Content-Based Image Retrieval Using Color Space Transformation and Wavelet Transform

Te-Wei Chiang¹  Tienwei Tsai²

¹ Department of Information Networking Technology, Chihlee Institute of Technology, No. 313, Sec. 1, Wunhua Rd., Banciao City, Taipei County 220, Taiwan, R.O.C.

² Department of Information Management, Chihlee Institute of Technology, No. 313, Sec. 1, Wunhua Rd., Banciao City, Taipei County 220, Taiwan, R.O.C.

Abstract

Retrieval by image content has received great attention in the last decades. In this paper, a content-based image retrieval method based on the wavelet transform is proposed. Due to the superiority in multiresolution analysis and spatial-frequency localization, the wavelet transform is applied to extract low-level features from the images. In the image database establishing phase, each image is first transformed from the standard RGB color space to the YUV space; then Y component of the image is further transformed to the wavelet domain. In the image retrieving phase, the system compares the most significant wavelet-subbands of the Y component of the query image and those of the images in the database and find out good matches. The feature sets derived from the color space transformation and the wavelet transform can reduce the processing time without sacrificing the retrieving accuracy. We have performed experiments on a thousand images database and our results show a high retrieval accuracy.

Keywords: Content-based image retrieval, query-by-example, wavelet transform, color space.
1. Introduction

Images have been used for a variety of purposes, including industrial, military, medical, civil (e.g., traffic), scientific, business, and entertainment. With the rapid spread of computer networks and further development of multimedia technology, digital contents can now be accessed more easily, thereby resulting in urgent need of image retrieval. Generally, image retrieval procedures can be roughly divided into two approaches: query-by-text (QbT) and query-by-example (QbE). In QbT, queries are texts and targets are images; in QbE, queries are images and targets are images. For practicality, images in QbT retrieval are often annotated by words, such as time, place, or photographer. To access the desired image data, the seeker can construct queries using homogeneous descriptions, such as keywords, to match these annotations. Such retrieval is known as annotation-based image retrieval (ABIR). ABIR has the following drawbacks. First, manual image annotation is time-consuming and therefore costly. Second, human annotation is subjective. Furthermore, some images could not be annotated because it is difficult to describe their content with words. On the other hand, annotations are not necessary in a QbE setting, although they can be used. The retrieval is carried out according to the image contents. Such retrieval is known as content-based image retrieval (CBIR) [1]. CBIR becomes popular for the purpose of retrieving the desired images automatically. Smeulder et al. [2] reviewed more than 200 references in this field.

In QbE, the retrieval of images basically has been done via the similarity between the query image and all candidates on the image database. To evaluate the similarity between two images, one of the simplest ways is to calculate the Euclidean distance between the feature vectors representing the two images. To obtain the feature vector of an image, some transform type feature extraction techniques can be applied, such as wavelet [3], Walsh, Fourier, 2-D moment, DCT [4-6], and Karhunen-Loeve. In our image retrieval scheme, the wavelet transform is used to extract low-level texture features due to its superiority in multiresolution analysis and spatial-frequency localization. We hope that the feature sets derived from the wavelet transform can reduce the processing time without sacrificing the retrieving accuracy.

In this paper, we propose a content-based image retrieval method based on the wavelet transform. In the image database establishing phase, each image is first transformed from the standard RGB color space to the YUV space; then Y component of the image is further transformed to the wavelet domain. In the image retrieving phase, the system compares the most significant wavelet subbands of the Y component of the query image and those of the images in the database and find out good matches.

The remainder of this paper is organized as follows. The next section is the related works. Section 3 presents the proposed image retrieval system. Experimental results are shown in Section 4. Finally, conclusions are drawn in Section 5.

2. Related Works

Content-based image retrieval is a technology to search for similar images to a query based only on the image pixel representation. However, the query based on pixel information is quite time-consuming because it is necessary to devise a means of describing the location of each pixel and its intensity. Therefore, how to choose a suitable color space and reduce the data to be computed is a critical problem in image retrieval.

