Visualization & Layout for Personal Photo Libraries[1]

Baback Moghaddam[1], Qi Tian[2], Neal Lesh[1], Chia Shen[1], Thomas Huang[2]
[1]Mitsubishi Electric Research Labs (MERL), Cambridge, MA 02139 {baback, lesh, shen}@merl.com
[2]University of Illinois at Urbana-Champaign, Urbana, IL 61801 {qitian, huang}@ifp.uiuc.edu

Abstract. In this paper, we present visualization and layout algorithms that can enhance informal storytelling using personal digital data such as photos in a face-to-face social setting. In order to build a more intuitive browser for retrieval, navigation and story-telling, we introduce a novel optimized layout technique for large image sets which respects (context-sensitive) mutual similarities as visualized on a shared 2-D display (a table-top). The experimental results show a more perceptually intuitive and informative visualization of traditional CBIR-based retrievals, providing not only a better understanding of the query context but also aiding the user in forming new queries. A framework for user-modeling is also introduced and tested. This allows the system to adapt to the user's preferences and relevance feedback.

Keywords. User-guided layout, content-based information visualization and retrieval, multi-person interactive informal storytelling, table top display, visual navigation.

1. Introduction

We often find ourselves using photos to "re-connect" with people, whether it be with our family who we have not seen all day, a friend or colleague whom we have not seen in a year, or our parents who live across the country. In this paper, we present visualization and layout algorithms that can enhance informal storytelling using personal digital data such as photos, audio and video in a face-to-face social setting. These algorithms are designed in the context of the Personal Digital Historian (PDH) system, which supports multi-person interactive informal storytelling by combining and extending research in largely two areas: (1) human-computer interaction and interface (the design of the shared-display devices, user interface for story-telling and on-line authoring, and story-listening), (2) content-based information visualization and retrieval (user-guided image layout, data mining and summarization).

Figure 1. PDH Table.

[1] Appear in *International Workshop on Content-Based Multimedia Indexing*, University of Brescia, Italy September 19-21, 2001

2. Motivation and Interface Design

The research in PDH is guided by the following principles:

(1) The display device should enable natural face-to-face conversation – not forcing everyone to face in the same direction (desktop) or at their own separate displays (hand-held devices).

(2) The physical sharing device must be convenient and customary to use – helping to make the computer disappear.

(3) Easy and fun to use across generations of users – minimizing time spent typing or formulating queries.

(4) Enabling interactive and exploratory storytelling – blending authoring and presentation.

Current software and hardware do not meet our requirements. Most existing software in this area provides users with either powerful query methods or authoring tools. In the former case, the users can repeatedly query their collections of digital content to retrieve information to show someone [1]. In the latter case, a user experienced in the use of the authoring tool can carefully craft a story out of their digital content to show or send to someone at a later time. Furthermore, current hardware is also lacking. Desktop computers are not suitably designed for group, face-to-face conversation in a social setting, while handheld story-telling devices [2] have limited screen sizes and can be used only by a small number of people at once.

We designed our system to work on a touch-sensitive, circular table-top display, as shown in Figure 1. The layout of the entire table-top display, as shown in Figure 2, consists of (1) a large story-space area encompassing most of the table-top until the perimeter, and (2) one or more narrow arched control panels.

There are a number of ways to implement this design. The physical sharing device can be either a touch sensitive table surface, or a rear projected or top projected horizontal table surface with infrared cameras to track hand motions [3]. A portable device with projection and sensing capabilities can also be created so that the virtual table-top can travel with people to places. Our first prototype is constructed as a top projection table with a whiteboard as the table surface. This implementation provides a high resolution as well as a sturdy surface, which serves well as both a display device and a table to set familiar objects such as tea cups on. We will add vision capabilities to the current table set up to enable both object avoidance in image display and multiple simultaneous control.

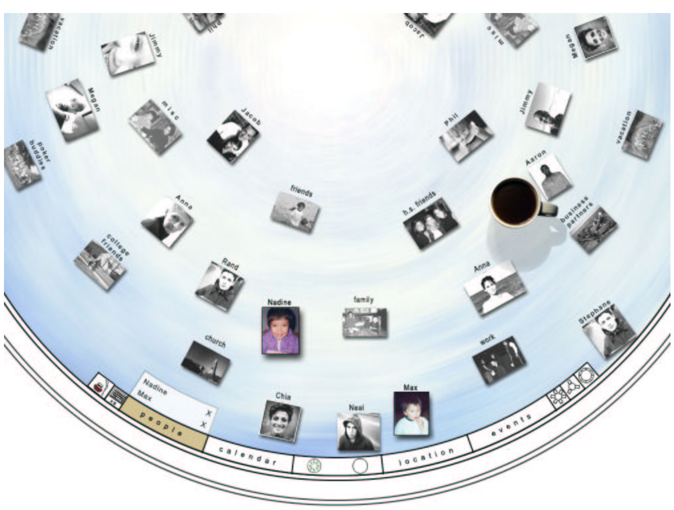


Figure 2: Close-up of "people-view" on the PDH table.

