Spatial Visualization for Content-Based Image Retrieval

Baback Moghaddam
Qi Tian
Thomas Huang

TR2001-53   February 2002

Abstract

In traditional content-based image retrieval (CBIR), the retrieved images are displayed in order of decreasing similarities from the query and can be considered as a 1-D display. In this paper, a novel optimized technique is proposed to visualize the retrieved images not only in order of their decreasing similarities but also according to their mutual similarities visualized on a 2-D screen. Principle Component Analysis (PCA) is first performed on the retrieved images to project the images from the original high dimensional feature space to 2-D screen. The result of PCA analysis is denoted as a PCA Splat. To minimize the overlap between images, a constrained nonlinear optimization approach is used. The experimental results show a more perceptually intuitive and informative visualization of the retrieval results. The proposed technique not only provides a better understanding of the query results but also aids the user in forming a new query.

Appears in: International Conference on Multimedia & Expo (ICME'01), Tokyo, Japan, 2001

SPATIAL VISUALIZATION FOR CONTENT-BASED IMAGE RETRIEVAL

Baback Moghaddam[1], Qi Tian[2], Thomas S. Huang[2]

[1]Mitsubishi Electric Research Laboratories, Cambridge, MA 02139, USA
[2]Beckman Institute, University of Illinois, Urbana-Champaign, IL 61801, USA

ABSTRACT

In traditional content-based image retrieval (CBIR), the retrieved images are displayed in order of decreasing similarities from the query and can be considered as a 1-D display. In this paper, a novel optimized technique is proposed to visualize the retrieved images not only in order of their decreasing similarities but also according to their mutual similarities visualized on a 2-D screen. Principle Component Analysis (PCA) is first performed on the retrieved images to project the images from the original high dimensional feature space to 2-D screen. The result of PCA analysis is denoted as a PCA Splat. To minimize the overlap between images, a constrained nonlinear optimization approach is used. The experimental results show a more perceptually intuitive and informative visualization of the retrieval results. The proposed technique not only provides a better understanding of the query results but also aids the user in forming a new query.

1. INTRODUCTION

With the advances in technology to generate, transmit and store large amounts of digital images and video, research in content-based image retrieval (CBIR) has gained more attention recently. In CBIR, images are indexed by their visual contents such as color, texture, etc. Many research efforts have been made to extract these low level features [1, 2, 3], evaluate distance metrics [4, 5] and look for efficient searching schemes [6, 7].

Traditional image retrieval systems display the returned images as a list, sorted by decreasing similarity to the query. The traditional display has one major drawback. The images are ranked by similarity to the query, and relevant images can appear at separate places in the list. Often the user would like to have a global view of the returned images in a way that reflects the relations among the images in the returned set. Only having an idea of surroundings can offer an indication of where to go next. The wider the horizon, the more efficient the new query will be formed. In [8], Rubner proposed a 2-D display technique based on multi-dimensional scaling (MDS) [9]. A global view of the images is achieved that reflects the relations among the images in the retrieved images.

MDS is a nonlinear transformation that minimizes the stress between high dimensional feature space and low dimensional display space. However, MDS is rotation invariant, non-repeatable, and especially slow to implement. These drawbacks make MDS unattractive for real time browsing or visualization of images.

In this paper, we propose an alternative 2-D display scheme based on Principle Component Analysis (PCA) [10]. Moreover, a novel optimized technique is proposed based on the PCA result, which provides a more perceptually intuitive and informative visualization of the retrieved images.

The rest of paper is organized as follows. Our content-based image retrieval system will be discussed in Section 2 and the optimized content-based visualization technique will be presented in Section 3. Results will be shown in Section 4. Summary will be given in Section 5.

2. OUR SYSTEM

Multimedia Analysis and Retrieval System (MARS) is an integrated relevance-feedback architecture for content-based image retrieval. [11, 12]

There are three visual features used in our system: color moments [1], wavelet-based texture [2], and water-filling edge feature [3]. The color space we use is HSV because of its de-correlated coordinates and its perceptual uniformity [1]. We extract the first three moments (mean, standard deviation and skewness) from the three-color channels and therefore have a color feature vector of length $3 \times 3 = 9$.

