Implementación de un sistema de reconocimiento de imágenes por contenido usando algoritmos genéticos

Juan Villegas-Cortez[1], Yolanda Pérez-Pimentel[2], Ismael Osuna-Galán[2]

[1] Departamento de Electrónica, Universidad Autónoma Metropolitana, Azcapotzalco, D.F., México

[2] Ingeniería en Mecatrónica, Universidad Politécnica de Chiapas, Tuxtla Gutiérrez, Chiapas, México

jvillegas@gmail.com,{ypimentel,iosuna}@upchiapas.edu.mx

Resumen. La recuperación de imágenes basada en su contenido propio (CBIR: Content–Based Image Retrieval), describe una serie de técnicas y métodos que utilizan contenidos visuales extraídos de las propias imágenes a estudiar, para buscar o clasificar imágenes de acuerdo con los intereses de un usuario o sistema; ésta ha sido un área de investigación con notables progresos en la investigación teórica y ya ha sido aplicada en sistemas de análisis de bases de datos de imágenes via software. En la actualidad, conviven gran cantidad de métodos y se han integrado técnicas más complejas con el fin de hacer sistemas CBIR más eficientes. En este artículo se presenta el desarrollo y la implementación de un sistema CBIR de análisis estadístico local de textura sobre LabVIEW, orientado a implementarse en hardware, dicho sistema emplea algoritmos genéticos para realizar la clasificación de las imágenes.

Palabras clave: Recuperación de imágenes basada en contenido, CBIR, algoritmos genéticos, K-Means, LabVIEW.

1. Introducción

La recuperación de información es el proceso de convertir una solicitud de información en un conjunto significativo de referencias. Los primeros trabajos sobre la recuperación de información aplicada a imágenes se puede remontar a finales de 1980. Desde entonces, el potencial de aplicación de las técnicas en bases de datos ha atraído la atención de los investigadores[6]. Las primeras técnicas no eran, en general, sobre la base de las características visuales, sino en metadatos extraídos en las propiedades de las imágenes tales como anotación al margen o el nombre de las imágenes, a estas técnicas se le llaman "de etiquetado". En otras palabras, las imágenes eran localizadas y caracterizadas utilizando un enfoque basado en texto con sistemas de gestión de etiquetas organizadas en bases de datos tradicionales.

A principios de 1990, como resultado de los avances en la Internet y las nuevas tecnologías en cámaras digitales, la creación de bases de datos de imágenes

digitales de gran volumen ha crecido enormemente, de la mano de la masificación de uso de las cámaras digitales de bolsillo. Las dificultades que enfrento la recuperación basada en texto se hizo más difícil y producía resultados insatisfactorios al principio, pero fue perfeccionándose hasta estar prácticamente agotada al día de hoy, obteniendo muchas veces resultados irrelevantes dado que a sólo usa el etiquetado, que no siempre contempla el contenido total de la imágen, o bien no contempla detalles finos, por ello la necesidad de clasificar o realizar la búsqueda de imágenes fue la fuerza motriz detrás de la aparición de las técnicas de recuperación de imágenes basadas en contenido o CBIR (*Content-Based Image Retrieval*) por sus siglas en inglés. La investigación sobre los métodos de extracción, organización e indexación de información visual han aumentado del mismo modo que una creciente demanda para fines comerciales [6].

Actualmente los sistemas CBIR se apoyan en diversas disciplinas para realizar la clasificación. Por un lado se requiere una forma "inteligente" para identificar las imágenes, y también una forma para realizar la clasificación en grupos o clases afines. Para realizar la tarea de identificación, una técnica es la de algoritmos genéticos (AG), hay varias aplicaciones de AG para la recuperación de información con especial énfasis en la de retroalimentación proveniente del usuario. En este caso el AG aplicado es un algoritmo matemático que transforma un conjunto de propiedades de las imágenes de forma individual, considerándola como una propiedad dinámica con respecto al tiempo y usando operaciones modeladas de acuerdo al principio Darwiniano de evolución, contemplando la reproducción y supervivencia del individuo, como una posible solución al problema, más apto. Cada uno de estos objetos matemáticos suele ser una matriz de propiedades que se ajusta al modelo de las cadenas de cromosomas, y se les asocia con una cierta función que refleja su desempeño o aptitud. Estos individuos con las características ideales se almacenan para la comparación de los datos extraídos de las imágenes y tomar una decisión.

