Image Description and Matching Scheme Using Synthetic Features for Recommendation Service

Won-Keun Yang, Ayoung Cho, Weon-Geun Oh, and Dong-Seok Jeong

This paper presents an image description and matching scheme using synthetic features for a recommendation service. The recommendation service is an example of smart search because it offers something before a user's request. In the proposed extraction scheme, an image is described by synthesized spatial and statistical features. The spatial feature is designed to increase the discriminability by reflecting delicate variations. The statistical feature is designed to increase the robustness by absorbing small variations. For extracting spatial features, we partition the image into concentric circles and extract four characteristics using a spatial relation. To extract statistical features, we adapt three transforms into the image and compose a 3D histogram as the final statistical feature. The matching schemes are designed hierarchically using the proposed spatial and statistical features. The result shows that each feature is better than the compared algorithms that use spatial or statistical features. Additionally, if we adapt the proposed whole extraction and matching scheme, the overall performance will become 98.44% in terms of the correct search ratio.

Keywords: Image description, image matching, recommendation service, spatial feature, statistical feature.

Manuscript Feb. 19, 2010; revised Nov. 8, 2010, accepted Dec. 6, 2010.

This work was supported by the IT R&D program of MKE/MCST/IITA, Rep. of Korea [2011-KI001797, Development of the Rich UCC Technology].

Won-Keun Yang (phone: +82 32 860 7415, email: aida@inha.edu) and Dong-Seok Jeong (email: dsjeong@inha.ac.kr) are with the School of Electronic Engineering, Inha University, Incheon, Rep. of Korea.

Ayoung Cho (email: ayoung.cho@lge.com) is with the Mechatronics & Storage R&D Lab., LG Electronics Inc., Seoul, Rep. of Korea.

Weon-Geun Oh (email: owg@etri.re.kr) is with the Contents Research Laboratory, ETRI, Daejeon, Rep. of Korea.

doi:10.4218/etrij.11.1510.0023

I. Introduction

A search is performed to find something that a user wants. The size of the search database has increased considerably as people can now produce content easily. Search service users obtain a variety of content from the large database. Further, whenever they want or wherever they are, they can obtain the content rapidly because of developed communication infrastructures, such as optic LAN, Wi-Fi, and WiBro. This environment makes people demand more developed search services, and this, in turn, leads to the evolution of the search service.

The search service has to satisfy some requirements. One such requirement is efficiency. An efficient search is a fast and accurate search. Basically, a search is performed in a collected database; therefore, a database should keep accumulating more information in order to offer accurate information to the user. However, the expansion of the database decreases the search speed. Some tools, such as a database management system, can help to improve the search speed.

Another requirement is that the search be smart. Smart means the search service provides useful information to the user before any such request is actually made. This can help to widen the search range and perform a more convenient search. A representative example of a wise search is a recommendation service. A recommendation service finds some content that is related to the search results in terms of keywords, meaning, or the content's point of view and offers the additional related content to the user. In general, the user finds the content desired from the search results, but also, discovers additional interesting content in the recommended content. Therefore, many search service providers adopt a recommendation service to help users.

In this paper, we present a recommendation service model. A link about the content owner is offered instead of a content recommendation in this service model. By following this link, a user can use a variety of content that the recommended content owner has. The supposition of this model originates from the fact that several people who have related content are interested in the same field. To offer the presented recommendation service, the service provider first has to find the related content in the database. We assume that the content is images and that the related content is an identical image.

The easiest method to find related images is a keyword-based method. Several text keywords are tagged to the images when the images are registered in the database. Also, some combinations of the word are used for finding suitable images. This method ensures a fast search, but it sometimes leads to certain problems. One such problem caused by a keyword is an unintended search result. Because the keywords are usually subjectively generated by an image creator, it is difficult to perform an accurate search using the keyword-based search method. Therefore, a new search technology such as a content-based search has been suggested to complement the insufficiency of the keyword-based search [1]-[3].

Because the content-based method extracts some features from the image itself and finds related images on the basis of extracted features, it is not affected by the image keywords. However, the improvement in search accuracy depends on the kind of feature used for the image description. There are two types of features used in the image description method: local features and global features. A local feature extracts interest points or regions in the image and describes each point or region [4]-[6]. The advantage of using this feature is that it can find a matched image even if only a part of the image is used for searching. However, the time consumption of this method in a large database is a serious problem. Alternatively, a fast search is possible through the use of the global feature because this feature is extracted in the same form from every image [7], [8]. The global feature is also divided into two types, spatial features and statistical features, according to how it uses image information to create a descriptor.

An image is described by the relation between overlapped and non-overlapped subregions in the spatial feature. Kim developed the spatial feature using a discrete cosine transform and an ordinary measure for copy detection [9]. The image is partitioned into 8×8 blocks, and the average intensity of each block is extracted. Then, after performing the discrete cosine transform to the partitioned image, it applies the ordinal measure to the AC coefficients. Wnukowicz proposed the still image copy detection algorithm using a trajectory of subblocks [10]. This method cuts a 256×256 area from a size-normalized image, and it is divided into 31×31 overlapping blocks. Then, the features for mean luminance, mean energy, and singular energy of each block are calculated. Each block is grouped according to the distance from the center block. The mean and standard deviation of three block features in each group are calculated. Finally, the image is described by the variation of each group feature following the trajectory.

The spatial feature is robust to illumination change or noise effects; however, it is not robust in terms of a geometric transformation like a viewpoint change. An image is also described through a statistical analysis of the information of each pixel using a statistical feature approach. A histogram is a representative method of statistical features [3], [11]. Swain and Ballard adapted the color histogram for a color indexing algorithm that is used in a robot [12]. The color structure method of MPEG-7 uses a histogram that is made on the basis of the color value and the arrangement of colors for image retrieval [13]. Histograms are invariant to translation and rotation about the viewing axis and are slightly and slowly changed only when there is a change in the view angle, scale change, and occlusion [12].

