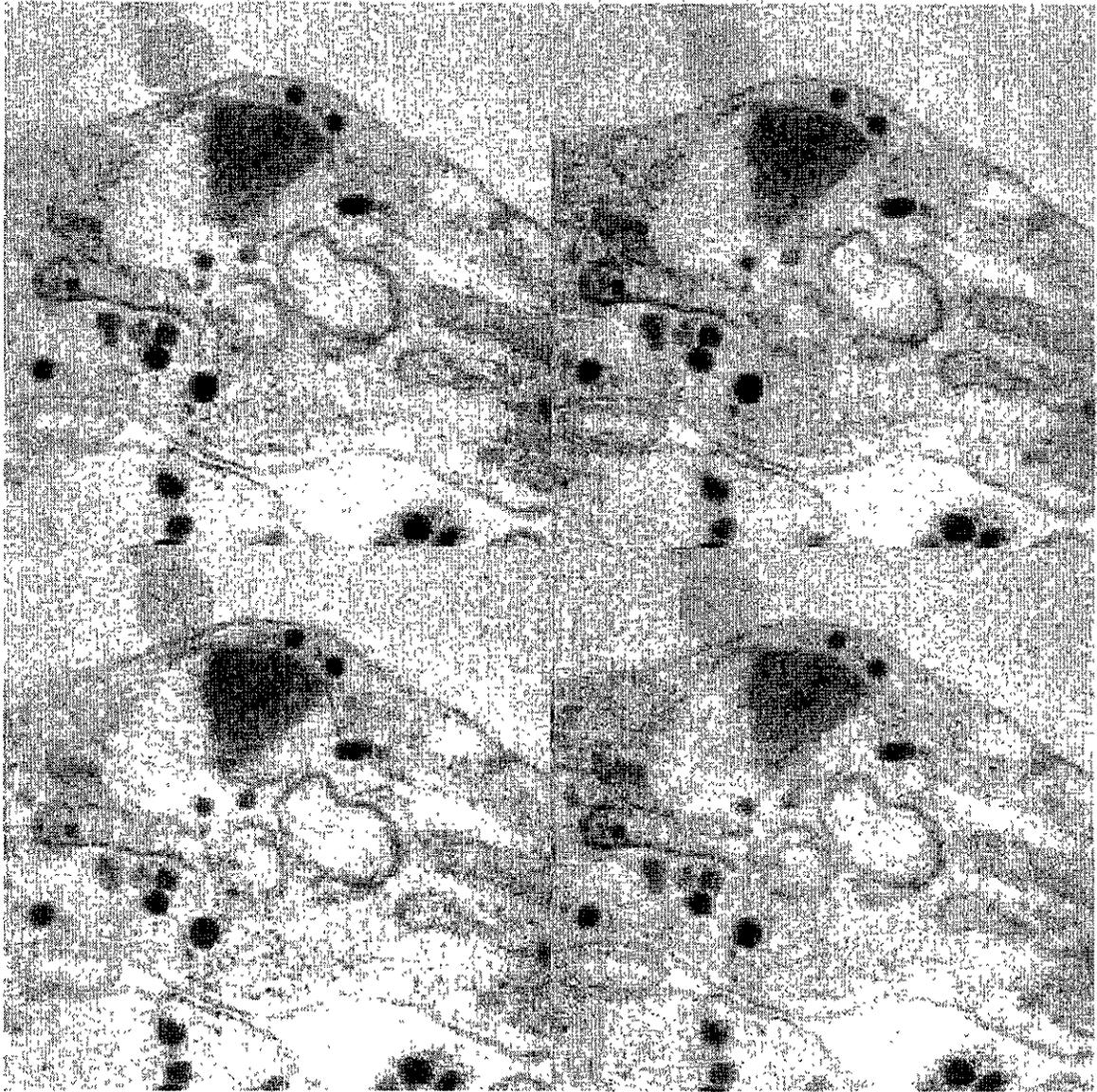


Figura 6.48. Imágenes del cuerpo laminar reconstruidas con HRPR.

En las Figuras 6.49, 6.50 y 6.51 se tiene las gráficas donde se puede observar la manera en la cual la energía de la imagen decrece de acuerdo al valor empleado para el paso de cuantización en cada uno de los modelos utilizados y conforme se van calculando los átomos de correspondencia.

Finalmente, en las gráficas de las Figura 6.52, 6.53, 6.54 y 6.55 se observa la manera en la cual la energía residual de la imagen decae en cada uno de los modelos cuando operan con las mismas condiciones de cuantización, esta medida se efectúa también con relación al número de átomos que se obtienen.

