RESEARCH REPORT

SGLD-BASED FEATURE REPRESENTATION OF IMAGES AND ITS MAPPING TO INDIVIDUAL'S IMAGERY

Ying Dai and Dawei Cai

Iwate Pref. University, Takizawa, Iwate, 020-0193, Japan

Abstract - In this paper, with the objective to retrieving digital collections more intuitively and flexibly, we proposed a scheme of visual perception-based image feature representation and its retrieval, which adopt individual user's imagery by the multi-form query, such as query-by-sample, query-by- objective perception words, or query-by- subjective impression words. For this, the eigen SGLD (Space Gray Level Dependence) matrices were used to extract the low-level textural features of images, and the performance of association between the extracted features and the visual attributes was analyzed. On this basis, the method how to map the textural features to the individual's visual perception, including the objective, subjective, and similarity perception, was proposed, and the application of the individual perception criteria was discussed because of the variety of individual's imagery. Furthermore, the scheme of flexible image retrieval based on user's imagery was designed by the flexible query types, including query-by-sample, query-by- objective perception words, or query-by- subjective impression words. The user satisfaction-based evaluation showed that the scheme proposed was effective. By adjusting the values of parameters regarding individual's visual perception criteria, the retrieved images were adaptable to the various imageries of subjects.

Keywords: *multiform query, Eigen SGLD matrices, individual perception criteria, flexible image retrieval, imagery*

1 INTRODUCTION

In the last few years, with the development of IT technology, digital library have give rise to the interests of many people from all over the world. The digital library plays an important role in enabling people to explore the collections for research, inspiration, learning, and enjoyment easily. With this objective, the study on digital collection retrieval based on people's imagery is indispensable.

Recently, a variety of related papers of digital collection retrieval have become available [1] [2] [3] [4] [5] [6][7] [8] [9] [10] [11]. In [1] [4] [5] [7], content-based image retrieval with the query image was discussed by using the features of color, texture, or shape of images. But, using the features of color, texture, and shape of images respectively to retrieve images seems tedious, because the texture features include the color primitives and their spatial layout. On the other hand, these methods ignored the gap existing between the high level semantics in the human minds (whether expressible by words or not) and the low level features computable by machines when the users searched images by a sample query. As a resolving approach, in [2] [3] [9], by asking the user to give feedbacks regarding the positive or

negative relevance of the current outputs of the system, machine learn associations on-line with user in the loop. Although the proposed methods are somewhat efficient, "A picture is worth a thousand words", that is, the dynamic interpretation of images under various circumstances are so abundant that there is a doubt that on-line learning is usable. In [8], the impression word-based image retrieval system was proposed. As a query, impression words were interpreted into color features to search images which match the semantics in user mind. But, it is not enough to embody the semantics of users under various circumstances only using impression words, because the imagery of people in mind can not always expressed obviously only by words. Otherwise, those approaches motioned above only take one query type, such as content-based, or impression word-based, for the retrieval of image data. Same as above, it is not efficient to retrieve images for embodying the semantics in the human minds under various circumstances.

In [10] [11], the flexibility of query type was taken into account based on the subjective perception models. In [10], concepts of the meta-layer took into account the

visual attributes of the paintings, and their manipulations which support a wider query base. But, the automatic mapping of low-level features to meta-level concepts was not focused on. In [11], authors showed the framework on the modeling of visual perception processes. On this basis, a 3D-object retrieval, a content-based similarity retrieval, and a trans-categorical similarity retrieval method was proposed using individual subjective criteria. But, this paper ignored that the user's subjective criteria was changeable under various circumstance, and the retrieval results did not expressed that the individual subjective criteria was used.

In fact, when a person searches digital collections, his/her intention is often imagery and ambiguous, because a picture is worth a thousand words. He/she would like to retrieve images by the query of samples, or key words, such as pretty, cool, red, two colors, etc. Maybe he/she is interested in the style, color arrangement or the composition of collections, or the full or partial similarity of images to the samples. Moreover, in the procedure of image retrieval, people's interests in collections, or his query types to collections are often multiple. Sometimes, he maybe look up images by impression words such as warm, soft, etc, then retrieve more suitable images by the query of sample, or the perception words, such as "two kinds of hues with the dominant one of red". On the other hand, some images with non-relevance of his/her previous imagery may become interesting to be used as a new query to search more images.

