DOI: 10.2298/CSIS120716035L

Content-based Image Retrieval using Spatial-color and Gabor Texture on a Mobile Device

Yong-Hwan Lee[1], Bonam Kim[2] and Sang-Burm Rhee[3]

[1] Department of Applied Computer Engineering, Dankook University
152, Jukjeon-ro, Suji-gu, Yongin-si, Gyeonggi-do, 448-701, Korea
hwany1458@empas.com

[2] Division of Electrical and Computer Engineering, Chungnam National University
99, Daehak-ro, Yuseong-gu, Daejon, 305-764, Korea
kimbona@cnu.ac.kr**

[3] Department of Applied Computer Engineering, Dankook University
152, Jukjeon-ro, Suji-gu, Yongin-si, Gyeonggi-do, 448-701, Korea
sbrhee@dankook.ac.kr

Abstract. Mobile image retrieval is one of the most exciting and fastest growing research fields in the area of multimedia technology. As the amount of digital contents continues to grow users are experiencing increasing difficulty in finding specific images in their image libraries. This paper proposes a new efficient and effective mobile image retrieval method that applies a weighted combination of color and texture utilizing spatial-color and second order statistics. The system for mobile image searches runs in real-time on an iPhone and can easily be used to find a specific image. To evaluate the performance of the new method, we assessed the Xcode simulations performance in terms of average precision and F-score using several image databases and compare the results with those obtained using existing methods such as MPEG-7. Experimental trials revealed that the proposed descriptor exhibited a significant improvement of over 13% in retrieval effectiveness, compared to the best of the other descriptors.

Keywords: Mobile Content-Based Image Retrieval, Image Representation and Recognition, Image Descriptor, Spatial-Color, Gabor Texture

1. Introduction

The mobile phones that we use in our everyday life have become popular multimedia devices, and it is not uncommon to observe users capturing photos on their mobile phones, instead of using dedicated digital cameras or video cameras. This ease of use means that the number of digital images will only continue to rise with the rapid development of mobile devices, and individual users rapidly accumulate very large image repositories including both their personal photo archives and image from wider-access digital libraries. At present, users generally browse their personal multimedia repositories on mobile devices by

** Corresponding author : Bonam Kim

scrolling through image libraries or by manually creating a series of folders to fit their needs, and browsing through these folders. However, as the amount of digital content continues to increase end-users are beginning to suffer difficulties in locating specific images. As a result, research into more effective image retrieval techniques is currently receiving a great deal of attention, and image retrieval is now one of the most exciting and fastest growing areas in the field of multimedia technology [11].

There are two main approaches to image retrieval: text-based retrieval and content-based retrieval [22]. The popular text-based method requires images to be annotated with one or more keywords that can then be easily searched. However, this method involves a vast amount of labor and tends to be colored by personal subjectivity; the resulting lack of clarity often leads to mismatches in the retrieval process. In particular, this approach runs into critical problem concern the possibility of mismatches in a personal photo database. The alternative content-based method indexes images in a database by identifying similarities between them based on lower-level visual features such as color, texture, shape and spatial information [22] [23]. Although some systems are designed for a specific domain such as medical image retrieval or personal identification [19], a CBIR (Content-based Image Retrieval) system typically requires the construction of an *image descriptor*, which is characterized by two primary functions [23]. One is an extraction process that encodes the image into feature vectors, and another is a similarity measure that compares two images. The image descriptor D is formulated into 2-tuples as $(F_D,\, S_D)$, where $F_D : \{I\} \rightarrow R^n$ is an extraction function that extracts a feature vector $\boldsymbol{f}$ from image $\boldsymbol{I}$, and $S_D : R^n \times R^n \rightarrow R$ is a distance measure function that computes the similarity between two feature vectors corresponding to images.

This paper proposes a new and more efficient mobile image descriptor that utilizes a weighted combination of color and texture features based on spatial-color and second-order statistical texture. This paper is an extension of work first presented in [9], here we provide a more thorough experimental comparison, and demonstrate much improved performance[4].

This paper is organized as follows: Section 2 reviews the related research on image retrieval and mobile image searching. In Section 3, we provide full details of our proposed descriptor. Section 4 presents some experimental results obtained using the proposed approach and assesses its performance compared to the methods currently used. Lastly, Section 5 concludes the paper by summarizing the study and discussing possible directions for future research.

[4] These improvements are due to three changes: First, a better algorithm is applied in the pre-processing stage to enhance computational time and memory efficiency. These involve checking the SubjectArea tag of EXIF metadata and reducing the size of the query image, either with main area or down sampling. Second, we utilize the Haar wavelet filter to analyze the image. Third, image databases were prepared more carefully and reasonably, focusing on the type of natural images that would be commonly found in personal photo libraries.

2. Related Work

This section summaries recently published research on content-based image retrieval including a consideration of the features commonly used in image searches and the issues involved in feature extraction from a color image.

2.1. Content-based Approaches

Many general-purpose image retrieval systems have been developed and proposed by both industrial and academic research laboratories, and it is not practicable to attempt to survey all of these in the limited space available. Hence, we focus on those works that deal specifically with standardization, especially those that feature in SC29/WG1 JPSearch. The standard approach involves two steps [10]:

```
Algorithm 1: RETRIEVAL finds out the relevant images from database
    Input: A query image
    Output: Relevant images in the database
    $Extracting\ image\ features\ to\ a\ distinguishable\ extent$
    $Matching\ these\ features\ to\ yield\ a\ result\ that\ is\ visually\ similar$
    return $relevant\ images$
```

Fig. 1 shows the approaches traditionally used to search digital images. The type of image search and retrieval systems shown schematically on the left hand side of the figure require each image to be associated with one or more keywords entered by a human operator, while those on the use an image as a query and then attempts to retrieve other images which are similar. This represents the current state-of-the-art in CBIR systems described in SC29/WG1 JPSearch, International Standardization [10].