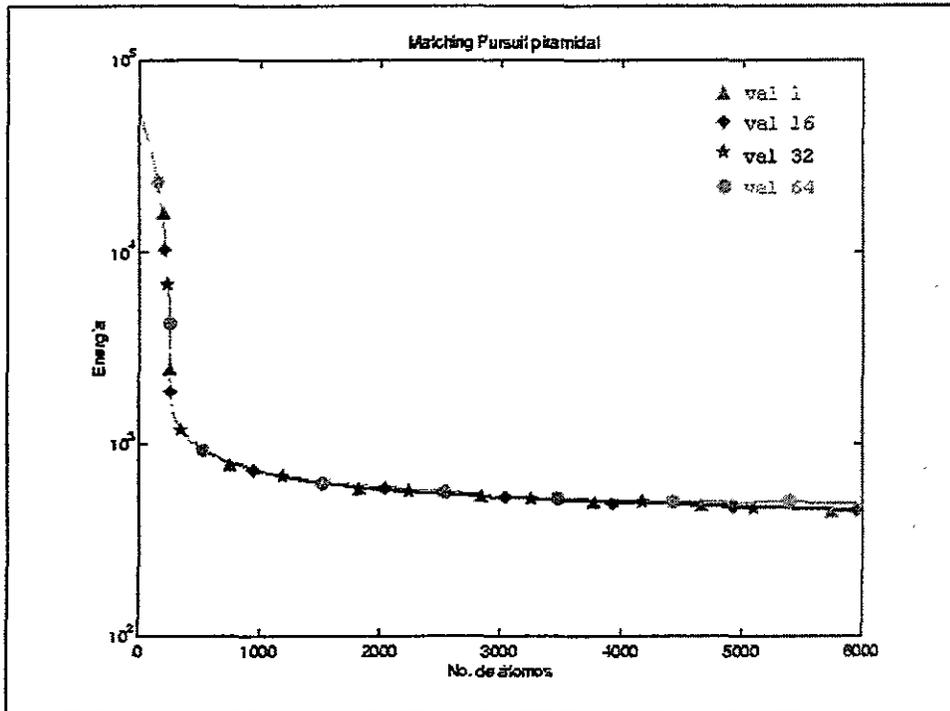


Figura 6.49. Efectos producidos por los diferentes valores del paso de cuantización en MPP para la imagen laminar.

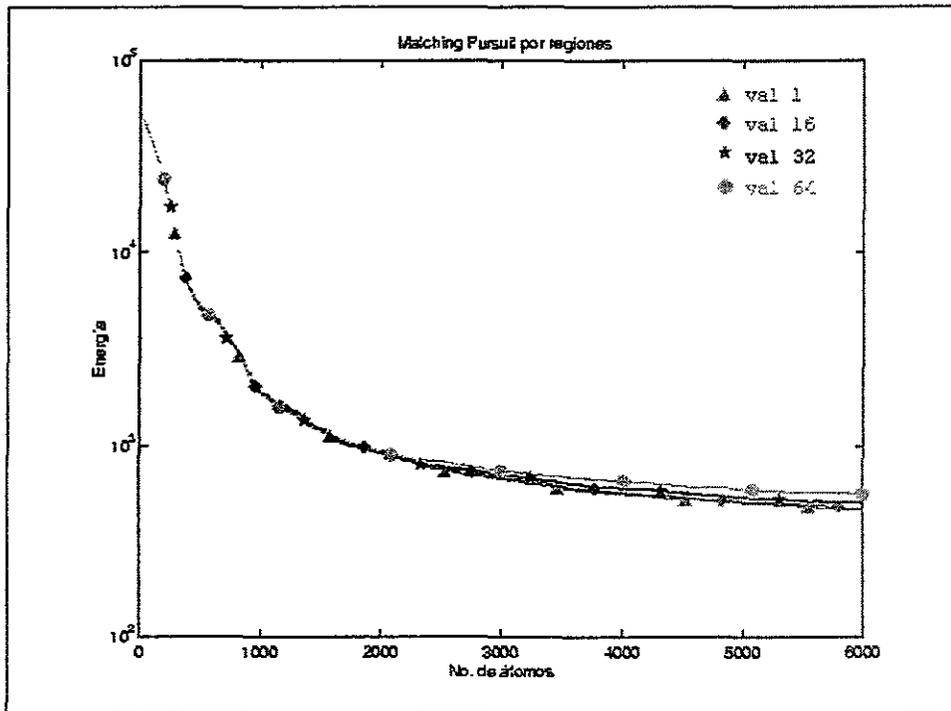


Figura 6.50. Efectos producidos por los diferentes valores del paso de cuantización en MPR para la imagen laminar.

