

Estudio de dos estructuras neuronales feed-forward para la compresión de imágenes digitales

Study of two neural feed-forward structures for digital image compression

*Andrés Eduardo Gaona Barrera**, *Néstor Andrés Lugo Currea*, *Álvaro Fernando Roldán Hernández*

Laboratorio de Automática, Microelectrónica e Inteligencia Computacional, LAMIC. Facultad de Ingeniería. Universidad Distrital Francisco José de Caldas. Carrera 7 N.º 40 – 53. Bogotá D.C. Colombia.

(Recibido el 03 de mayo de 2012. Aceptado el 6 de noviembre de 2012)

Resumen

Este documento muestra y explica el proceso para comprimir imágenes, en escala de grises y a color, usando dos topologías neuronales: función de imagen y embudo. Para el análisis de los esquemas neuronales se consideran el número de neuronas y capas, tipo de imagen, tamaño y cantidad de bloques durante el entrenamiento; con el fin dar soporte experimental a las arquitecturas neuronales.

También se analizan criterios de calidad de la imagen obtenida, como relación pico señal a ruido (PSNR) y tasa de compresión. Se evidencia la importancia de la selección de parámetros evaluados en calidad y tiempo de compresión. El proceso de experimentación muestra que la arquitectura tipo embudo permite lograr valores superiores a 35dB en términos de PSNR y 2 bits por pixel en imágenes en grises o 3 bpp en imágenes a color, con tiempos inferiores a 3 segundos para imágenes menores a 1 mega pixel. Finalmente, se realizan algunas recomendaciones basadas en la metodología empleada cuando se desee realizar compresión de imágenes con redes feed-forward entorno a la selección de parámetros de la arquitectura, al pre-procesamiento de la imagen y al entrenamiento de la red.

----- *Palabras clave:* Compresión con pérdidas, compresión digital de imágenes, redes neuronales feedforward, relación pico señal a ruido, tasa de compresión.

* Autor de correspondencia: teléfono: +57 + 1 + 344 01 15, correo electrónico: aegaona@udistrital.edu.co (A. Gaona)

Abstract

This paper shows and explains the process implemented for the development of feed-forward neural networks with the aim of compress digital image color. It sets out some traditional techniques and develops two topologies to implement feed forward.

During the development of networks, the items that are considered: number of layers, number of neurons, image type, size and number of blocks to train, to optimize performance during final training.

It also discusses the standard quality of the image obtained, as peak signal noise relation (PSNR) and compression, which typically obtain values above 35dB in terms of PSNR and 2 bits per pixel in gray or 3 bpp color images, with maximum time of 3 seconds for images less than 1 mega pixel. Finally out some drawbacks and presents conclusions of this type of compression.

----- *Keywords:* Compression rate, digital image compression, lossy compression, feedforward neural networks, peak signal noise relation

Introducción

Las imágenes digitales están compuestas por píxeles y cada uno se representa con 24 bits en el formato RGB o YCbCr, por lo que se tienen 16 millones de posibles colores en una imagen estándar. Los algoritmos de compresión representan esa información con menos datos, si la imagen después de comprimirla no es exacta a la original, se denomina con pérdidas o no reversible, de lo contrario se conoce como compresión sin pérdidas o reversible [1].

El principal inconveniente de las imágenes digitales es la cantidad de datos para su representación y así el tamaño de los archivos generados. Al usar técnicas de compresión, se aumentan las imágenes que se pueden almacenar en disco, disminuye el tiempo y la posibilidad de error en la de recepción por un canal ruidoso. Estas características generan interés en la evolución de las técnicas de compresión [1].

Técnicas tradicionales como JPG y PNG han sido usadas en la compresión de imágenes por su efectividad, pero hoy en día son mayores las demandas de compresión, en tiempos razonablemente bajos y manteniendo la calidad visual de la imagen. Las redes neuronales

artificiales (RNA) pueden satisfacer estas necesidades, ya que permiten recibir y procesar extensas cantidades de información en tiempo real, lo que las convierte en una alternativa eficaz. Cuando se emplean RNA, el problema se extiende a la búsqueda de parámetros de la topología de los cuales no existe un conocimiento a priori y que influya en las medidas de desempeño de la imagen tales como el PSNR o la tasa de compresión.

Por ejemplo, en [2] se usa el algoritmo Back propagation Al-Alaoui con error cuadrático medio, en el que mejoran la tasa de convergencia y obtienen un rendimiento en términos de PSNR superior en 2 dB. En [3] se obtienen bloques con la transformada wavelet para reducir la representación de la imagen e ingresarlos a la red. En [4] presentan una síntesis de varias técnicas de compresión como mapas auto-organizados y redes back-propagation, así mismo las comparan con técnicas tradicionales.

La transformada wavelet y ecualización de histograma se implementa en [5] y se aplica en el pre-procesamiento de mamografías. En [6] mejoran la tasa y el tiempo de compresión mediante una función de distribución acumulativa, con la que estiman los valores de píxeles vecinos. En [7]

usan una capa oculta, con la detección de bordes y umbralización reducen el tamaño de la imagen inicial, esta es ingresada en forma de bloques a la red. En [8] se entrenan varias redes neuronales basadas en la entropía de cada bloque, variando la tasa de compresión en cada una. En [9] implementan topología back-propagation variando la tasa de compresión así como el tamaño de los bloques.

En los anteriores trabajos se aborda la compresión de imágenes eligiendo un tipo de red y utilizando alguna técnica de preprocesamiento de la imagen, tales como la transformada Wavelet, entropía y detección de bordes, entre otras. El objetivo de estos trabajos se centra principalmente en mejorar el desempeño de la compresión pero no son consideradas las características topológicas de la RNA y sus posibles implicaciones sobre la compresión de la imagen. Este trabajo cambia el enfoque de emplear RNA en compresión, analizando los efectos de la arquitectura neuronal sobre el PSNR y la tasa de compresión de la imagen.