In view of statements above, with the objective to retrieving digital collections more intuitively and flexibly, we proposed a scheme of visual perception-based image feature representation and its retrieval, which adopt individual user's imagery by the multiform query, such as query-by-sample, query-by-perception-words, or query-by-impression-words. For this, the eigen SGLD (Space Gray Level Dependence) matrices were used to extract the low-level textural features of images, and the performance of association between the extracted features and the visual attributes was analyzed. On this basis, the method how to map the textural features to the individual's visual perception was

proposed, and the framework of flexible image retrieval based on user's imagery was designed.

This paper is organized as follows. In section 2, statistical and textural feature set of images in HSV color space, which is used to represent images, is presented. In section 3, the associations of the feature set with the semantic features of images based on human visual perception are analyzed. In section 4, the methods how to map the textural features to the individual's visual perception are described. Then in section 5, the framework of flexible image retrieval based on user's imagery is introduced, and user satisfaction-based evaluations are given.

2 SGLD EIGEN AND DIFFERENCE MATRICES FOR FEATURE EXTRACTION

2.1 SGLD (Spatial Grey Level Dependence) Eigen Matrices

As the name suggested, the SGLD [12] is constructed from the image by estimating the pairwise statistics of pixel intensity. Each element (m,n) of the matrix represents an estimate of the probability that two pixels with a specified separation having grey levels m and n . The separation is usually specified by a displace, d and an angle, θ . That is

$$SGLD, \Phi(d, \theta) = [\hat{f}(m, n | d, \theta)] \quad (1)$$

Afterwards, μ, σ are calculated as the means and standard deviations of row sums of SGLD matrix, respectively.

Then, let $R = \Phi \Phi'$, and λ_i, U_i are eigen values and eigen vectors of $R(i \in [1, I])$, we define the eigen SGLD matrices as

$$E^i = (e_{mn}^i) = \lambda_i U_i U_i', \quad i \in [1, I]. \quad (2)$$

Furthermore, we define the difference SGLD matrices as

$$D^i = \begin{cases} -e_{mn}^i, & \text{if } e_{mn}^i < 0; \\ 0, & \text{if } e_{mn}^i \geq 0. \end{cases} \quad (3)$$

where, I is the rank of R , and $m \in [1, I], n \in [1, I]$. The i denotes the number of eigen or difference SGLD matrix. As the example, the E^1 indicates the first eigen SGLD matrix. Then, the total number of eigen and difference SGLD matrices is I , respectively.

Features are derived from the matrix. A large number of textural features have been proposed starting with the original fourteen features described in [12]. We only

use some of them. For the function $\hat{f}(m,n)$ of the elements regarding the matrix, the index of which correspond to the separation with the gray levels m and n , the textural features utilized in this paper are defined as

$$f_1 = \text{Energy} = \sum_{m,n} \hat{f}(m,n)^2 \quad (4)$$

$$f_2 = \text{Entropy} = -\sum_{m,n} \hat{f}(m,n) \log \hat{f}(m,n) \quad (5)$$

$$f_3 = \text{Correlation} = \sum_{m,n} m * n * \hat{f}(m,n) \quad (6)$$

$$f_4 = \text{Inverse} = \sum_{m,n} \frac{1}{1+(m-n)^2} \hat{f}(m,n) \quad (7)$$

$$f_5 = \text{Inertia} = \sum_{m,n} (m-n)^2 \hat{f}(m,n) \quad (8)$$

So, for any choice of d , θ , and i , we obtain a separate eigen textural feature set F_e , and a difference textural feature set F_d , respectively

$$F_e = \{fe_j(d, \theta, i) :$$

$$j = (1, 2, 3, 4, 5), d = 1, \theta = (0^\circ, 45^\circ, 90^\circ, 135^\circ), i \in [1, I]\} \quad (9)$$

and

$$F_d = \{fd_j(d, \theta, i) :$$

$$j = (1, 2, 3, 4, 5), d = 1, \theta = (0^\circ, 45^\circ, 90^\circ, 135^\circ), i \in [1, I]\} \quad (10)$$

Here, $\{fe_j(d, \theta, i)\}$ are the textural features computed from the i th eigen SGLD matrix E^i based on the expressions of (4), (5), (6), (7), (8). And $\{fd_j(d, \theta, i)\}$ are those computed from the i th difference SGLD matrix D^i .