In this paper, we introduce the image description scheme that is synthesized by using the spatial and statistical features for the recommendation service. We use a 3D histogram composed of three pixel characteristics as the statistical feature and a descriptor based on concentric circle partitioning as the spatial feature. A matching scheme is also proposed because the entire descriptor is composed of two different global features.

The rest of this paper is composed as follows. We explain the image recommendation service model in section II, and the extraction scheme for the spatial and statistical features is described in section III. Further, we explain the matching and selection scheme using the syntheses of two features in section IV. Section V presents some experimental results and discussion, and we conclude the paper in section VI.

II. Image Recommendation Service Model

The main goal of a search is to quickly and accurately find information that the user wants. The user may enter a combination of words as a query to find information; however, this is not sufficient for an accurate search. Therefore, the search service provider cannot offer sufficient information to the user. To overcome this limitation, some other techniques are required. Many service providers adopt a recommendation service. The recommendation service offers some useful or interesting content or information to the user before any request is made. So, it is one of application examples for smart search.

In this paper, we introduce a recommendation service model. In general, the recommended content or links include a variety of information (daily issued, most hit, or some commercial).

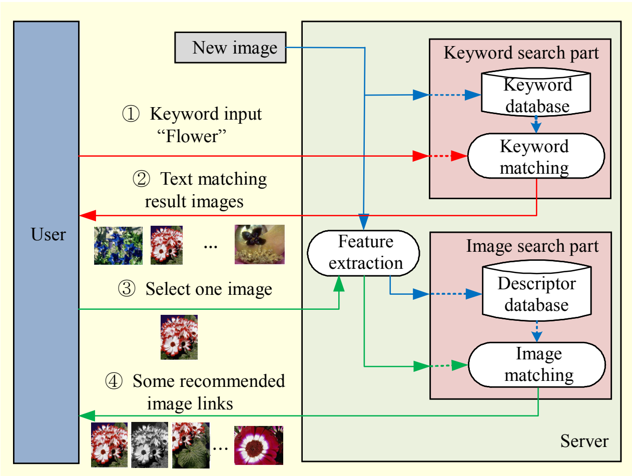

Fig. 1. Flowchart of proposed image recommendation service model.

However, our recommendation model offers a link to information about content owners instead of the content itself because this helps to not only offer useful content but also form a social network. Figure 1 shows the diagram of the proposed image recommendation service model.

At first, people who have new images register the images with the database to post them. If interested in the posted images, first, the user searches the images using the search service by using keywords. Then, the search engine finds the images that have the same keyword, arranges them, and shows the results to the user. If the image that the user wants does not exist in the search result, the user tries to search for it using another keyword. Otherwise, the user selects one image. Then, image matching is carried out using the selected image as a query. Basically, image matching is performed using a descriptor. The descriptor is a small piece of binary data that represents the image. When image matching begins, a query image descriptor is extracted, and it is compared with the pre-extracted and stored descriptors in the database.

Because many types of images are registered in the database, it may include identical images (the same images, transformed images, or similar images). The search engine finds these identical images and provides the link to information about the person who owns these images. Even if the search fails, the user can find some interesting images by using the recommended link.

Let us consider an example. Person A registers all the Disney animation character images that the person has with a database. It includes Mickey Mouse, Goofy, and Donald Duck. Another person, B, wants to find a particular Goofy image. However, the user does not know the character's name is Goofy, having only seen this character with Mickey Mouse. In this case, the user searches for Mickey Mouse images first using the keyword-based search. If person B selects one image, the selected image is used as a query, and image matching is performed. The link to person A who has Goofy images is offered by the recommendation service, and finally, the user, person B, can obtain the desired Goofy image by following the offered link.

III. Feature Extraction Scheme

We extract two types of features (spatial and statistical features) from the input image. In general, the image is divided into subregions to extract the spatial feature. Some algorithms compose subregions as $N\times N$ square blocks [14]. Our previous research also divided the image into the $N\times N$ square blocks to extract spatial features [15]. However, the feature based on square block partitioning is weak with respect to the rotation transform. Alternatively, if we divide an image into concentric circles, we can obtain a rotation invariant feature. The spatial feature that is used in the proposed extraction scheme is related to our previous research [16]. In our previous research, the image is described on the basis of concentric circle partitioning.

The image is normalized into a 256×256 block after gray conversion in the preprocessing process. This size normalization is performed to overcome the scaling problem. To extract the spatial feature, the image is partitioned into concentric circles, and four types of characteristics are extracted. Finally, the extracted characteristics go through hashing, and the spatial feature is completed.

Meanwhile, the proposed statistical feature is composed by using a 3D histogram. To compose the histogram, various characteristics are extracted from each pixel. The extracted characteristic values are quantized so that they can be used as an axis of a 3D histogram. In general, texture [17], [18], shape [19], and color [12], [17], [20]-[22] are the well-known characteristics of the statistical feature. The proposed statistical

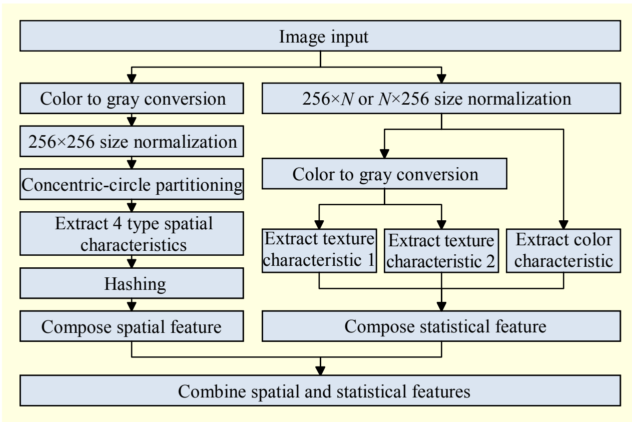

Fig. 2. Block diagram of proposed feature extraction scheme.

feature includes two texture characteristics and one color characteristic. The extraction of the statistical feature also begins from image resizing. The resizing is performed by maintaining a certain width-height ratio and letting the shorter length be 256 pixels. Two texture characteristics are extracted from the gray-converted image, and the color characteristic is directly extracted from the resized image. We compose a 3D histogram using these three characteristics, and it is used as the statistical feature. Figure 2 is the complete flowchart of the proposed feature extraction scheme.