En gran parte de la bibliografía consultada, la imagen es dividida en bloques para luego ser procesada e ingresada a la RNA. Este tipo de segmentación también es adoptada para el desarrollo de este trabajo y permite la comparación de los resultados obtenidos bajo condiciones similares de entrada con los trabajos arriba mencionados y con el formato de compresión JPG. Los experimentos se realizan con dos topologías de RNA tipo feed-forward. En el primer caso, la red actúa de manera similar a un sistema de regresión, donde las entradas son los valores de los píxeles y la salida es el valor de la función. En el segundo caso se implementa un embudo o cuello de botella, que consiste en tener una capa con menor cantidad de neuronas con respecto a la capa de entrada.

Este documento se organiza así: la sección 2 ilustra las dos topologías abordadas y la variación de parámetros de entrenamiento, en los que se seleccionan los datos de entrada a la red, igualmente el análisis de la topología de la función de la imagen y embudo. Los resultados de los experimentos planteados se muestran

en la sección 3, posteriormente en la sección 4 se enuncian las conclusiones de este trabajo y recomendaciones que tengan lugar sobre el uso de RNA en compresión de imágenes.

Experimentación

Compresión por función de la imagen

Definición

Existen varias maneras de representar imágenes digitales, una es ver los colores mediante sus píxeles, otra es a través de una matriz de intensidades [10], en la que cada elemento está relacionado con el color del píxel, esta matriz se puede escalar en forma de vector y mostrar como una función.

Una primera aproximación a la compresión de imágenes usando la técnica de regresión con redes neuronales, se realiza a través de la función de la imagen, que tiene como variable independiente la posición del píxel (X,Y) y dependiente el valor de intensidad en dicho punto ($f(X,Y)$), esta función es pre-procesada para ingresar a la RNA, tal como se muestra en la figura 1. Con el objetivo de lograr una representación lo más fiel posible a la imagen original, es evidente que cada función es particular, lo cual representa un inconveniente si el objetivo es obtener un sistema capaz de generalizar.

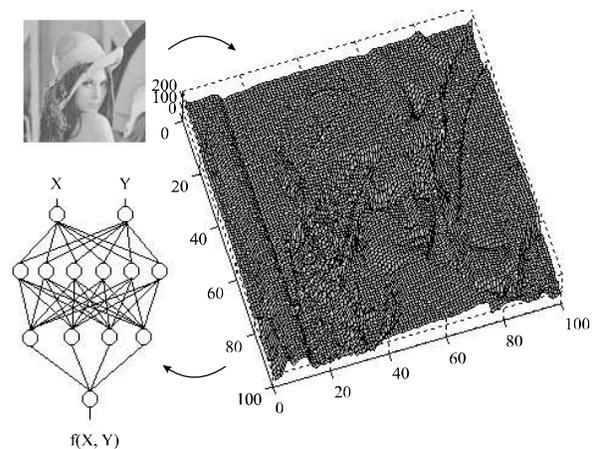


Figura 1 Modelo función de la imagen. Imagen de Lena tomada de [10]

El pre procesamiento necesario para que la red pueda trabajar con imágenes, consiste en normalizar las intensidades para que sean comparables con los pesos iniciales de la red, luego se hace ecualización de histograma como se muestra en la figura 2, lo que permite una distribución más uniforme de los tonos y mejora el desempeño de la red, pues no existen cambios tan abruptos entre pixeles adyacentes.

Para imágenes a color se hace una transformación al espacio YCbCr (Luminancia, diferencia de croma del azul y rojo), pues la representación en RGB (Red-Green-Blue) no es ideal para la compresión, al no aprovechar las características del ojo humano, que percibe más la luminancia que otros componentes, el espacio YCbCr aprovecha esta característica [5], por lo que es el usado en este trabajo.

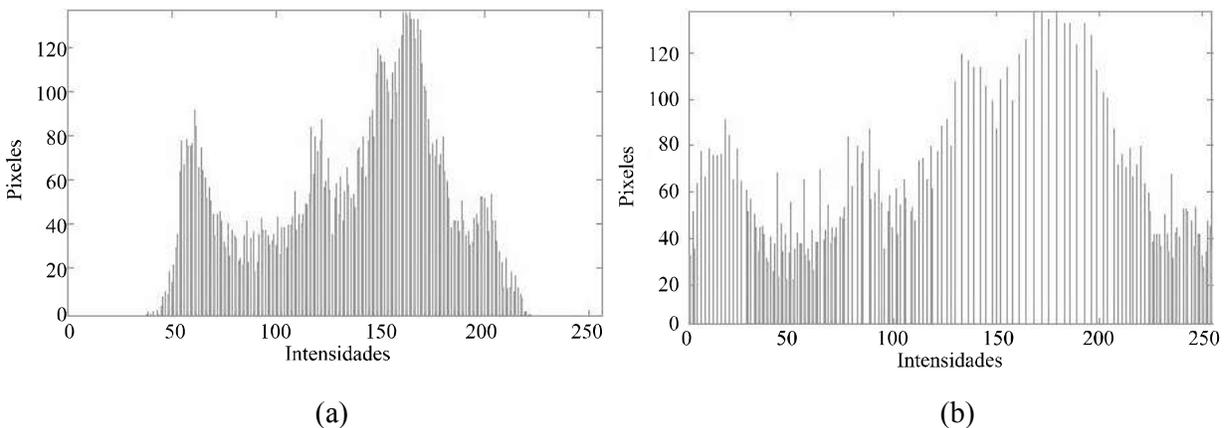


Figura 2 Ecualización de histograma para Lena. a) Histograma b) Ecualización

La relación pico señal a ruido (PSNR, *Peak Signal Noise Relation*) es un parámetro relativo de calidad y se calcula de acuerdo a (1), que permite comparar la imagen original con la reconstruida basado en el error cuadrático medio (MSE, *Mean Squared Error*) para imágenes, mostrado a través de (2).