There already exist many research systems which apply image retrieval techniques to mobile search. IDeixis is an image-based mobile search system which combines image retrieval method and text-based search techniques. It uses CBIR methods to search the Web repository and/or other databases for matching images, and their result pages are based on current users location to find relevant images [26]. But, they still have the possibility of mismatching with text-based method. Reference [1] presents a content-based multimedia retrieval system designed for mobile platforms running Symbian-based OS. It is built on client-server architecture, and the system basically focusing on server-side application, while the client-side consists of the user interface and controllers. MOSIR is also a CBIR system for mobile phones, which facilitates instantaneous search on mobile phones for images similar to photos taken via phone cameras [18]. To find similar images, edge-based and color-layout features are used. It also enables region-based queries by detecting salient regions and extracting their features. They utilized image segmentation techniques to image

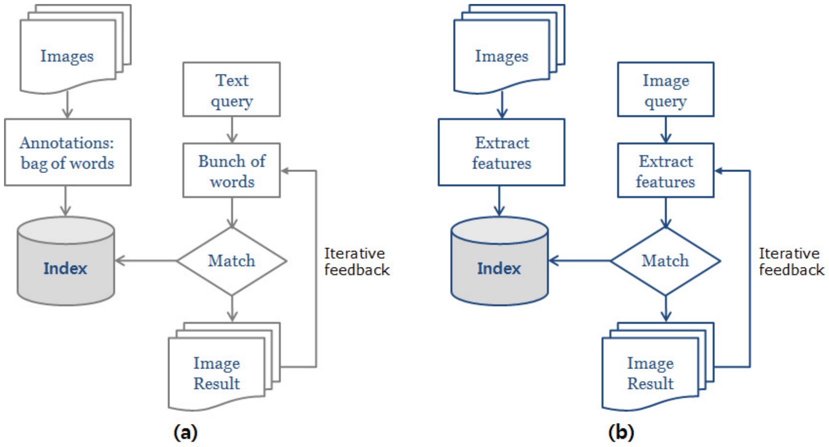


Fig. 1. Naive system view of image search and CBIR [10].

filtering processes prior to image analysis, while we use EXIF data, which gives a better way to find approximate position within image. In [24], they present mobile search systems which support image queries and audio queries, covering typical design for mobile visual and audio search. But, their work flow of matching core logic is running on the server, and client-side module is only working as interface for the user.

2.2. Conventional Features Relevant to the Current Research

This subsection explains the conventional features used in the proposed retrieval method, namely *auto-correlograms* for color features and *gray-level co-occurrence metrics* as texture features.

A histogram is a graphical display of frequencies that represents the total distribution in a digital image [4]. The histogram for color c_i of image $\boldsymbol{I}$ is formulated as follows:

$$H_{c_i}(I) \ = \ probability_{p \in I}[p \in I_{c_i}] \tag{1}$$

Since the histogram simply corresponds to the probability of there being any pixels of color c_i in an image, this feature does not take into account the spatial distribution of color across different areas of the image. A correlogram characterizes not only the color distributions of pixels, but also the spatial correlations of pairs of colors (c_i, c_k) [20]. This feature describes the probability of finding a pixel p_2 of special color c_k at a distance d for a pixel p_1 of given color c_i. The correlogram for a color pair (c_i, c_k) is formulated as follows:

$$C^d_{c_i,c_k}(I) \ = \ probability_{p_1 \in I_{c_i}, p_2 \in I}[p_2 \in I_{c_k} || p_1 - p_2| = d] \tag{2}$$

With all possible combinations of color pairs, the size of the correlogram is inevitably very large so a formulation, known as an auto-correlogram, is generally used. An auto-correlogram gives the probability of capturing the spatial correlation between identical colors only, and this is effectively a simplified subset of the correlogram, signified by $\Gamma_c^d(I) = C_{c_i,c_k}^d(I)$. Thus, the auto-correlogram is formulated as follows:

$$\Gamma_c^d(I) \ = \ probability_{p_1 \in I_c, p_2 \in I}[p_2 \in I_c || p_1 - p_2| = d] \tag{3}$$

Gray Level Co-occurrence Matrices (GLCM) contain information about the positions of pixels having similar gray level values [6]. The GLCM is defined by calculating how often a pixel with the intensity value i occurs in a specific spatial relationship to a pixel with the value j [12]. That is, the GLCM is created by first specifying a displacement vector $d = (d_x, d_y)$ and then counting all the pairs of pixels separated by d having gray levels i and j. The GLCM G is therefore computed over an $n \times m$ image, parameterized by an offset $(\Delta x, \Delta y)$ as follows:

$$G_d[i,j] = \sum_{p=1}^{n} \sum_{q=1}^{m} \begin{cases} 1, \text{ if } I(p,q) = i \text{ and } I(p + \Delta x, q + \Delta y) = j \\ 0, \text{ otherwise} \end{cases} \tag{4}$$

These spatial relationships can be specified in a number of different ways, but the default is that between a pixel and its immediate neighbor to its right. However, it is possible to specify this relationship with different offsets and angles, as described in Section 3.