2.2 Feature Set of Images in HSV

It is known that the images are displayed in their composition and color style. The color style, which is represented by the hue and tone, is often described by the color name and the number of colors that the images have, or the impression words, such as warm, bright, etc. Composition is embodied by the spatial layout of color. On the other hand, as we know that the texture can be described by its color primitives and their spatial layout. The spatial layout can be periodic, quasi-periodic or random. It is natural that we can associate that the relations between the style of images and their texture features, or the composition of images and their texture features may be found by analyzing the texture performance of color primitives and their spatial layout. Furthermore, we can extract the associations between the visual attributes of images and the texture features,

because the human visual perception is mainly influenced by the style and the composition of images. Based on the above statements, we use the H, S, V components of images as our analyzing objects due to the perceptual uniformity of HSV color space.

For the H, S, V components (where, $c=1$: H components, $c=2$: S components, $c=3$: V components), let $k = j \times c$, the separate feature sets of the eigen and difference SGLD matrices are given by

$$F_e = \{fe_k(d, \theta, i) :$$

$$k \in [1, 15], d = 1, \theta = (0^\circ, 45^\circ, 90^\circ, 135^\circ), i \in [1, I]\} \quad (11)$$

and

$$F_d = \{fd_k(d, \theta, i) :$$

$$k \in [1, 15], d = 1, \theta = (0^\circ, 45^\circ, 90^\circ, 135^\circ), i \in [1, I]\} \quad (12)$$

Furthermore, let $\mu(c), \sigma(c)$ ($c = (1, 2, 3)$) the means and the standard deviations of row sums of SGLD matrices, respectively. Let λ_i the i th eigen value of SGLD matrix of H components, μ_i the mean of row sums of i th eigen SGLD matrix of that. As the whole, the feature set of the image can be represented by the following parameters of F set.

$$F = \{\mu(c), \sigma(c), \lambda_i, \mu_i, F_e, F_d, i \in [1, I], c = (1, 2, 3)\} \quad (13)$$

Fig. 1 shows a sample of image, its first and second eigen SGLD matrices and its second difference SGLD matrix meshes regarding H components, when $d = 1, \theta = 0^\circ$.

(a) A sample

(b) First eigen matrix

(c) Second eigen matrix

(d) Second difference matrix

Figure 1: A sample and its eigen and difference SGLD matrix meshes.

Fig. 1 (b) is the mesh of the first eigen SGLD matrix of the sample showed in (a), which is the function of the elements regarding the separation with the gray levels m and n . From (b), it is observed that the element with the value close to the peak were concentrated on the index (2,2), which indicates the separation with the gray levels

2 and 2 of H components. The remaining elements of the matrix were almost equal to 0. However, the gray level 2 is close to the value of the hue of the dominant object (sunflower) which is orange. It means that the first eigen SGLD matrix extracted the dominant hue component of the sample, and reflected the distribution of that. The index with which the element had the peak value corresponded to the gray level of the dominant hue.

Fig. 1 (c) presents the second eigen SGLD matrix of H components. From (c), it is observed that the element the value of which was close to the peak values were with the separation of (10,10), which are close to the gray level of blue. Because the sample's background was nearly blue, the above observation just expressed the fact that the second eigen matrix extracted the sub-dominant hue, and reflected the distribution of that. The index with which the elements had the larger value corresponded to the gray level of the sub-dominant hue components.