$$PSNR = 10 \log_{10} \left(\frac{255^2}{MSE} \right) \quad (1)$$

$$MSE = \frac{1}{mn} \sum_{y=1}^m \sum_{x=1}^n [s - \tilde{s}(x, y)]^2 \quad (2)$$

Donde m y n son las dimensiones de la imagen, $s(x, y)$ es la imagen original y $\tilde{s}(x, y)$ es la imagen después de comprimirla y descomprimirla. En una imagen se puede llegar a tener 255 tonos, que es el valor máximo que puede tomar un pixel en la imagen.

Típicamente un valor mayor a 35dB (PSNR) tiene una buena calidad visual y solo ojos entrenados pueden detectar errores. Un valor aceptable es 30dB, donde algunos errores se pueden detectar.

Metodología

En la figura 3 se muestra la clásica imagen de Lena [10], representada de dos maneras distintas, ambas contienen la misma información pero la naturaleza de los datos es distinta. La figura 3.b. indica la variación de intensidad de pixeles de la imagen en escala de grises mostrada en la figura 3a.

La función la de imagen, posición de pixeles, cantidad de capas y neuronas, iteraciones, funciones de activación e inicialización de pesos, entre otros condicionan el desempeño de la red.

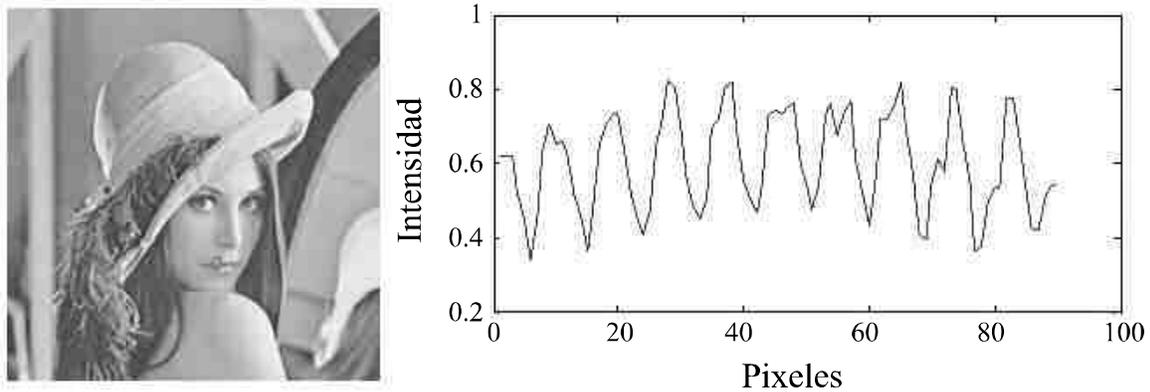


Figura 3 Lena con detalle de la boca. (a) Imagen, (b) función. Imagen de Lena tomada de [10]

Las entradas a la red son las posiciones de los píxeles, por lo que se diseña una red con dos neuronas en la primera capa, entre ambas reciben la posición (x,y) , la salida de la red es de una neurona y es equivalente a la intensidad que varía entre cero y uno.

Para lograr la compresión, se guardan los pesos de la red en un archivo e información de control como ancho y alto de la imagen. Para recuperarla se simula una red con los pesos guardados y se obtiene la función de la imagen, luego se organiza en forma de matriz para visualizarla de forma gráfica.

Las pruebas para este tipo de compresión consisten en realizar variaciones de la cantidad de capas y del número de neuronas de la red, para evaluar el impacto en términos de calidad visual y tasa de compresión.

Compresión por embudo

Definición

El proceso de compresión y descompresión abordado para una RNA tipo embudo se ilustra en la figura 4. La imagen original es pre-procesada tomando bloques de píxeles, obteniendo un vector de entrada que se propaga a las salidas de la red. El resultado es un vector de menor tamaño, por lo que se tiene un archivo que representa la imagen comprimida, la fase de descompresión realiza el procedimiento inverso. El objetivo se centra en lograr la mejor configuración del vector de entrada y parámetros de la red, así como una respuesta adecuada en el tiempo de procesamiento de una imagen.