For wavelet-based texture, the original image is fed into a wavelet filter bank and is decomposed into 10 de-correlated sub-bands. Each sub-band captures the characteristics of a certain scale and orientation of the original image. For each sub-band, we extract the standard deviation of the wavelet coefficients and therefore have a texture feature vector of length 10.

For water-filling edge feature vector, we first pass the original images through an edge detector to generate their corresponding edge map. We extract eighteen (18) elements from the edge maps, including *max fill time, max fork count, etc.* For a complete description of this edge feature vector, interested readers are referred to [3].

Figure 1 shows an example of the retrieved images in MARS system before relevance feedback. The database is a collection of 534 images. The 1st image (building) is the query. The other 9 relevant images are ranked in 2nd, 3rd, 4th, 5th, 9th, 10th, 17th, 19th and 20th places, respectively.
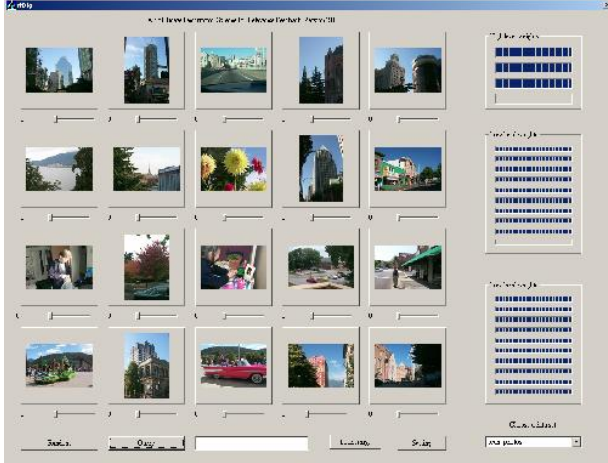
Figure 1 The retrieved images in MARS system (ranked from left to right, from top to bottom, the top left 1 is the query)

3. CONTENT-BASED VISUALIZATION

3.1 PCA Splats

The purpose of our proposed content-based visualization of the retrieved images is augmenting a user's perception so as to grasp a large information space that cannot be perceived by traditional sequential display in rank order of their visual similarities. Traditional sequential display is a simple 1-D visualization. In this section, we propose an alternative technique to MDS in [8] that displays mutual similarities on a 2-D screen based on visual features extracted from images. The retrieved images are displayed not only in ranked order of similarity from the query but also according to their mutual similarities, so that similar images are grouped together rather than being scattered along the entire returned 1-D list of images.

To create such a display, Principle Component Analysis (PCA) [10] is first performed on the retrieved images to project the images from the high dimensional feature space to 2-D screen. Image thumbnails are placed on the screen so that the screen distances reflect as closely as possible the similarities between the images. If the computed similarities from the high dimensional feature space agree with our perception, and if the resulting feature dimension reduction preserves these similarities reasonably well, then the resulting display should be informative and intuitive.

In our experiments, the 37 visual features (9 color moments, 10 wavelet moments and 18 water-filling features) are pre-extracted from the image database and stored off-line. The 37 visual features can be formed into a single feature vector and projected to the 2-D screen based on Principle Component Analysis (PCA). PCA is a very fast linear transformation that achieves the maximum distance preservation from the original high dimensional feature space to 2-D screen among all the linear transformations [10]. This layout is denoted as a PCA Splat.

Figure 2(a) shows an example of PCA Splat for the top 20 retrieved images shown in Figure 1. Clearly the relevant images are close to each other in the PCA Splat while these images are