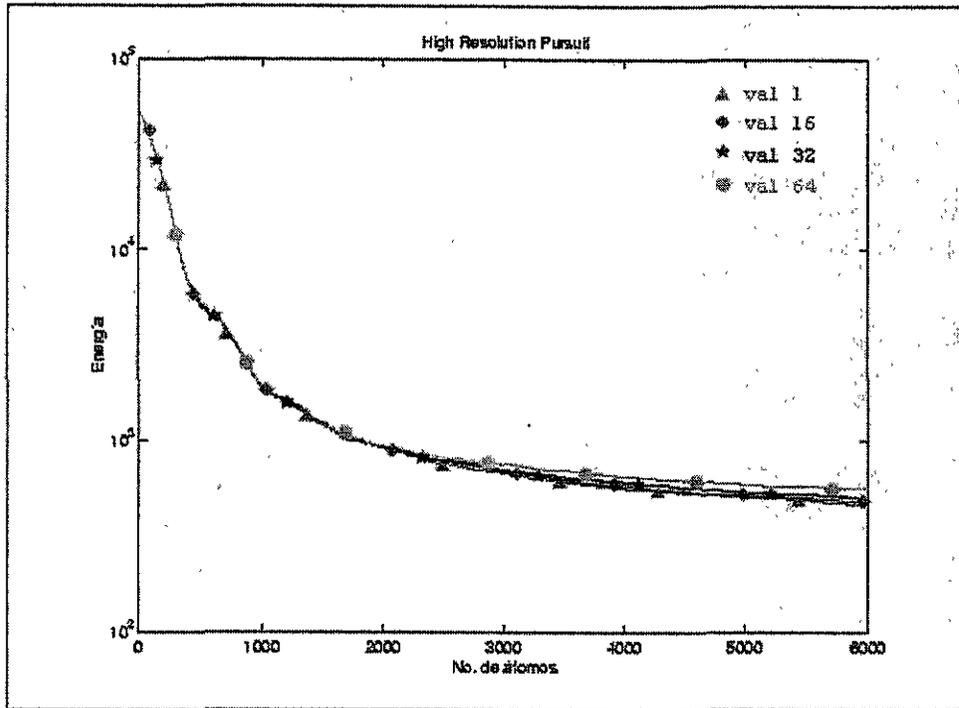


Figura 6.51. Efectos producidos por los diferentes valores del paso de cuantización en HRPR para la imagen laminar.

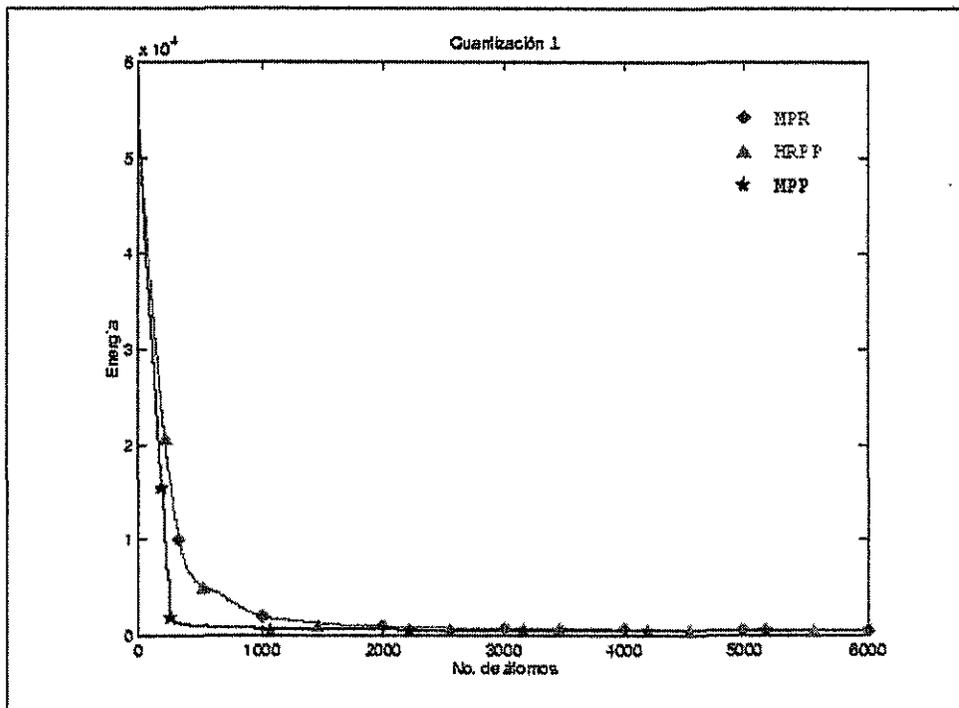


Figura 6.52. Efectos producidos cuando el valor del paso de cuantización en los tres modelos de codificación es 1.