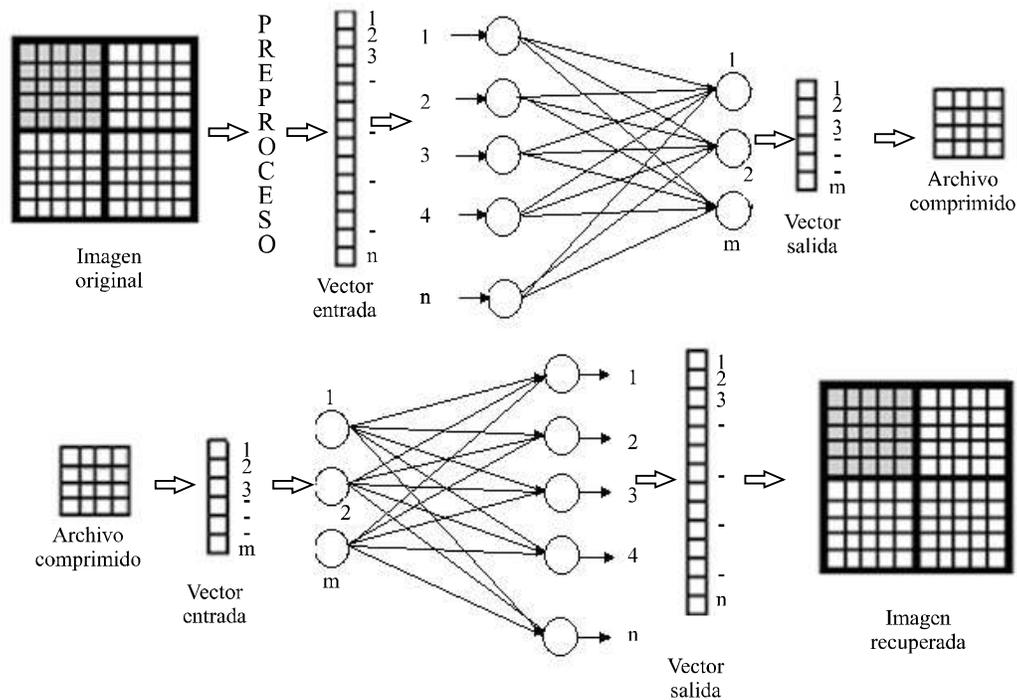


Figura 4 Topología tipo embudo para compresión y descompresión

El término embudo se refiere a una topología de redes feed-forward, cuya principal característica es que las capas ocultas cuentan con una cantidad menor de neuronas que las capas de salida y de entrada, también se conoce como cuello de botella [3]. La figura 5 ilustra la topología descrita.

la capa de entrada o de salida y la capa oculta [11]. El entrenamiento de estas redes tiene más parámetros, como el tamaño y la cantidad de subbloques de la imagen, tipo de imagen y tasa de compresión.

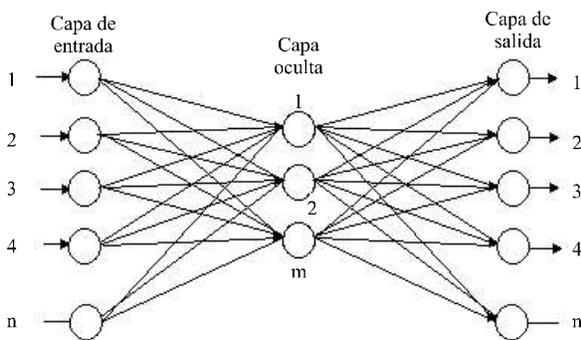


Figura 5 Red neuronal con topología embudo

El desarrollo de una red tipo embudo de tres capas permite generalizar la compresión. La tasa de compresión (CR) teórica mostrada en (3), se define como la relación entre las neuronas de

$$CR = \frac{\text{neuronas capa entrada}}{\text{neuronas capa intermedia}} \quad (3)$$

Metodología

La experimentación con las RNA tipo embudo inicia con el pre-procesamiento de la imagen, para que la entrada a la red favorezca los resultados. Así mismo se varían de manera controlada parámetros como la probabilidad de tomar un bloque de la imagen, el tamaño de estos, el tipo de imagen que se va a comprimir, la cantidad de neuronas en la capa oculta y la cantidad de capas que componen la red.

En el pre-procesamiento se divide la imagen original en bloques cuadrados de píxeles, luego se reduce a vectores de n posiciones, donde n son los píxeles en el bloque, posteriormente se

normaliza y se lleva a la red como entrada y salida, las neuronas en estas capas son n (puede variar de 2 en adelante).

Los pesos se inicializan en un rango $[-1,1]$ y se realiza el entrenamiento con imágenes en formato PNG, para obtener un PSNR lo más alto posible. Una vez entrenada la red se guardan los pesos y se realiza validación sobre un conjunto de imágenes. La compresión se logra al convertir el vector de entrada en uno de menor dimensión, que idealmente representa la misma información.

Se elige el formato PNG para aplicar compresión, las principales razones para esta elección es que es de uso libre, no tiene pérdidas en su compresión y soporta 24 bits, por lo que se convierte en un formato ideal para las imágenes de entrenamiento al momento de guardar el archivo comprimido.

Entrenamiento por sub-bloques

En una red tipo embudo usada para la compresión de imágenes digitales, cada neurona de la capa de entrada representa un pixel de la imagen original, estas neuronas reciben un bloque de pixeles a la vez, los valores típicos de este bloque abarcan el rango de 2×2 pixeles hasta 16×16 . Una iteración corresponde al barrido de todos los bloques de la imagen.

Típicamente en una imagen digital se identifican diferentes zonas formadas por una extensión delimitada y grande de pixeles, cuya principal característica es una baja variación de tonalidad. La probabilidad de que existan varios bloques de pixeles con intensidades similares es alta, lo que lleva a la posibilidad de realizar el entrenamiento con la totalidad de bloques de la imagen o con un porcentaje inferior.

En la figura 6 se muestra un ejemplo de la redundancia de pixeles adyacentes descrita anteriormente, para el caso se identifican algunas zonas con esta característica (carretera, montaña, zonas verdes cielo entre otros) y se detallan dos bloques similares en dos recuadros.

En esta parte del estudio se realizan pruebas variando la cantidad de bloques de la imagen que

se toman para entrenar con el fin de evaluar la hipótesis que se maneja; es decir, se comprueba si el hecho de reducir el número de bloques tomados al azar, afecta la calidad de la imagen recuperada teniendo en cuenta la similitud de intensidad existente entre muchos de los pixeles que conforman una imagen.



Figura 6 Selección de bloques con intensidades similares a. Imagen resaltando dos bloques de 7×7 pixeles b. Detalle de los bloques. Imagen tomada de [11]

Para determinar el comportamiento de las redes, se entrena con la totalidad de la imagen, seguido del 70%, 50% y 30% del total de la imagen, estos bloques se toman al azar y cada bloque es del porcentaje mencionado.