En este artículo mostramos el desarrollo e implementación de un sistema CBIR, orientado a imágenes de escenarios naturales. Dicha implantación se realizó en LabVIEW y la misma está orientada a trabajar sobre hardware (FPGA). El trabajo se desarrolla proporcionando el Estado del arte, para posteriormente en la sección de Desarrollo explicar la metodología de la extracción de las características sobre las imágenes digitales, a partir de descriptores estadísticos; posteriormente se analiza el proceso de agrupación de las imágenes por medio del algoritmo K-means, con ello se proporcionan los resultados experimentales obtenidos y finalmente de comparten las conclusiones y el trabajo futuro de este trabajo.

2. Estado del arte

A principios de los años 90's, aparece por primera vez mencionada la técnica CBIR en un trabajo de T. Kato, el término CBIR fue utilizado para describir a un sistema que recuperaba imágenes de una base de datos basándose en el color y la forma, pero parece que hasta inicios de la década del 2000 se tiene precisión en

el concepto orientado hacia la extracción automática de características, así como a la representación de los datos. Las técnicas CBIR usan características de bajo nivel, e.g. texturas, color y forma para representar a las imágenes.

Por otro lado, el gran reto actualmente es el reconocimiento y la clasificación de imágenes en grandes volúmenes, y el mejor ejemplo de esto es la Internet, que se ha convertido en un lugar de confluencia de diversos tipos de imágenes (rostros, coches, dispositivos electrónicos, mapas, paisajes, flora, fauna, etc.). Cada tipo de imagen tiene sus retos a resolver en las dos tareas de fondo, el reconocimiento y la clasificación, más allá del propósito final del usuario de la consulta, de ahí que generalmente las aplicaciones de CBIR se aterrizan hacia un tipo de imágenes, siendo las de escenarios naturales un tipo muy complicado por la mezcla de colores y formas no regulares. La forma como tradicionalmente de atacó el problema de la organización de imágenes fue a partir de la generación de "etiquetas" asociadas a cada una. Es así que se generaban bases de datos de registros de etiquetas ligadas a cada imagen, pudiendo una imagen tener tantas etiquetas como conceptos se le puedan asociar. Esta técnica se ha automatizado en motores de búsqueda usando autómatas, tal que al analizar una página web con imágenes, se hace un análisis estadístico de las palabras contenidas en texto de la página, y de forma automática se asocian las palabras con mayores frecuencias, como etiquetas de las imágenes de esa página. Esta técnica ha sido muy probada y ampliamente usada, pero se tienen resultados irrelevantes, i.e. que al buscar imágenes de un concepto o palabra el buscador nos regresa imágenes no tienen relación con lo buscado. Otra desventaja directa de la técnica de etiquetado es el idioma de las etiquetas, teniendo que usar un proceso adicional de traducción de las mismas a otros idiomas.

Uno de los pioneros en aplicación de CBIR fue IBM, que patentó el sistema Query By Image Content en el año 1995. Con el paso delos años, fueron apareciendo más sistemas con diferentes variantes, entre los cuales se destacan Photobook realizado en el MIT, Blobworld desarrollado en UC Berkeley, SIMBA, FIRE entre otros [6]. Con el auge y masificación de las imágenes digitales en la Internet aparecieron buscadores de imágenes como Webseer en el año 1996 y Webseek en el año 1997. A su vez, hubo desarrollos importantes en motores de búsqueda de bases de datos relacionales como IBM DB2 y Oracle, donde se incluyeron herramientas a sus productos para facilitar la recuperación de imágenes por contenido visual, acercando el área de CBIR al ámbito de la industria.