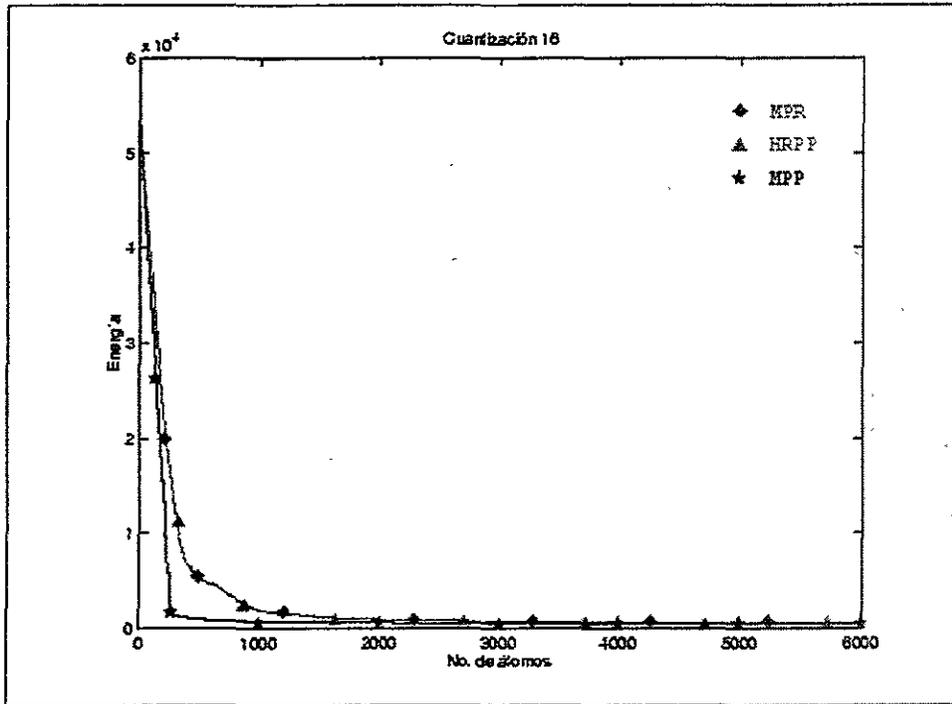


Figura 6.53. Efectos producidos cuando el valor del paso de cuantización en los tres modelos de codificación es 16.

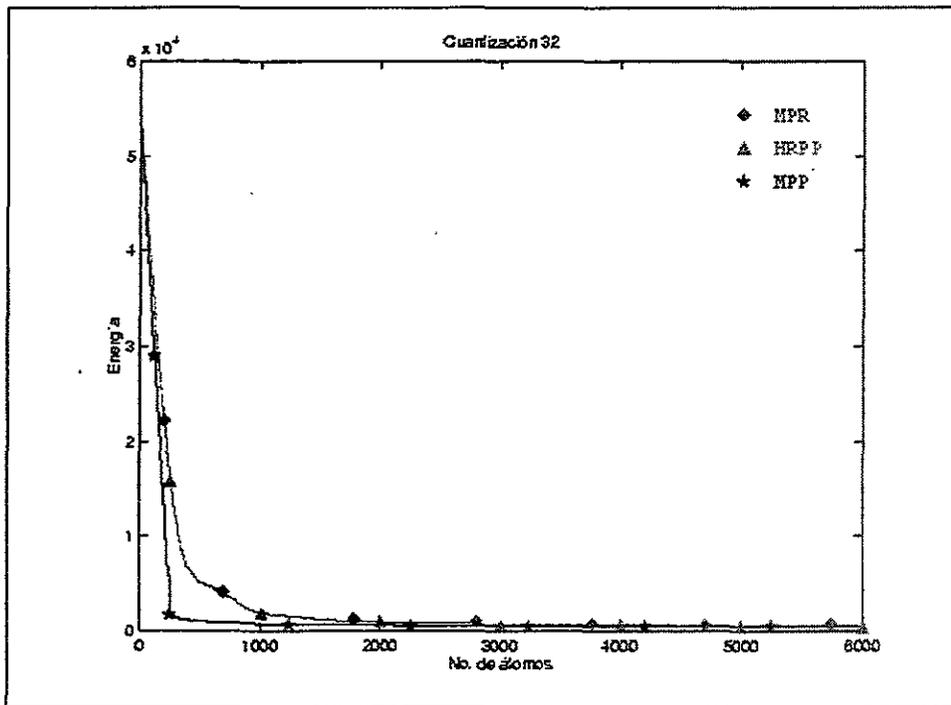


Figura 6.54. Efectos producidos cuando el valor del paso de cuantización en los tres modelos de codificación es 32.