Some of the systems employ color histograms. The histogram measures are only dependent on summations of identical pixel values and do not incorporate orientation and position. In other words, the histogram is only statistical distribution of the colors and loses the local
information of the image. Therefore, we propose an image retrieval scheme to retrieve images from their transform domain, which tries to reduce data and still retains their local information.

In this paper, we focus on the QbE approach. The user gives an example image similar to the one he/she is looking for. Finally, the images in the database with the smallest distance to the query image will be given, ranking according to their similarity. We can define the QbE problem as follows. Given a query image $Q$ and a database of images $X_1, X_2, \ldots, X_n$, find the image $X_i$ closest to $Q$. The closeness is to be computed using a distance measuring function $D(Q, X_i)$ which will be defined in Section 3.3. In the next section we shall introduce our image retrieval scheme.

### 3. The Proposed Image Retrieval System

#### 3.1 System Architecture

Figure 1 shows the system architecture of our wavelet-based QbE system. This system contains two major modules: the feature extraction module and the similarity measure module. The details of each module are introduced in the following subsections.

#### 3.2 Feature Extraction

Classifying an unknown input is a fundamental problem in pattern recognition. A common method is to define a distance function between the features of patterns and find the most similar pattern in the reference set. That is, features are functions of the measurements performed on a class of patterns that enable that class to be distinguished from other classes in the same general category [7]. To have an effective retrieval, we have to extract distinguishable and reliable features from the images.

During the feature extraction process, the images have to be converted to the desired color space. There exist many models through which to define the valid colors in image data. Each
of the following models is specified by a vector of values, each component of that vector being valid on a specified range. This presentation will cover the following major color spaces definitions [8]: RGB (Red, Green, and Blue), CMYK (Cyan, Magenta, Yellow, and Black Key), CIE (Centre International d’Eclairage), YUV (Luminance and Chroma channels), etc. In our approach, the RGB images are first transformed to the YUV color space.

3.2.1 RGB Color Space

A gray-level digital image can be defined to be a function of two variables, \( f(x, y) \), where \( x \) and \( y \) are spatial coordinates, and the amplitude \( f \) at a given pair of coordinates is called the intensity of the image at that point. Every digital image is composed of a finite number of elements, called pixels, each with a particular location and a finite value. Similarly, for a color image, each pixel \((x, y)\) consists of three components: \( R(x, y) \) (red), \( G(x, y) \) (green), and \( B(x, y) \) (blue), each of which corresponds to the intensity of the red, green, and blue color in the pixel, respectively.

3.2.2 YUV Color Space

Originally used for PAL (European "standard") analog video, YUV is based on the CIE Y primary, and also chrominance. The Y primary was specifically designed to follow the luminous efficiency function of human eyes. Chrominance is the difference between a color and a reference white at the same luminance. The following equations are used to convert from RGB to YUV spaces:

\[
Y(x, y) = 0.299 \times R(x, y) + 0.587 \times G(x, y) + 0.114 \times B(x, y),
\]

\[
U(x, y) = 0.492 \times (B(x, y) - Y(x, y)),
\]

\[
V(x, y) = 0.877 \times (R(x, y) - Y(x, y)).
\]

After converting from RGB to YUV, the features of each image can be extracted by the wavelet transform.