Para realizar la variación de un parámetro, se dejan todos los otros constantes con el fin de evaluar el impacto en el desempeño de la red. Para este caso, el tamaño del bloque es 4×4 , 5 neuronas en la capa intermedia, inicialización aleatoria de los pesos, cantidad de iteraciones, tamaño y tipo de la imagen, funciones de aprendizaje y activación en las neuronas.

Se usan 5 imágenes de entrenamiento con 1600 bloques cada una y se entrena durante 1250 iteraciones, la función de activación es tangencial logarítmica y la de aprendizaje es gradiente descendiente. Se entrenan 16 RNA con las que se puede llegar a establecer el mejor criterio de bloque de entrada.

Elección del tipo de imágenes de entrenamiento

Una vez determinado uno de los parámetros de entrada a la red, se estudia el impacto en la capacidad de generalización con imágenes en las

que se presenten modificaciones considerables en la cantidad de colores, formas, detalles y cambios de tonalidad. La variación de estas cualidades se puede observar en la figura 7, en cuatro clases de imágenes: médicas, que contengan texto, rostros y paisajes.

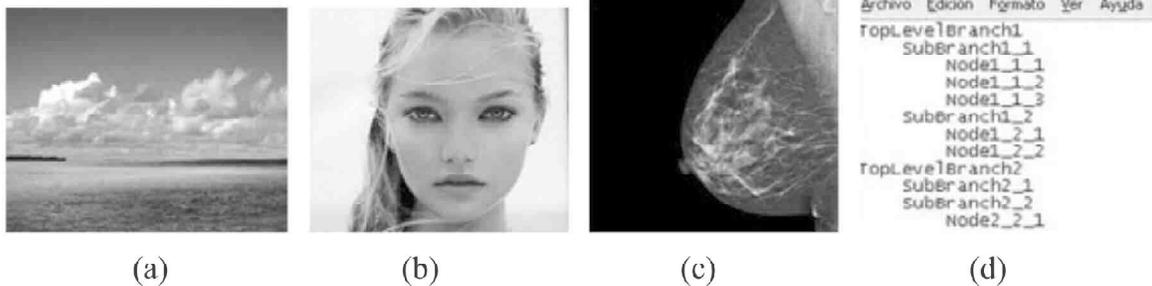


Figura 7 Imágenes de las categorías a) Paisaje, b) Rostros, c) Médicas, d) Texto. Tomadas de [11].

Las principales características en imágenes médicas, son que se limita a 256 tonos de gris, tienen poco detalle y los cambios de tonalidad son constantes; en las que contienen texto, los cambios de tono son abruptos, con límites bien definidos y de pocos colores; en rostros, hay bastante detalle y bordes definidos; e imágenes en general (otro tipo de imágenes), tienen muchos colores, detalles, formas y cambios de tono. Para determinar la mejor red que presenta el mejor PSNR en una categoría, se mantienen todas las variables constantes y se valida el entrenamiento con 20 imágenes por categoría.

Variación del tamaño de bloque de entrada

En esta sección se establece la cantidad de neuronas en la primera capa de la red que permita el mejor PSNR, se exploran las posibilidades de tomar desde 4 hasta 36 neuronas o bloques de 2x2 hasta 6x6 píxeles. La tasa de compresión permanece fija en 25%, excepto las redes con 25 y 9 neuronas en la entrada, en las que no se puede implementar y se desarrolla 24% y 22%, respectivamente.

El principal motivo para excluir el desarrollo de topologías con más de 36 neuronas en la capa inicial es el costo computacional. En la figura 8

se muestra el aumento exponencial del tiempo de entrenamiento para una red en 250 iteraciones, para obtener la red final entrenada con 10 imágenes hacen falta 2500.

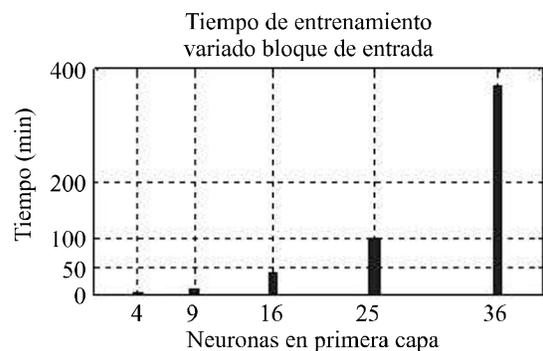


Figura 8 Tiempo de entrenamiento al variar el tamaño del bloque de entrada

Variación de la tasa de compresión

Una vez definida la cantidad de neuronas de la capa de entrada que presenta el mejor PSNR, se analizan las tasas de compresión (CR) que se pueden implementar. Esto se realiza a través de la variación del número de neuronas de la capa oculta en la red, con la variación de este parámetro se determina la configuración con la que se obtiene el mejor desempeño para una red de tres capas.

La variación de la cantidad de neuronas en la capa oculta, abarca el intervalo de 1 neurona hasta una cantidad igual a las de la primera capa menos 1; sin olvidar las limitaciones del equipo de cómputo.

Se espera que al aumentar las neuronas en la capa oculta, el PSNR mejore y el tiempo de entrenamiento aumente, lo que es equivalente a mejor calidad con mayor coste computacional.

Capas de la red

Es claro que no se pueden trabajar menos de tres capas, sin embargo no hay restricción para aumentarlas, con la excepción de las características de la máquina en la que se realiza el entrenamiento, por lo que se decide aumentar a cinco capas donde la tercera es la de menor cantidad de neuronas y la topología tiene simetría, por lo que se tienen la misma cantidad de neuronas en la primera y última capa, así como en la segunda y cuarta. Se decide mantener la proporción de neuronas entre una capa y la adyacente, para obtener cambios graduales en las diferentes capas. Con esta implementación se espera encontrar mayor correlación entre bloques y píxeles, además el cambio entre dos capas es más suave.