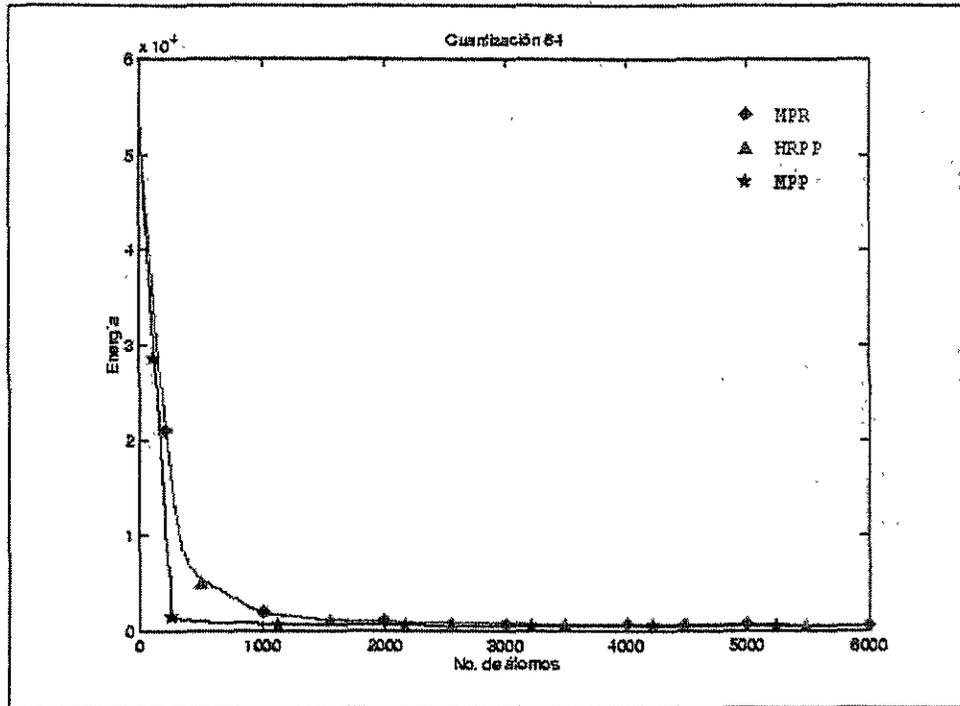


Figura 6.55. Efectos producidos cuando el valor del paso de cuantización en los tres modelos de codificación es 64.

En primer lugar definiremos las características principales de la imagen del cuerpo laminar de la manera más sencilla. Estas características son:

Energía de la imagen: 53,000.

Textura: 0 %

Número perceptual de regiones visibles: 25.

De la Tabla 6.11 se tiene:

Mayor incidencia en mejores resultados: MPP.

A partir de las Figuras 6.49, 6.50 y 6.51 se nota:

MPP: +6,000 átomos.

MPR: 2,200 átomos.

HRPP: 2,200 átomos.

Finalmente, analizando las gráficas de las Figuras 6.52, 6.53, 6.54 y 6.55 se encuentra:

Mayor incidencia en mejores resultados: MPP.

Referencias.

- [1] F. Bergeaud y S. Mallat, Matching Pursuit: adaptive representations of images and sounds, *Computational and applied mathematics*, Vol. 15, No. 2, pp. 97-109, 1996.
- [2] R L. Daubechies, Ten Lectures on wavelets, *SIAM*, 1992.

Conclusiones y perspectivas.

A partir de los resultados que se han obtenido, y con el análisis efectuado a dichos resultados, se está en la posibilidad de emitir una serie de conclusiones que permitan evaluar el trabajo realizado. Como se ha visto, en este trabajo se han propuesto un par de modelos de codificación basados en técnicas de selección adaptable, ambos emplean la técnica "Matching Pursuit", y uno de ellos, además, la reemplaza por "High Resolution Pursuit" para de esta manera, poder establecer una diferencia entre las dos técnicas de selección adaptable al realizar la misma labor.

Como se mencionó en el primer capítulo, estos modelos contribuyen al estudio de la codificación de imágenes, y pasan a formar parte de la gran cantidad de técnicas ya existentes. Además, debido a la originalidad de estos trabajos, se está presentando una de las miles de formas con la cual, utilizando un poco de ingenio, se pueden acondicionar y utilizar las bases teóricas que los científicos de la física y matemáticas proporcionan, para desarrollar una aplicación útil y práctica.

Adicionalmente, en este capítulo se mencionan las perspectivas y el trabajo futuro que se prevé para estos modelos de codificación.

7.1 Conclusiones.

En esta sección, culminante para este trabajo, al momento de formular todas las conclusiones a las que se llegan, se presentan diferentes perspectivas y puntos de vista. Esta diversidad se debe a que las conclusiones que se emiten pueden verse desde un punto de vista particular, general, de aplicación, comparativo, basado en la experiencia, ó a futuro, entre otros más. De ahí la dificultad de emitir conclusiones absolutas, no sólo para este trabajo, sino para cualquier otro trabajo de investigación que se llegue a desarrollar. Porque, así como un problema puede atacarse desde diferentes ángulos, las soluciones que se encuentren no llegan a ser únicas, y mucho menos concluyentes. Indiscutiblemente,

pueden existir excelentes soluciones, unas mejores que otras, pero esa excelencia no deja de ser de carácter relativo.

En las primeras etapas de este trabajo, con la literatura consultada, el objetivo que se tenía en mente era simplemente desarrollar un modelo de codificación de imágenes basado en la técnica de selección "High Resolution Pursuit", dicho modelo debería ser novedoso y proporcionaría muy buenos resultados. Se había considerado el uso de "High Resolution Pursuit" debido a que esta técnica es novedosa, de reciente creación, y en primera instancia, parecía mejor que cualquier otra técnica conocida, específicamente, mejor que "Matching Pursuit" y "Basis Pursuit". Sin embargo, el problema principal que se encontró al comenzar a trabajar con esta técnica, era un excesivo tiempo de cómputo, quizás esta técnica no es muy compleja desde el punto de vista computacional, sin embargo, su consumo de recursos de cómputo sí lo es.

En ese momento era necesario replantear el trabajo que se tenía en mente. Previamente se tenía previsto trabajar con "Basis Pursuit", sin embargo, con base en el trabajo y los estudios desarrollados para esta técnica en [1], "Basis Pursuit" fue descartada por los mismos motivos que ahora se presentaban: un excesivo tiempo de cómputo, el cual se considera poco práctico en la codificación de imágenes, e incluso en cualquier trabajo.