3. Proposed Mobile Image Descriptor

Fig. 2 depicts a block diagram of the proposed retrieval approach. A single feature may lack sufficient discriminatory information to permit the retrieval of relevant images [21], so for this study we opted to use multiple features utilizing a combination of color and texture features that have been extracted separately. In addition, local features have been proved to be effective in image analysis [7]. When a query image I_Q is entered into the retrieval system, it must first be preprocessed by reducing the size of the querying image and by performing color space conversion and color channel separation. Each of the channels is then wavelet decomposed into a wavelet image, after which color features $\boldsymbol{f}_c$ and texture features $\boldsymbol{f}_t$ are extracted from the transformed image. Next, the system combines two features with appropriate weighting to generate the query vector $\boldsymbol{F}$, which is applied to the extraction process used to encode images into feature vectors. In order to respond to the user's query, the system then computes the similarity between the query vector $\boldsymbol{F}_Q$ and each of the target vectors $\boldsymbol{F}_T$ in

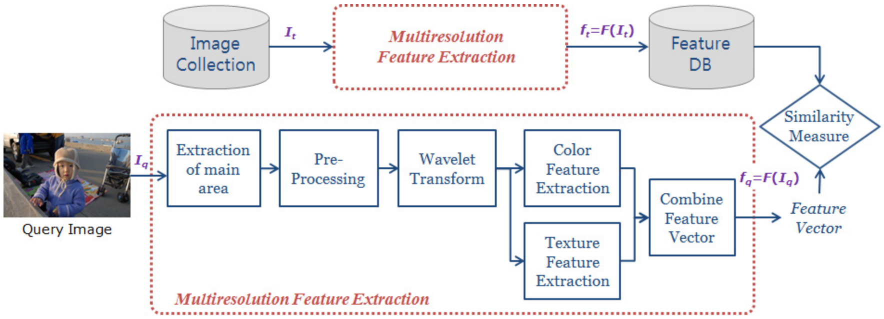


Fig. 2. Block Diagram of the Proposed Mobile Image Retrieval Method.

the database. Finally, it returns similar target images from the image database according to their similarity rankings.

Details of how the proposed algorithm extracts the color features using a wavelet spatial-color correlogram and the texture features using second-order statistical data for the texture based on Gabor wavelet transforms are provided in the following subsection.

3.1. Extraction of Spatial-color Based on Main Focus Region

The procedure for extracting the color features from an input image is shown in Fig. 3. Given a RGB query image, the SubjectArea tag of EXIF is first checked for resizing the entered image. The tag indicates the location and area of the main subject in the overall scene [5] [3], enabling the main area of the image to be extracted from the original if the tag is set as the region of interest (ROI). The main area is decided by computing the value of the tag, and by choosing the largest areas of intersection, as shown in Fig. 4. Alternatively the whole image can be reduced using bi-directional down-sampling algorithm such as YCbCr $4:2:0$ Co-sited of the JPEG standard [5]. Then, the RGB color space of the reduced image is converted into HSV space, presented as I^c, where $c \in \{H, S, V\}$. When extracting a color feature with a correlogram, HSV color space is known to provide better correspondence with human perceptions of color similarities than other color spaces [20].

Next, each of the three channels is wavelet transformed into two consecutive levels using a Haar filter, which is a good compromise between computational time and performance [13], denoted as $W^C_{s,l}$ where s indicates the four orientations of the sub-bands $s \in \{LL, LH, HL, HH\}$, l is the level of wavelet decomposition, and C represents the color channels. Thereafter, wavelet coefficients are quantized into $Q^C_{s,l}$ with different levels for each scale and sub-band. The number of quantization levels for each sub-band is weighted, according to the ratio $LL : LH : HL : HH = 2 : 1 : 1 : 0$. The correlogram for the HH sub-band cannot be computed because wavelet coefficients corresponding to HH have no significant spatial correlation. In order to reduce the computational

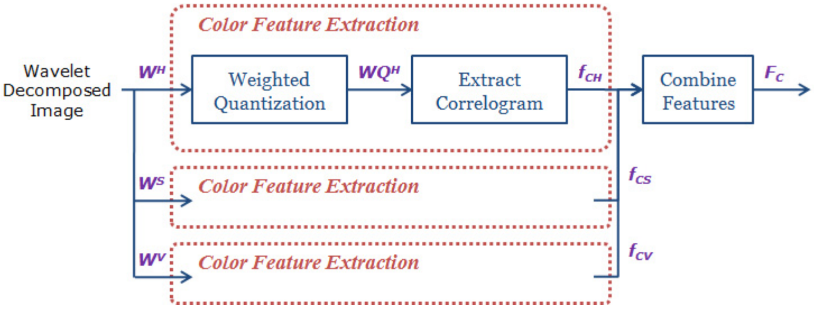


Fig. 3. Procedure for Weighted Wavelet Spatial Color Extraction.

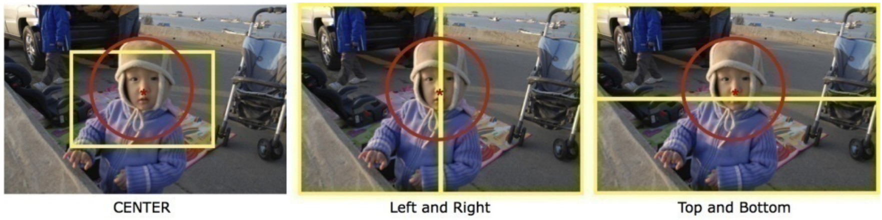


Fig. 4. Five Types of Main Area of Query Image with SubjectArea Tag (Red point and circle).

time required for the extraction of the feature vectors a quantized color *codebook* was created for the proposed algorithm that functions as a color lookup table. Next, the horizontal, vertical and both directional correlograms for the quantized coefficients are calculated for the LH, HL and LL sub-bands in each scale. The correlogram of image I, which comprises the pixels $p(x, y)$ is then re-formulated from the definition of equation (X):

$$\Gamma_c^d(I) = \frac{|\{p(x,y) \mid I(x,y) = c_i; I(x \pm d, y \pm d) = c_i\}|}{|\{p(x,y) \mid I(x,y) = c_i\}|} \tag{5}$$

where c_i is the distinct value of each color and d is the fixed distance of the correlation.