1. Spatial Feature Extraction

Firstly, the coordinate system conversion from Cartesian coordinates to polar coordinates is executed on the preprocessed image. This coordinate system conversion is performed for convenience of average intensity calculation. Each pixel (x, y) is set at a distance r from the image center (x_0, y_0) and at the angle θ according to the coordinate system conversion. Further, we compose a polar coordinate map using the calculated values for r and θ. Figure 3 shows the original image and its polar coordinate map.

As shown in Fig. 3, the fan-shaped area of the original image is represented by a square block area in the polar coordinate map after performing the coordinate system conversion. The composed polar coordinate map is partitioned into $M\times N$ blocks. That is, the original image is partitioned into fan-shaped sectors with M-radius and N-angle levels. Then, we extract four types of spatial characteristics from the partitioned map. We use $M=32$, 16 and $N=36$.

The first characteristic is the average intensity level distribution of each ring region. Its mathematical representation is written as

$$\mathbf{C}=\{C_0,C_1,\cdots,C_{M-1}\},\quad C_i=\frac{1}{N}\sum_{j=0}^{N-1}S_{(i,j)},\tag{1}$$

where i is the index of the i-th ring subregion and j is the index of the j-th angle subregion. The first characteristic vector $\mathbf{C}$ is an array of the average intensity value C_i in each ring-shaped subregion, and C_i is calculated by the mean operation of the i-th ring region S_i.

The second characteristic is the difference distribution of the first characteristic. This step is as follows:

$$\mathbf{V}=\{V_0,V_1,\cdots,V_{M-2}\},\quad V_i=\left|C_{i+1}-C_i\right|.\tag{2}$$

The second characteristic vector $\mathbf{V}$ is an array of the absolute difference of the first characteristic between adjacent ring regions.

The third characteristic is the symmetrical difference distribution. This characteristic is calculated as

$$\mathbf{T}=\{T_0,T_1,\cdots,T_{M-1}\},\quad T_i=\frac{1}{N/2}\sum_{j=0}^{N/2-1}\left|S_{(i,j)}-S_{(i,j+N/2)}\right|.\tag{3}$$

The third characteristic vector $\mathbf{T}$ is an array of the symmetrical difference T_i between two symmetrically positioned subregions. The absolute difference between the subregions is calculated as T_i. Figure 4(a) depicts an example pair of the third characteristic. Then, we average the difference value of each ring. The third characteristic T_i is extracted using this average value.

The last characteristic is the circular difference distribution. Mathematically, it is written as

$$\mathbf{E}=\{E_0,E_1,\cdots,E_{M-1}\},\quad E_i=\frac{1}{N}\sum_{j=0}^{N-1}\left|S_{(i,j)}-S_{(i,j+1 \bmod N)}\right|.\tag{4}$$

This characteristic vector $\mathbf{E}$ is an array of the difference value between two adjacent subregions. The last characteristic also extracts one average value per ring region like the third characteristic, but it computes the absolute difference between each subregion and its neighborhood subregion in a counterclockwise direction instead of the symmetrical position. Figure 4(b) shows the subregion pair that is used in the circular difference calculation.

The extracted spatial feature is hashed for obtaining a small data size and fast matching. The hashing is a process to convert a scalar value into a binary value. Our hashing is very simple.

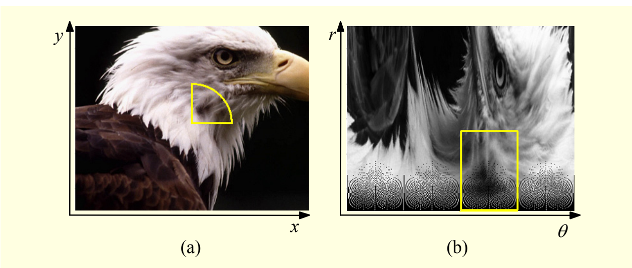

Fig. 3. Representation of image in Cartesian coordinate system and polar coordinate system: (a) original image and (b) its polar coordinate map. Fan-shaped region of original image is represented as square block in polar coordinate map.

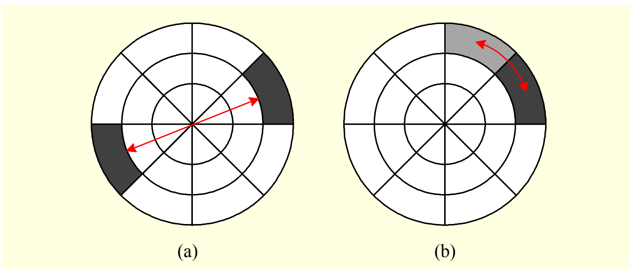

Fig. 4. Subregion pairs to extract third and fourth characteristics: (a) symmetrically positioned subregion pair and (b) neighborhood-positioned subregion pair.

Each extracted characteristic value is arranged from the inner ring region to the outer ring region. The hashing is processed by comparing the characteristic values of the neighboring region. If the characteristic value of the outer region is greater than that of the current region, we assign the hash value to 1. In the opposite case, we assign the hash value to 0.