Después de reflexionar un poco, pero ahora con una mayor cantidad de literatura consultada y de pruebas realizadas, se decide seguir adelante, desarrollando un par de modelos de codificación: piramidal y por regiones, en los cuales, se emplearía "Matching Pursuit" por su eficiencia computacional y, además, en aquel modelo que permitiera un mejor control y desempeño práctico, se desarrollaría "High Resolution Pursuit" para fines comparativos. El modelo de codificación por regiones resultó ser el adecuado para esta tarea. Se pensó además que estos modelos deberían involucrar técnicas de cuantización, de selección de la "mejor" base de descomposición (el modelo piramidal MPP), y de búsqueda de energía (los modelo por regiones MPR y HRPR) para complementarlos.

Por otro lado, la literatura insinuaba que un buen diccionario de descomposición para imágenes era aquel constituido por una gama de funciones de Gabor, argumentando, entre otros aspectos de carácter biológico, que una función de Gabor es equivalente a la respuesta que presenta la corteza visual de los mamíferos [2, 3, 4]. Sin embargo, este tipo de diccionario, durante las pruebas que se realizaron con MPR, introducía áreas de energía no deseadas que complicaban el desempeño del algoritmo. Este fenómeno ya era conocido y había sido estudiado en [5]. Por esa razón, fue necesario limitar este diccionario, conservando solamente aquellas funciones que en la práctica no introducían esas áreas de energía no deseadas. Se observó que los átomos que se conservaron se asemejaban a funciones "Spline", de modo que, una vez definido el diccionario a emplear para el modelo por regiones, se consideró el empleo de funciones "Spline" para el modelo piramidal MPP.

Finalmente, con base en las observaciones y la experiencia obtenida durante el desarrollo de este trabajo, se puede decir que la realización y desempeño global de los modelos desarrollados, es equivalente, con las mismas dificultades y complejidad. Decir

que un modelo es mejor que el otro desde esta perspectiva está fuera de contexto. Sin embargo, considerando el modelo por regiones, indudablemente, el desempeño logrado con MPR es mucho mejor que el obtenido con HRPR, esta aseveración se realiza considerando simplemente el factor tiempo de procesamiento, en el cual se presenta una diferencia notable entre ellos, no así en la complejidad computacional, cuyas diferencias son despreciables.

Los modelos presentados en este trabajo, fueron desarrollados en Matlab™.

Por otra parte, con base en los resultados obtenidos en el capítulo anterior, se puede decir que el modelo MPP, en términos generales es el que mejor resultados proporciona. Sin embargo, entrando a detalle, vemos que MPR y HRPR llegan a superar a MPP en alguna ocasión. Por ejemplo, para imágenes con las características presentes en la imagen de fluorescencia, se observa que tanto MPR como HRPR presentan mejores resultados que MPP.

Todo esto se puede ver de la siguiente manera: considerando los resultados obtenidos, tenemos que si la energía de la imagen es mayor a 20,000, MPP es mejor modelo, pero si la energía de la imagen es menor a 20,000, MPR y HRPR superan a MPP.

Desde otra perspectiva, si el número de regiones en la imagen es mayor a 10, MPP es mejor modelo, pero con un número de regiones menor a 10, nuevamente MPR y HRPR son mejores. La textura en este caso, no llega a ser factor influyente en primera instancia para el desempeño y comparación de los modelos propuestos.

Considerando la velocidad con la cual los modelos extraen energía de la imagen, no se puede establecer la supremacía de alguno de los modelos. Pero, si se desea saber a que modelo afecta en mayor medida el valor del paso de cuantización, en apariencia MPR es el modelo más sensible, seguido por HRPR.

Otro punto de estudio se encuentra al determinar el valor adecuado del paso de cuantización. En esta ocasión, en términos generales, utilizar un valor $val = 16$, ó $val = 32$ parece una buena opción ya que presentan buenos resultados.

Como se manejan diferentes variables, tales como el tipo de diccionario, el paso de cuantización y el criterio de paro, poder establecer como constante alguna de estas variables no es del todo trivial. Por ejemplo, se requiere en principio un estudio detallado de las características de la imagen para definir adecuadamente el diccionario que utilizaría el modelo, y esto se puede lograr, tal y como en algún momento se mencionó, empleando una estrategia de entrenamiento, o bien, empleando funciones de costo similares a las presentadas en 5.3.1.1. Y a continuación, se realizaría un estudio a los valores de los coeficientes obtenidos para determinar el valor del paso de cuantización adecuado.

Ahora bien, la velocidad con la cual decrece la energía de la imagen podría ser otro factor para determinar que modelo presenta un mejor desempeño, pero esta medida varía

tanto entre modelos de codificación como entre los valores de los pasos de cuantización y las imágenes empleadas.