Thus, the correlogram of wavelet coefficients for LL is computed as follows:

$$\alpha_{c_i}^d(W_{LL}) = \frac{|\{(x,y) \mid W_{LL}(x,y) = c_i; W_{LL}(x \pm d, y \pm d) = c_i\}|}{8 \times d \times |\{(x,y) \mid W_{LL}(x,y) = c_i\}|} \tag{6}$$

where W_{LL} is the wavelet decomposed image of LL sub-band, c_i is the quantized color, and d is the correlation distance.

Wavelet coefficients for LH correspond to the low pass filter and the high pass filter in the horizontal and vertical directions, respectively. Correlogram calculations on the LH sub-band can logically proceed only in a horizontal direction (low pass filtering), so the horizontal correlogram of the LH coefficients is computed as follows:

$$\alpha_{c_i}^{d}(W_{LH}) \;=\; \frac{|\{(x,y) \mid W_{LH}(x,y)=c_i; W_{LH}(x,y\pm d)=c_i\}|}{2\times d\times |\{(x,y)\mid W_{LH}(x,y)=c_i\}|} \tag{7}$$

Similarly, the vertical correlogram of the HL coefficients is computed using only the vertical direction, as described in equation (8):

$$\alpha_{c_i}^{d}(W_{HL}) \;=\; \frac{|\{(x,y) \mid W_{HL}(x,y)=c_i; W_{HL}(x\pm d,y)=c_i\}|}{2\times d\times |\{(x,y)\mid W_{HL}(x,y)=c_i\}|} \tag{8}$$

Fig. 5 represents the neighboring pixels of point p with distance $d = 1$, used in the proposed approach.

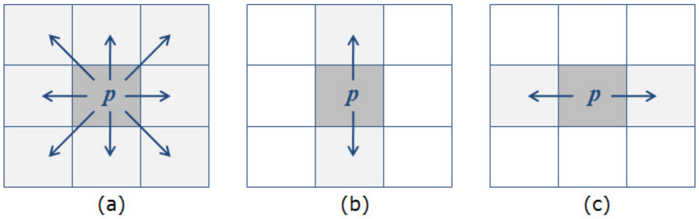


Fig. 5. Neighboring Pixels of Point p with Distance $d = 1$ (a) 8-directions for LL, (b) 2-directions for LH, and (c) 2-directions for HL.

Next, the wavelet color-spatial feature is combined with different weights for the sub-bands in the wavelet transform for each color channel, as follows:

$$f_c(C) \;=\; \lfloor \omega_{LL}\times\alpha_{LL}^{d},\; \omega_{LH}\times\alpha_{LH}^{d},\; \omega_{HL}\times\alpha_{HL}^{d} \rfloor \tag{9}$$

where the large C indicates the channel of the color image that satisfies the condition $C \in \{H, S\}$, and ω is the weighted value for LL, LH and HL sub-bands.

In this extraction process the results of the feature vectors inherit the multi-scale and multi-resolution properties from the wavelet and the translation invariant property from the correlogram.

3.2. Extraction of Texture Feature Using GLCM

The second part of the proposed descriptor consists of the texture feature extraction, which is shown in Fig. 6.

In order to extract a texture feature from the transformed domain, a Gabor wavelet filter is commonly used, as this is known to outperform tree-structured wavelet transforms, pyramid-structured wavelet transforms and multi-resolution simultaneous auto-regressive models [16] [27]. In the first step, the image is

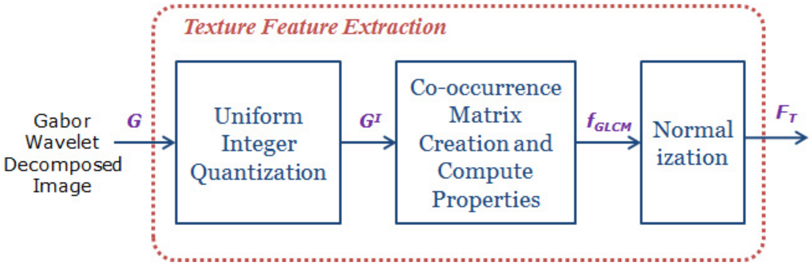

Texture Feature Extraction
Gabor Wavelet Decomposed Image
G
Uniform Integer Quantization
G^I
Co-occurrence Matrix Creation and Compute Properties
f_GLCM
Normalization
F_T

Fig. 6. Texture Feature Extraction Procedure in the Gabor Wavelet Domain.

converted to a gray-scale image I_G and a Gabor filter with two scales and four orientations is constructed. Gabor wavelet decomposition of the converted image is then performed, after which a GLCM is generated with five displacements ($0, 45, 90, 135$ and 315, as shown in Fig. 7) after performing integer quantization.