2. Statistical Feature Extraction

The first texture characteristic is the degree of symmetry obtained by the modified generalized symmetry transform (MGST) in some area. The original symmetry transform is developed as an attention operator by Reisfeld [23]. Further, Park designed a noise-tolerant phase weight function for the backlight unit inspection of LCD [24]. We apply Park's phase weight function to the original symmetry transform to compute how symmetric the region is. The symmetry value of one pixel is calculated by using the sum of the symmetry contribution value between the perimeter pixel pairs in MGST. The transform is started by

$$\gamma_k = \log_2\left(1+\|\nabla q_k\|\right),\quad \phi_k = \arctan\left(\frac{\partial}{\partial x}q_k \Big/ \frac{\partial}{\partial y}q_k\right),$$
$$\nabla q_k = \left(\frac{\partial}{\partial x}q_k, \frac{\partial}{\partial y}q_k\right). \tag{5}$$

We assume the pixel pair to be q_i, q_j. Then, we can calculate the gradient magnitudes γ_i, γ_j and gradient directions $\square_i$, $\square_j$ of the pair according to (5). Further, the phase weight function P_{ij} is defined using gradient directions, and the angle value α_{ij} between a straight line that is composed by two pixels q_i, q_j and the horizontal line is measured.

$$P_{ij} = \sin\left(\frac{\phi_j+\phi_i}{2} - \alpha_{ij}\right)\times\sin\left(\frac{\phi_j-\phi_i}{2}\right). \tag{6}$$

The first term of the phase weight function indicates that the weight is at maximum value when the gradients at q_i and q_j are oriented in the same direction towards each other. The second term means that the weight is at the maximum value when each gradient direction at q_i and q_j is oriented 90° from the straight line between q_i and q_j. The transform defines another weight function. It is called the distance weight function and depends on the distance between two pixels. The distance weight function is defined as

$$D_{ij} = \frac{1}{\sqrt{2\pi\sigma}}\exp\left(-\frac{\|i-j\|}{2\sigma}\right), \tag{7}$$

where σ is the standard deviation of the Gaussian function decided by the area of the region where the symmetry transform is applied. The symmetry contribution value of the pixel pair O is measured by multiplying the distance weight function D, phase weight function P, and gradient magnitudes at two pixels as

$$O_{ij} = D_{ij}P_{ij}\gamma_i\gamma_j. \tag{8}$$

Finally, the amount of symmetry for one pixel S with area R is the sum of the symmetry contribution values that are measured by using the perimeter pixel pairs as

$$S = \sum_{i,j\in R} O_{ij}. \tag{9}$$

The calculated symmetry value is quantized to five levels and used as an axis of statistical features. The quantization function is designed according to the probability density function of the symmetry value derived from the database of 135,609 different images. The composed probability density function follows a Laplace distribution. Therefore, we design a quantization function where each level has the same portion.

The second texture characteristic depends on the average intensity of some regions. For the rotation invariant characteristic, we set up a circular region centered at one pixel. Next, we measure the variation of the set region. The set region is divided into two symmetrical half-circle subregions. Then, we calculate the average intensity of each subregion individually and the absolute difference between the two average intensity values of the subregions. The baseline to divide a subregion is rotated by 18°, and the absolute difference is calculated recursively. After selecting the baseline angles that have the maximum and the minimum absolute difference values, we inspect whether the absolute difference of two values is higher than some threshold or not. If the value is lower than the threshold, we regard this region as a flat area and extract the second characteristic using the average intensity of the area. Otherwise, we use the maximum absolute difference value of the area as the second characteristic. The radius of the circular region is set as 12 pixels, and the threshold is set as 12 in our approach. The second characteristic has 10 quantization levels after going through the quantization procedure.

A color characteristic is extracted by an adaptation of the color space conversion. We adapt the hue color channel in the HSI color space because the hue channel is invariant to the light intensity change and shift [25]. An average RGB value of some area is calculated. Then color quantization to extract the color characteristic is performed. Color quantization refers to the process in which some colors are combined and represented by some levels.

The color quantization method that is adapted in the extraction of the color characteristic is as follows. We calculate the average RGB value of the 3×3 area. The calculated RGB value has a resolution of 0.5. Each color value is mapped to a new RGB value according to the multiplication by 2 and modulo 256 operations. Then, we perform the RGB-to-HSI conversion using the new RGB value and calculate the hue value. The calculated hue value is quantized to 11 levels. Ten levels are assigned for the hue value, and one level is assigned for the case that the hue value does not exist, in other words, in case of a gray pixel.

IV. Feature Matching and Selection Scheme

The image matching scheme exploits the two types of features extracted above. Image matching is the process that measures the similarity between two images using the descriptors extracted from the images. The proposed image descriptor takes the combined form of the spatial feature and the statistical feature. As the two features are independent from each other, we calculate the similarity of two images individually.

Our spatial feature is composed of a bit string of 192 bits. Therefore, the distance between two images is calculated in terms of the Hamming distance. The Hamming distance is calculated using the exclusive OR operation.

We use the sum of the absolute difference for measuring the statistical feature distance between two images because the statistical feature is represented as a vector. The proposed statistical feature consists of a 550-bin histogram and is divided into five subgroups according to the symmetry value level. Therefore, we calculate each subgroup distance individually and calculate the entire distance by averaging the subgroup distances. The statistical feature distance is calculated as

$$Dist_{\text{statistical}} = \frac{\sum_{i=0}^{4} nd_i}{5},$$
$$nd_i = \frac{\sum_{n=0}^{9}\sum_{m=0}^{10} \left| R_{\text{statistical}}(m,n) - Q_{\text{statistical}}(m,n) \right|}{\sum_{n=0}^{9}\sum_{m=0}^{10} \left(R_{\text{statistical}}(m,n) + Q_{\text{statistical}}(m,n) \right)}, \quad (10)$$

where R and Q is the statistical feature of the reference image and the query image.

We have two different selection schemes. According to one scheme, we should select from the database only one image that is most closely related to the query. In this case, we select one image that has the shortest distance from the query using the spatial feature. If the spatial feature distance between the selected image and the query image is smaller than the predefined threshold, the choice is final. However, if the spatial distance is greater than the threshold, we choose an image that has the shortest statistical distance from the query.

