Imagery-based Image Retrieval Considering Individual's Subjective Perception

YING DAI DAWEI CAI
Iwate Prefectural University
Faculty of Software and Information Science
Takizawa, Iwate
JAPAN

Abstract: - In this paper, we proposed an approach of imagery-based image retrieval considering individual's subjective perception. By using eigen SGLD (Space Gray Level Dependence) matrices, the statistical and textural features of images were extracted, the associations between these features and semantic features based on human visual perception were analyzed, and the mapping models were designed to bridge the gap between the feature representation and the individual use's imagery. On this basis, the scheme of imagery-based flexible image retrieval by multiform query was proposed. Furthermore, the user satisfaction-based evaluation was made.

Keyword: - flexible query, imagery-based image retrieval, individual perception

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. The previous 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, 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. For the other methods, 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 in the human minds under various circumstances.

In order to overcome the defects of above methods in image retrieval, the flexibility of query type was taken into account based on the subjective perception models. In [1], 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 [2], 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 semantics 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 primary 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 flexible, we proposed a scheme of visual perception-based image feature representation and its retrieval, which can adapt 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.

2 SGLD Eigen Matrices and the Visual Perception-based Feature Extraction

2.1 SGLD (Spatial Grey Level Dependence) Eigen Matrices

As the name suggested, the SGLD [3] is constructed from the image by estimating the pairwise statistics of pixel intensity. Each element (*i, j*) of the matrix represents an estimate of the probability that two pixels with a specified separation have grey levels *i* and *j*. The separation is usually specified by a displace, d and an angle, ϑ. That is

$$SGLD, \Phi(d,\vartheta) = [\hat{f}(i,j\,|\,d,\vartheta)]$$

$\mu_i, \mu_j, \sigma_i, \sigma_j$ are calculated as the means and standard deviations of row and column 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

$$E^i = (e^i_{mn}) = \lambda_i U_i U^{'}_i, \quad i \in [1, I].$$

Furthermore, we define the difference SGLD matrices as

$$D^i = \begin{cases} -e^i_{mn}, & if\ e^i_{mn} < 0; \\ 0, & if e^i_{mn} \geq 0. \end{cases}$$

where, I is the rank of *R,* and $m \in [1, I], n \in [1, I]$. The *i* indicates the eigen or difference SGLD matrix number. 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.

In texture classification, features are derived from the matrix. A large number of textural features have been proposed starting with the original fourteen features described in [5]. We only use some of them. If $\hat{f}(i,j)$ is the value of element in the matrix, the textural features utilized in this paper are defined as

$$f_1 = Energy = \sum_{i,j} \hat{f}(i,j)^2 \quad (1)$$

$$f_2 = Entropy = -\sum_{i,j} \hat{f}(i,j) \log \hat{f}(i,j) \quad (2)$$

$$f_3 = Correlation = \sum_{i,j} i * j * \hat{f}(i,j) \quad (3)$$

$$f_4 = Inverse = \sum_{i,j} \frac{1}{1+(i-j)^2} \hat{f}(i,j) \quad (4)$$

$$f_5 = Inertia = \sum_{i,j} (i-j)^2 \hat{f}(i,j) \quad (5)$$

So, for any choice of d, ϑ, eigen SGLD matrix and difference SGLD matrix, we obtain a separate eigen textural feature set $\{fe_j(d,\theta,i)\}$, and a difference textural feature set $\{fd_j(d,\theta,i)\}$. The separate feature set of eigen and difference SGLD matrices used here is given by

$$F_e = \{fe_j(d,\theta,i): \\ j = (1,2,3,4,5), d = 1, \theta = (0^o, 45^o, 90^o, 135^o), i \in [1, I]\}$$

and

$$F_d = \{fd_j(d,\theta,i): \\ j = (1,2,3,4,5), d = 1, \theta = (0^o, 45^o, 90^o, 135^o), i \in [1, I]\}$$

2.2 Textural Feature Set of Images in HSV

As described as in [5], we use the H, S, V components of images as our analyzing objects due to the perceptual uniformity of HSV color space.

If we use the n_λ of eigen SGLD and the (n_λ*-1)* of difference SGLD matrices, 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 set of eigen and difference SGLD matrices for the image are given by

$$F_e = \{fe_k(d,\theta,i): k \in [1,15], d=1, \theta=(0^o,45^o,90^o,135^o), i \in [1,n_\lambda]\}$$

and

$$F_d = \{fd_k(d,\theta,i): k \in [1,15], d=1, \theta=(0^o,45^o,90^o,135^o), i \in [2,n_\lambda]\}$$

The meaning of n_λ will be described afterwards.