not consecutive to each other in the 1-D display in order of the decreasing similarities in Figure 1. PCA Splats also convey mutual distance information about all $\binom{n}{2}$ similarities between images while ranked 1-D display in Fig. 1 does not provide this perceptually informative visualization to the user; neither does the 2-D random display in Figure 2(b).
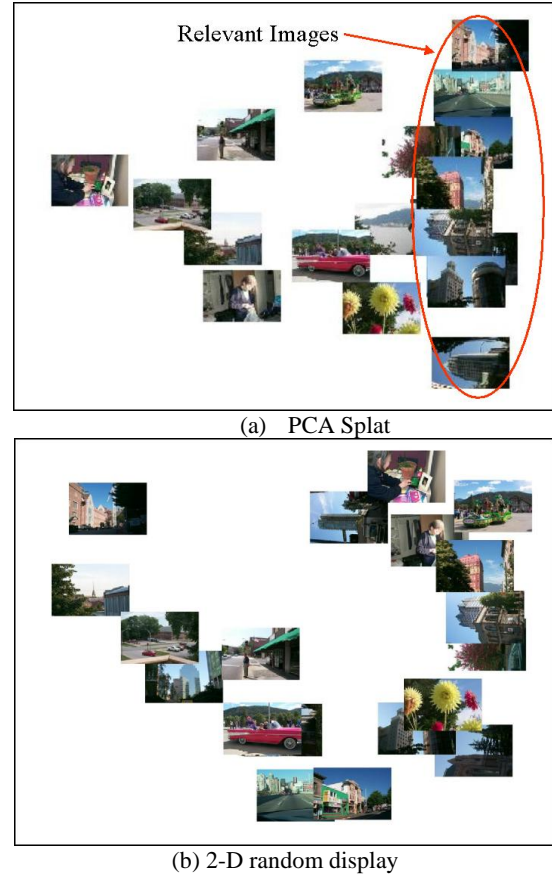

(a) PCA Splat

(b) 2-D random display

Figure 2. Different display schemes

3.2 Optimization

However, one drawback of PCA Splat is that the some images are partially or totally overlapped which makes it difficult to view all the images at the same time. The overlap will be even worse when the number of retrieved images becomes larger, e.g. larger than 50. To solve the overlapping problem between the retrieved images, a novel optimized technique is proposed in this section.

Given the sets of the retrieved images and their corresponding sizes and positions, our optimizer tries to find a solution that places the images at the appropriate positions while deviating as little as possible from their initial PCA Splat positions. Assume the number of images is N. The image positions are represented by their center coordinates (x_i, y_i), $i = 1, \cdots, N$, and the initial

image positions are denoted as (x_i^o, y_i^o), $i = 1, \cdots, N$. The minimum and maximum coordinates of the 2-D screen are $[x_{\min}, x_{\max}, y_{\min}, y_{\max}]$. The image size is represented by its radius r_i for simplicity, $i = 1, \cdots, N$ and the maximum and minimum image size is $r_{\max}$ and $r_{\min}$ in radius, respectively. The initial image size is r_i^o, $i = 1, \cdots, N$.

To minimize the overlap, one can move the images away from each other to decrease the overlap between images, but this will increase the deviation of the images from their initial positions. Large deviation is certainly undesirable because the initial positions provide important information about mutual similarities between images. So there is a trade-off problem between minimizing overlap and minimizing deviation. Without increasing the overall deviation, an alternative way to minimize the overlap is to simply shrink the image size as needed, down to a minimum size limit. The image size will not be increased in the optimization process because this will always increase the overlap. For this reason, the initial image size r_i^o is assumed to be $r_{\max}$.

The total cost function is designed as a linear combination of the individual cost functions taking into account two factors. The first factor is to keep the overall overlap between the images on the screen as small as possible. The second factor is to keep the overall deviation from the initial position as small as possible.

$$J = F(p) + \lambda \cdot Sca \cdot G(p) \tag{1}$$

where $F(p)$ is the cost function of the overall overlap and $G(p)$ is the cost function of the overall deviation from the initial image positions, Sca is a scaling factor which brings the range of $G(p)$ to the same range of $F(p)$ and Sca is chosen to be $(N-1)/2$. λ is a weight and $\lambda \geq 0$. When λ is zero, the deviation of images is not considered in overlapping minimization. When λ is less than one, minimizing overall overlap is more important than minimizing overall deviation, and vice versa for λ is greater than one.