Compresión imágenes a color

Este análisis se puede extender a la aplicación de imágenes a color, en las que a diferencia de las escala de grises en las que se tienen 256 tonos (8 bits), se llegan a tener 16.7 millones de posibles colores (24 bits). Estas se pueden representar con 3 planos que contienen distinta información, la estándar es mostrar RGB, en el que los colores se dividen en 3 componentes de rojo, verde y azul.

La implementación de YCbCr sirve para comprimir más los planos Cb y Cr, a los que el ojo es menos sensible. De esta manera el primer plano se comprime más comparado con los otros dos.

Resultados

Función de la imagen

En la figura 9 se muestra un entrenamiento para el aprendizaje de la función de la imagen, en 2500 iteraciones y una topología de 4 capas con 2-12-20-1 neuronas en la respectiva capa, es evidente la mejora hasta 500 iteraciones, sin embargo de este punto en adelante el PSNR aumenta 3% en 2000 iteraciones. En todos los entrenamientos realizados se observa un comportamiento similar, a partir de la iteración 1000 el desempeño de la red tiende a permanecer estable.

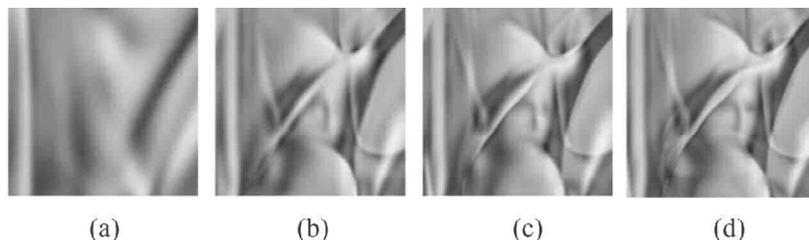


Figura 9 Entrenamiento Lena con función de imagen, topología cuatro capas (a) 200, (b) 500, (c) 1000, (d) 2500 iteraciones

El PSNR en la figura 9 es 10,87, 11,55, 11,76 y 11,92dB respectivamente, lo que es insuficiente para aplicar a compresión, pues se observan los bordes principales deteriorados y una aproximación en el tono de los píxeles, sin llegar a tener una calidad que permita distinguir entre dos imágenes parecidas. Al observar constantemente

este comportamiento, se decide implementar redes de 5 capas, variar la inicialización de los pesos, las funciones de activación y las neuronas por capa, siendo los resultados similares a los mostrados en la figura 9.

Al implementar 5 capas se elaboran redes con 2-10-15-25-1 neuronas por capa, así como varios

experimentos donde se varía la distribución de las neuronas por capa como 2-10-25-15-1 y 2-25-15-10-1. No se muestran las imágenes luego de la compresión debido a que su calidad visual es deficiente.

En las pruebas con aprendizaje de función se tienen resultados poco fiables en PSNR y se incrementa el tiempo de entrenamiento, para obtener la figura 9 (d) tarda 100 minutos, a cambio de esto el archivo comprimido tiene un tamaño de 970 bytes, sin embargo la relación calidad/compresión es mala.

Embudo

Los resultados del entrenamiento variando la porción de la imagen tomada para entrenar se muestran en la figura 10. Es posible observar que conviene elegir la totalidad de bloques de píxeles al momento de entrenar una red neuronal cuya función sea la compresión, toda vez que los resultados muestran que de esta forma se logra conseguir un PSNR mayor que cuando se emplea un porcentaje inferior.

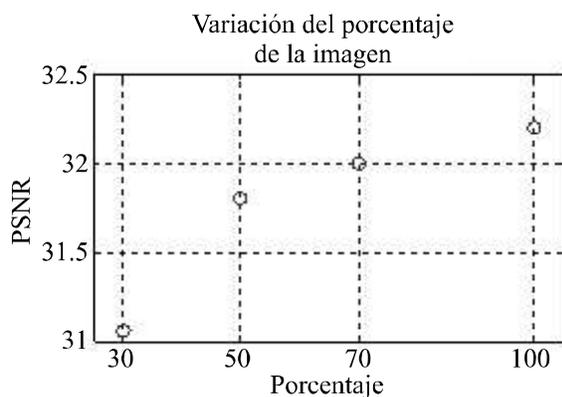


Figura 10 PSNR promedio en 20 imágenes versus probabilidad de tomar un bloque para entrenar

Al entrenar con la totalidad de la imagen, el tiempo de entrenamiento es mayor y proporcional a la cantidad de bloques para entrenar, por tanto,

al tomar toda la imagen el tiempo es el doble comparado con el 50%, vale la pena recordar que en la mayoría de aplicaciones prima la calidad de la imagen.

Posteriormente, el estudio se centra en obtener la mejor red luego de entrenar con una base de datos de las 4 clases de imágenes y enseñar a cada red cinco millones de bloques de píxeles y así comprobar, si existe alguna correlación entre el tipo de imagen en el entrenamiento y la generalización, que permita mejorar el desempeño de acuerdo al tipo de imagen a comprimir.

Los resultados de las pruebas se sintetizan en la tabla 1, donde \bar{x} es el promedio del PSNR y σ es la desviación estándar del mismo, siempre sobre un conjunto de 20 imágenes.

Tabla 1 Resultados entrenamiento con las 4 categorías de imágenes abordadas

Imagen entrena- miento	Imagen generalización							
	Texto		Rostro		Médica		Paisaje	
PSNR	\bar{x}	σ	\bar{x}	σ	\bar{x}	σ	\bar{x}	σ
Texto	18,87	2,8	18,64	2,7	16,19	1,2	17,24	1,9
Rostro	18,16	2,6	19,23	2,5	16,98	0,9	18,74	1,7
Médica	22,61	5,4	34,08	2,7	33,24	6,3	29,91	4,8
Paisaje	23,24	5,3	34,61	2,7	34,24	6,4	30,47	4,6

Las redes entrenadas con paisajes presentan el mejor PSNR, mostrando un 2% de mejora en el PSNR comparado con las médicas.