Furthermore, let λ_j the jth eigen value of SGLD matrix of H components, and let μ_j the means of row sums of jth eigen SGLD matrix of H components. As the whole, the statistical and textural features of the image can be composed by the following parameters.

$$F = \{\mu(c), \sigma(c), \lambda_j, \mu_j, F_e, F_d, j \in [1,n_\lambda], c=(1,2,3)\} \tag{6}$$

3 Visual attribute analysis regarding feature set *F*

So called visual attributes here indicate the things as color parameter, composition parameter, object layout, etc. In following, the performance of association between the extracted features 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 [5], when $\sigma(1) < 2$, with the increasing of $\mu(1)$, the primary hue of images is varied in the range of hue circle, as red, orange,…, blue,…, purple. When $\sigma(1)=2$ or $\sigma(1)=3$, the images appear a little complex color layout. When $\sigma(1) > 3$, the images have contrast hue components. But, when $\sigma(1) > 6$, the image appears the simple color layout again, because of the Hue circle.

3.2 For $\mu(2), \sigma(2), \mu(3) and \sigma(3)$ in tone distribution

Similar to 3.1, $\mu(2), \sigma(2)), \mu(3) and \sigma(3)$ are the mean and standard deviation for S, V components respectively. These four parameters reveal the tone distribution of images. The larger the values of $\mu(2), \mu(3)$, the bright 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 $\mu(2), \sigma(2)), \mu(3) and \sigma(3)$ are shown in Fig. 1. These examples verified the above analysis that the images with lager values of $\mu(2), \mu(3)$ are bright; and those with larger values of $\sigma(2), \sigma(3)$ have the abundant tones.

(a) $\mu_i(2)=10, \mu_i(3)=15, \sigma_i(2)=2, \sigma_i(3)=1$
(b) $\mu_i(2)=14, \mu_i(3)=14, \sigma_i(2)=2, \sigma_i(3)=2$
(c) $\mu_i(2)=10, \mu_i(3)=14, \sigma_i(2)=6, \sigma_i(3)=4$
(d) $\mu_i(2)=12, \mu_i(3)=6, \sigma_i(2)=4, \sigma_i(3)=6$

Fig. 1 Examples with different values of $\mu(2), \sigma(2), \mu(3), \sigma(3)$

3.3 For eigen value λ_i in hue distribution

As described above, in the case of $1 < \sigma(1) \le 6$, the images appears a little complex spatial layout of color. In this case, it is difficult to interpret the style of images well only by using the measurement of $\mu(1)$. So, we use the use the eigen values of $\{\lambda_i\}$ of H components to measure the hue distribution of images.

Let the rate of eigen values is defined as

$$r_i = \frac{\lambda_i}{\sum_{i=1}^{I} \lambda_i},$$

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

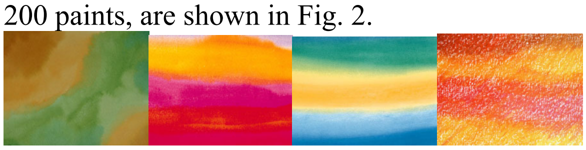
$$n_\lambda = numberof\{r_i > T\}.$$

Because the first eigen value λ_1 of SGLD matrix regarding H components represented the primary hue distribution of images, and the second eigen value λ_2 does the sub-primary hue distribution, it can be deduced that the hue distribution of images becomes complex with the increase of n_λ.

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

$$d_1 = \sum_{i=1}^{I} (r_i - r_i^q)^2, \tag{7}$$

where r_i, and r_i^q indicate the rate of eigen values of the two images, respectively. The smaller the d_1, the more similar the images in hue distribution. Based on formula (7), 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. 2.



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

Fig.2 Similar images in hue distribution

From fig.2, it is observed that the retrieved images and the sample have the similar hue distributions that the primary hue occupied about the half region of images, respectively.

3.4 For μ_j in hue

Based on above, we can deduced further that the means of row sums of first eigen SGLD matrix of H components μ_1 reveals the primary hue that image have. $\mu_1 = 1$ represents that the primary hue of the image is 1, which corresponds to the color near to red of hue circle. The following are some image samples which have the different values of μ_1.

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

Fig.3 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, when $1 < \sigma \le 6$, we give the similarity measurement of two images regarding F_e, F_d by the following distance.

$$d_2 = \sum_{i=1}^{n}\sum_{k=1}^{15} w1_i((fe_k(1,0^o,i) - fe_k^q(1,0^o,i))^2 + \sum_{i=2}^{n}\sum_{k=1}^{15} w2_i(fd_k(1,0^o,i) - fd_k^q(1,0^o,i))^2 \quad (8)$$

Where $fe_k, fd_k, fe_k^q, fd_k^q$ are the textural features regard with the eigen and difference SGLD matrices for the two images. $n = \max\{n_\lambda, n_\lambda^q\}$. $\{w1_i\}$ are the weights regarding the textural features of the eigen SGLD matrices of H, S, V components. The smaller the d_2, the more similar the image is to the other. Furthermore, we let $w1_1 = 10^x$, $w1_i = 1 (i > 1)$, and

$$w2_i = \begin{cases} w1_1 & if w1_1 > w1_i, \\ 1 & if w1_1 \le w1_i, \end{cases} \quad i > 1.$$

As described in [5], by adjusting $w1_1 = 10^x$, that is, adjusting the weight of first and other eigen SGLD matrices, the images which are particularly similar to foreground or background can be stably ranked upwards.

