

Recuperación de imágenes con base en el color utilizando descriptores visuales del estándar MPEG-7

Jorge Ochoa Somuano ^{1*} & José Manuel Gómez Alavez

Resumen

La recuperación de imágenes basada en el contenido (CBIR, por sus siglas en inglés) pertenece al procesamiento digital de imágenes (PDI) y su interés ha aumentado considerablemente en los últimos años en el campo de la investigación científica. En este trabajo de investigación se ha realizado un programa de recuperación de imágenes con base en el contenido. El proceso consiste en extraer características basadas en el color utilizando descriptores visuales establecidos en el estándar MPEG-7. A partir de una imagen de entrada que se proporciona al sistema como ejemplo, se realiza una búsqueda en la base de imágenes que se tiene como referencia, la búsqueda se realiza con base en el color, de tal forma que se muestra una serie de imágenes ordenadas de mayor a menor similitud. La primera imagen que se muestre como resultado será la misma que se esté buscando, si ésta se encuentra en la base de imágenes, en caso contrario, se mostrará aquella que se parezca más con la imagen de ejemplo y, así sucesivamente hasta encontrar la que menos relación tenga con la imagen de ejemplo. Después de realizar algunos experimentos, se puede asegurar que los resultados obtenidos con el programa informático que se implementó son muy buenos; el resultado es el esperado.

Palabras clave: descriptores, características basadas en el color, recuperación de imágenes, MPEG-7, CBIR.

Recibido: 03 de septiembre de 2018

Abstract

Content-based image retrieval belongs to Digital Image Processing, and the interest towards it has increased considerably in recent years in the field of scientific research. In this paper, a content-based image retrieval program has been tested. The process consists of the selection of color features using visual descriptors of the MPEG-7 standard. From an input image provided to the system as an example, a search is performed in the image database as a reference and, the search done so that a series of ordered images of greater to lesser similarity is obtained. The first image displayed as a result will be the same one for which you are searching, if it is in the image database; otherwise, it will show the one that looks the most like the example image and so on, until the least similar image is found. After performing various experiments, it is certain that the results obtained with the computer program that was implemented are very good; the result is as expected.

Key words: descriptors, features based on color, image retrieval, MPEG-7, CBIR.

Aceptado: 04 de octubre de 2018

¹ Instituto de Industrias, Universidad del Mar campus Puerto Escondido. Ciudad Universitaria, Vía Sola de Vega km 1.5 Carretera Puerto Escondido-Oaxaca. San Pedro Mixtepec-Juquila, Oaxaca, México, 71980.

* Autor de correspondencia: ochoa@zicatela.umar.mx (JOS)

Introducción

En la actualidad es común subir a Internet una enorme cantidad de imágenes con una tasa de crecimiento exponencial. La mayoría de las personas con acceso a internet tiene una gran facilidad para subir fotos a redes sociales como Facebook, Twitter, Instagram, entre otras. Lo anterior ayuda en gran medida a dicho crecimiento. Derivado de lo anterior y a la gran cantidad de imágenes que circulan por la web se tiene una gran demanda en el desarrollo de aplicaciones para la edición y administración de las mismas. Motores de búsqueda como Google trabajan arduamente en nuevas tecnologías que permitan realizar búsquedas de imágenes en la web, debido a que buscar una imagen en específico, en ocasiones, puede ser un trabajo muy tedioso y tardado, sobre todo si no se sabe con claridad qué tipo de imagen se está buscando.

Existe una variedad de software comercial especializado en la recuperación de imágenes por contenido, dicho software es de gran utilidad en áreas como la medicina la cual tienen la función de ayudar en el diagnóstico de ciertas enfermedades como la osteoporosis o para determinar la gravedad de una fractura de hueso en un paciente, también tienen aplicaciones en la milicia por medio de la implementación de sistemas para la detección de armas ocultas en el cuerpo humano. Cabe mencionar que a este tipo de software se le conoce en el área de PDI como *Content Based Image Retrieval* (CBIR, por sus siglas en inglés).

La recuperación de imágenes lleva un proceso inmerso, la descripción, mismo que caracterizará a las imágenes con base en las características físicas visuales, éstas pueden ser de color, forma o textura; dicho proceso se lleva a cabo mediante algoritmos que permiten realizar extracción de descriptores.

