MVA '98 IAPR Workshop on Machine Vision Applications, Nov. 17-19, 1998, Makuhari, Chiba, Japan

13—5

An Interface for Visualizing Feature Space in Image Retrieval

Yoshinori Musha†, Atsushi Hiroike†, Yasuhide Mori‡, and Akihiro Sugimoto†
† Information-Base Functions Hitachi Laboratory, RWCP
c/o Advanced Research Laboratory, Hitachi, Ltd.
e-mail: {sha, he, sugimoto}@harl.hitachi.co.jp
‡ Information-Base Functions Tsukuba Laboratory, RWCP
e-mail: ymori@trc.rwcp.or.jp

Abstract

We have developed a visualization method and an interface for image retrieval. In our method, principal component analysis is dynamically applied to the retrieved images in order to determine their eigenspace and the retrieved images are displayed in that space. We also experimentally evaluated our method from a statistical point of view. And we found that our method effectively decentralizes the retrieved images over the two-dimensional space.

1 Introduction

The remarkable advance in hardware technology has enabled us to store vast digital still images as a database. This has led to much interest in developing image-retrieval systems. In a classical image-retrieval system, we manually attach tags, such as keywords, to the images in the database, and the system retrieves images from the database on the basis of the tags. In other words, the keywords-search method used for the text database plays a central role in the system. When the number of images in a database becomes large, however, the cost of manually attaching keywords to the images becomes very expensive. Furthermore, because people retrieve images from large collections in a variety of different contexts and with a wide range of purposes, it is doubtful that keywords attached to the images beforehand can always correctly suit the users' retrieval contexts.

Content-based image retrieval is recently becoming an active research field where the similarity between images is evaluated at the pattern level (for example, [1, 2, 3, 5, 8, 9]). To be more precise, features are first extracted from an image in a database in order to construct a histogram which represents the image, and then histograms representing the images are matched in order to evaluate the similarity between the images.

On the other hand, pattern information such as images has long been studied in pattern recognition literature. This research is aimed at giving computers our ability to discriminate. Nevertheless, the discrimination ability that computers have is greatly limited and far below ours. That is why we should take an alternative approach to image retrieval even though discrimination ability is indispensable in image retrieval. That is, we should develop an image-retrieval system from the viewpoint that we use computers to assist us in retrieving images.

We are aiming at establishing a system [6] that assists users in retrieving images by visualizing the retrieved images for the users. Along these lines, we present, in this paper, a visualization method where principal component analysis is dynamically applied to the retrieved images in order to determine their eigenspace and the retrieved images are effectively displayed in that space. We also present a quantitative evaluation of our method from a statistical point of view. We found that our method effectively decentralizes the retrieved images over the two-dimensional space. This result indicates that users can easily capture the special characteristics of the retrieved images and, therefore, our visualization method assists users in retrieving images.

2 What are missing in conventional image retrieval interfaces?

2.1 Why is visualization important?

In an information-retrieval system, it is usual for retrieved results to be aligned based on their similarity rank. This means that the discrimination results of the system are displayed. When the retrieved results correctly reflect users' retrieval contexts, this kind of display is preferable. In most cases, however, this is not the situation. That is, the retrieved results are not in harmony with our retrieval contexts or our contexts themselves are ambiguous. In such cases, the conventional display forces users to accept the discrimination results of the system.

From the information retrieval point of view, a system should visualize the retrieved results as much as possible and provide users with an interface from which they can discriminate the retrieved results.

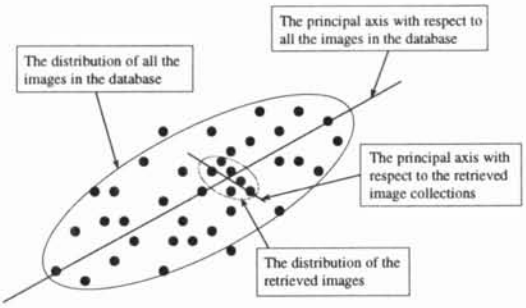


Figure 1: Principal axes of image collections (all the images in the database v.s. retrieved images).

Here visualization of the retrieved results should assist users to easily and intuitively understand the characteristics of the retrieved results. This visualization helps users to sharpen their retrieval contexts and, at the same time, helps them in selecting the next queries.

2.2 Why should statistical analysis be applied to the retrieved images?

It is straightforward, for vast image collections, to use the statistical properties of image features for effective visualization so that users can intuitively capture the characteristics of the image collections. In fact, principal component analysis is applied to image features of all the images in a database in order to recognize objects in the images [7]. To analyze the statistical properties of image features, factor analysis is applied to all the images in a database, which enables us to visualize them on a plane [4]. Kohonen's self-organization map allows all the images in the database to be visualized in a three-dimensional space [10]. These studies statistically analyze image features of all the images in the database in order to characterize the image collections. This approach, however, is not suitable for assisting users to retrieve images. This is because the principal axes for all the images in the database are, in general, not identical to those for the retrieved images (Fig. 1).