$w1_1 = 10^x$, as a adjusting parameter, certainly play a role in bridging the gap between the human visual attributes in spatial layout of objects and the textural feature set F_e, F_d. Based on (8), by increasing the value of $w1_1$, the images which are more similar to the sample on the foreground can be ranked upwards; contrary, by decreasing the value of $w1_1$, the images which are similar to the query on the background can be ranked upwards. By adjusting $w1_1$, the images which match the user's imagery in mind further can be retrieved stably.

4 Application of individual perception criteria

So called individual perception here means the perception of a person on things that he/she is doing. For enjoying collections, that 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 performance of associations between the textural feature set F and the visual attributes were analyzed. In this section, on this basis, we describe the methods of mapping the textural feature representation to the semantic feature representation of user's imagery using the individual's perception criteria. The individual's perception here includes the objective, subjective, and similarity perception.

- Objective perception

The human objective visual perception on images corresponds to the words as red, two kinds of color, complex color layout, etc. Such objective perception is associated with the parameters n_λ and $\{\mu_j\}$ in feature set F, because in section 3, we know that n_λ reveals the hue distribution of images, and the values of $\{\mu_j\}$ are associated with the values of hue circle. The mapping models of n_λ and μ_1 to the objective visual perception-based semantic features are given in the following.

Table 1 Mapping model regarding objective perception

Key words	n_λ	μ_1
complex	$n_\lambda > to_1$	
red	$n_\lambda \le to_1$	$1 \le \mu_1 < to_2$
yellow	$n_\lambda \le to_1$	$to_2 \le \mu_1 < to_3$
green	$n_\lambda \le to_1$	$to_3 \le \mu_1 < to_4$
blue	$n_\lambda \le to_1$	$to_4 \le \mu_1 < to_5$
purple	$n_\lambda \le to_1$	$to_5 \le \mu_1 < to_6$

Here, the parameter set $TO = \{to_i, i \in [1,6]\}$ is adjustable according to the individual user's objective perception criteria, that is, the user's acknowledgement regarding the color and color layout. The method how to adjust the parameters $\{to_i\}$ will be explained in next section.

- Subjective perception

The human subjective visual perception on images corresponds to the words as pretty, elegant, classic, clear, etc. Furthermore, such words can be mapped to the 3D space of warm/cool, soft/hard, clear/grayish, which are associated with the color hues and color tones [4]. Generally, the feeling of warm is associated with the color of red, pink, cool with the blue, and calm with the other colors. The feeling of soft is associated with the abundant tones and brightness, and hard with the darkness. Accordingly, these visual attributes are associated with the parameters of n_λ , μ_1 , $\mu(2), \sigma(2), \mu(3) and \sigma(3)$ of feature set F on the basis of section3. On the other hand, the definite decision criteria for the subjective perception are different for the different individuals. So, the mapping models of the subjective visual perception-based semantic features to the feature representation are given in the following.

Table 2 Mapping model regarding subjective perception

Key words	n_λ	μ_1
Warm	$n_\lambda < ts_1$	$1 \le \mu < ts_2, ts_5 \le \mu \le 16$
Calm	$n_\lambda < ts_1$	$ts_2 \le \mu < ts_3, ts_4 \le \mu \le ts_5$
Cool	$n_\lambda < ts_1$	$ts_3 \le \mu < ts_4$

(a)

Key words	$\sigma(2)$	$\mu(2)$	$\mu(3)$
soft	$\sigma(2) \ge ts_6$	$ts_7 \le \mu(2) \le 16$	
hard	$\sigma(2) < ts_6$	$1 \le \mu(2) \le ts_7$	
clear		$ts_8 \le \mu(2) \le 16$	$1 \le \mu(3) < ts_9$
grayish		$1 \le \mu(2) \le ts_8$	$ts_9 \le \mu(3) \le 16$

(b)

Here, the parameter set $TS = \{ts_i, i \in [1,9]\}$ is adjustable according to the individual user's subjective perception criteria, that is, the user's imagery regarding the color and color layout.

- Similarity perception

The similarity perception here indicates the object similarity of images.
The mapping model of object similarity to the textural feature set F is based on formula (8). The images are ranked based on d_2. The top images are more similar to the sample in the object arrangement. By adjusting the weight wl_1, the images which are more similar to the sample on the primary object(foreground) or the other objects(background) can be ranked upwards.