3.2.3 Wavelet Transform

An image of size \( M \times N \) can be decomposed into its wavelet coefficients by using Mallat’s pyramid algorithm [9]. Mathematically, it can be described as the following recursive equations [10]:

\[
LL^{(k)}(m, n) = \left[ [LL_{\text{rows}}^{(k-1)} \ast H]_{2^{k}} \right]_{\text{columns}} \ast \overline{H},
\]

\[ m = 1, \ldots, M / 2^{k}; \quad n = 1, \ldots, N / 2^{k}, \]

\[
LH^{(k)}(m, n) = \left[ [LL_{\text{rows}}^{(k-1)} \ast \overline{H}]_{2^{k}} \right]_{\text{columns}} \ast G,
\]

\[ m = M / 2^{k} + 1, \ldots, M / 2^{k-1}; \quad n = 1, \ldots, N / 2^{k}, \]

\[
HL^{(k)}(m, n) = \left[ [LL_{\text{rows}}^{(k-1)} \ast G]_{2^{k}} \right]_{\text{columns}} \ast H,
\]

\[ m = 1, \ldots, M / 2^{k}; \quad n = M / 2^{k} + 1, \ldots, N / 2^{k-1}, \]
Here $LL$, $LH$, $HL$, and $HH$ represent four subimages of the image being decomposed. $L$ and $H$ are used to indicate low- and high-frequency components. $\overline{H}$ and $\overline{G}$ correspond to the low-pass and the high-pass filters, respectively. Expression $2 ↓ 1(1 ↓ 2)$ denotes sampling along column (row), and $k$ is the level of wavelet decomposition. Equations (4)-(7) indicate that any image signal can be decomposed in a specific wavelet domain. The wavelet decomposition is illustrated in Figure 2. $LL^{(0)}$ is the original image. The output of high-pass filters $LH^{(1)}$, $HL^{(1)}$, $HH^{(1)}$ are three subimages with the same size as low-pass subimage $LL^{(1)}$, and presenting different image details in different directions. After wavelet decomposition, the object image energy is distributed in different subbands, each of which keeps a specific frequency component. In other words, each subband image contains one feature. Intuitively, the feature at different subbands can be distinguished more easily than that in the original image.

3.3 Distance Measures

To decide which image in the image database is the most similar one with the query image, we have to define a measure to indicate the degree of similarity (or dissimilarity). In our approach the sum of squared differences (SSD) is used for this purpose. Since the lowest frequency wavelet subband, $LL^{(k)}$, is the most significant subband for $k$th-level wavelet decomposition, we can retrieve the desired image(s) by comparing the $LL^{(k)}$ subband of the candidate images with that of the query image. Assume that $LL^{(k)}_q(m,n)$ and $LL^{(k)}_{s_n}(m,n)$ represent the wavelet coefficients of the query image $Q$ and image $X_n$ under $LL^{(k)}$, respectively. Then the distance between $Q$ and $X_n$ under the $LL^{(k)}$ subband can be defined as

$$D_{LL^{(k)}}(Q, X_n) = \sum_m \sum_n \left( LL^{(k)}_q(m,n) - LL^{(k)}_{s_n}(m,n) \right)^2.$$  

(8)
4. Experimental Results

In this preliminary experiment, 1000 images downloaded from the WBIIS [11] and the SIMPLicity [12] database are used to demonstrate the effectiveness of our system. The user can query by an external image or an image from the database. The difference between these two options is that when an external image is used its features need to be extracted while if the image is already in the database, its features are already extracted and stored in the database along the image. Therefore, when a query submitted using an image from the database, only the index of the image in the database is transferred to the server. In both cases the features used are imposed by the database selected at the start. In this experiment, we only use the images from the database as query images. To evaluate the retrieval efficiency of the proposed method, we use the performance measure, the precision rate as shown in Equation (9),

\[
\text{Precision rate} = \frac{R_r}{T_r}, \tag{9}
\]

where \( R_r \) is the number of relevant retrieved items, and \( T_r \) is the number of all retrieved items.