En este trabajo presentamos un sistema CBIR, basado en la extracción de características de bajo nivel desarrollado por [6], que realiza de forma automática la detección y clasificación de las imágenes sin la necesidad de un etiquetado, y brindando un enfoque estadístico de análisis de la textura local de ventanas que se abren a partir de puntos de la imagen a estudiar. Otros enfoques de estudio de la imagen, local y global, se han realizado y se pueden analizar su diferentes niveles de desempeño a partir del tipo de imagen, grado de iluminación, rotación, etc. Una detallada revisión de los avances en los últimos años sobre CBIR puede ser consultada en [2], así como en [6].

3. Desarrollo

Para poder desarrollar un sistema de reconocimiento por contenido de imágenes, es necesario contar con bases de datos de imágenes de un tipo o propósito para el entrenamiento del sistema. En ésta fase pueden utilizarse bases de datos estándar y de libre uso como la de Torralba y Vogel. Realizado el entrenamiento se puede proceder con las pruebas, primero con las mismas imágenes con que fue entrenado el sistema, y posteriormente probando con nuevas imágenes del tipo o propósito deseado.

3.1. Extracción de características

El proceso de extracción de características desarrollado en el presente trabajo se divide en cuatro subprocesos, todo visto como una metodología: *(i)* extracción de Información de la imagen en espacio de color RGB, *(ii)* conversión de los datos al espacio de color HSI (para tener más información a bajo nivel), *(iii)* selección de muestras en ventanas de 10 × 10 pixeles, y *(iv)* cálculo valores estadísticos asociados a las texturas seleccionadas (la media (μ), desviación estándar (σ) y *homogeneidad*); tal como se muestra en la Fig. 1.

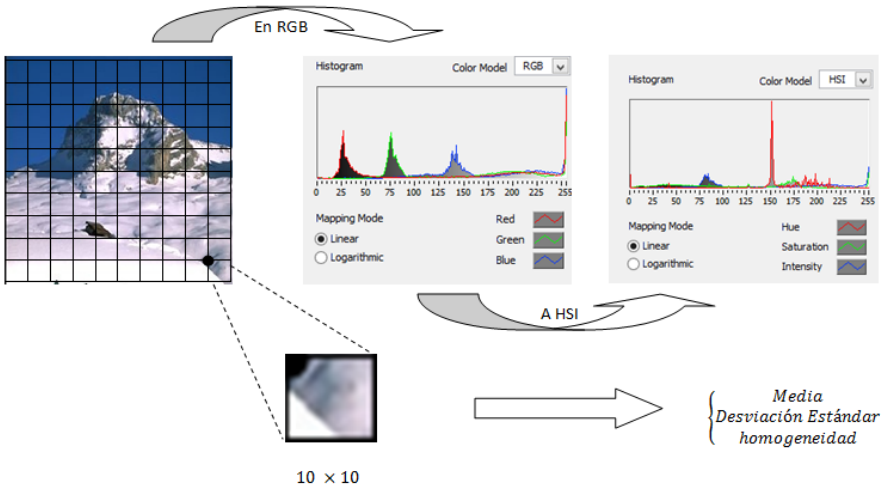

Fig. 1. Extracción de características de la imagen.

Proporcionando más detalle de la metodología, una vez determinada una base de datos con la cual trabajar, se realiza la extracción de información de cada imagen en valores de espacio RGB. Las imágenes de las bases de datos procesadas tienen un tamaño de 256 × 256 pixeles. La información que se obtiene de cada imagen se trabaja en matriz–vector, teniendo un vector unidimensional

de 196608 elementos. En éste vector, se encuentran tres datos por cada pixel, uno para el canal R, uno para el G y otro para el B, iniciando con el $pixel_0$ hasta el $pixel_n$ de la imagen. Para facilitar el manejo de los datos, se redimensiona dicho vector a una matriz de 65536 $\times$ 3 elementos, teniendo en cada fila los datos de los canales RGB correspondientes a cada $pixel_i$.

En esta parte, los valores de la imagen aún están expresados en RGB, pero para obtener mejores resultados, deben ser convertidos a valores HSI *(Hue, Saturation,Intensity)*.