Se han realizado algunos trabajos en relación a la extracción de características, clasificación e identificación de objetos en imágenes digitales, tal es el caso del trabajo realizado por Ochoa-Sommano *et al.* (2013) quienes desarrollaron un software clasificador de objetos. Para la extracción de las características de

los objetos en las imágenes se emplearon los momentos invariantes de *Hu* los cuales son invariantes a rotación, escala y traslación. Para la clasificación de estos mismos objetos se utilizó el clasificador *K-means*. En las pruebas que se realizaron para la validación de este software se utilizaron: números, letras, herramientas, tornillos, figuras geométricas y llaves con monedas. Como resultado de sus pruebas, los autores reportan un porcentaje de certeza de 97.76 % y un porcentaje de error de 2.24 % bajo condiciones controladas del ambiente. Una característica de este trabajo de investigación es que los objetos presentes en las imágenes de pruebas son similares en cuanto a tamaño, color y forma.

Por otro lado, Reyes *et al.* (2013) propuso un sistema de identificación de escenas, definiendo como escena aquellos fotogramas que comparten características visuales en un mismo contexto, utilizando una cámara capaz de capturar video. Una vez capturado el video, el sistema selecciona algunos *frames*, de los cuales se extraen descriptores basados en color, textura y forma, permitiendo efectuar una clasificación de las escenas más relevantes del video. Para ello se hace uso del estándar JPEG2000 consiguiendo de esta forma comprimir dichas imágenes y al mismo tiempo obtener información valiosa de las imágenes, es decir, características fundamentales que ayuden a describirlas.

Uno de los trabajos que ha utilizado algunos de los descriptores del estándar MPEG-7, es el reportado por Dvorak *et al.* (2012), en el cual utilizó dos descriptores: color dominante y la forma del contorno. Los descriptores que utilizan describen la región de la imagen, su color y su forma. Sin llegar a reportar porcentajes de efectividad, los autores señalan que los resultados son muy alentadores y, los casos de éxito que pueden destacar son aquellos en los que el cielo y el agua están presentes.

Además, Zhao *et al.* (2011) proponen un esquema de identificación de vehículo que combina la extracción del objeto del vehículo basado en el método de Frame Difference (DF) y la coincidencia de imagen basada en el Color Layout Descriptor (CLD) del MPEG-7 con el

objetivo de buscar una solución de equilibrio. De acuerdo con los experimentos que realizaron los autores con datos reales de videos de vigilancia de la oficina de la administración de transporte de Shanghai, muestran que el algoritmo propuesto podría aumentar la velocidad de coincidencia de la imagen en gran medida y obtener un rendimiento tan bueno como con el método clásico de CLD.

Finalmente, en este artículo se muestran algunas características del estándar MPEG-7 desarrollado por MPEG (Moving Picture Expert Group) y aprobado por ISO en el 2001. El nombre oficial es *Multimedia Content Description Interface* y esta versión se desarrolló estrictamente para describir el contenido multimedia. El MPEG-7 provee descripción para audio y para video. Sin embargo, el trabajo que se reporta en el presente artículo se basa en los descriptores de video: color, forma, textura y movimiento.

Material y métodos

Área de estudio

El trabajo tuvo una temporalidad de ocho meses durante el año 2017 en la Universidad del Mar campus Puerto Escondido ubicada en Puerto Escondido, Oaxaca, México.

Para realizar el proceso de recuperación de imágenes con base en el color se realiza una serie de pasos, tal como se muestra en la figura 1.

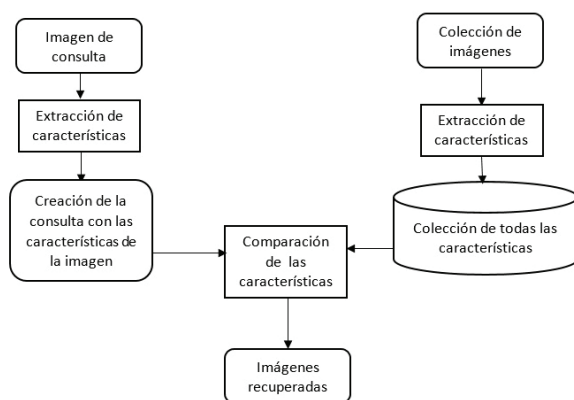


Figura 1. Modelo general de un sistema CBIR (Rorissa, 2007).