To verify the effectiveness of the color space transformation and the wavelet transform, we conducted a series of experiments using different type of feature sets, i.e., Y-component image, Y-component \( LL^{(1)} \) subimage, and Y-component \( LL^{(2)} \) subimage. A butterfly is used as the query image. The retrieved results are ranked in the top ten in similarity. Fig. 3 - Fig. 8 show the retrieved results, where the items are ranked in the ascending order of the distance to the query image from the left to the right and then from the top to the bottom. Figure 3 shows the retrieved results by using the SSD distance between the original RGB bitmap of the query image and those of the images in the image database. Figure 4 shows the retrieved results by using the SSD distance between the Y-component of the query image and those of the images in the image database. Note that the number of features of the Y-component image is one-third of that of the original RGB image; therefore, there exists a two-third reduction of the size of feature set. Figures 5 and 6 show the retrieved results (RGB images and wavelet-based image) by using the SSD distance between the \( LL^{(1)} \) subimage of the Y-component of the query image and those of the images in the image database. Note that the number of features of the \( LL^{(1)} \) subimage of the Y-component image is one-fourth of that of the Y-component image; therefore, there exists a three-fourth reduction of the size of feature set. Figures 7 and 8 show the retrieved results (RGB images and wavelet-based image) by using the SSD distance between the \( LL^{(2)} \) subimage of the Y-component of the query image and those of the images in the image database. We can find that the retrieved results are similar for different type of features in terms of the precision rate, but the size of feature set can be reduced tremendously by using more elaborate feature sets. Table 1 shows the precision rate and reduction rate of the test image over different comparison methods. Therefore, the feature

<table>
<thead>
<tr>
<th></th>
<th>Original RGB image</th>
<th>Y-Component image</th>
<th>( LL^{(1)} ) subimage</th>
<th>( LL^{(2)} ) subimage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Precision rate</td>
<td>0.7</td>
<td>0.7</td>
<td>0.7</td>
<td>0.7</td>
</tr>
<tr>
<td>Reduction rate</td>
<td>0</td>
<td>0.667</td>
<td>0.917</td>
<td>0.979</td>
</tr>
</tbody>
</table>

Table 1. Precision rate and the reduction rate of the test image under different type of features.
sets derived from the color space transformation and the wavelet transform can improve the reduction rate and hence reduce the processing time without sacrificing the retrieving accuracy.

5. Conclusions

Content-based image indexing and retrieval has been a very active research area. In this paper, a content-based image retrieval method based on the color space transformation and the wavelet transform is proposed. Each image is first transformed from the standard RGB color space to the YUV space; then Y component of the image is further transformed to the wavelet domain. The wavelet transform is applied to extract low-level features from the images due to its superiority in multiresolution analysis and spatial-frequency localization. To achieve QbE, the system compares the most significant wavelet-subbands of the Y-component of the query image and those of the images in the database and find out good matches.

Since only preliminary experiment has been made to test our approach, a lot of works should be done to improve this system:

Since there is no feature capable of covering all aspects of an image, the discrimination performance is highly dependent on the selection of features and the images involved. For each type of feature we will continue investigating and improving its ability of describing the image and its performance of similarity measuring. Then, since several features may be used simultaneously, it is necessary to develop a scheme that can integrate the similarity scores resulting from the matching processes.

A long-term aim is combining the semantic annotations and low-level features to improve the retrieval performance. That is, the retrieved images should be somehow related to the objects contained in the scenes. For the analysis of complex scenes, the concept that provide a high amount of content understanding enable highly differentiated queries on abstract information level. The concept is worthy of further study to fulfill the demands of integrating semantics into CBIR.

Reference


Figure 3. Retrieved results via the comparison of the original RGB images.

Figure 4. Retrieved results via the comparison of the Y-component images.
Figure 5. Retrieved results (RGB images) via the comparison of the Y-component $LL^{(1)}$ subimages.

Figure 6. Retrieved results (wavelet-based images) via the comparison of the Y-component $LL^{(1)}$ subimages.
Figure 7. Retrieved results (RGB images) via the comparison of the Y-component $LL_{(2)}$ subimages.

Figure 8. Retrieved results (wavelet-based images) via the comparison of the Y-component $LL_{(2)}$ subimages.