El primer paso para la conversión consiste en normalizar los valores RGB,

$$r = \frac{R}{R+G+B}, g = \frac{G}{R+G+B}, b = \frac{B}{R+G+B}$$

Realizada la normalización, los componentes HSI se obtienen mediante:

$$h = cos^{-1}\left\{\frac{0.5[(r-g)+(r-b)]}{[(r-g)^2+(r-b)(g-b)]^{1/2}}\right\} h \in [0,\pi] \text{ for } b \leq g \tag{1}$$

$$h = 2\pi - cos^{-1}\left\{\frac{0.5[(r-g)+(r-b)]}{[(r-g)^2+(r-b)(g-b)]^{1/2}}\right\} h \in [\pi,2\pi] \text{ for } b > g \tag{2}$$

$$s = 1 - 3 \cdot min(r,g,b) \quad s \in [0,1] \tag{3}$$

$$i = \frac{R+G+B}{3 \cdot 255} \quad i \in [0,1] \tag{4}$$

Por conveniencia para valores de colores neutros tales como negro, blanco y gris, en dónde los valores para R, G y B sob iguales, se considera $H = 0$. De igual manera, los valores H, S e I se convierten en los siguientes rangos:

$$H = \frac{h \times 180}{\pi} \tag{5}$$

$$S = s \times 100 \tag{6}$$

$$I = i \times 255 \tag{7}$$

El siguiente paso, es la toma de muestras. Para el procesamiento de la imagen no se requiere utilizar el total de los datos extraídos de cada imagen. En vez de eso, se decide tomar 100 muestras con una distribución uniforme de la imagen. Por ello, la imagen se trata como si fuera una cuadrícula uniforme de 10 $\times$ 10. En el origen de cada cuadro, se abre una ventana de 10 $\times$ 10 pixeles que conforma la muestra y se integra a una nueva matriz cuyo tamaño final es de 10000 $\times$ 3, la cual se encuentra ya expresada en valores HSI. Para cada ventana se obtienen, la media, desviación estándar y la homogeneidad (de la matriz de co-ocurrencia), la cuál está dada por:

$$\sum_{i,j=0}^{N-1} \frac{P_{i,j}}{1+(i-j)^2} \tag{8}$$

Ahora, se tiene de cada imagen, una matriz de 100 × 9, en los que cada renglón corresponde a los rasgos característicos de cada pixel correspondiente a una muestra de de 10 × 10 pixeles, conformado de la siguiente manera,

$$\{H_\mu, H_\sigma, H_{homogeneidad}, S_\mu, S_\sigma, S_{homogeneidad}, I_\mu, I_\sigma, I_{homogeneidad}\}$$

3.2. Agrupación

Habiendo obtenido la matriz de características de cada imagen se procede a la elección de los centroides. El centroide no es sino un representante que por sus características es el adecuado para cada clase, es decir para cada grupo. Para la definición de los centroides se utiliza un AG estándar de estado estacionario con una estrategia de selección elitista, en donde los elementos con mejores candidatos de cada iteración se mantienen a la siguiente iteración y es invariante bajo permutaciones. La operación de mutación se aplica como es habitual en los candidatos elegidos al azar, como se muestra en la Fig. 2.

Fig. 2. Selección de Centroides usando AG

3.3. Algoritmo genético

La primera generación se elige aleatoriamente, los K elementos elegidos de los grupos $\{\mathcal{C}_1, \mathcal{C}_2, ..., \mathcal{C}_K\}$ respectivamente, definidos en la base de datos de entrenamiento. Consideremos la j-ésima iteración $\mathcal{P}_j = C_1, C_2, ..., C_K$ donde

C_i es un centroide que correspone a la matriz de características de una imagen de la base de datos. Se toma una matriz B correspondiente a una imagen en la base de datos de entrenamiento, se calcula la matriz $D_i^B = |C_i - B|$ para $i = 1, 2, ..., K$ y se define la distancia

$$d(C_i, B) = \frac{\sum d_{ij}}{m \times n}$$

donde $D_i^n = (d_{ij})$ con tamaño $m \times n$.

Se considera al índice

$$i^* = \{\, i \mid \text{el valor } d(\mathcal{C}_i, B) \text{ sea mínimo para } i = 1, 2, ..., K\}$$

Este valor indica que la imagen con matriz de características B pertenece al grupo $\mathcal{C}_{i^*}$.

