MVA2000 IAPR Workshop on Machine Vision Applications, Nov. 28-30, 2000, The University of Tokyo, Japan

3—19 Visualization System Displaying Retrieved Images in a 2-D Semantic Space

Yoshinori Musha * sha@crl.hitachi.co.jp

Atsushi Hiroike * he@crl.hitachi.co.jp

Information-Base Functions Hitachi Laboratory, RWCP
http://www.rwcp.or.jp/lab/ib-hitachi/home-e.html

Abstract

We have developed a similarity-based image retrieval system that represents retrieved images as a scatter diagram in a *semantic space*. An axis of the space shows the suitabilities of a keyword assigned to the images. The suitabilities are estimated by a linear transformation of the image features of retrieved data, and the coefficients of the transformation are learned by a multiple regression analysis. The system can store multiple key-images and can retrieve images by using the center of gravity of the key-image feature vectors. The system can effectively assist a user in retrieving images through the use of semantic visualization and this center of gravity retrieval based on similarity. The system performs as follows. First, a user retrieves images. Then, the system presents the user with some keywords that provide large variances of the estimated suitabilities from among the retrieved results. Next, the user can choose two axes from the given keywords. Finally, the system displays the retrieved images in the *semantic space* that is spanned by the axes. We examined our method quantitatively by using a large number of samples, each of which was a pair made up of an image and its assigned keywords.

1 Introduction

Recently, against the background of demand from image content providers and companies that process multimedia information, similarity-based image retrieval has become an active field of research where the similarity between images is being evaluated at the pattern level[1, 5]. Many researchers have been working to improve the retrieval ability of systems by developing retrieval methods that reflect a user's intentions. However, there is an important problem with these methods in that a user's intentions are sometimes indeterminate. In the cases, a user tends to view as many images as possible in the user's effort to find targets that fulfill the user's need. Therefore a retrieval system needs to present users with many images and many variations of similarity among the images by using image visualization, and to assist the users in their trial and error search. It is the purpose of our researches. Operating from this point of view, we developed a visualization system for similarity-based image retrieval[2, 3]. Although one of our system's functions allows a user to select two axes in a visualization space from many axes constructed automatically, it is difficult to motivate the user to select the axes. This is because it is not always easy for a user to understand the meanings of the axes. Therefore, to assist the user, it is very important to assign some meanings to the axes. To provide the user with many variations of similarity among images, a semantic visualization is also very important. To meet this challenge, we developed a similarity-based image retrieval system, which represents retrieved images as a scatter diagram in a semantic space[4]. Because the keyword suitabilities are estimated from the features of retrieved images, It isn't necessary for each image in a database to have any keyword. The system can retrieve images by using the center of gravity among features of multiple key-images. The system can effectively assist a user in retrieving images through the use of semantic visualization and this center of gravity retrieval.

*Address: c/o Hitachi Ltd.,Central Research Laboratory, 1-280, Higashi-Koigakubo, Kokubunji-shi, Tokyo 185-8601, Japan.

2 Similarity-based Image Retrieval

In the system, when a user retrieves images, the system first receives an example image set from an arbitrary location the user has specified in a database. The image set is displayed in a space, the axes of which are two keywords that the image set is effectively visualized. The user can change the arrangement of the image set by selecting other keywords and the user can select a key-image for the next retrieval. Next, the system searches, and displays the retrieved images in two keywords' space, and then prompts the user to select a key-image. The system extracts image feature vectors that reflect similarities from each image in a database in advance and gathers images that are near to a key-image based on distances from a key-image vector to the image vectors. Thus, the similarity-based image retrieval is performed. The system can store multiple key-images selected by a user and it can retrieve images by using the center of gravity of the key-image feature vectors. By using this center of gravity retrieval, we can expect retrieved images to have different properties for the multiple key-images and an image retrieval process that emphasizes the similar properties of multiple key-images. We used a color feature and a differential orientation feature for our retrieval features. In the color feature, we prepared 4 x 4 x 4 boxels by dividing an RGB color space beforehand. The system sorts all the pixels in an image into the 64 boxels, according to the quantized color vectors for the pixels, makes a histogram using the 64 boxels counts, and finally normalizes the histogram. In the differential orientation feature, the system transforms color images to gray scale images, calculates the orientation and power of the changes of brightness, adds the powers of all the pixels in the image to one of 8 bins according to the quantized orientation of each pixel, makes a histogram using the 8 bins counts, and finally normalizes the histogram. The system applies the above feature extractions to each 4 x 4 area and each of four resolutions. We compressed the feature dimensions by applying principal component analysis to the features in order to speed up retrieval.