Fig. 1 (d) present the second difference SGLD matrix of the sample. From (d), it is observed that the elements with the larger values were with the separation of (2,10), or that of (2,5), which were close to the gray levels of orange, blue, and green respectively. Because the major objects in the image corresponded to these hues respectively, the value of element (2,10), or that of element (2,5) revealed the information of the major edges. That the value of element (2,10) is larger than that of element (2,5) meant the boundary shaped by the orange object and the blue object was longer than that done by the orange object and the green object, which just accorded with the observation that the boundary shaped by the sunflower and the sky was longer than that done by the sunflower and the grass. Accordingly, the second difference extracted the information of the major edges of the image, and reflected the distribution of the main edges.

In fact, besides this sample, the eigen SGLD matrices and the difference SGLD matrices of the other 100 samples, randomly extracted from the Sozaijiten-Image Book 1, published in Japan, revealed the similar tendency regarding the distribution of the objects.

3 VISUAL ATTRIBUTE ANALYSIS REGARDING FEATURE SET F

The term "visual attributes" is defined in this text as color parameter, composition parameter, object layout, etc. In the followings, the association between the extracted feature set F and the visual attributes is analyzed. H,S,V components are scaled into 16 levels.

3.1 For $\mu(1)$ and $\sigma(1)$ in Hue Distribution

As described in [15], $\sigma(1)$ is the deviation of the SGLD matrix of H components, which reflects the hue distribution of the images. When $\sigma(1) < 2$, the images appear the simple color layout--one hue layout. When $\sigma(1) = 2$ or $\sigma(1) = 3$, the images appear a little complex color layout. When $\sigma(1) > 3$, the images include the hues which are complementary. But, when $\sigma(1) > 6$, the images appear the simple color layout again, because of the Hue Circle.

3.2 For $\mu(2), \sigma(2), \mu(3), \sigma(3)$ in Tone Distribution

Similar to 3.1, $\mu(2), \sigma(2), \mu(3), \sigma(3)$ are the means and deviations of SGLD matrices of S, V components respectively. These four parameters reveal the tone distribution of images. The larger the values of $\mu(2), \mu(3)$, the brighter the images appear; contrary, the images with the smaller values of $\mu(2), \mu(3)$ appear darker. On the other hand, the images with the larger values of $\sigma(2), \sigma(3)$ appear the more abundant tones.

Some examples of images with different values of $\mu(2), \sigma(2), \mu(3), \sigma(3)$ are shown in Fig. 2. These examples verified the above analysis that the images with larger values of $\mu(2), \mu(3)$ are bright; and those with larger values of $\sigma(2), \sigma(3)$ have the abundant tones.

Figure 2: Examples with different values of $\mu(2), \sigma(2), \mu(3), \sigma(3)$

Besides these samples, the $\mu(2), \sigma(2), \mu(3), \sigma(3)$ of other samples, randomly extracted from the Sozaijiten-Image Book 1, published in Japan, showed the similar property regarding the tone distribution.

3.3 For Eigen Value λ_i in Hue Distribution

As described above, in the case of $1 < \sigma(1) \leq 6$, the images appear a little complex spatial color layout. So, we use the eigen values $\{\lambda_i\}$ of the SGLD matrix regarding H components to measure the hue distribution of the images.

Let the rate of eigen values be defined as

$$r_i = \frac{\lambda_i}{\sum_{i=1}^J \lambda_i}, \quad (14)$$

where, λ_i indicates the i th eigen value. Then the number of eigen values, the rate of which is larger than a threshold, is calculated as

$$n_\lambda = \text{numberof}\{r_i > T\}.$$

The larger eigen value represents that its corresponding eigen matrix has the larger variance for the SGLD matrix. Accordingly, based on the explanation of eigen SGLD matrices regarding the hue distribution in Sec. 2.2, it can be deduced that the first eigen value λ_1 reflects the dominant hue distribution of images, and the second eigen value λ_2 does the sub-dominant hue distribution. Moreover, the hue distribution of images becomes complex with the increasing of n_λ .

Furthermore, in order to compare the similarity of the hue distribution of the two images, we give the distance measurement between the two images as

$$d_1 = \sum_{i=1}^J (r_i - r_i^q)^2, \quad (16)$$

where r_i , and r_i^q indicate the rate of eigen values of the two images, respectively. Based on the expression (16), a sample and its top 3 similar images which were extracted from the Sozaijiten Image Book 1, vol. 9, published in Japan, which includes 200 paints, are shown in Fig. 3.