The cost function of overall overlap is designed as

$$F(p) = \sum_{i=1}^{N} \sum_{j=i+1}^{N} f(p) \tag{2}$$

$$f(p) = \begin{cases} 1 - e^{-\frac{u^2}{\sigma_f}} & u > 0 \\ 0 & u \leq 0 \end{cases} \tag{3}$$

where $u = r_i + r_j - \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}$, u is a measure of overlapping. When $u \leq 0$, there is no overlap between the ith image and the jth image, thus the cost is 0. When $u > 0$, there is partial overlap between the ith image and the jth image. When $u = 2 \cdot r_{\max}$, the ith image and the jth image are totally overlapped. σ_f is a curvature-controlling factor.

Figure 3 shows the plot of $f(p)$. With the increasing value of u ($u > 0$), the cost of overlap is also increasing.
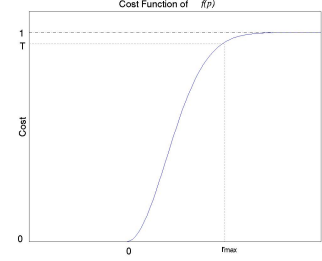

Figure 3. Cost function of overlap function $f(p)$

From Fig. 3, σ_f in Eq. (3) is calculated by setting $T = 0.95$ when $u = r_{\max}$.

$$\sigma_f = \frac{-u^2}{\ln(1-T)} \Big|_{u=r_{\max}} \tag{4}$$

The cost function of overall deviation is designed as

$$G(p) = \sum_{i=1}^{N} g(p) \tag{5}$$

$$g(p) = 1 - e^{-\frac{v^2}{\sigma_g}} \tag{6}$$

where $v = \sqrt{(x_i - x_i^o)^2 + (y_i - y_i^o)^2}$, v is the measure of deviation of the ith images from its initial position. σ_g is a curvature-controlling factor. (x_i, y_i) and (x_i^o, y_i^o) are the optimized and initial center coordinates of the ith image, respectively, $i = 1, \cdots, N$.

Figure 4 shows the plot of $g(p)$. With the increasing value of v, the cost of deviation is also increasing.
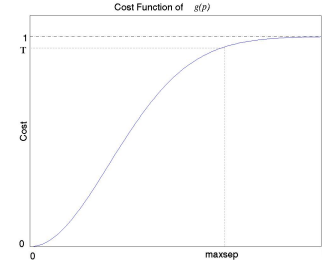

Figure 4. Cost function of function $g(p)$

From Fig. 4, σ_g in Eq. (6) is calculated by setting $T = 0.95$ when $v = maxsep$. In our work, $maxsep$ is set to be $2 \cdot r_{\max}$.

$$\sigma_g = \frac{-v^2}{\ln(1-T)} \Big|_{v=maxsep} \tag{7}$$

The optimization process is to minimize the total cost J by adjusting the size and positions of the images until a global minimum cost is reached. The images will be redisplayed based on the new optimized sizes and positions.

4. RESULTS

Figure 5 (a) shows the optimized PCA Splat for Fig. 2(a). Clearly, the overlapping is minimized while the relevant images are still close to each other to allow a global view. With such a display, the user can see the relations between the images, better understand how the query performed, and subsequently express future queries more naturally.

An improved visualization can also be achieved when this technique is applied to the random display of Fig. 2(b) simply because the overlapping is minimized as shown in Figure 5(b).
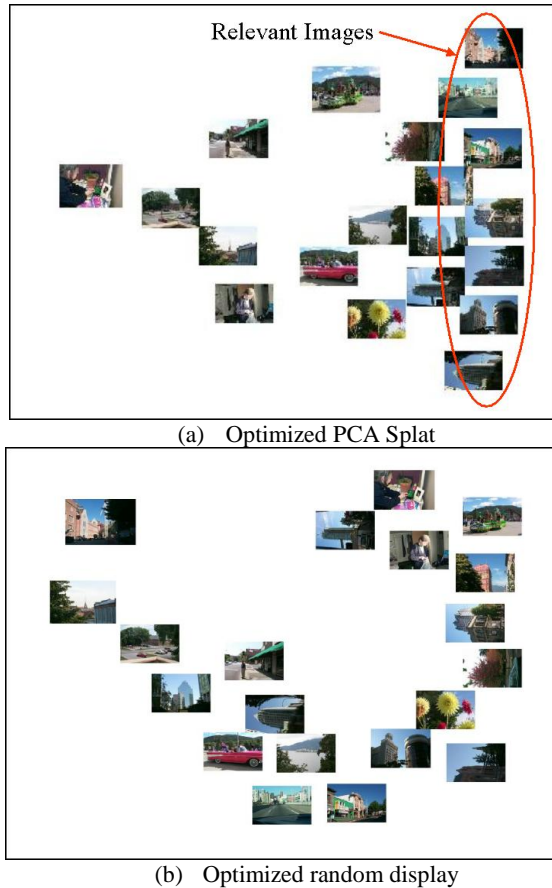