According to the other scheme, we should select several related images. In this case, if we want to select L images, we need to first select images with a spatial distance smaller than the given threshold. If the selected image count is smaller than L, we select more images that have a shorter distance by using the statistical feature to find L images. However, if the image count is greater than L, we choose L images that have a small spatial distance in the ascending order.

V. Experimental Results and Analysis

The performance of the proposed image description and the matching scheme is evaluated by two types of experiments. The first one is a robustness test under various transformations. The robustness is measured using the correct search ratio (CSR). We calculate CSR as follows. First, we prepare a number of original images A and their transformed images. Then, we select one image as a query from the original images A and find the most-related image from the transformed images using the image matching and selection scheme. If the chosen image is the transformed version of the query image, we increase the search count K. After repeating the same process for all original images A, we obtain the final CSR as

$$CSR = \frac{K}{A}. \quad (11)$$

We obtained 10,000 original images from the Art Explosion 800,000 CD. We tested the robustness under a non-geometrical transformation, such as brightness change, added Gaussian noise, blur, image enhancement via autolevel, color reduction, monochrome, and histogram equalization, as well as geometrical transformation, such as flip, crop, translation, rotation, aspect ratio change, skew, perspective, and a combination of crop and scaling. We compared our algorithm with Kim's algorithm, the trajectory-based algorithm (well-known spatial features), and the color structure descriptor of MPEG-7 (a well-known statistical feature). Table 1 shows the CSR of three algorithms and the proposed image description and matching scheme for various transformations.

The experimental result shows that Kim's algorithm, the trajectory-based algorithm, and the proposed spatial feature are very robust against all non-geometrical transformations. Further, Kim's algorithm shows good performance with respect to the aspect ratio change, and the trajectory-based algorithm shows good performance in the cases of histogram equalization and simple rotation. The proposed algorithm shows good performance in the cases of histogram equalization, simple rotation, and aspect ratio change. The next is a comparison between the proposed statistical feature and the color structure descriptor. Two algorithms are robust against almost all geometrical transformations. However, the color structure descriptor is especially weak in the cases of noise,

Table 1. Average CSR (%) of proposed algorithm and some spatial and statistical features: (a) proposed synthetic feature, (b) proposed spatial feature, (c) Kim's algorithm, (d) trajectory-based algorithm, (e) proposed statistical feature, and (f) MPEG-7 color structure descriptor.

Transformation	(a)	(b)	(c)	(d)	(e)	(f)
Gaussian noise (std. dev.: 4.0)	99.94	99.95	99.99	99.99	99.90	83.48
Gaussian noise (std. dev.: 8.0)	99.96	99.97	99.98	99.98	98.99	46.04
Gaussian noise (std. dev.: 16.0)	99.93	99.94	99.95	99.97	97.36	29.91
Blur 3×3	99.99	99.99	99.98	99.98	99.97	99.93
Blur 5×5	99.98	99.98	99.99	99.97	99.97	99.60
Blur 7×7	99.99	99.99	99.98	99.97	99.97	99.21
Bright +10%	99.91	99.95	99.95	100.00	94.25	99.73
Bright +20%	99.42	99.89	99.77	99.98	67.24	95.92
Bright +25%	98.87	99.83	99.65	99.98	53.94	91.05
Color reduction (24 bits → 8 bits)	99.75	99.90	99.95	99.94	75.16	0.72
Color reduction (24 bits → 16 bits)	99.98	99.96	99.97	99.99	99.84	99.89
JPEG compress (quality 80)	100.00	100.00	100.00	100.00	100.00	100.00
JPEG compress (quality 60)	99.97	99.98	99.97	100.00	99.94	99.99
JPEG compress (quality 30)	99.98	99.99	99.95	99.99	99.05	99.82
Monochrome	99.98	99.99	99.94	100.00	0.16	0.00
Autolevel	100.00	100.00	99.93	100.00	76.18	76.57
Histogram equalization	89.04	98.68	89.49	99.74	2.07	0.72
Scale –10%	100.00	100.00	99.96	100.00	100.0	100.00
Scale –30%	99.99	99.99	99.97	99.98	100.00	99.98
Scale –50%	99.96	99.96	99.93	99.97	100.00	100.00
Flip	99.94	99.94	100.00	99.97	100.00	100.00
Simple rotation 90°	99.96	99.95	1.13	99.97	100.00	100.00
Simple rotation 180°	99.93	99.92	99.98	99.97	100.00	100.00
Simple rotation 270°	99.94	99.94	1.13	99.98	100.00	100.00
Random crop 90% (remained)	99.60	2.41	32.41	22.79	99.80	99.95
Random crop 77% (remained)	99.07	0.23	0.74	0.79	99.30	99.62
Random crop 60% (remained)	84.39	0.12	0.09	0.12	84.53	94.77
Perspective 4°	99.71	3.09	57.84	8.64	99.92	99.98
Perspective 6°	99.59	1.63	13.98	2.84	99.83	99.96
Perspective 10°	99.24	0.65	1.52	1.41	99.50	99.83
Aspect ratio change (4:3 → 16:9)	100.00	100.00	99.96	24.18	99.93	99.54
Aspect ratio change (4:3 → 6:3)	99.99	99.99	99.98	11.75	99.88	99.51
Skew 4°	99.95	99.11	99.16	99.34	100.00	100.00
Skew 6°	99.96	92.48	94.80	86.60	100.00	100.00
Skew 10°	99.91	65.25	67.93	48.48	99.99	100.00
Rotation 10°	99.97	98.99	13.50	1.05	99.99	99.99
Rotation 25°	99.82	48.86	0.29	0.44	99.93	100.00
Rotation 45°	99.75	10.18	0.12	0.26	99.89	99.97
Translation 10%	99.36	0.27	0.27	0.51	99.63	99.89
Translation 20%	96.41	0.12	0.13	0.25	96.56	97.84
Translation 30%	69.17	0.06	0.06	0.16	69.37	87.05
Combination (crop 95%, scaling 80%)	99.82	19.96	91.43	85.81	99.96	99.96
Combination (crop 90%, scaling 65%)	99.67	2.37	33.35	22.84	99.80	99.77
Combination (crop 84%, scaling 50%)	99.37	0.76	5.63	3.54	99.55	99.84
Average	98.44	69.19	65.99	64.12	91.17	88.64

Table 2. Average number of correct matches for proposed algorithm and other features under different upper limitation.