Una vez asignados todos los elementos de la base de datos de entrenamiento a sus respectivos grupos, se verifica que se cumpla la condición

$$\#(\mathcal{C}_i) \in [m - d, m + d]$$

donde $m = \frac{N}{K}$ y $d = \frac{3N}{K^2}$ para $i = 1, 2, ..., K$. En caso de no cumplir con esta condición, se procede a modificar el conjunto de centroides para la siguiente iteración del AG.

En el momento en el que el sistema deja de iterar se toman el vector de centroides para ser usados en la fase de pruebas.

Para el trabajo actual $N =$ No. de elementos en la base de datos y $K = 10$. Cada elemento del vector de centroides es una matriz de características de 100×9.

```
Algoritmo Genético
Begin AG
   Obtener población inicial P_0 = {C_1, C_2, ..., C_K} aleatoriamente.
   Mientras (PARO) no se cumpla:
    Para i = 1, 2, ..., K
      Calcular d(C_i, B) para cada B en la base de datos.
    Fin
    Calcular i*.
    Asignar B al grupo C_{i*}.
     Para j = 1,2,...,K.
      Si #(C_j) ∈ [m − d, m + d] entonces se conserva el elemento C_j.
      caso contrario mutar el elemento C_j.
     Fin.
    Si no hubo mutación entonces PARO.
Fin.
```

4. Resultados experimentales

El sistema fue implementado usando el lenguaje de programación LabVIEW. Se realizaron diversas subrutinas para integrar dos módulos principales: Entrenamiento y Pruebas. Sólo en este último se tiene la visualización de imágenes por cuestiones de rendimiento y velocidad de procesamiento.

Prueba 1 Se realizaron diversas corridas de pruebas, iniciando con una pequeña base de 36 imágenes de bosques de la Base de Torralba, aunque las imágenes correspondían a escenarios de bosques específicamente, se seleccionaron aquellas que presentaran mayores diferencias entre ellas. Los resultaron obtenidos fueron bastante alentadores ya que el sistema devuelve imágenes bastante similares a la solicitada, como se aprecia en la Fig. 3. En caso de que la imagen que se presenta al sistema corresponde a un centroide, el sistema responde con un 100 % de exactitud, es decir, devuelve la misma imagen.

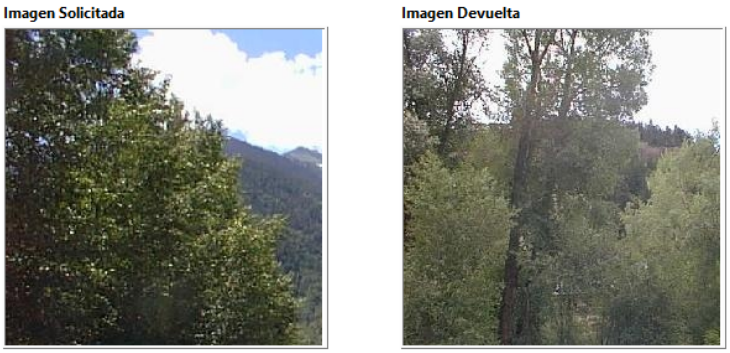

Fig. 3. Resultados con base de 36 imágenes.

Prueba 2 Para la segunda prueba, se utilizaron 328 imágenes de bosques de la misma base de imágenes de la Base Torralba para lograr un mejor entrenamiento que se reflejara en una mayor similitud en el momento de devolver la imagen, lo que efectivamente se consiguió.

Una de las principales interrogantes fue: ¿Qué imagen devolvería el sistema si se le presenta una imagen que no corresponda a un bosque? Así que entre las imágenes de entrada, se le presentaron imágenes pertenecientes a otras categorías como montañas, costas y praderas, en los que se observó un par de cosas interesantes,

1. En primer lugar, el tiempo que el sistema se tomaba para devolver una imagen fue hasta diez veces más que cuando se le presentaron imágenes de bosques solamente.
2. En segundo lugar, tal cómo se esperaba, la imagen devuelta tiene una alto grado de similitud en textura con la imagen solicitada como lo muestra la Fig. 4, aunque corresponde un bosque, lo anterior demuestra que los clusters del sistema han sido formados de manera correcta.