3 Visualization System Using Keywords as Axes in the Space

A visualization in this paper means the representations of an image set made by arranging images in a space according to characteristics of the images. The visualization actively uses the intrinsic property of an image, the characteristics of which are represented as a view of the image. This visualization serves to advance a user's intuitive understanding of the similarities among images. Because the system uses keywords as the axes in the space, it's desirable for the keywords to reflect a view of the images. We proposed a method where the system estimates the keyword suitabilities for each image by transforming the features of the image and uses the suitabilities as the image coordinates. By using this method, the system can effectively present the visualization of an image set and the variations of similarity among the images as semantic information to the user. To make this semantic information for the user, the system uses mainly the following functions: (I) Estimates the keyword suitability for each retrieved image. (II) Presumes representative keywords, which are commonly included in the retrieved images. (III) Presumes visualization keywords, which express the differences in the retrieved images and are supplied for a user to set the keywords to the axes.

4 Keyword Suitability Estimation

It's necessary that the above functions, (I), (II), and (III), are performed whenever retrieved images are determined. To keep quick responses of the user interface and complete the function calculations in the interval between retrieval and result output, the keyword suitability estimation must be performed quickly. Therefore, for keyword suitability estimation, we used a multiple regression analysis in which the system can perform a fast transaction because of the linear calculation. To presume representative and visualization keywords, we designed the system to calculate the mean and variance of the suitabilities in the images and then sort the keywords in mean and variance order, respectively, considering as follows. In representative keywords, the mean value of estimated suitabilities tends to be high, because all the retrieved images tend to include the keywords, thus, each of which makes any image highly suitable. In visualization keywords, the variance value of estimated suitabilities tends to be high, because the keywords tend to divide the retrieved images into two groups evenly and then scatter the images depending on their suitabilities. When the system calculates the mean and variance in images, the linear transaction enables the system to calculate a mean vector and a variance-covariance matrix of the image features and then apply them to all the keywords. We can reduce the calculation steps by using the above method.

A learning method and an estimation method of keyword suitability are shown below. Let $\mathbf{x}_i$ be a feature vector of an image i and $y_i(k)$ be a criterion variable given by the k'th keyword,

$$y_i(k) = \begin{cases} Y_e(k)\,, & \text{if the keyword is attached} \\ Y_n(k)\,, & \text{otherwise .} \end{cases} \quad (1)$$

We used $\{Y_e, Y_n\} = \{1 - N(k)/N, -N(k)/N\}$, where N is the total number of learning samples, and $N(k)$ is the number of samples that the k'th keyword is attached to. It is known that using multiple regression analysis with this definition is equivalent to the linear discriminant analysis of two categories. The partial regression coefficients are calculated from the statistics of learning samples,

$$b_0(k) = \overline{y}(k) - \overline{\mathbf{x}}^T\mathbf{b}(k)\,, \quad \mathbf{b}(k) = \mathbf{C}_{xx}^{-1}\mathbf{C}_{xy}(k)\,, \quad (2)$$

where $\mathbf{C}_{xx}$ is a variance-covariance matrix of $\{\mathbf{x}_i\}$, $\mathbf{C}_{xy}(k)$ is a covariance vector between $\{\mathbf{x}_i\}$ and $\{y_i(k)\}$, and $\overline{\mathbf{x}}$ and $\overline{y}(k)$ are the means of $\{\mathbf{x}_i\}$ and $\{y_i(k)\}$, respectively. The estimated value $\widehat{y}_i(k)$ for the above stated function(I) is given as

$$\widehat{y}_i(k) = b_0(k) + \mathbf{x}_i^T\mathbf{b}(k)\,. \quad (3)$$

Eq.(4) and Eq.(5) are used for calculating the representative keyword(II) and the visualization keyword(III), respectively. If a set of images is given, the mean and the variance of $\widehat{y}_i(k)$ are calculated as

$$\tilde{y}(k) = b_0(k) + \tilde{\mathbf{x}}^T\mathbf{b}(k) \quad (4)$$

$$\tilde{\sigma}_y^2(k) = (\mathbf{b}(k))^T\,\widetilde{\mathbf{C}}\,\mathbf{b}(k)\,, \quad (5)$$