Limitation / Feature	10	20	30	40	50	60	70	80	90
Proposed scheme (10 ppm)	9.9996	19.9987	29.9761	39.9482	44.7429	44.8222	44.8652	44.8919	44.9131
Proposed spatial feature	9.9997	19.9986	29.9191	32.3047	32.7096	32.9546	33.1300	33.2871	33.4141
Kim	9.9994	19.9980	29.4565	31.5753	32.1525	32.518	32.7821	32.9904	33.1689
Trajectory-based	9.9995	19.9987	29.0267	30.2422	30.6011	30.8429	31.0254	31.1778	31.3025
Proposed statistical feature	9.9986	19.9984	29.9948	39.3270	41.1485	41.5545	41.7007	41.7633	41.8663
Color structure	9.9998	19.9981	29.9861	38.8828	39.4112	39.5178	39.5758	39.6210	39.6553

Table 3. Threshold value for each false positive ratio and corresponding average CSR.

False positive ratio (ppm)	0.01	0.05	0.1	0.5	1	5	10	50	100
Threshold	16	20	22	26	28	33	36	41	43
Average CSR (%)	96.54	97.20	97.47	97.91	98.08	98.36	98.44	98.29	98.05

color change, and histogram equalization. Further, the proposed statistical feature is weak in the cases of contrast change, color change, and histogram equalization. When we adapted the proposed synthetic feature and the proposed matching and selection scheme, we obtained a very high CSR value of 98.4%.

Another performance measure is recall-precision. It is mainly used for evaluating the retrieval algorithm. For this test, we prepared 10,000 original images and made 44 transformed images from each original. The complete image database includes 450,000 images. Further, we selected one original image and performed image retrieval using this image as the query. In this retrieval step, we limited the number of retrieved images. The number of relevant images was 45 in this experiment because the database includes 44 modified images of each original image. This retrieval step was performed again after changing the limitation of the number of retrieved images. Then, we repeated this whole retrieval step 10,000 times with 10,000 original images and found the average number of correct matches under different limitation of the number of retrieved images. Table 2 is the result of this experiment. After that, we calculated the average recall and precision values according to

$$recall = \frac{\#\,of\ relevant\ and\ retrieved\ images}{\#\,of\ relevant\ images},$$
$$precision = \frac{\#\,of\ relevant\ and\ retrieved\ images}{\#\,of\ retrieved\ images}. \qquad (12)$$

Figure 5 is the recall-precision graph of the competing algorithms. According to Fig. 5, we know that the proposed spatial feature is better than any other spatial features and the

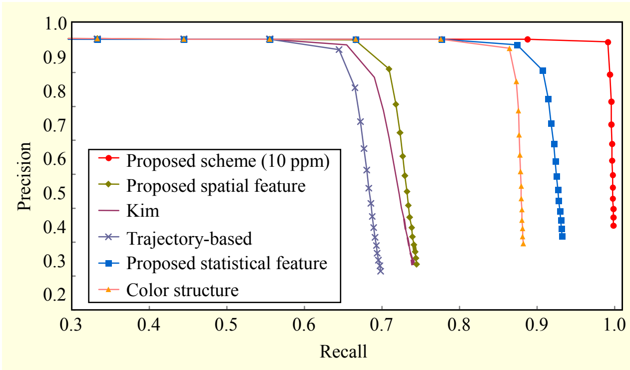

Fig. 5. Recall-precision graph of competing algorithms.

proposed statistical feature is better than the color structure descriptor from the retrieving point of view. The combination of the proposed spatial and statistical features results in good performance.

The proposed matching and selection scheme uses the spatial feature distance threshold. The use of a suitable threshold value leads to improved search ability. A low threshold value reduces the false positive probability, and a high threshold value improves CSR in general. To find a suitable threshold value, we measure the search performance under various threshold values.

The candidate threshold value is decided as follows. We prepared 135,609 different images and composed all possible pairs. Then, we calculated the spatial distance between all the possible pairs and made a distance histogram. The candidate threshold value was the distance positioned at 0.01, 0.05, 0.1, 0.5, 1, 5, 10, 50, and 100 parts per million (ppm). Table 3 shows the threshold value and the average CSR of each false