Los procesos de la parte derecha de la figura 1 se consideran generales. La *colección de imágenes* está compuesta por la totalidad de las imágenes. El proceso de *extracción de características* se realiza a todas las imágenes que se tienen disponibles. La *colección de todas las características*, son todas las características extraídas en el paso anterior. Los procesos de la parte izquierda de la figura 1 se consideran particulares. La *imagen de consulta* es sólo una imagen y es aquella que se desea utilizar para localizar la misma imagen de entre todas las que se tienen, o bien, encontrar la que más se le parezca. La *extracción de características* procesa únicamente a la imagen de consulta. La *creación de la consulta con las características de la imagen* prepara la información extraída del paso anterior para proceder a la comparación. Finalmente, los procesos de la parte central se encargan de realizar la *comparación de las características* que corresponden a la imagen de consulta con todas aquellas que se obtuvieron en la extracción de características de todas las imágenes, dependiendo del grado de similitud se obtendrá un listado de imágenes de mayor a menor porcentaje de similitud. Para realizar el proceso de extracción de características se utiliza el estándar MPEG-7.

MPEG-7. Es un estándar para descripción de audio y video perteneciente a ISO / IEC (International Standardization Organization/ International Electrotechnical Commission, por sus siglas en inglés). Al estándar MPEG-7, Martínez *et al.* (2002) lo denomina formalmente como Multimedia Content Description Interface. De acuerdo con Manjunath *et al.* (2002), el estándar MPEG-7 se estructura en ocho partes generales:

1. Systems. Indica las herramientas que serán necesarias para especificar los descriptores.
2. Description Definition Language. Especifica el lenguaje para definir nuevas descripciones de esquemas.
3. Visual. Especifica los descriptores para describir la parte visual del material.
4. Audio. Especifica los descriptores para describir la parte de audio del material.

5. Generic Entities and Multimedia Description Schemes (MDS). Especifica los descriptores con la descripción genérica multimedia.
6. Reference Software. Se especifican algunas herramientas experimentales de software para llevar a cabo la descripción del contenido.
7. Conformance Testing. Define guías para probar la conformidad con MPEG-7.
8. Extraction and Use of MPEG-7 Descriptions. Provee información sobre la extracción y uso de las herramientas para descripción.

Descriptores visuales. Se dividen en color, forma, textura y movimiento. Uno de ellos y que se utiliza para la presente investigación es el *color layout descriptor (CLD)*. Sergi (2009) indica que el CLD es un descriptor para la representación de color muy compacto y con una alta velocidad de recuperación de imágenes, este descriptor ha sido diseñado para representar de manera efectiva la distribución espacial del color en una imagen. La principal característica del CLD es que con su correcta aplicación es posible comparar imágenes a una gran velocidad con una gran precisión. El proceso de extracción de este descriptor consiste en cuatro fases: división de la imagen, detección del color representativo, conversión DCT (Discrete Cosine Transform) y el escaneo en zigzag. En la figura 2 se puede observar la representación del proceso de extracción de este descriptor.

División de la imagen. La imagen se divide en una matriz de 64 bloques con la finalidad

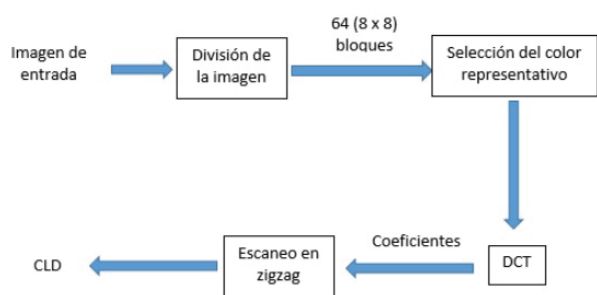


Figura 2. Esquema general de la extracción del descriptor CLD.

de garantizar que el cambio en la resolución de la imagen no se vea afectada, en la figura 3 se muestra un ejemplo de cómo se vería tal división.