(a) a sample (b) Image 192 (c) Image 186 (d) Image 196

Figure 3: Similar images in hue distribution

From fig.3, it is observed that the retrieved images and the sample have the similar hue distributions that the dominant hue occupied about the half region of images, respectively.

3.4 For μ_i in Hue Distribution

μ_1 is the mean of row sums of the first eigen SGLD matrix of H components, which reveals the average gray level of that. On the basis of the explanation in Sec. 2.2, the first eigen SGLD matrix extracts the dominant hue component, and reflects the distribution of that. Accordingly, it can be deduced that μ_1 is close to the gray level of the dominant hue. $\mu_1 = 1$ expresses that the gray level of the dominant hue is close to 1, which corresponds to the color near to red. The following are some samples which had the different values of μ_1 that represented the different gray levels of the dominant hues.

$\mu_1 = 1$ $\mu_1 = 3$ $\mu_1 = 5$ $\mu_1 = 8$

Figure 4: Samples with different values of μ_1

3.5 For F_e, F_d in Object Similarity

In order to analyze the similarity of objects in images, we give the similarity measurement of two images regarding F_e, F_d by the following distance.

$$d_2 = \sum_{i=1}^2 \sum_{k=1}^{15} w1_i ((fe_k(d, \theta, i) - fe_k^q(d, \theta, i))^2 + \sum_{k=1}^{15} w2 (fd_k(d, \theta, 2) - fd_k^q(d, \theta, 2))^2 \quad (17)$$

Where $fe_k, fd_k, fe_k^q, fd_k^q$ are the features regard with the eigen and difference SGLD matrices of the two images respectively, calculated by (4)-(8). In the expression (17), the features of the first and the second eigen SGLD matrices, and the second difference SGLD matrix are used. $d=1$, and $\theta=0^\circ$. $\{w1_i\}$ are the weights of features regarding the eigen SGLD matrices. We let $w1_1 = 10^x$, $w1_2 = 1$. Furthermore, we let

$$w2 = \begin{cases} w1_1 & \text{if } w1_1 > w1_2 \\ 1 & \text{if } w1_1 \leq w1_2 \end{cases}$$

The smaller the d_2 , the more similar the image is to the other. In the case of $x = -1$, the weight of features of the first eigen matrix is weak, and that of the second eigen matrix is relatively strong. In the case of $x = 1$, the weight of first eigen matrix becomes relatively strong. With varying the values of x , the performance of associations between F_e, F_d and the object layout

can be analyzed by observing the ranking changes of images in similarity.

Here is an example which demonstrated the change tendency of ranking in object similarity regarding a reference. With varying $wl_1 = 10^x$, the change tendency of the images in rank 1, and these images are shown in Fig. 5. Image set used here comes from the Sozaijiten-Image Book 1, Vol. 20, published in Japan, which includes 200 wild flowers.

Figure 5: Change tendency of rank 1 with varying wl_1

Image 9 is an image in rank 1 when $-1 \leq x \leq 0.05$, Image 162 is that when $0.05 < x \leq 0.1$, Image 106 is that when $0.1 < x \leq 1$. From Fig.5, we can observed that the retrieval image in rank 1 is stable in a certain scope of x . Furthermore, although the listed images in Fig.5 are all similar to the query image to some extent, it can be seen that Image 9, which was in rank 1 when the value of x was small, is more similar to the query in background (green tree); Image 106, ranked in first with the larger value of x , are more similar to the query in foreground (the flower). Accordingly, we can see that the increasing of x weights the similarity of the dominant object; by decreasing the value of x , the similarity to the other objects is stressed. So, by adjusting $wl_1 = 10^x$, that is, adjusting the weight of first and the second eigen SGLD matrices, the images which are similar to a particular object (foreground or background) can be ranked upwards.

Besides this sample, analyses to other samples also revealed the similar rank change tendency of candidates when wl_1 was varied. Furthermore, besides the image set above, the image database of Sozaijiten-Image Book

1, Vol. 2, Vol. 9, and Vol. 20 was also tested. Each volume includes 200 images of papers & fabrics, paints & pastels, or wild flowers. Analysis to these images showed the similar results of above.