Con la variación del bloque de entrada se evidencia que el PSNR promedio de las redes con 25 neuronas en la capa de entrada es mayor que el ofrecido por las demás. Teniendo en cuenta la característica de mayor relevancia, para el caso se busca la calidad de la imagen recuperada, por lo que la mejor opción son las redes 25-6-25, lo que se observa en la figura 11.

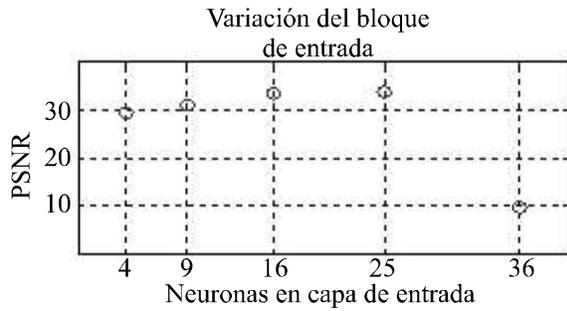


Figura 11 PSNR promedio versus neuronas en capa de entrada

La respuesta obtenida en términos de PSNR para las tasas de compresión implementadas del 4%, 8%, 12%, 16%, 20%, 24%, 36% y 48%, se muestran en la tabla 2. Este valor es teórico debido a que al guardar esta matriz en formato png, se incluye compresión adicional. El PSNR se puede estimar mediante (4), donde x es la tasa de compresión teórica.

Tabla 2 Desempeño en las redes de 3 capas comparada con 5 capas

CR (%)	25	12,5	8,3	6,25	5	4,2	2,8	2,1
Topología (25-x-y-z-25)	5-1-5	7-2-7	9-3-9	10-4-10	11-5-11	12-6-12	15-9-15	17-12-17
3 capas	25,891	27,604	30,255	31,247	32,435	33,829	35,979	38,642
5 capas	24,974	26,307	28,714	29,518	30,375	31,462	32,820	Fuera de memoria

$$PSNR = -42x^2 + 50x + 24 \quad (4)$$

La última variación realizada es aumentar a 5 capas la red neuronal, sin embargo se observa un aumento en el tiempo que en algunos casos llega a ser del 100%, y para la red 25-17-12-17-25 la memoria disponible no es suficiente, lo

que ya es un inconveniente para esta topología. Aparte de tener una calidad de imagen inferior, el desempeño de la red a través de las iteraciones mejora de una forma mucho más lenta. La tabla 2 muestra el resultado promedio de las redes de 3 y 5 capas, donde se evidencia una calidad más baja en las de 5 capas.

Tabla 3 Desempeño en imágenes en escala de grises

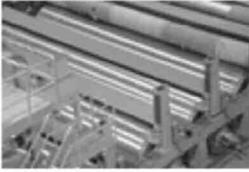
Nombre imagen	Mejor calidad visual				Mejor tasa de compresión			
	Escala de grises		Color		Escala de grises		Color	
	PSNR	CR (bpp)	PSNR	CR (bpp)	PSNR	CR (bpp)	PSNR	CR (bpp)
Lena	37,29	2,16	38,57	2,12	22,84	0,26	29,4	0,7
Cameraman	40,59	1,72	-	-	24,45	0,2	-	-
Airplane	38,45	2,45	39,31	2,46	23,9	0,26	26,4	0,69
Peppers	35,62	2,09	37,51	2,26	26,09	0,22	28,9	0,75
Mandrill	35,59	2,49	28,82	2,72	20,43	0,27	22,33	0,82
Boat	33,63	2,09	-	-	24,15	0,24	-	-

En la tabla 3 se muestra el desempeño; medido con el PSNR y CR, de algunas imágenes estándar [11] [12] en escala de grises y a color para un tamaño de 512*512 píxeles para la mejor calidad visual o tasa de compresión. Es posible de esta forma tener un rango de selección de calidad o de magnitud de la compresión a partir de los parámetros de la arquitectura de la red tipo

embudo o del tamaño del bloque de píxeles en el pre-procesamiento.

Para ilustrar visualmente el comportamiento de la compresión usando topología de red neuronal tipo embudo, en la tabla 4 se muestran algunas imágenes con variaciones en tono, textura y brillo, con PSNR superiores a 35dB y CR de orden de 2bpp.

Tabla 4 Desempeño en imágenes particulares. Imágenes tomadas de [11]

<i>Imagen</i>				
PSNR	42,46	42,32	38,28	40,42
CR (bpp)	2,04	1,55	2,36	2,27

Conclusiones

En el presente trabajo se abordó el problema de la compresión de imágenes digitales desde la perspectiva de la inteligencia computacional mediante el empleo de RNA feedforward, para lo cual fueron implementados algoritmos de compresión y descompresión de imágenes. Se abordaron las posibilidades de realizar la compresión mediante el aprendizaje de la función de la imagen y el uso de embudo. Con respecto a la primera, se puede afirmar que sus resultados no fueron satisfactorios debido a que después de 2500 iteraciones en el entrenamiento, su respuesta en PSNR no fue superior a 15 dB, haciendo que la imagen recuperada no pudiese ser reconocible por el ojo humano. De igual manera, de acuerdo con los resultados este método no es fiable debido al alto tiempo de entrenamiento, la imposibilidad de generalización y la incertidumbre entorno a convergencia del entrenamiento. Por lo anterior, no es recomendable la implementación de este tipo de redes FF para aplicar en técnicas de compresión de imágenes como las estudiadas.