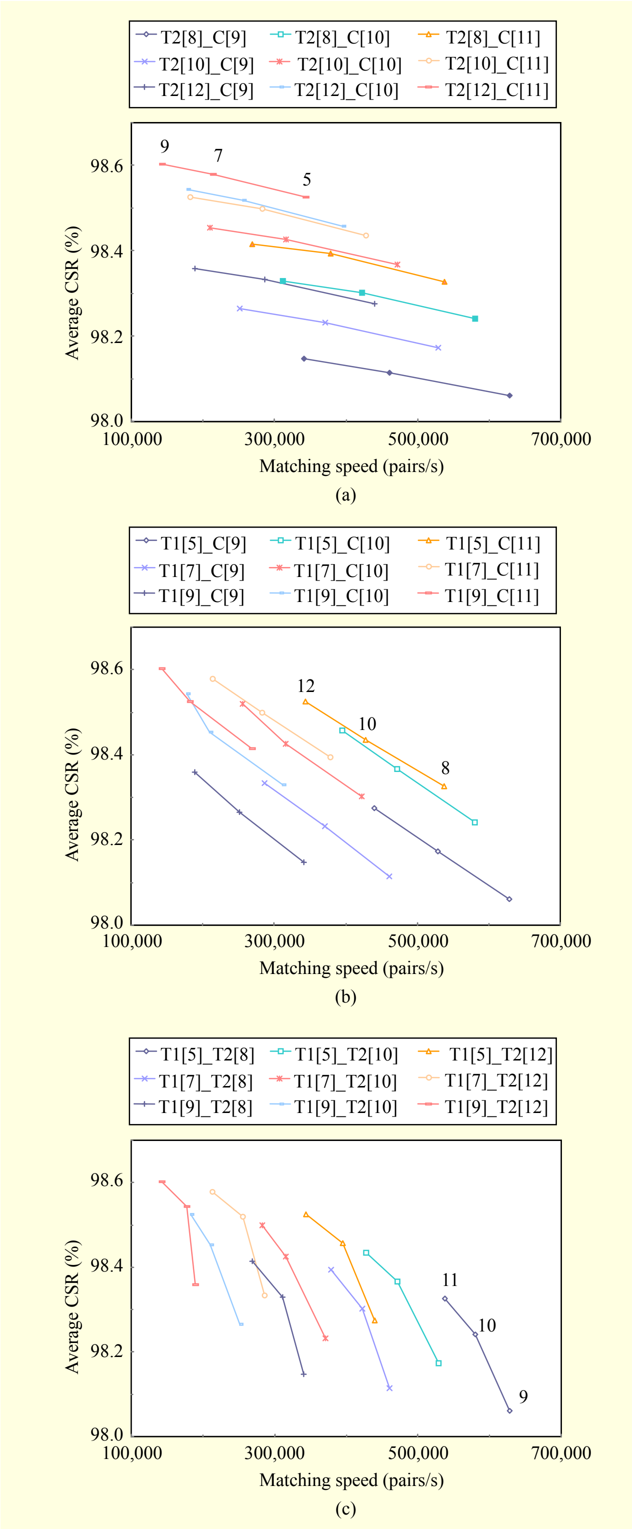

Fig. 6. Relation graph between search accuracy and speed for various quantization levels of each extracted characteristic: (a) by varying quantization levels (5, 7, and 9) of texture 1 characteristic, (b) by varying quantization levels (8, 10, and 12) of texture 2 characteristic, and (c) by varying quantization levels (9, 10, and 11) of color characteristic.

positive ratio. We found that a threshold value of 10 ppm is the most suitable value.

Matching speed is another very important issue for the introduced recommendation service because image matching has to be carried out in real-time for a commercial service. The spatial feature can ensure a fast matching speed because the distance is calculated by using an exclusive OR operation. However, for the statistical feature, the matching speed depends on the number of bins. Furthermore, the number of bins in the statistical feature affects the search accuracy. Having many bins results in high search accuracy but a low matching speed, and having only a few bins leads to low search accuracy but a high matching speed. To decide the suitable number of bins, we measured search accuracy and matching time according to the quantization levels of three characteristics that comprise the proposed statistical feature. Figure 6 is the relation graph between the search accuracy and the matching time according to the quantization levels of each characteristic. Figure 6(a) shows the relation graph with various quantization levels of texture 2 and color characteristics. Each mark of graph means the quantization level of the texture 1 characteristic (5, 7, and 9). In the same way, Fig. 6(b) is about the texture 2 characteristic (each mark means the quantization level 8, 10, and 12), and Fig. 6(c) is about color characteristic (each mark means the quantization level 9, 10, and 11).

The average CSR decreases slowly as the quantization level of texture 1 characteristic increases, as shown in Fig. 6(a). In this case, we select five levels for the high efficiency. Choosing a small level can guarantee a fast matching speed. Alternatively, the average CSR is decreased rapidly, as shown in Fig. 6(c). This implies that we should use many quantization levels for the color characteristic. Therefore, the derived best choice is 5 levels for the texture 1 characteristic, 10 levels for the texture 2 characteristic, and 11 levels for the color characteristic.

VI. Conclusion

In this paper, we proposed an image description and matching scheme using synthetic features for a recommendation service. The recommendation service is a very useful technology because it provides an efficient and diverse search. For this service, the image description method had to be robust to varying transformations and fast when matching. To satisfy these requirements, we introduced an image description scheme that was a synthesis of spatial and statistical features. The spatial feature is designed to increase the discriminability by reflecting delicate variations. Also, the statistical feature is designed to increase the robustness by absorbing small variations. Our spatial feature is based on concentric circle partitioning and extracts spatial information from the image. Our statistical feature is based on three different characteristics of the image and is formed into a 3D histogram. Further, we introduced a matching and selection scheme using two independent features. From the various experiments, we proved that both our spatial and statistical features exhibited a higher performance than the competing algorithms. We also confirmed that the proposed matching and selection scheme using synthetic features had a high performance under various transformations.

Much work is required to extend our technology. Speed has to be improved by introducing a filter in the preprocessing process. The development of geometrical transformation invariant features is another topic for future work. The proposed method is robust against almost all geometrical transformations except the heavy cropping and translating transformation because of the loss of image information. To extend the query from the entire image to the region of interest, this weakness should be covered. A combination with a local feature like SIFT can be one solution. Another topic to be explored is the extension of this method to videos. A considerable amount of content has moved from images to videos. Therefore, the recommendation service should be extended to videos. Because videos are sequences of images, the proposed image description and matching scheme can be easily extended to handle videos.

References

[1] Y.-S. Kim et al., "Development of Content-Based Trademark Retrieval System on the World Wide Web," *ETRI J.*, vol. 21, no. 1, Mar. 1999, pp. 39-53.

[2] A.W.M. Smeulders et al., "Content-Based Image Retrieval at the End of the Early Years," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 12, Dec. 2000, pp. 1349-1380.