From the analysis results, we can deduce that expression (17) play a role in bridging the gap between the objects and the feature set F_e, F_d . Based on (17), by increasing the value of wl_1 , the images which are more similar to the dominant object of the sample can be ranked upwards; contrary, by decreasing the value of wl_1 , the images which are similar to the other objects of the sample can be ranked upwards. By adjusting wl_1 , the images which match the user's request in mind further can be retrieved.

3.6 For F_d in Composition Similarity

Based on the definition of F_d , F_d possesses the orientation property because of the parameter of angle θ . So, we use the F_d to measure the composition similarity of images. Let

$$r_{kl} = \frac{|fd_k(1, l+1, 2) - fd_k(1, l, 2)|}{\sum_{l=1}^3 |fd_k(1, l+1, 2) - fd_k(1, l, 2)|} \quad (18)$$

Where, $l = \{1, 2, 3, 4\}$ denotes the angle of $0^\circ, 45^\circ, 90^\circ, 135^\circ$, respectively. Then, let r_{ikl} and r_{qkl} denote the r_{kl} of a certain candidate C_i and a reference respectively, the similarity between the candidates and the reference in composition can be calculated by the following measurement.

$$Sub1_i = \sum_{k=1}^{15} \sum_{l=1}^3 4 * |r_{ikl} - r_{qkl}| \quad (19)$$

The candidate C_i which is similar to the reference in composition can be extracted form the database by the following expression. T1 is a threshold.

$$Sub1_i < T1 \quad (20)$$

Moreover, let

$$s = \sum_{k=1}^{15} r_{kl}, \text{ and} \\ Sub2_i = s_i - s_q, \quad (21)$$

where, s_i and s_q denotes the s of the candidate and the sample respectively. Same as above, the following expression

$$Sub2_i < T2 \quad (22)$$

is used as the other condition to extract the images which are similar to the reference in composition. T2 is also a threshold.

Fig. 6 shows the extracting performance regarding (20) (22) in image composition similarity by changing $T1$ or $T2$. A image set used [here comes from](#) the Sozaijiten-Image Book 1, Vol. 20, published in Japan, which includes 200 wild flowers. The x axis indicates the change values of $T1$ or $T2$, and y axis does the image ID of the extracted candidates, which meet the condition of (20) or (22), and are in the rank of top 40 in similarity based on the following distance.

$$d_3 = \sum_{k=1}^{15} (fd_k(d, \theta, i) - fd_k^q(d, \theta, i))^2 \quad (23)$$

where, fd_k , fd_k^q are the features of the second difference SGLD matrices regarding the two images, respectively. Furthermore, $d = 1$, $\theta = 0^\circ$, and $i = 2$.

(1) A query image

(2) Candidates when $T2$ changed

(3) Images of top 4 when $T2=3$
Extracted candidates when $T2$ changed

Figure 6: Extracted candidates in composition similarity

From Fig. 6, it is observed that in a certain scope of $T1$ or $T2$, the extracted candidates were stable. For example, in the scope of $T1=3$ and $T2=\{2,3,4,5\}$, the extracted candidates were almost unchangeable. When $T1$ or $T2$ was becoming larger, with the increasing of the relevant images, the images which seem not to be relevant to the sample were also retrieved.

On the other hand, besides this image set, the image database mentioned in 3.2 was also used to examine the extraction performance of the candidates in composition similarity. The results were similar.

Based on the analysis above, we can see that $T1=3$ and $T2=3$ can be used as the initial thresholds to extract images which are similar to the sample in composition. Furthermore, by adjusting the parameter $T1$ or $T2$, the images which accord with the user's request further for composition similarity can be extracted.

4 APPLICATION OF INDIVIDUAL PERCEPTION CRITERIA

The term of "individual perception" here means the perception of a person on things that he/she is doing. For enjoying images, "individual perception" means the felling of the person to the style of images, described as warm, soft, red, et al, and the understanding to the composition of those, described as similar to that in shape, et al.