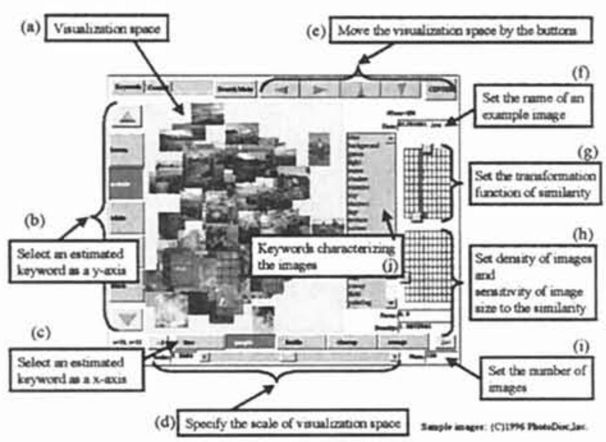

Figure 1: Visualization interface using semantic axes

where $\widetilde{\mathbf{x}}$ and $\widetilde{\mathbf{C}}$ are the mean vector and the variance-covariance matrix of the feature vectors within the given set. By collecting keywords having large values of $\tilde{y}(k)$, we can obtain keywords that can be assumed to characterize the set. And keywords having large values of $\tilde{\sigma}_y^2(k)$ can be assumed to clarify the differences among the set.

5 Visualization Interface

The system is constructed of a server and clients. The server retrieves images and calculates the statistics analysis. The client has a graphical user interface(GUI), as described by JAVA^{TM}, and is capable of displaying images and keywords to users and forwarding user requests to the server. The system has three main functions, which correspond to the aforementioned (I),(II), and (III).

Figure 1 shows the parts of the GUI. The parts (a),(b),(c), and (j) are the characteristic functions of the system, defined as follows: (a) shows a visualization space where the retrieved images are arranged according to the keyword suitabilities of each image (I). The user can select some images in the space, and then add them to a key-image set in which the center of gravity feature is applied to the retrieval. (b) and (c) show user-selectable buttons which present 50 visualization keywords to the user (III) and allow the user to select favorite keywords from them as the axes in the space. (j) shows a text area that displays 20 representative keywords to the user (II). The other parts show the functions of (e)moving, (d)expanding and contracting a range of the space, and (g),(h)changing a relation between image similarity and image size, and so on.

6 Data Flow in the System

Figure 2 shows the data flow which creates the interface in the system. An image feature database is made by extracting features from all the images.

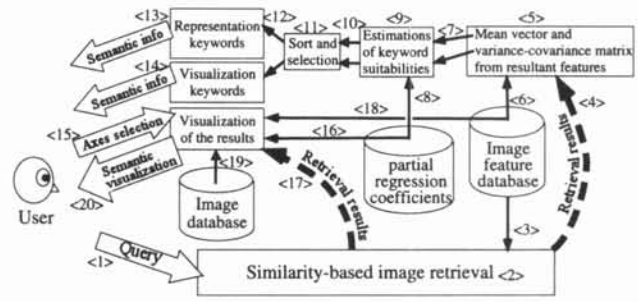

Figure 2: The data flow in the system

A partial regression coefficient database is made beforehand by applying Eq.(2) to the image feature database and each registered keyword. The system gets a query<1> from a user and performs a similarity-based image retrieval<2>. Then the system outputs retrieved results<4> after scanning an image feature database<3>. A mean vector and variance-covariance matrix<5> are calculated from the image features<6> of the retrieved results. The system obtains the coefficients of a keyword from the keyword partial regression coefficient database<8> and applies Eq.(4) and Eq.(5)<9> to the vector and matrix<7>. Then, the system outputs a mean and variance<10> of a keyword suitability estimation. By repeating the above operations, the system collects two sets of the first N_k keywords in mean and variance order, respectively<11>. The set, each of which has large mean value, is constructed of the representative keyword and the other set, each of which has large variance value, is constructed of the visualization keyword <12>. The system provides semantic information to the user by displaying the two keyword sets <13,14>. The user selects two keywords as the X- and Y-axes<15> from the visualization keywords<14>. The system calculates the keyword suitability, Eq.(3), by using the partial regression coefficients<16> of the two keywords and the image features of the retrieved result<17>, and then it arranges a thumbnail image<19> on the coordinates which are the two keyword suitabilities. Thus, the system displays the images visualized by the keywords to the user<20>.