5 Imagery-based digital collection Retrieval

5.1 Imagery-based digital collection Retrieval Scheme

Based on the above discussion, we propose a scheme of imagery-based flexible digital collection retrieval by multiform query as in Fig. 4.

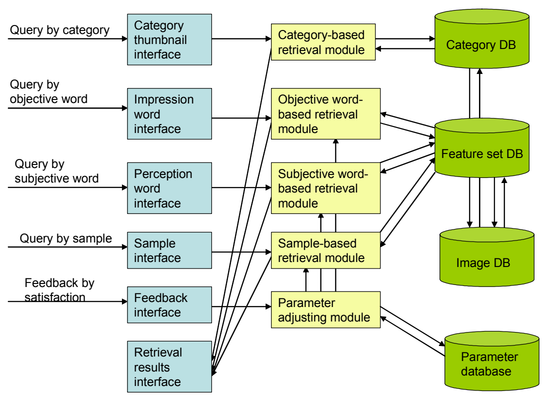


Fig. 4 image retrieval scheme

This scheme realizes the flexible image retrieval by the multiform queries, which consider the individual users` particular interests in image's style, object arrangement, composition, or the combination of them. Based on this scheme, if user has not the sample collections in hand, the images can be queried by some objective or subjective words. Here, the related objective and subjective words for retrieving images were shown in Fig. 6 and Fig. 7. If the user has the samples in hand, the more suitable images can be retrieved by the flexible query types and the individual's feedback about the satisfaction.
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 $\{to_i, i \in [1,6]\}$, $\{ts_i, i \in [1,9]\}$,and wl_1 , by increasing or decreasing the values of those

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 $\{to_i\}$ are adjusted if the object word-based retrieval module is executed. The increasing of to_1 results in the decreasing of extracted images with the complex hue layout.

For the other example, the parameter wl_1 is adjusted if sample-based module is executed. If the user lays particular stress on the primary object of the sample, increasing the value of wl_1 ($wl_1 = 10^x$) can adhere to user's intention, that is, the images which are similar to the primary object of sample are ranked upwards; similarly, decreasing the value of wl_1, the images which adhere to user's particular stress on the other objects can be ranked upwards.

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

5.2 User satisfaction-based evaluation

Based on the scheme proposed above, we made a prototype of imagery-based image retrieval by flexible query types and execute the user satisfaction-based evaluation testing of the image retrieval. Evaluation items include: 1) the initial ranking of retrieved images is satisfy; 2) the modified ranking is satisfy; 3) the times that you found the images matched to your imagery from the top 20 retrieved images for the first ranking; 4) the times that you found the images matched 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 utilized in image retrieval process, including query by objective words, query by subjective words, and query by sample, were as the subject's pleasure. Table 3 showed the test results. Item a) indicates the average percent of positive answers of 1), item b) does the average percent of positive answers of 2), item c) is the average times of evaluation item 3), and item d) is the average times of 4).

Table 3 Test results of user satisfaction

Number of images	**Image database**	**Number 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 of user satisfaction testing shown in Table 3, we can know that the scheme of imagery-based digital collection retrieval is effective, and most of retrieving results satisfied testers' requirements. Furthermore, by adjusting the values of parameters $\{to_i\}$, $\{ts_i\}$, wl_1 *T1,* or *T2*, the modified ranking could satisfy testers' requirements further.

6 Conclusion

In this paper, firstly, we defined the statistical and textural feature set of digital collections in HSV components by using eigen SGLD matrices, which was used to index image database. Then, based on the human visual perception, the associations of the high level semantic features with the low level statistical and textural features were analyzed. Then, the methods how to map the textural features to the individual's visual perception were described. On this basis, the scheme of flexible image retrieval based on user's imagery was proposed, and user satisfaction-based evaluations were executed.

For the next work, we will do more tests to confirm that the user satisfy the ranking results of digital collection retrieval based on his individual imagery in mind.

7 References

[1] M. Yelizaveta, et al. Representation and retrieval of paintings based on art history concepts, Proc. Of IEEE ICME'04, CDROM, Taiwan, 2004.

[2] T. Kato, Trans-category retrieval based on subjective perception process models, Proc. Of IEEE ICME'04, CDROM, Taiwan, 2004.

[3] R. M. Haralick, Textural features for image classification, IEEE Trans. SMC SMC-3(6), 1973, pp.610-621.

[4] S. Kobayashi, Color System (Koudan Publication, 1999).

[5] Y. Dai, Intention-based image retrieval with or without a query image using Eigen SGLD matrices, Proc. of IEEE IRI 2003,USA, pp. 266-272.