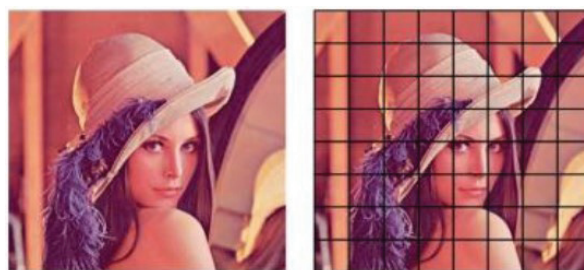


Figura 3. Imagen dividida en una matriz de 64 bloques (8x8), Sergi (2009).

Selección del color representativo. Una vez que la imagen ha pasado por la fase de división, un único color representativo es seleccionado de cada bloque de la matriz. Para la selección del color representativo se puede elegir la metodología del color dominante o cualquier otra metodología similar que sea capaz de extraer el color representativo de una imagen. Pero el estándar MPEG-7 recomienda el uso del promedio de colores de los píxeles en cada bloque, es decir, se obtiene el color de cada píxel en el bloque y se calcula el promedio de ese bloque, de esta manera se obtiene el color representativo de cada bloque.

El resultado de la selección del color representativo de la imagen es una pequeña imagen de dimensiones 8x8, es decir, la imagen se agrupa en bloques de 8x8, 64 datos (coeficientes) que resultan ser muy representativos para las técnicas de procesamiento digital de imágenes. En la figura 4 se puede apreciar la transformación de la imagen original a su equivalente en color representativo.

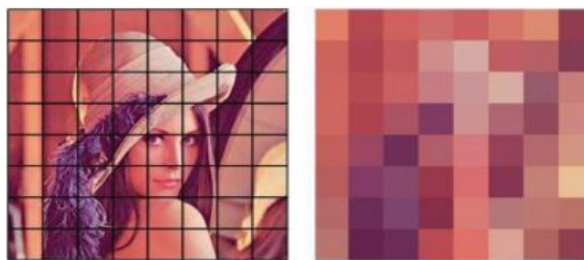


Figura 4. Selección del color representativo de cada bloque de la imagen, Sergi (2009).

Antes de realizar el siguiente proceso, se debe hacer una transformación de color de RGB a YCrCb, el objetivo es separar la luminancia de la crominancia. En la figura 5 se muestra la expresión utilizada para realizar la transformación.

$$\begin{pmatrix} Y \\ C_r \\ C_b \end{pmatrix} = \begin{pmatrix} 0.299 & 0.587 & 0.114 \\ -0.16875 & -0.33126 & 0.5 \\ 0.5 & -0.41869 & -0.08131 \end{pmatrix} \begin{pmatrix} R \\ G \\ B \end{pmatrix}$$

Figura 5. Transformación de formato RGB a YCrCb.

Transformación a DCT. En esta fase, la luminancia (Y), la crominancia roja (Cr) y la crominancia azul (Cb) son transformadas en tres matrices DCT de dimensión 8x8, de esta manera se obtienen tres conjuntos de 64 coeficientes de DCT. Para calcular el DCT en un arreglo bidimensional se utilizan las fórmulas recomendadas por Majidpour *et al.* (2015) que se indican en las ecuaciones 1, 2 y 3.

$$(1) \quad B_{pq} = \alpha_p \alpha_q \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} A_{mn} \cos \frac{\pi(2m+1)p}{2M} \cos \frac{\pi(2n+1)q}{2N},$$

$$0 \leq p \leq M-1, 0 \leq q \leq N-1$$

$$(2) \quad \alpha_p = \begin{cases} \frac{1}{\sqrt{M^3}} & p = 0 \\ \sqrt{\frac{2}{M^3}} & 1 \leq p \leq M-1 \end{cases}$$

$$(3) \quad \alpha_q = \begin{cases} \frac{1}{\sqrt{N^3}} & q = 0 \\ \sqrt{\frac{2}{N^3}} & 1 \leq q \leq N-1 \end{cases}$$

Donde:

B representa la matriz resultante con los valores de la transformada discreta del coseno.

p representa el índice para cada una de las filas de B.

q representa el índice para cada una de las columnas de B.

m representa el índice para cada una de las

filas de la imagen a procesar.