7 Experiment of Keyword Suitability Estimations

To evaluate keyword suitability, we used 24,728 images from PhotoDisc,Inc., each of which had 15 to 30 keywords. We divided the images into two sets, that is, a sample image set and a test image set. There were 7,091 kinds of keywords assigned to at least two images of both these sets. Prior to the evaluation, we made a coefficient database for the linear transformation used in the estimation, by applying a multiple regression analysis to all image features and each keyword. Then we performed our evaluation using the sample image set as follows. First, we

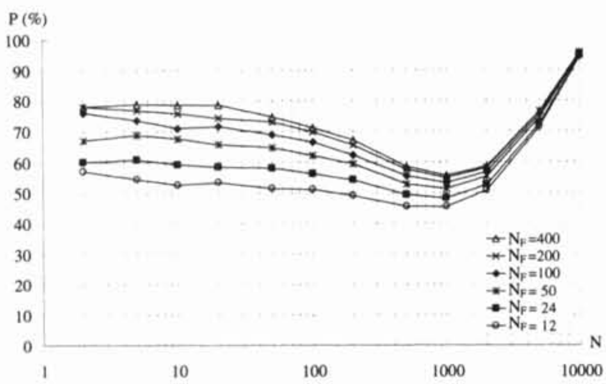

Figure 3: Evaluation of keyword suitability estimations: P is the ratio including correct answer images within N ranking and is taken average of the ratios of first 50 keywords in suitability variance order.

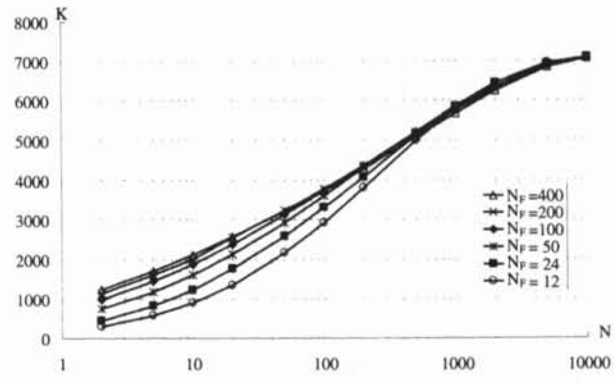

Figure 4: The number of effective keywords: K is the number of keywords, each of which is correctly assigned to at least one image within N ranking in suitability order.

estimated the suitability of each keyword and counted the number of the images that appeared with the keyword within the first N ranking in suitability order. Next, we calculated the ratio of this number to the max number of images that were capable of appearing most often with the keyword within the first N ranking. In the system, the user selects two from the first 50 keywords, the suitabilities of which have large variances in the retrieved images. Therefore, in the Fig. 3, we plotted the ratios averaged in the first 50 keywords in suitability variance order. The ratios were at least 50 percent within the first 100 ranking in the cases of the features having 12 elements. The ratios tended to get larger; they were calculated by taking an average of the smaller number of keywords from the top. To get the high ratios in case of many keywords, we need to improve the estimation ability of the system.

Figure 4 shows the number of keywords used in the ability of keyword suitability estimation. Within the top 100 ranking, there were 3,775 keywords in the features of $N_F = 400$. Increasing the keywords according to the number of feature elements causes saturation. We thought the other keywords included those that could not be learned well enough and ones whose suitability could not be intrinsically estimated using linear calculation. They never appeared within the first 100 ranking in the test data. To speed up the system and save the memory, we can delete the uncontributed data from the keyword partial coefficients database. The figure represents the design policy of our system.

8 Conclusion

We have developed the visualization system to be used as a graphical user interface for a similarity-based image retrieval. The system uses two keywords as axes in semantic space. The system dynamically shows a user representative keywords and visualization keywords as soon as the user retrieves images. Then, the system displays the retrieved images as a scatter diagram in a 2-D semantic space, two axes of which are selected by the user from the visualization keywords. Because the keywords are presumed by transforming features of the retrieved images, the image data don't have to use any text information. The system can effectively assist the user in retrieving images through the use of semantic visualization and similarity-based image retrieval using the center of gravity among features of key-images. We reported on our evaluation of the efficacy of the estimation of keyword suitability. We were able to obtain a design policy of the system. We need to improve the estimation ability to have a dependable system.

References

[1] M. Flickner, and H. Sawhney, *et al.*: "Query by Image and Video Content: The QBIC System," IEEE Computer, 28, 9, pp. 23–32, 1999.

[2] A. Hiroike, Y. Musha, A. Sugimoto, and Y. Mori: "Visualization of information spaces to retrieve and browse image data," Third International Conference on Visual Information Systems, pp. 155–162, 1999.

[3] Y. Musha, A. Hiroike, and A. Sugimoto: "Visualization interface of image feature space in similarity-based image retrieval(in Japanese)," IEICE D-II, J82-D-II, 10, pp. 1626–1633, 1999.

[4] Y. Musha and A. Hiroike: "Visualization interface displaying the retrieved images in a 2-D semantic space(in Japanese)," IPSJ Symposium Series (Interaction 2000), 2000, 4, pp. 205–212, 2000.

[5] M. Szummer and R. W. Picard: "Indoor-Outdoor Image Classification," Proc. of Content-Based Access of Image and Video Databases, pp. 42–51, 1998.