In the above section, the associations of the feature representation set F of the images with the visual attributes regarding the images were analyzed. In this section, on this basis, we describe the models of mapping the feature set F to the user's imagery based on the individual's visual perception criteria. The individual's visual perception here includes the individual user's objective, subjective, and similarity perceptions, because it is considered that the human's imagery is influenced on these perceptions. Of course, other kinds of individual's perceptions are considered to be expanded if necessary.

● Objective visual perception

The human objective visual perception on images corresponds to the words as red, two kinds of color, etc. Such objective perception is associated with the parameters n_λ and μ_1 in feature set F , because in section 3, we know that the n_λ reveals the hue distribution of images, and the value of μ_1 is associated with the gray level of the dominant hue. The mapping model of n_λ and μ_1 to the objective visual perception is given in the followings.

include the objective or subjective words. The related objective and subjective words are those listed in Fig. 7 and Fig. 8. These words represent the user's imagery space. The other subjective words, such as pretty, elegant, etc, are interpreted by the words of warm/cool, soft/hard, or clear/grayish. Furthermore, the images can be queried by the sample. The images which are similar to the reference in the objects or the composition are retrieved. By the combination of these query means and the parameters' adjusting of the relevant retrieval module according to user's feedback, the images which are closer to the individual user's imagery can be retrieved flexibly.

The parameter adjusting module responds to the user's feedback about the satisfaction by applying the individual perception criteria, which were described in section 4. This module adjusts the parameters, including $\{t_{o_i}, i \in [1,6]\}$, $\{t_{s_i}, i \in [1,8]\}$, $w1_1, T1$, and $T2$, according to the user's feedback. Which parameters are adjusted is determined according to the retrieval module that the user is executing.

For example, the parameters $\{t_{o_i}\}$ are adjusted if the object word-based retrieval module is executed. The variation of parameter t_{o_1} means the different acknowledge of the users on the complexity of the color layout of the images.

For the other example, the parameters $w1_1$, $T1$, and $T2$ are adjusted if sample-based module is executed. If the user lays particular stress on the dominant object of the sample, increasing the value of $w1_1$ ($w1_1 = 10^x$) can adhere to user's intention, that is, the images which are similar to the dominant object of sample are ranked upwards; similarly, decreasing the value of $w1_1$, the images which adhere to user's particular stress on the other objects can be ranked upwards. By adjusting $T1$ and $T2$, the numbers of retrieved images similar in composition can be controlled.

This parameter adjusting procedure is continued until the feedback of satisfaction from the user returns.

5.2 User Satisfaction-based Evaluation

On the basis of the scheme proposed above, we made a prototype of imagery-based image retrieval by flexible

query types. Differently from the general evaluation method of the image retrieval, which use the precision and recall to evaluate the performance of the proposed methods, we conducted the user satisfaction-based evaluation for the proposed methods, because it embodies the evaluation of the usability in the imagery-based image retrieval. On the other hands, we have not found any publications that presented such a system retrieving images by the combination of the multiform query, which embodied the individual user's imagery.

Evaluation items include: 1) the initial ranking of retrieved images is satisfactory; 2) the modified ranking is satisfactory; 3) the number of times that you found the images matching to your imagery from the top 20 retrieved images for the initial ranking; 4) the number of times that you found the images matching to your imagery from the top 20 retrieved images for the modified ranking. The samples for query were selected randomly from the image database of Sozaijiten, Image Book 1, mentioned above. The rest images were used as the candidate images for retrieving.

Three persons as test subjects were asked to answer the above items. They executed the prototype 10 times for 10 independent imagery-based queries, respectively. The query types including query by objective words, query by subjective words, and query by sample, depended on the preference of each subject. If the initial retrieval ranking was not satisfied by a subject, the subject adjusted the threshold of the relevant parameters of $\{t_{o_i}\}$, $\{t_{s_i}\}$, $w1_1$, $T1$, or $T2$ by GUI, and obtained a modified retrieving ranking (How to adjust parameters easily will be presented in another paper). Table 1 showed the test results. Item a) indicates the average percent of positive answers of 1) regarding the number of queries, which were 10 in the experiment. Item b) does the average percent of positive answers of 1) and 2). Item c) is the average times regarding the evaluation item 3). Item d) is the average times of 3) and 4).