En este trabajo se presentan solamente tres modelos, sin embargo se mencionó que estas técnicas son muy flexibles, y con ayuda del ingenio personal, se pueden crear muchos más modelos, basta consultar la literatura reciente donde se encuentran algunos otros modelos de codificación que emplean selección adaptable [2, 3, 4, 6, 7, 8, 9]. Los modelos referidos son contemporáneos de los que aquí se presentan, es decir, en este momento, la aplicación de las técnicas de selección adaptable para la codificación de imágenes no tiene más de un lustro de inicio, y los modelos presentados en los trabajos referidos comenzaron a desarrollarse al mismo tiempo que los que aquí se presentan.

Lamentablemente, los resultados obtenidos con los modelos presentados en la literatura no permiten realizar una comparación efectiva con los obtenidos en este trabajo. Esos otros trabajos simplemente presentan buenos resultados, y además, se puede notar que emplean imágenes de características diversas a las utilizadas en este trabajo, o hacen uso de otras imágenes.

Por ejemplo, en [2, 3, 4] se emplea MPI ("Matching Pursuit of Images") con la misma imagen Lenna que se ha utilizado en este trabajo, pero sólo presentan las reconstrucciones con un número determinado de átomos sin hacer mención de alguna medida de calidad, simplemente se habla de una buena calidad visual en la reconstrucción. En la Tabla 7.1 se muestran los resultados que presentan los modelos de codificación de imágenes que emplean las técnicas de selección adaptable hallados en la literatura.

En [6] se presenta el modelo PP ("Perceptual Pruning"), y PP con USQ ("Uniform Scalar Cuantizer"). En [7] se tiene al modelo PP-QT ("Projection Pursuit Quadtree") y PP-BST ("Projection Pursuit Binary Space Partitioning tree"), mientras en [8] se introducen los modelos MP-QT ("Matching Pursuit Quadtree"), y BP-QT ("Basis Pursuit Quadtree"). Finalmente, en [9] se introduce al modelo SOMP ("Segmented Orthogonal Matching Pursuit").

Desde esta perspectiva, se puede decir que no caben comparaciones directas con los modelos de codificación que emplean selección adaptable encontrados en la literatura, y los modelos propuestos en este trabajo. Principalmente, porque, mas allá de los resultados, es primordial conocer los tiempos de procesamiento utilizados para obtener dichos resultados, las características de los diccionarios empleados y algún otro detalle que en general, no llega a ser mencionado, como las características físicas del equipo utilizado ó la plataforma de desarrollo. Sin embargo, con la finalidad de tener una referencia, en la Tabla 7.2 se presentan los valores de la tasa de compresión obtenidos con los modelos propuestos y JPEG. Estos valores se obtuvieron para resultados estrictamente iguales desde el punto de vista de calidad objetiva. Se sabe que JPEG emplea un tiempo de procesamiento muy inferior a los modelos propuestos, sin embargo, observamos que los modelos propuestos presentan una tasa de compresión superior.

Modelo de codificación.	Imágenes de prueba.	Tasa de compresión.	Calidad objetiva.	Observaciones.
MPI [2, 3, 4]	Lenna, 256.	1.5 bpp	NA	Buena calidad.
PP [6]	Lenna, región de 256	0.57 bpp 0.36 bpp	SNR = 27.45 NA	Mejor que JPEG.
PP-USQ [6]	Lenna, región de 256	0.57 bpp	SNR = 27.26	Mejor que JPEG.
PP-QT [7]	Lenna 512	0.125 bpp	PSNR = 27.25	Mejor que JPEG.
PP-BST [7]	Lenna 512	0.125 bpp	PSNR = 27.65	Mejor que JPEG.
MP-QT [8]	Lenna 512	0.125 bpp	PSNR = 30.49	Mejor que JPEG y EZW.
BP-QT [8]	Lenna 512	0.125 bpp.	PSNR = 30.58	Mejor que JPEG y EZW.
SOMP [9]	Peppers	NA	NA	Visualmente mejor que JPEG.

Tabla 7.1 Resultados presentados por modelos de codificación con selección adaptable.

Imagen	HRPR	JPEG	MPR	JPEG	MPP	JPEG
Lenna	1.0188	4.3316	1.0380	4.0249	1.0000	3.4815
Bárbara	0.9600	4.9662	0.9600	4.9642	0.9520	4.2402
Cameraman	0.9660	4.4965	0.9660	4.4977	0.8800	2.8544
FFM01	0.9600	1.5030	0.9660	1.5030	0.8800	1.4959
FFM02	1.0342	5.1567	1.0350	5.1553	0.9520	4.3850

Tabla 7.2 Tasa de compresión (bpp) imágenes reconstruidas.

Con respecto a algún otro modelo de codificación de imágenes, tenemos por ejemplo, que los resultados obtenidos en este trabajo, superan a los presentados en [10], no así a los presentados en [11]. Lenna fue empleada para las pruebas. Estas diferencias se pueden apreciar en la Tabla 7.3 y están referidas a la métricas PSNR y SNR.