To accurately reflect the characteristics of the retrieved image collections, we should directly use the statistical properties of the retrieved images. This approach enables the system to visualize the retrieved images so that the display accurately reflects their characteristics. Users can then intuitively capture the special characteristics of the retrieved image collections.

On the other hand, if we characterize the retrieved image collections in the eigenspace obtained by applying statistical analysis to all the images in the database, the number of required axes for the characterization tends to be large. If we characterize the retrieved image collections in the eigenspace obtained by applying statistical analysis to the collections, however, we can reduce the number of required axes. In other words, a smaller number of axes is enough for the characterization when eigenspace is obtained by applying statistical analysis to the retrieved image collections. Applying statistical analysis to the retrieved images is, therefore, effective from the point of view of information compression.

3 Our interface for visualizing feature space

3.1 Overview of the interface

Taking the previous discussion into consideration, we developed an interface in which the retrieved images are effectively visualized so that users can easily understand the special characteristics of the image collections.

In our interface (Fig. 2), the system dynamically applies principal component analysis to the retrieved images and then decentralizes them over the visualization plane whose axes are determined by two eigenvectors. The upper-right part in Fig. 2 conducts specifying the conditions in retrieval, such as image features, while the upper-left part displays query images. The lower-left part is a conventional interface which aligns the retrieved images. The lower-right part is our interface which displays the retrieved images on the visualization plane. As shown in Fig. 3, our interface has functions to

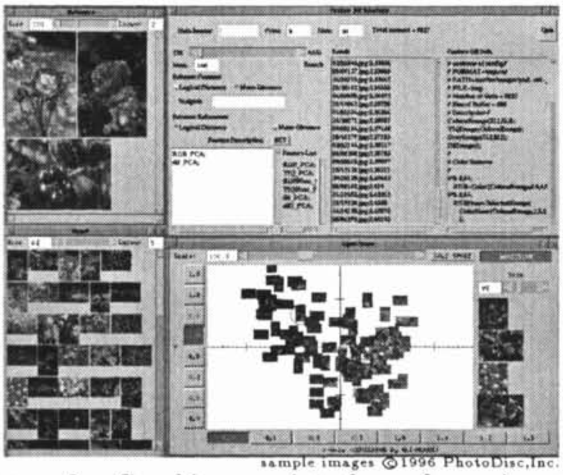
1. decentralize the retrieved images over the visualization plane by selecting two eigenvectors, which are determined by principal component analysis, as the axes of the plane ((4),(5), and (6) in Fig. 3).
2. change the display of the retrieved images continuously between the rough overview and the detailed overview by specifying the scale of the visualization plane ((2) in Fig. 3).
3. make the selected images larger or smaller ((7) in Fig. 3).
4. set query images by selecting some of the images on the visualization plane ((3),(8), and (9) in Fig. 3).
5. set query images inside the region that is selected on the visualization plane ((1) and (3) in Fig. 3).

3.2 Experiments

3.2.1 Methods

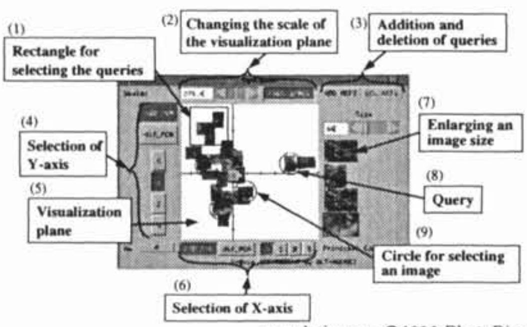
We constructed an image database in which 9807 still images are stored. We then constructed the image feature database where each image is represented as a vector (we call it a *feature vector*) of dimension 1024 on the basis of its color. (Each image is divided into 4×4 regions and the RGB feature space constructed by each region is divided into 64 boxels.) Next, we applied principal component analysis to the feature vectors and reduced their dimensions to 100 to facilitate retrieval and statistical analysis.

We randomly selected one image from the database to use as a query for image retrieval. For 100 retrieved images, we computed the variance $\sigma_t^2(d)$ of the dth entries of the feature vectors



sample images ©1996 PhotoDisc,Inc.

Figure 2: Graphic user interface of our image-retrieval system.



sample images ©1996 PhotoDisc,Inc.

Figure 3: Functions of our visualization panel.