Table 1: Test results of user satisfaction

Num. of images	Image database	Num.of queries	a)	b)	c)	d)
200	Vol. 2	10	56%	80%	6	8
200	Vol. 9	10	50%	66%	6	7
200	Vol. 10	10	66%	77%	7	8

200	Vol. 20	10	40%	63%	5	7
-----	---------	----	-----	-----	---	---

From the results shown in Table 1, we can see that the scheme of imagery-based digital collection retrieval is effective. More than 50% of the initial retrieving satisfied the subjects' requests of imagery on the images, and for the 10 imagery-based queries, the average times that the subjects found the images matching to their imagery from the top 20 retrieved images was more than 5 for the initial ranking. Furthermore, by adjusting the thresholds of parameters $\{to_i\}$, $\{ts_i\}$, $w1_i$, $T1$, or $T2$, the retrieval results were improved further. Adding the number of positive answers after the adjustment, the average of about 70% of the retrieving results satisfied the subjects' requests on imagery, and the average times that the subjects found the images matching to their imagery from the top 20 retrieved images became more than 7 times.

6 CONCLUSION

In this paper, firstly, we define the feature set for representing images in HSV components by using the eigen and difference SGLD matrices. Then, the associations of the low level feature set with the high level human visual attributes are analyzed. The method how to map the feature set to the individual's visual perception is also described. On this basis, the scheme of flexible image retrieval based on user's imagery is proposed, and user satisfaction-based evaluations are executed.

For the next work, on the basis of the proposed scheme of imagery-based image retrieval, we will develop a perception interface regarding the individual user's imagery, and evaluate the usability of our system furthermore.

REFERENCES

[1] W. Y. Ma, and H. J. Zhang; Content-based image indexing and retrieval, Handbook of multimedia computing; CRC, 227-254 (1999).
 [2] X. S. Zhou, and T. S. Huang; Small sample learning during multimedia retrieval using BiasMap, Proc. Of IEEE CVPR'01, Vol. I; Hawaii, 11-17 (2001).
 [3] Y. Rui, and T. S. Huang; Optimizing learning in image retrieval, Proc. Of IEEE CVPR'00; 236-243 (2000).
 [4] T. Gevers, et al.; Color constant ratio gradients for image segmentation and similarity of texture objects, Proc. Of IEEE CVPR'01; Vol. I, Hawaii, 8-25, (2001)

[5] S. Belongie, et al.; Color- and texture-based image segmentation using EM and its application to content-based image retrieval, Proc. Of IEEE ICCV 1998; India, (1998).
 [6] Proc. of IEEE workshop on content-based access and video libraries; CVPR (1997).
 [7] B. V. Funt, et al.; Color constant color indexing, Trans. Of IEEE PAMI; 17(5), 522-529 (1995).
 [8] K. Yoshida, et al. Image retrieval system using impression words, Proc. Of IEEE SMC'98; 2780-2784 (1998).
 [9] G. D. Duo, et al.; Learning similarity measure for natural image retrieval with relevance feedback, Proc. Of IEEE CVPR'01; Vol. I, Hawaii, 731-736 (2001).
 [10] M. Yelizaveta, et al.; Representation and retrieval of paintings based on art history concepts, Proc. Of IEEE ICME'04; CDROM, Taiwan, (2004).
 [11] T. Kato; Trans-category retrieval based on subjective perception process models, Proc. Of IEEE ICME'04; CDROM, Taiwan, (2004).
 [12] R. M. Haralick; Textural features for image classification, IEEE Trans. SMC; SMC-3(6), 610-621 (1973).
 [13] Y. Dai, et al.; Face-texture model based on SGLD and its application in face detection in a color scene, Pattern Recognition; Vol. 29; No. 6, 1007-1017 (1996).
 [14] S. Kobayashi; Color System; Koudan Publication, Japan, (1999).
 [15] Y. Dai; Intention-based image retrieval with or without a query image using Eigen SGLD matrices, Proc. of IEEE IRI 2003; Las Vegas, USA, 266-272 (2003).