Modelo	bpp	PSNR (dB)	SNR (dB)
WT#5 [10]	0.50	30.23	NA
WT#8 [10]	0.50	23.19	NA
MPP	0.50	32.50	25.25
MPR	0.50	31.00	22.60
HRPR	0.50	30.25	21.40
TG [11]	0.44	NA	29.20
TG-DCT [11]	0.44	NA	30.00
MPP	0.44	32.30	25.00
MPR	0.44	30.90	22.50
HRPR	0.44	29.90	21.20

Tabla 7.3 Valores PSNR y SNR para Lenna a 0.44 bpp y 0.50 bpp.

Con relación a los resultados obtenidos en este trabajo, después de realizar algunas comparaciones, podemos decir que los modelos propuestos presentan resultados que caen dentro de un rango aceptable.

Finalmente, simplemente observemos que se ha ido más allá del objetivo inicialmente planteado para este trabajo: se han presentado tres modelos de codificación de imágenes novedosos, lo más completos posible, los cuales, con una alta tasa de compresión, producen reconstrucciones de muy buena calidad visual, con muy buenas perspectivas.

7.2 Perspectivas.

Los modelos presentados pueden mejorarse con un conocimiento previo de las propiedades de la imagen que nos permita fijar, esencialmente, lo mejor posible el diccionario a emplear, el valor del paso de cuantización y el criterio de paro.

Actualmente se han desarrollado funciones de costo aplicables a la codificación. El introducir dichas funciones de costo adecuadamente en las técnicas de selección adaptable daría más flexibilidad a los modelos propuestos y produciría, probablemente, mejores resultados.

Para efectos de transmisión, la distorsión juega un papel importante. Una buena opción para estos modelos sería introducir una etapa de retroalimentación que permita controlar dicha distorsión.

Por otro lado, si se observa el comportamiento que presenta la energía residual durante el proceso, este comportamiento muchas veces es deseado para efectuar tareas de control. Al fin y al cabo, las técnicas de selección adaptable se usan, y se han usado para procesamiento de todo tipo de señales, y estas técnicas no están casadas con un área de conocimiento en específico, sino por el contrario, dado que las señales se pueden encontrar en casi todas las áreas de conocimiento del hombre, estas técnicas se pueden utilizar y adaptar lo mejor posible para solucionar los problemas que se enfrentan en algún otra área determinada.

Finalmente, los modelos presentados son pensados para emplearse en codificación de imágenes desde un punto de vista general, es decir, para todo tipo de imágenes. Sin embargo, si se emplearan estos modelos para un tipo de imagen en específico, se podrían fijar de manera particular y definitiva algunas de las variables que intervienen en el modelo para obtener mejores resultados.

Las perspectivas dependen de la imaginación. Todas las herramientas, aunque parezcan hechas para realizar una tarea específica, pueden emplearse para otras tareas, y esta virtud depende meramente de quien las utilice. La historia del hombre está llena de este tipo ejemplos, basta echar un vistazo al pasado.

Referencias.

- [1] R. Montufar-Chaveznavia y F. García-Ugalde, Image Analysis by Basis Pursuit, Proc. of SPIE, Vol. , 1997.
- [2] F. Bergeaud y S. Mallat, Matching pursuit of images, *IEEE Int. Symp. on TFTS Analysis*, pp. 330-333, 1994.
- [3] F. Bergeaud y S. Mallat, Processing images and sound with matching pursuits, *SPIE*, Vol. 2569, pp. 2-13, Septiembre 1995.
- [4] F. Bergeaud y S. Mallat, Matching pursuit: adaptive representations of images and sounds, *Computational and applied mathematics*, Vol. 15, No. 2, pp. 97-109, 1996.
- [5] S. Jaggi, Multiscale geometric feature extraction and object recognition, *Tesis Doctoral del MIT*, LIDS-TH-2384, Febrero 1997.
- [6] M. J. Horowitz y D. L. Neuhoff, Image coding by matching pursuit and perceptual pruning, *IEEE 0-8186-8183-7/97*, pp. 654-657, 1997.
- [7] H. R. Rabiee, R. L. Kashyap y S. R. Safavian, Multiresolution segmentation-based image coding with hierarchical data structures, *IEEE ICASSP'96, Atlanta GA*, pp. 1870-1873, 1996.
- [8] H. R. Rabiee, R. L. Kashyap y S. R. Safavian, Adaptive multiresolution image coding with matching and basis pursuits, *IEEE ICIP'96, Laussane, Suiza*, pp. 273-276, 1996.
- [9] H. R. Rabiee, R. L. Kashyap y S. R. Safavian, Adaptive image representations with segmented orthogonal matching pursuit, *IEEE 0-8186-8821-1/98*, pp. 233-236, 1998.
- [10] J. D. Villasenor, B. Belzer y J. Liao, Wavelet filter evaluation for image compression, *IEEE Trans. on Image Processing*, Agosto 1995.
- [11] H. Wang y H. Yan, Efficient implementation of Gabor transforms for image compression, *Electronic letters*, Vol. 28, No. 8, pp. 870-871, 23 de Abril de 1992.