M es el número total de filas de la imagen a procesar.

n representa el índice para cada una de las columnas de la imagen a procesar.

N es el número total de columnas de la imagen a procesar.

A representa la imagen a procesar.

α_p es un factor que se calcula utilizando (2) con base en el valor actual de p.

α_q es un factor que se calcula utilizando (3) con base en el valor actual de q.

Escaneo en zigzag. Una vez realizado el proceso de transformación a DCT se realiza un escaneo en zigzag con los tres arreglos bidimensionales de coeficientes. En la figura 6 se muestra un ejemplo.

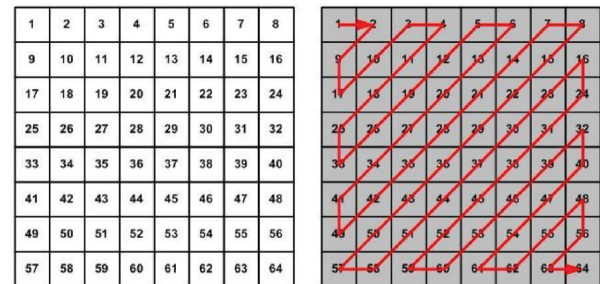


Figura 6. Escaneo en zigzag, Sergi (2009).

El objetivo del escaneo en zigzag es la agrupación de los coeficientes de baja frecuencia, como lo indica en estos coeficientes se acumula la información de las imágenes reales. Esto se debe a que se ha pasado de una representación espacial en pixeles a una representación de componentes de frecuencia.

Comparación de imágenes. Para realizar la comparación entre dos elementos y determinar si estos son iguales, se utiliza la distancia euclídea (ver la ecuación 4). Con base en los resultados del CLD se considera el par de CLD como: (DY, DCb, DCr) y (DY', DCb', DCr'), la letra D se antepone a las letras del modelo YCbCr para indicar que se trata de los Descriptores.

(4)

$$D = \sqrt{\sum_i w_{yi}(DY_i - DY'_i)^2} + \sqrt{\sum_i w_{bi}(DCb_i - DCb'_i)^2} + \sqrt{\sum_i w_{ri}(DCr_i - DCr'_i)^2}$$

Donde:

w_{yi} representa el cuadrado de la diferencia de la componente DY del elemento 1 con la componente DY' del elemento 2.

w_{bi} representa el cuadrado de la diferencia de la componente DCb del elemento 1 con la componente DCb' del elemento 2.

w_{ri} representa el cuadrado de la diferencia de la componente DCr del elemento 1 con la componente DCr' del elemento 2.

D representa la suma de las raíces cuadradas de cada una de las sumatorias de las diferencias de las componentes de los modelos YCbCr del elemento 1 y las componentes de los modelos Y'Cb'Cr' del elemento 2.

Resultados

Se muestra un ejemplo de recuperación de imágenes con base en el color, en la figura 7 se puede apreciar de lado izquierdo la imagen de consulta, es decir, la imagen que se desea buscar. De lado derecho de la misma figura se pueden observar las imágenes que se han recuperado por orden de similitud, de mayor a menor. Es fácil determinar visualmente que efectivamente la primera imagen de lado derecho corresponde exactamente con la

imagen de consulta, por lo tanto, se tiene una similitud del 100%. Por otro lado, las imágenes sucesivas comparten características con la imagen de consulta, sin embargo la similitud es decreciente. En caso de que la imagen de consulta no se encuentre en la base de imágenes donde se está realizando la búsqueda, aun así mostrará un resultado que empezará con la imagen que mayor parecido tenga con la imagen de consulta. La única forma de obtener un resultado donde no se muestren imágenes es que no se tenga ni una sola imagen en la base de imágenes.