Prueba 3 Partiendo de los resultados de la prueba anterior, se integró una base de datos mixtas con imágenes extraídas de la base de datos de Torralba, pero en ésta ocasión mixta, usando 25 imágenes de costas, 25 de bosques, 25 de praderas y 25 de montañas.

La Fig. 5, muestra una de las imágenes con la que se probó el sistema con la base de datos mixta.

Prueba 4 Teniendo los resultados preliminares, se realizaron pruebas utilizando imágenes que no se encontraban en las bases de datos de entrenamiento. En ésta cuarta prueba, se tomaron imágenes escenarios naturales del Ejido Emiliano Zapata, municipio de Jiquipilas, Chiapas.

Una de las principales diferencias que se observó fue el tamaño, se hizo necesario escalarlas para que la información obtenida al procesarlas fuera representativa. El tamaño de éstas imágenes quedó en (256×192), y al presentarlas al sistema los resultados fueron similares a los obtenidos usando las imágenes de la misma base de datos usada en el entrenamiento, como se ve en la Fig. 6.

Finalmente, se realizaron pruebas con las diferentes clases de imágenes por el contenido asociado–detectado, y estas se muestran en la Tabla 1.

Tabla 1. Resultados obtenidos con matriz mixta

No	$Vector_1$		$Vector_2$		$Vector_3$	
	Presentada	Devuelta	Presentada	Devuelta	Presentada	Devuelta
1	costa	costa	costa	costa	costa	montaña
2	bosque	bosque	costa	costa	costa	montaña
3	bosque	bosque	costa	costa	costa	costa
4	montaña	bosque	bosque	bosque	bosque	bosque
5	pradera	montaña	bosque	bosque	bosque	bosque
6	pradera	pradera	bosque	pradera	montaña	montaña
7	costa	costa	montaña	montaña	montaña	montaña
8	costa	costa	pradera	bosque	montaña	montaña
9	montaña	montaña	pradera	pradera	montaña	pradera
10	montaña	montaña	pradera	pradera	pradera	pradera
Exactitud	80 %		90 %		70 %	

5. Conclusiones

Con respecto a las diferentes técnicas de CBIR, se eligió estudiar la implementación de los métodos de K-means y AG's, partiendo del hecho que se desea un equilibrio, entre el tiempo de procesamiento y eficiencia en la tarea de caracterización de imágenes en base de imágenes de gran volumen. El software creado fue comparado para medir la calidad de recuperación contra los módulos de MATLAB. Con el fin de mostrar avances en cuanto al software, se realizó la

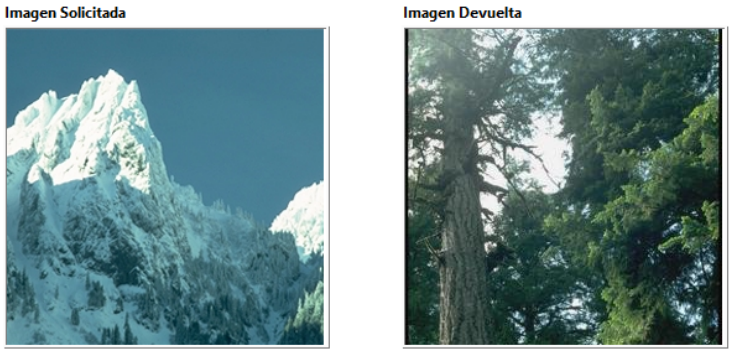

Fig. 4. Resultado solicitando imagen de montaña con base de datos de bosques.