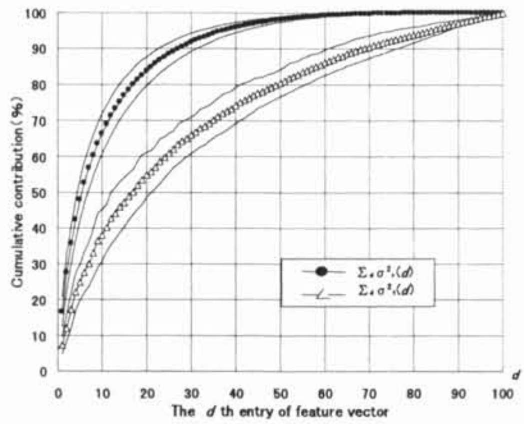


Figure 4: Cumulative contribution ratios.

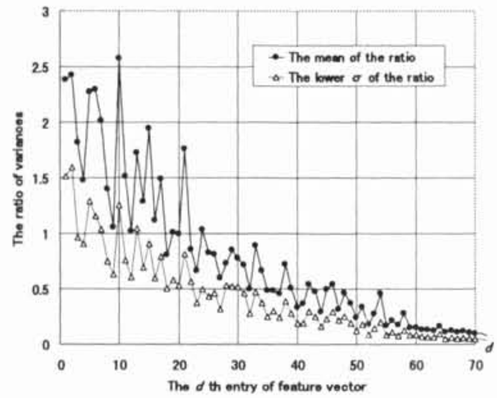


Figure 5: Geometric means and the standard deviations of $\sigma_{\rm r}^2(d)/\sigma_{\rm t}^2(d)$.

($d = 1, 2, \ldots, 100$). We dynamically applied principal component analysis to the 100 retrieved images to compute the variance (eigenvalues) $\sigma_{\rm r}^2(d)$ of the dth entries of the feature vectors. Next, we iterated retrieval 700 times, and for $d = 1, 2, \ldots, 100$, we calculated the means of cumulative contribution ratios for the 700 iterated retrievals (Fig. 4). We also computed the geometric means and the standard deviations of $\sigma_{\rm r}^2(d)/\sigma_{\rm t}^2(d)$ for the 700 iterated retrievals (Fig. 5). Note that we did not plot the upper standard deviations in Fig. 5. We also note that $\sigma_{\rm r}^2(d)$ or $\sigma_{\rm t}^2(d)$ was almost zero in one retrieval among the 700 where $d = 73, 74, \ldots, 100$. This is why the length of the horizontal axis is 72.

3.2.2 Results and discussions

Figure 4 shows that the cumulation of $\sigma_{\rm r}^2(d)$ becomes larger at a faster pace than that of $\sigma_{\rm t}^2(d)$. This indicates that when the first several entries of the feature vector are selected as the axes of the visualization plane, dynamically applying principal component analysis to the retrieved images is more effective in decentralizing them.

To evaluate this observation quantitatively, we computed the ratio of the cumulation of $\sigma_{\rm r}^2(d)$ to that of $\sigma_{\rm t}^2(d)$ over the three largest eigenvalues (Table 1). Table 1 shows that dynamically applying principal component analysis to the retrieved images is about twice as effective in decentralizing them as applying principal component analysis in advance to all the images in the database. It also shows that for a given cumulative contribution, the dimensions for $\sigma_{\rm t}^2(d)$ have to be accumulated about three times more than those for $\sigma_{\rm r}^2(d)$. This indicates that our method has the ability to compress information about three times more than the method where principal component analysis is applied in advance to all the images in the database.

On the other hand, Fig. 5 shows that $\sigma_{\rm r}(d)^2/\sigma_{\rm t}(d)^2$ is 2.38, 2.42, and 1.82, respectively, for $d = 1, 2$, and 3. This indicates that when the first k entries of feature vectors are used as the axes of a k-dimensional visulization space ($k = 1, 2$ and 3), our method decentralizes the retrieved images about twice as well as the method where principal component analysis is applied in advance to all the images in the database. Accordingly, users can more easily capture the special characteristics of the retrieved image collections. We see that the mean of $\sigma_{\rm r}(d)^2/\sigma_{\rm t}(d)^2$ is at least 1.0 for $d = 1, 2, \ldots, 17$, while for $d = 18, \ldots, 72$ (except for $d = 19, 20, 21$, and 24), it is smaller than 1.0. Since $\sigma_{\rm r}(d)^2/\sigma_{\rm t}(d)^2 \geq 1.0$ implies that the retrieved images are more decentralized over the visualization plane, our method displays the retrieved image collections more effectively on the plane when its axes are selected from the first 17 entries of the feature

Table 1: Comparison of the cumulative contribution between σ_r^2 and σ_t^2 over the three largest eigenvalues.

cumulative dimensions k	1	2	3
$\sum_{d=1}^{k} \sigma_r^2(d) / \sum_{d=1}^{k} \sigma_t^2(d)$	2.38	2.37	2.15
cumulative dimensions for σ_t^2	3	6	9