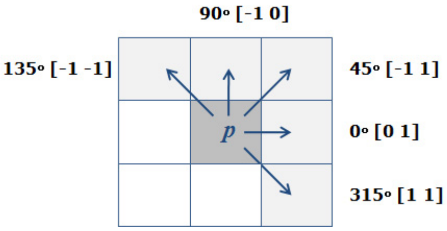


Fig. 7. Spatial Relationship used for Computing GLCM.

Five common texture features, namely contrast, correlation, energy, entropy and homogeneity are then calculated with the GLCM, as shown below:

$$\begin{aligned}
Contrast &= \Sigma\Sigma(i-j)^2 \cdot P(i,j) \\
Correlation &= \Sigma\Sigma\frac{(i-\mu_i)(j-\mu_j)\cdot P(i,j)}{\delta_i\delta_j} \\
Energy &= \Sigma\Sigma P(i,j)^2 \\
Entropy &= \Sigma\Sigma P(i,j)\cdot\log P(i,j) \\
Homogeneity &= \Sigma\Sigma\frac{P(i,j)}{1+|i-j|}
\end{aligned} \tag{10}$$

where $P(i,j)$ is the $(i,j)_{th}$ entry in the co-occurrence matrix P, and $\delta_i\delta_j$ and $\mu_i\mu_j$ are the mean and standard deviation of P, respectively.

Since this similarity measure does not consider rotation invariance, relevant texture images with different orientations may be missed by the searching process, as they would be assigned a low rank. Many papers using the Gabor texture feature either fails to consider rotation invariance or consider shifting feature elements in every direction to find the best match between the query image and images in the database [25]. However, both these approached require expensive calculations.

In this paper, we propose implementing a simple circular shift on the feature map to solve the rotation variance problem associated with Gabor texture features. The total energy for each orientation is calculated and then the orientation with the highest total energy is deemed to be the dominant orientation. The feature elements in the dominant element are then shifted to become the first element in the feature vector ft and the other elements are circularly shifted accordingly. For example, if the original feature vector is $\boldsymbol{f} = [\epsilon_1, \epsilon_2, \cdots, \epsilon_n] = [1, 3, 2, 5, 2, 3]$ and ”5” is the dominant orientation, the circularly shifted feature will be $\boldsymbol{f}_{CSF} = [5, 2, 3, 1, 3, 2]$.

3.3. Combination With Both Visual Features

The final step is to combine the two feature vectors. The color and texture features must first be normalized to reduce the effect of different feature dimensions and variances of the feature components. The normalized multiple feature F_D is computed as follows:

$$F_D \;=\; \lfloor \omega_c \times \frac{f_c}{N_c \times \delta_c \mu_c},\; \omega_t \times \frac{f_t}{N_t \times \delta_t \mu_t} \rfloor \tag{11}$$

where N_c and N_t are the dimensions of the color and texture feature vectors, $\delta_c \delta_t$ and $\mu_c \mu_t$ are the mean and standard deviations for color and texture, respectively, and ω_c and ω_t indicate the weights of color and texture, over the ranges $0 \leq \omega_c, \omega_t \leq 1$ and $\omega_c + \omega_t = 1$.

3.4. Similarity Measure

Once the features of the image have been extracted the retrieval results are obtained by measuring the similarity between the features of the query image and the pre-extracted features of the images in the database.

One of the most important parts of the matching process is the similarity function, because this decides how similar two features are. There are two methods commonly used to perform this function: the Minkowski-form metric and the Quadratic-form metric [4]. While the former compares only the corresponding bins between the histograms, the latter also considers the cross-relationships between the bins.

For the similarity measure, we can compute a distance that consists of the sum of the normalized distances for the visual feature:

$$S_D(F^Q, F^T) \;=\; \sum_{i=0}^{n-1} \frac{|F_i^Q - F_i^T|}{1 + F_i^Q + F_i^T} \tag{12}$$

where n is the number of feature dimensions, and F^Q and F^T are the query feature vector and target feature vector, respectively.

4. Experiments and Results

4.1. Datasets

To evaluate the performance of the proposed descriptor, we selected three datasets of images. Two of the datasets were the Corel photo gallery and the MPEG-7 common color dataset (CCD), both of which are widely used in the field of image retrieval. The third dataset included natural photos obtained from the website www.freeimages.co.uk. Each collection has images with a range of resolutions (e.g., $320 \times 240, 384 \times 256, 640 \times 420, 768 \times 512$ and $1,600 \times 1,200)$ formatted as JPEGs and various kinds of images, including humans, flowers, vehicles, structures, fruits, materials, and so on. We used a subset of three datasets consisting of 2,200 images belonging to 85 classes of different kinds of images, chosen to estimate the effectiveness of image retrieval. To identify the SubjectArea of the EXIF tag, each of the images was manually generated to highlight a ROI. Fig. 8 shows just sample images from the MPEG-7 CCD, representing 50 image classes.

The ground truth sets (GTS) for evaluation were provided with the class in the experiments, but this was only used to calculate the effectiveness of the proposed approach. Retrieved images were considered to be relevant if they belonged to the same class as the query image.

Fig. 8. Sample Images from MPEG-7 CCD, from 50 Image Classes.

4.2. Experimental Results

The most common evaluation measures used in information retrieval (IR) are precision and recall, usually presented as a precision-recall curve [4]. Precision denotes the ratio of retrieving an image that is relevant to the query, and recall indicates the ratio of the relevant images being retrieved, calculated as follows:

$$\begin{aligned} precision &= \frac{no.ofrelevantimageretrieved}{no.ofrelevantimagesincollection} = \frac{a}{a+b} \\ recall &= \frac{no.ofrelevantimageretrieved}{no.ofimagesretrieved} = \frac{a}{a+c} \end{aligned} \tag{13}$$

where a is the number of relevant images retrieved, b is the number of irrelevant images retrieved, and c is the number of relevant images that were not retrieved.