Discusión

En la última década se ha incrementado exponencialmente la cantidad de imágenes que se utilizan en internet, principalmente en las redes sociales. Anteriormente se agregaban metadatos a las imágenes, es decir, información de carácter relevante para su identificación, como podían ser: tipo de imagen; interior o exterior, contenido de la imagen; playa, bosque, edificios, etc., entre otro tipo de información. Con la información anterior se podían realizar búsquedas sobre esos metadatos y podíamos obtener imágenes con los criterios que se estaban buscando. Sin embargo, colocar toda esa información a cada una de las imágenes se volvió un trabajo muy complicado y agotador. Por ello es que surgen alternativas para buscar imágenes con base en su contenido, sin tener la necesidad de ingresar información adicional a cada imagen.



Figura 7. Ejemplo de la recuperación de imágenes utilizando el software CIBR..

La recuperación de imágenes con base en su contenido, es una alternativa muy buena, ya que puede utilizar una variedad de características para encontrar imágenes con base en una imagen de ejemplo. Para el desarrollo del presente trabajo se optó por considerar un descriptor basado en la distribución del color (CLD). Una de las principales características del CLD es la gran velocidad de comparación entre imágenes y su gran precisión. Respecto a la comparación con otras tecnologías, se considera como parte de los trabajos futuros ya que por el momento la solución que se implementó es *offline* (sin conexión).

Conclusiones

La recuperación de imágenes con base en su contenido es una buena alternativa si se tienen limitantes con el tiempo, ya que los resultados se pueden obtener en un tiempo relativamente corto, dicha velocidad se consigue gracias a las características con las que cuentan las herramientas para procesamiento digital de imágenes del estándar MPEG-7. Como se puede apreciar en la sección de resultados, la precisión en la recuperación de imágenes en caso de contar con la imagen de consulta, es de un 100%. Sin embargo cabe señalar que en caso de que la imagen de consulta no esté en la base de consulta, la lista de las imágenes resultantes iniciará con aquella imagen que más parecido tenga con la imagen de consulta, motivo por el cual siempre se tendrá un resultado.

Agradecimientos

A un revisor anónimo que realizó valiosos comentarios que ayudaron a enriquecer el presente trabajo.

Referencias

- Dvorak, P. & M. Zupal. 2012. Description of Objects in Images Using MPEG-7 Descriptors, *Elektro Revue*, vol. 3, no. 1, ISSN: 1213-1539, pp. 26-31.
- Majidpour, J., Kheari, E., Hassansade, H. & K. S. Mohammend. 2015. Interactive tool to improve the automatic image annotation using MPEG-7 and multi-class SVM, *International Conference on Information and Knowledge Technology*. Irán: pp.

1-7, DOI: 10.1109/IKT.2015.7288777.

- Manjunath, B.S., Salembier, P. & T. Sikora. 2002. *Introduction to MPEG-7 Multimedia Content Description Interface*, John Wiley & Sons Inc. USA, ISBN: 978-0471486787, pp. 396.
- Martínez, J.M., R. Koenen, & F. Pereira. 2002. MPEG-7: the generic Multimedia Content Description Standard, *IEEE MultiMedia*, 9 (2): pp. 78-87. DOI: 10.1109/93.998074.
- Ochoa-Sommano, J., Pérez, C. Toscano, J.H. & C. G. Pereyra. 2013. Clasificación de objetos rígidos a partir de imágenes digitales empleando los momentos invariantes de Hu, *X Congreso Internacional Sobre Innovación y Desarrollo Tecnológico CIINDET*, Cuernavaca, Morelos, ISBN: 978-607-95255-4-5.
- Reyes, S. 2013. Sistema de identificación de escenas en JPEG2000 y sus aplicaciones al indexado MPEG7. Tesis de Maestría, Universidad Nacional Autónoma de México, Facultad de Ingeniería, México.
- Rorissa, A. 2007. *Image Retrieval Benchmarking Visual Information Indexing and Retrieval*. American Society for Information Science and Technology. USA. 2007.
- Sergi, L.V. 2009. *Color Based Image Classification and Description*. Tesis de maestría. Universidad Politécnica de Cataluña, España.
- Zhao, R., Wu, Y., Zhu, J. & Z. Wei. 2011. Efficient vehicle identification using MPEG-7 Color Layout Descriptor. *International Conference on Business Management and Electronic Information (BMEI)*, Guangzhou, vol. 5, pp. 128-131. DOI: 10.1109/ICBMEI.2011.5914443.