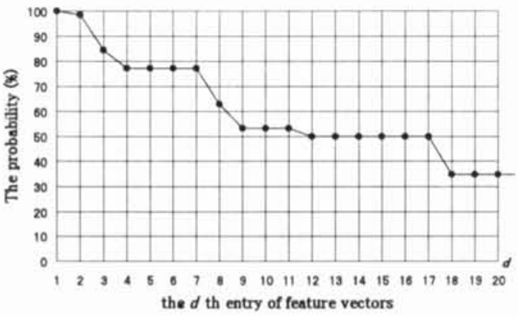


Figure 6: The probability that our method is superior with respect to the first d entries of the feature vectors.

vector. As a result, our method effectively visualizes the retrieved image collections for users.

When we shift our attention to the lower standard deviations of $\sigma_r(d)^2/\sigma_t(d)^2$ in Fig. 5, we see that the lower standard deviation is greater than 1.0 for $d = 1, 2$, though it is smaller than 1.0 for $d \geq 3$ (except for $d = 5, 6, 7, 10$ and 13). If we regard the cumulative contribution ratio as a random variable and suppose that it is subject to the normal distribution, this observation indicates that our method is superior with a probability greater than 84% for $d = 1$ and 2.

By countering the cases of $\sigma_r(d)^2/\sigma_t(d)^2 > 1.0$ for the 700 iterated retrievals, we computed the probability that our method is superior with respect to the first d dimensions (Fig. 6). Fig. 6 shows that, for example, if any two are selected from the first 7 dimensions as the axes of the visualization plane, our method decentralizes the retrieved images with a probability of at least 77% more effectively than the other method.

4 Conclusion

We presented a method for displaying retrieved images in image retrieval where the system dynamically applies principal component analysis to the retrieved image collections and then decentralizes them over a visualization plane whose axes are selected from eigenvectors. By using statistics, we compared our method with the method where principal component analysis is applied in advance to all the images in the database, and we found that:

- the cumulative contribution ratio of our method is about twice more than that of the other method.
- our method can compress information about three times as much as the other method.
- we only have to provide users with the first 17 (2) dimensions so that our method is superior to the other method with a probability of at least 50% (84%). (In this case, users can select any two of the provided dimensions as the axes of the visualization plane.)
- we computed the probability that our method is superior with respect to the first d dimensions ($d = 1, 2, \ldots, 20$) (Fig. 6). (This computation enables us to provide the axes of the visualization plane with the probability shown in Fig. 6. Any selection of the two axes among them makes our method superior to the other method.)

These results indicate that our method decentralizes retrieved images better than the other method and that our method is more effective in users' understanding of the special characteristics of the retrieved image collections.

In our experiments, we used one given image database. This implies that the obtained results depend on the database. To make the results independent of the database, we have to evaluate many databases and investigate the statistical properties of the results. Though this should be done in our future work, analyzing statistics of the retrieved results even for one large-scale database is significant. Improving the functions and operations of our interface is also included in our future work.

References

[1] D. S. Faulus and R. T. Ng: An Expressive Language and Interface for Image Querying, *Machine Vision and Applications*, 10 (1997), pp. 74–85.

[2] M. Flickner, H. Sawhney, *et al.*: Query by Image and Video Content: The QBIC System, *IEEE Computer*, 28 (1995), 9, pp. 23–32.

[3] A. Gupta and R. Jain: Visual Information Retrieval, *Communications of the ACM*, 40 (1997), 5, pp. 71–79.

[4] S. Harada, H. Tanaka, *et al.*: On Constructing Pictorial Feature Space For Image Retrieval (in Japanese), *Technical Report of IEICE*, NLC95-31, pp. 7–12, 1995.

[5] T. P. Minka and R. W. Picard: Interactive Learning Using a 'Society of Models', *Procs. of CVPR*, pp. 447–452, 1996.

[6] Y. Musha, Y. Mori and A. Hiroike: Visualizing Feature Space for Image Retrieval (in Japanese), *Procs. of the 3rd Symposium on Intelligent Information Media*, pp. 301–308, 1997.

[7] Y. Sakurai, H. Iwasa, *et al.*: Face Recognition by Eigen-space Method Using Cylindrical Range Data (in Japanese), *Technical Report of IEICE*, PRU95-193, pp. 22–28, 1996.

[8] A. Stricker and A. Dimai: Spectral Covariance and Fuzzy Regions for Image Indexing, *Machine Vision and Applications*, 10 (1997), pp. 66–73.

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

[10] Y. Uehara and T. Kakimoto: 3D Space Browsing for Multimedia Information Retrieval (in Japanese), *Procs. of the 2nd Symposium on Intelligent Information Media*, pp. 221–228, 1996.