Since precision and recall are not always the most appropriate measures for evaluating IR, precision and recall scores are often combined into a single measure of performance, known as the *F-score* [2]. Higher values of the F-score are obtained when both precision and recall are higher. The formula for calculating the F-score is:

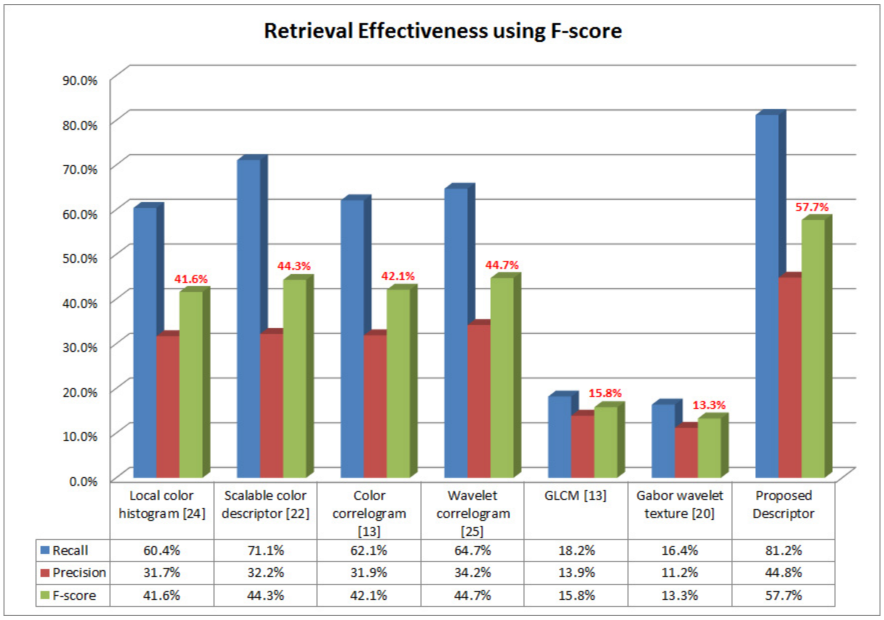
$$Fscore = \frac{2 \times precision \times recall}{precision + recall} \tag{14}$$

The following experimental approach was adopted to evaluate the search results and quantify any improvement in the retrieval performance. Leaveone-out cross validation (LOO-CV) performance was applied to obtain more reliable estimates compared to previous experiments whose results were based on a small number of queries, for example MPEG-7 [25]. Thus, each image in the database was selected in turn as the query image, and queried against the remaining images.

Fig. 9 shows the results for the comparison of retrieval effectiveness over the entire query. The source codes of C++ for other image descriptors [14], [6], [25], [20], [16], [17] are called within Objective-C++, to compare the effectiveness of the retrieval results. The values shown are computed in terms of recall and precision after the top 50 images have been retrieved, denoted as P(50) and R(50), respectively. The other methods included for comparison included a higher proportion of irrelevant images during the search, as indicated by their low precision and high recall. Based on the average from all queries for the F-score, 57.7% of all relevant images were retrieved by the new algorithm, which compares favorably to the best of the comparable methods, which retrieved just 44.7% of the relevant images in the image collection and was much better than the worst case, which retrieved only 13.3%. SCD, which is one of representative descriptor of MPEG-7, achieved 44.3% of relevant images, as shown in experimental results.

Experimentally the proposed method achieved an overall retrieval result score of 57.7%, markedly better than the 47.3% achieved by reducing query image through down sampling without checking main area of the image tag.

These results clearly indicated that the retrieval results achieved by the proposed approach achieved a higher ranking than any of the other methods tested based on its ability to cope with different scales and resolutions in the dataset of relevant images. Thus, this test shows that the proposed descriptor offers a more efficient way to conduct multi-resolution and multi-scale image retrieval.



	Local color histogram [24]	Scalable color descriptor [22]	Color correlogram [13]	Wavelet correlogram [25]	GLCM [13]	Gabor wavelet texture [20]	Proposed Descriptor
🟦 Recall	60.4%	71.1%	62.1%	64.7%	18.2%	16.4%	81.2%
🟥 Precision	31.7%	32.2%	31.9%	34.2%	13.9%	11.2%	44.8%
🟩 F-score	41.6%	44.3%	42.1%	44.7%	15.8%	13.3%	57.7%

Fig. 9. Comparisons of Retrieval Effectiveness using F-Score.

Dimensionality of the feature vector is one of the most important factors affecting not only the amount of storage space needed, but also the retrieval accuracy and computational time [15]. Although the retrieval accuracy generally tends to improve as the dimension increases the amount of stage space and the computation time also increase. Thus, it is very important to choose the optimum dimension of the feature vector that will result in acceptable retrieval accuracy without incurring an excessive amount of storage space and computation time in CBIR. Since the ways of composing feature vectors in the search methods are quite different, it is necessary to fix all their feature vector dimensions to the same value for a fair comparison. Here, we chose the vector dimensions of image color features to be around 100, close to the dimension of the extracted vector for the proposed descriptor.