(a)   Optimized PCA Splat

(b)   Optimized random display

Figure 5. Optimized displays

5. SUMMARY

In this paper, we proposed an optimized content-based visualization technique to generate a 2-D display of the retrieved images for content-based image retrieval. The results show a more perceptually intuitive and informative visualization of the retrieval results is achieved by the proposed technique, which not only provides a better understanding for the query results but also helps user forming a new query for the next run.

The proposed content-based visualization method can be easily applied to project the images from high-dimensional feature space to a 3-D space for more advanced visualization and navigation. Features can be multi-modal, expressing individual visual features, e.g., color alone, audio features and semantic features, e.g., keywords, or any combination of the above. The proposed layout optimization technique is also a quite general approach and can be applied to avoid overlapping of any type of images, windows, frames and boxes.

For the future work, relevance feedback can be achieved based on the visualization of the optimized PCA Splat. By grouping the relevant images together, a new user-modeling technique will be proposed.

6. ACKNOWLEDGEMENT

7. REFERENCES

[1] M. Stricker and M. Orengo, "Similarity o Color Images", *Proc. SPIE Storage and Retrieval for Image and Video Databases*, 1995.

[2] J. R. Smith and S. F. Chang, "Transform Features for Texture Classification and Discrimination in Large Image Database", *Proc. IEEE Intl. Conf. on Image Proc.*, 1994.

[3]   S. X. Zhou, Y. Rui and T. S. Huang, "Water-filling algorithm: A novel way for image feature extraction based on edge maps", in *Proc. IEEE Intl. Conf. On Image Proc.*, Japan, 1999.

[4] S. Santini and R. Jain, "Similarity measures", *IEEE PAMI*, vol. 21, no. 9, 1999.

[5] M. Popescu and P. Gader, "Image Content Retrieval From Image Databases Using Feature Integration by Choquet Integral", in *SPIE Conference Storage and Retrieval for Image and Video Databases VII*, San Jose, CA, 1998.

[6] D. M. Squire, H. Müller, and W. Müller, "Improving Response Time by Search Pruning in a Content-Based Image Retrieval System, Using Inverted File Techniques", *Proc. of IEEE workshop on CBAIVL*, June 1999.

[7] D. Swets, J. Weng, "Hierarchical Discriminant Analysis for Image Retrieval", *IEEE PAMI*, vol. 21, no.5, 1999.

[8] Y. Rubner, "Perceptual metrics for image database navigation", Ph.D. dissertation, Stanford University, 1999.

[9] W. S. Torgeson, *Theory and methods of scaling*, John Wiley and Sons, New York, NY, 1958.

[10] Jolliffe, I.T., *Principal Component Analysis*, Springer-Verlag, New-York, 1986.

[11] T. S. Huang, S. Mehrotra, and K. Ramchandran, "Multimedia analysis and retrieval system (MARS) project", in *Proc. of 33rd Annual Clinic on Library Application of Data Processing – Digital Image Access and Retrieval*, 1996.

[12] Y. Rui, T.S. Huang, M. Ortega and S. Mehrotra, "Relevance Feedback: A Power Tool for Interactive Content-Based Image Retrieval", *IEEE Trans. On Circuits and Video Technology, Special Issue on Segmentation, Description, and Retrieval of Video Content*, vol. 8, no. 5, Sept, 1998.