[3] C. Carson et al., "Blobworld: Image Segmentation Using Expectation-Maximization and Its Application to Image Querying," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 8, Aug. 2002, pp. 1026-1038.

[4] D. Lowe, "Distinctive Image Feature from Scale-Invariant Keypoints," *Int. J. Comput. Vision*, vol. 60, no. 2, Nov. 2004, pp. 91-110.

[5] H. Zhang et al., "Local Image Representations Using Pruned Salient Points with Applications to CBIR," *Proc. ACM Multimedia*, vol. 14, 2006, pp. 287-296.

[6] D.-H. Kim et al., "Support Vector Machine Learning for Region-Based Image Retrieval with Relevance Feedback," *ETRI J.*, vol. 29, no. 5, Oct. 2007, pp. 700-702.

[7] E. Chang et al., "CBSA: Content-Based Soft Annotation for Multimodal Image Retrieval Using Bayes Point Machines," *EIEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 1, Jan. 2003, pp. 26-38.

[8] J.Z. Wang and J. Li, "Automatic Linguistic Indexing of Pictures by a Statistical Modeling Approach," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 9, Sept. 2003, pp. 1075-1088.

[9] C. Kim, "Content-Based Image Copy Detection," *Signal Process.: Image Commun.*, vol. 18, no. 3, Mar. 2003, pp. 169-184.

[10] K. Wnukowicz, G. Galiński, and R. Tous, "Still Image Copy Detection Algorithm Robust to Basic Image Modifications," *Proc. ELMAR*, 2008, pp. 455-458.

[11] R. Brunelli and O. Mich, "Histograms Analysis for Image Retrieval," *Pattern Recognition*, vol. 34, no. 8, Aug. 2001, pp. 1625-1637.

[12] M.J. Swain and D.H. Ballard, "Color Indexing," *Int. J. Computer Vision*, vol. 7, no. 1, Nov. 1991, pp. 11-32.

[13] ISO/IEC Std. 15938-3, "Multimedia Content Description Interface Part 3: Visual," International Standard, 2002.

[14] Q. Tian, Y. Wu, and T.S. Huang, "Combine User Defined Region-of-Interest and Spatial Layout of Image Retrieval," *Proc. ICIP*, vol. 3, 2000, pp. 746-749.

[15] W.-K. Yang et al., "Image Description and Matching Scheme for Identical Image Searching," *Proc. Computation World*, 2009, pp. 669-674.

[16] I.-H. Cho et al., "Very Fast Concentric Circle Partition-based Replica Detection Method," *Lect. Notes Comput. Sci.*, vol. 4872, Nov. 2007, pp. 905-918.

[17] B.S. Manjunath et al., "Color and Texture Descriptors," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, no. 6, June 2001, pp. 703-715.

[18] P.W. Huang and S.K. Dai, "Image Retrieval by Texture Similarity," *Pattern Recognition*, vol. 36, no. 3, Mar. 2003, pp. 665-679.

[19] H.A. Moghaddam et al., "Wavelet Correlogram: A New Approach for Image Indexing and Retrieval," *Pattern Recognition*, vol. 38, no. 12, Dec. 2005, pp. 2506-2518.

[20] C.-H. Wei et al., "Trademark Image Retrieval Using Synthetic Features for Describing Global Shape and Interior Structure," *Pattern Recognition*, vol. 42, no. 3, Mar. 2009, pp. 386-394.

[21] C.C. Chang and Y.K. Chan, "A Fast Filter for Image Retrieval Based on Color Features," *Proc. SEMS*, 2000, pp. 47-51.

[22] Y.-K. Chan and C.-Y. Chen, "Image Retrieval System Based on Color-Complexity and Color-Spatial Features," *J. Syst. Software*, vol. 71, no. 1-2, Apr. 2004, pp. 65-70.

[23] D. Reisfeld, H. Wolfson, and Y. Yeshurun, "Context Free Attentional Operators: The Generalized Symmetry Transform," *Int. J. Comput. Vision*, vol. 14, no. 2, Feb. 1995, pp. 119-130.

[24] C.-J. Park et al., "An Efficient Context-Free Attention Operator for BLU Inspection of LCD," *Proc. IASTED SIP*, 2000, p. 251.

[25] J. van de Weijer, T. Gevers, and A. Bagdanov, "Boosting Color Saliency in Image Feature Detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 1, Jan. 2006, pp. 150-156.

Won-Keun Yang received his BS in electronic engineering and his MS in information engineering from Inha University, Rep. of Korea, in 2004 and 2006, respectively. He is currently working toward a PhD in information engineering at the same university. His research interests include augmented reality, visual search, and image and video signature.

Ayoung Cho received her BS in electronic engineering and her MS in information engineering from Inha University, Rep. of Korea, in 2003 and 2005, respectively. She is currently working toward a PhD in information engineering at the same university. Her research interests include machine vision, watermarking, and image and video signature.

Weon-Geun Oh received his BS from Chungbuk National University, Rep. of Korea, in 1979, and his MS from Yeungnam University, Rep. of Korea, in 1981. He received his PhD from Osaka University, Japan, in 1988. He now works at ETRI as a principal research engineer. His research interests include computer vision, pattern recognition, and digital rights management.

Dong-Seok Jeong received his BSEE from Seoul National University, Rep. of Korea, in 1977, and his MSEE and PhD from Virginia Tech, USA, in 1985 and 1988, respectively. He became a member of the IEEE in 1983 and a senior member in 2000. He is also a member of SPIE and HKN. From 1977 to 1982, he was a researcher at the Agency for Defense Development of Korea. Since 1988, he has been a professor at the School of Electronic Engineering, Inha University, Rep. of Korea. He served as the president for the Institute for Information and Electronics Research, Inha University, from 2000 to 2004. His research interests include image and video processing, image and video signature, and forensic watermarking.