Después de realizar una gran variedad de experimentos, se encontró que el modo embudo de las redes feedforward ofrece mejores resultados en términos de calidad visual, tasa de compresión y tiempo de ejecución, frente al aprendizaje de la función de la imagen.

Al experimentar acerca de la porción de la imagen que al ser usada para entrenamiento presenta mejor respuesta en calidad visual, se halló que el tomar la totalidad de la imagen representa una mejora del 0,625% respecto a la elección del 70% de la imagen. Así mismo se estableció que al categorizar las imágenes en 4 grandes tipos dependiendo de sus características (médicas, rostros, paisajes y texto) se podría encontrar cuál(es) de ellas favorecería los procesos de entrenamiento y generalización de la red. En efecto, se encontró que las redes neuronales tipo feedforward presentan problemas al momento de generalizar imágenes de tipo texto, independientemente de la categoría con la que se haya entrenado, pues no se obtuvo un PSNR mayor de 23 dB, siendo difícilmente reconocibles

los caracteres en la mayoría de casos. También se estableció que para la compresión de imágenes médicas, rostros y paisajes, lo adecuado es realizar un entrenamiento con imágenes de paisajes, pues generan mejor calidad visual en la descompresión en un 1,53%, 2,92% y 1,83% respectivamente frente al entrenamiento con la categoría de imágenes médicas.

En cuanto a la segmentación previa de la imagen, existen técnicas adaptativas teniendo en cuenta parámetros como compacidad, redundancia y correlación entre píxeles, que favorecerían la calidad visual de la imagen descomprimida. Sin embargo, en este trabajo se decide tomar la segmentación por bloques para poder realizar una comparación efectiva con respecto al formato comercialmente más difundido en la actualidad, JPG. No obstante, los resultados y tipos de pruebas del presente trabajo pueden ser tomados como punto de partida para la implementación de otras técnicas de división de la imagen.

Teniendo en cuenta lo anterior, al realizar segmentación por bloques de la imagen, se recomienda realizar la en la que se entrena con bloques de 5x5 píxeles ya que de esta manera se obtiene una ganancia del 0,78% (PSNR) con respecto al realizado con bloques de 4x4 píxeles, además su compresión puede llegar a ser mayor en un 64%. También se aconseja usar redes de 3 capas, pues la implementación de 2 capas adicionales implica una disminución en el PSNR de la imagen descomprimida de aproximadamente un 5.85% en promedio.

Las referencias bibliográficas consultadas no abordan el tema de la compresión de imágenes digitales a color mediante el uso de RNA, por lo que los estudios en este tipo de imágenes y sus consecuentes resultados se consideran un aporte propio importante a la investigación referente a dicho tema.

La aplicación de otros tipos de segmentación usando técnicas de compacidad, redundancia y correlación de píxeles, permitiría ampliar el presente estudio a no sólo considerar aspectos topológicos de la arquitectura de la red

neural sino a analizar los efectos que sobre la compresión posee este tipo de preprocesamiento.

Cuando se utilizan las redes FF configuradas para obtener la más alta calidad visual, se obtiene una ventaja en PSNR del 21,59% frente a JPG. En redes FF, en la mejor tasa supera a JPG en 2,63%.

Referencias

1. Tecnología de los contenidos multimedia. *Edición y compresión de imágenes estáticas*. Programa oficial de postgrado: Master en comunicaciones, redes y gestión de contenidos. España. Disponible en: <http://ocw.innova.uned.es/mm2/tcm/contenidos/pdf/tema3.pdf>. Diciembre 2008.
2. R. Ferzli, M. Adnan. *Subsampling image compression using Al-Aloui backpropagation algorithm*. *Electronics, Circuits and Systems*. 14th IEEE International Conference. Arizona. 2007. pp. 1260-1263.
3. W. Gillespie. *Still image compression using neural networks*. Utah University. Logan. UT. 2005.
4. V. Rama, P. Vaddella, K. Rama. "Artificial neural networks for compression of digital images: a review". *International Journal of Reviews in Computing*. Vol 10. 2010. pp. 75-82
5. D. Alvarez, M. Guevara, G. Holguin. *Preprocesamiento de imágenes aplicadas a mamografías digitales*. *Universidad tecnológica de Pereira*. Scientia et Technica Año XII. No 31. Pereira. Colombia. 2006. pp. 1-6.
6. A. Durai, E. Saro. "Image Compression with Back-Propagation Neural Network using Cumulative Distribution Function". *World Academy of Science, Engineering and Tech*. Vol. 17. 2006. pp. 60-64.
7. S. Basu, K. Kayal, J. Sil. *Edge preserving image compression technique using adaptive feed forward neural network*. Proceedings of the ninth IASTED International Conference INTERNET AND MULTIMEDIA SYSTEMS AND APPLICATION. Switzerland. 2005. pp. 467-471.
8. H. Sahoozadeh, A. Suratgar. *Adaptive image compression using neural networks*. Setit, 5th International Conference: Sciences of Electronics Technologies of Information and Telecommunications. Tunisia. 2009. pp. 1-5.
9. I. Vilovic. *An experience in image compression using neural networks*. 48th International Symposium ELMAR Focused on Multimedia Signal Processing, Zadar Croatia. 2006. pp. 95-98.

10. Signal and image processing institute. *Ming Hsieh Department of Electrical Engineering. University of southern*. California. Disponible en: <http://sipi.usc.edu/database/database.php?volume=misc&image=12>
Consultado en Noviembre de 2011.
11. N. Lugo, F. Roldan. "Imágenes usadas en la compresión de imágenes con redes neuronales". Disponible en: https://rapidshare.com/files/3208725692/Imágenes_usadas_en_la_compresión_de_imágenes_con_redes_neuronales.rar. Publicado en Enero de 2012 y consultado Marzo de 2012.