We implemented all approaches using Objective-C and C++ (source code for other methods) with Xcode 4.4.1 on a MacBook Pro running OS X 10.7 (Lion). Table 1 shows the computational characteristics of each method in order to compare the dimension of the feature vector and computational time required for each. Each of the computational times was calculated by the averages through batch processing. The retrieval time for exhaustive search is the sum of two times: T_{sim} and T_{sort} [8]. T_{sim} is the time to calculate the similarity between the query and every image in the database, and T_{sort} is the time to rank all the images in the data according to their similarity to the query. However, the retrieval time is highly depends only on the measuring the similarity, especially on the time of extracting features. At this point, the proposed algorithm requires more computing time than others, enhancement of time is necessary for computing power and this is remained in future work.

Table 1. Computational time and dimensions of feature vectors for descriptors. Column (A) is the average time for extraction of visual features [sec/image], and (B) is the number of dimensions used for the feature vectors.

Method	(A)Extraction time	(B)Feature vector
Local color histogram	0.081	96
Scalable color descriptor	0.094	64
Color correlogram	0.127	96
Wavelet correlogram	0.131	96
GLCM	0.067	16
Gabor wavelet texture	0.098	48
Proposed Method	0.176	72

Fig. 10 depicts the user interfaces for our prototype system, which were used solely in extraction and retrieval mode for this study. Thus, users of the new system would only use the retrieval interface presented in Fig. 10(a) and 10(b). Fig. 10(c) shows how users can scroll through the retrieved results. However, the system does not yet support the re-query procedure, which remains for future work.

5. Conclusions

This paper proposes a new, more efficient mobile image descriptor that utilizes a combination of color features based on color-spatial information and texture features that make use of the Gabor texture of an image. In the preprocessing stage, the query image is resized, either by extracting the main area of the image or by down sampling to avoid a memory leak, taking into account the EXIF metadata. When using a correlogram for the color features, more computational time is required than for a histogram-based approach. For this reason, we incorporated a wavelet transform, whose coefficients provide information that is independent of the original image resolution, and appropriately weight the LL, LH and HL sub-bands. Also, the use of a color codebook helps to reduce the computational time needed.

The results of extensive experimental trials revealed that the proposed method produced a significant improvement of around 13% in retrieval effectiveness compared to the best of the other descriptors tested. However, the memory efficiency still needs further improvement, since limited memory resources remain a critical problem for mobile devices.

The main contribution of this paper lies in its weighted combination of color and texture for the use of mobile image retrieval based on spatial-color and

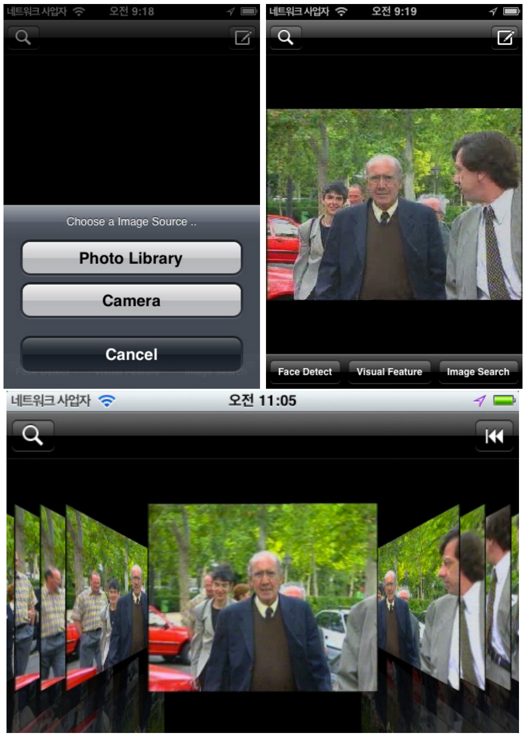


Fig. 10. Screenshots of the Prototype System: (a) User Interface for Selecting Query Image, (b) Query Image Chosen for Mobile search, and (c) Retrieved Images shown as Cover-flow on a iPhone Simulator, which is the Target Device used in Our Work.

second order statistics. As for future work, there are two main avenues for further development to enable the system to operate on smart phones such as the iPhone and Android. The first is the addition of an automatic procedure to identify the main area of an image, which had to be performed manually for these experiments. For example, automatic face detection and recognition would be particularly helpful. The second is the addition of textual or semantically related information such as geo-location and user events to the existing algorithm to enable users to search for photographs associated with specific features.

Acknowledgments. This research is supported by Ministry of Culture, Sports and Tourism (MCST) and Korea Creative Content Agency (KOCCA) in the Culture Technology (CT) Research and Development Program.

References

1. Ahmad, I., Gabbouj, M.: A generic content-based image retrieval framework for mobile devices. Multimedia Tools Applications (2010)