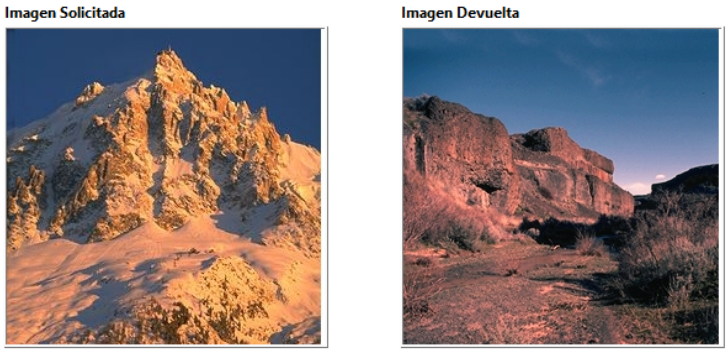

Fig. 5. Resultado solicitando imagen en Base de Imágenes Mixta

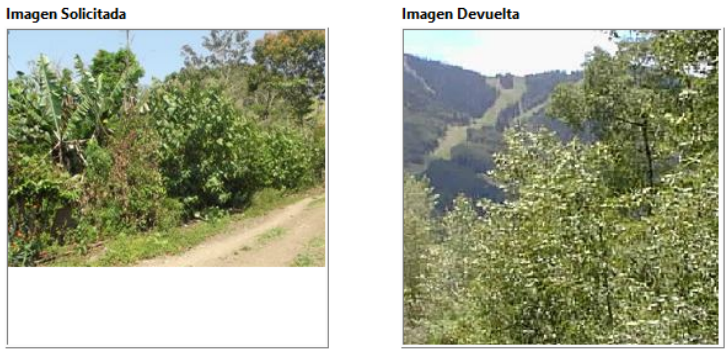

Fig. 6. Resultado solicitando imagen de paisaje de Chiapas.

programación completa de rutinas de K-means y AG. Es de destacar dichos módulos no se encuentran disponbles en LabVIEW a diferencia de MATLAB. Se uso el software LabVIEW porque en etapas futuras se desea la implementación del software en hardware especializado para tener un procesamiento en tiempo real. Este aspecto se considera importante, si se toma en cuenta que fueron usadas fotos personales ajenas a cualquier base de datos y se usan computadoreas comerciales. Se observaron tiempos de entrenamiento lineales con respecto a la cantidad de elementos en la base de datos de entrenamiento, de igual forma, se obtuvieron diferencias en tiempos de procesamiento, mostrando una reducción considerable en los tiempos de entrenamiento y prueba.

Los resultados fueron satisfactorios como se observa en la Tabla 1, logrando niveles de exactitud en recuperación hasta de un 90 %

Como trabajo futuro planteamos dos puntos inmediatos: *(i)* ampliar el número de imágenes de prueba, y *(ii)* realizar la implementación en hardware FPGA o dispositivo móvil del sistema CBIR optimizado, de tal manera que mediante una cámara del propio dispositivo se adquieran y procesen las imágenes, y sea capaz de clasificar el tipo de imagen por medio de sus características usando esta metodología aquí dispuesta.

Referencias

1. Bosch, A., Zisserman, A., Muoz, X. Scene classification using a hybrid generative/discriminative approach. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 30, 712–727 (2008)
2. T. Dharani and I. Aroquiaraj. A survey on content based image retrieval. In: International Conference on Pattern Recognition, Informatics and Mobile Engineering (PRIME), pp. 485–490 (Feb 2013)
3. Fukunaga, K. Introduction to statistical pattern recognition. (2nd ed.) San Diego, CA, USA: Academic Press Professional, Inc. (1990)
4. Li, J., Wang, J. Z. Real-time computerized annotation of pictures. Proceedings of the 14th annual ACM international conference on multimedia. New York, NY, USA: ACM. (pp. 911–920) (2006)
5. Liu, Y., Zhang, D., et al. A survey of content-based image retrieval with high level semantics. Pattern Recognition, 40, 262–282 (2007)
6. Serrano-Talamantes J.F. , Avilés-Cruz C., Villegas-Cortez J., and Sossa-Azuela J. H., Self organizing natural scene image retrieval, Expert Systems with Applications, vol. 40, no. 7, pp. 2398–2409 (2013)
7. Vogel, J., Schiele, B. Semantic modeling of natural scenes for content-based image retrieval. International Journal of Computer Vision, 72(2), 133–157 (2007)