2. Baeza-Yates, R., Ribeiro-Neto, B.: Modern Information Retrieval: The Concepts and Technology behind Search. 2nd Editions. ACM Press Books, USA (2011)
3. Camera, Association, I.P.: Exchangeable image file format for digital still cameras: Exif version 2.3 (2010), [Online]. Available http://www.cipa.jp/english/hyoujunka/kikaku/pdf/DC-008-2010E.pdf (current March 2013)
4. Castelli, V., Bergman, L.D.: Image Databases - Search and Retrieval of Digital Imagery. Wiley Inter-Science (2002)
5. Group, M.W.: Guidelines for handling image metadata version 2.0. Motorola Labs, Paris (2010), [Online]. Available http://www.metadataworkinggroup.org/pdf/mwg_guidance.pdf (current March 2013)
6. Howarth, P., Ruger, S.: Evaluation of texture features for content-based image retrieval. In: Lecture Notes of Computer Science. pp. 326–334 (2004)
7. Jiang, X., Sun, T., Fu, G.: Multi-scale image semantic recognition with hierarchical visual vocabulary. Computer Science and Information Systems (2011)
8. Krishnamachari, S., Abdel-Mottaleg, M.: Hierarchical clustering algorithm for fast image retrieval. In: Part of the IS&T/SPIE Conference on Storage and Retrieval for Image and Video Databases VII. pp. 427–435 (1999)
9. Lee, Y.H., Kim, B., Rhee, S.B.: Content-based image retrieval using wavelet spatial-color and gabor normalized texture in multi-resolution database. In: International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing. pp. 1–1. IMIS, Australia (2012)
10. Leong, M.K., Chang, W.: Framework and system components. ISO/IEC JTC1/SC29/WG1N3684 pp. 1–1 (2006)
11. Leong, M.K., Chang, W.: Iso/iec pdtr 24800-1: Jpsearch - part 1: Framework and system components. ISO/IEC JTC1/SC29/WG1N4203 pp. 1–1 (2007)
12. M.A., A., Maldague, Z., W.B., L.: A new color-texture approach for industrial products inspection. Journal of Multimedia pp. 44–50 (2006)
13. Makris, C.: Wavelet tress: a survey. Computer Science and Information Systems 9(2) (2012)
14. Malinga, B., Raicu, D., Frust, J.: Local vs. global histogram-based color image clustering. Technical Report TR06-010, School of Computer Science, De Paul University (2006)
15. Man, W.K.: Content-based Image Retrieval using MPEG-7 Dominant Color Descriptor. Master Thesis, Dept. of Electronic Engineering, City University of HongKong (2004)
16. Manjunath, B., Ma, W.: Texture features for browsing and retrieval of large image data. IEEE Transactions on Pattern Analysis and Machine Intelligence (Special Issue on Digital Libraries) 18(8), 837–842 (1996)
17. Moghaddam, H.A., Khajoie, T.T., Rouhi, A.: A new algorithm for image indexing and retrieval using wavelet correlogram. In: Proceedings of International Conference on Image Processing. pp. 497–500 (2003)
18. Nakagawa, A., Kutics, A., Phyu, K.H.: Mosir: Mobile-based segment and image retrieval. In: International Conference on Wireless Communications, Networking and Mobile Computing. pp. 1–4. WiCOM (2011)
19. Ogiela, M.R., Ogiela, L.: Personal identification based on cognitive analysis of selected medical visualization. Journal of Internet Services and Information Security 2(3/4), 148–153 (2012)

20. Ojala, T., Rautiainen, M., Matinmikko, E., Aittola, M.: Sementic image retrieval with hsv correlograms. In: Proceeding 12th Scandinavian Conference in Image Analysis. pp. 621–627 (2001)
21. Skulsujirapa, P., Aramvith, S., Siddhichai, S.: Development of digital image retrieval technique using autocorrelogram and wavelet based texture. In: IEEE International Midwest Symposium on Circuits and Systems. pp. 273–276 (2004)
22. Smeulders, A.W., Worring, M., Santini, S., Gupta, A., Jain, R.: Content-based image retrieval at the end of the early years. IEEE Transactions on Pattern Analysis and Machine Intelligence 22(12), 1349–1380 (2000)
23. Torres, R.S., Falcao, A.Z.: Content-based image retrieval: Theory and applications. In: Brazilian Symposium on Computer Graphics and Image Processing. pp. 161–185. SIBGRAPO (2006)
24. Xing Xie, Lie Lu, M.J.H.L.G.S.W.Y.M.: Mobile search with multimodal queries. In: Proceedings of the IEEE. pp. 589–601. WiCOM (2008)
25. Yamada, A., Ocallaghan, R., Kim, S.K.: Mpeg-7 visual part of experimentation model (xm) version 27.0. ISO/IEC JTC1/SC29/WG11/N7808 (2006)
26. Yeh, T., Tollmar, K., Grauman, K., Darrell, T.: A picture is worth a thousand keywords: Image-based object search on a mobile platform. In: Proceeding Conference Hyman Factors in Computing System (2005)
27. Zhang, D., Wong, A., Indrawan, M., Lu, G.: Content-based image retrieval using gabor texture features. IEEE Transactions PAMI pp. 13–15 (2000)

Yong-Hwan Lee is received his Ph.D. degree in Electronics and Computer Engineering and M.S.degree in Computer Science from Dankook University, Korea, in 2007 and 1995, respectively. He is a research professor in Department of applied computer engineering at Dankook University. His research interests are the area of Image/Video Representation and Retrieval, Face Recognition, Augmented Reality, Mobile Programming and Multimedia Communication.

Bonam Kim is received the Ph.D. degree in Computer Science and Software Engineering from the Auburn University, Alabama, USA in 2006. She is a research professor in division of electrical and computer engineering at Chungnam National University since 2010. She joined the School of Electrical and Computer Engineering, CNU in March 2007. Her current research interests are in the areas of wireless ad hoc and sensor networks, network security and MIPv6.

Sang-Burm Rhee is received the Ph.D. degrees in Electronics Engineering from Yonsei Univ. in 1986. Now he is a professor at Dankook Univ. since 1979. His research interests are the area of Microprocessor, SoC(System-On-Chip), Pattern Recognition, Multimedia Processing. They include topics such as Object-oriented Methods for Audio/Video Watermarking, Pattern Recognition and HDL for SoC..

Received: July 16, 2012; Accepted: January 18, 2013.