When people use the system they will cause different sets of their pictures to be displayed on the table. The pictures will be displayed in a radial-pattern in the story-space, so that pictures initially will be oriented towards the outer rim of the table. Users have the option of moving or re-orienting pictures using the touch-screen, in a rough simulation of how real pictures would slide along a table. By dragging one's hand along the outer rim of the table, the users can also spin the entire contents of the table, much like a lazy-susan.

The primary method of navigation is organized about four questions essential to storytelling: who?, when?, where?, and what?[2] Control panels located on the perimeter of the table contain buttons labeled "people", "calendar", "location", and "events", corresponding to these four questions. When a user presses the "location" button, for example, the display on the table changes to show a map of the world. Every picture in the database that is annotated with a location will appear as a tiny thumbnail at its location. The user can pan and zoom in on the map to a region of interest, which increase the size of the thumbnails. Similarly, by pressing one of the other three buttons, the user can cause the pictures to be organized by the time they were taken along a linear straight timeline, the people they contain, or the event-keywords which the pictures were annotated with.[3]

The user can form implicit Boolean queries, or filters, by selecting items or regions in the different navigational views. For example, if someone selects two friends in the people view, then (until this selection is retracted) only pictures containing one or both of these friends will be highlighted in subsequent navigation. If the user next selected "location", for example, then they would see where they have traveled with either of these friends by observing where the highlighted pictures appear on the map. If the user selected "calendar" instead, they would see when they have taken pictures with these friends. Another non-menu driven query metaphor used in PDH is invoked when the user presses and holds down a particular picture. The system then offers the user the ability to display pictures taken at a similar time, or a similar place, or with the same people, or at the same event as the selected picture.

At any point, the users can add or remove pictures from an initially blank "story-space". Users can add pictures based on their partial annotations or based on their similarity to other pictures, and can add them one at a time or in large groups. For example, users can use the above method of navigation to select all pictures that include a specified person over a specified time period and then press the "show" button to display all selected pictures in their story-space. In the following sections, we discuss how visualization and layout algorithms can help decide where to place images in the story-space, and can support the user's re-arrangement of pictures within this space.

3. Visualization

Image retrieval will be an important part of the prototype. 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. Rubner [4] proposed a 2-D display technique based on multi-dimensional scaling (MDS) [5]. 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 (or disparity) between high dimensional feature space and its low dimensional display space. However, MDS is valid up to a rotation, generally non-repeatable, computationally intensive and consequently slow to implement. These drawbacks make MDS unattractive for real time browsing or visualization of images.

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 easily 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 [4] 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.

---

[2] We do not currently support "why?, which is also useful for storytelling.
[3] We assume the pictures are *partially* annotated.

3.1 PCA Splats

In our experiments, the 37 visual features (9 color moments [7], 10 wavelet moments [8] and 18 water-filling features [9]) 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 [6]. This visual layout is denoted as a "PCA Splat."

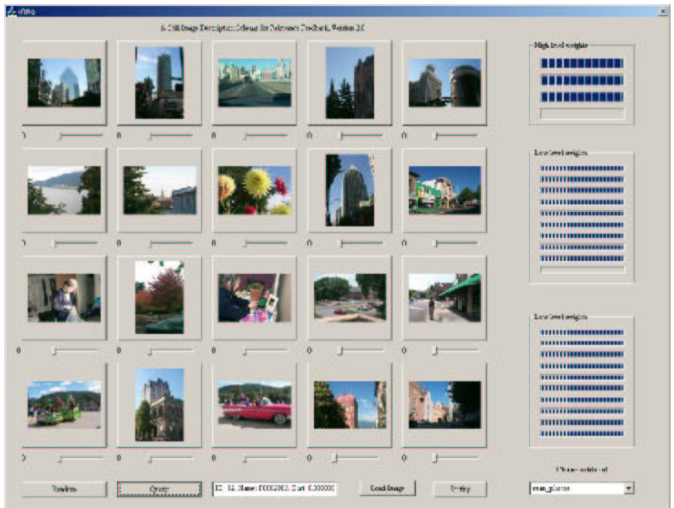
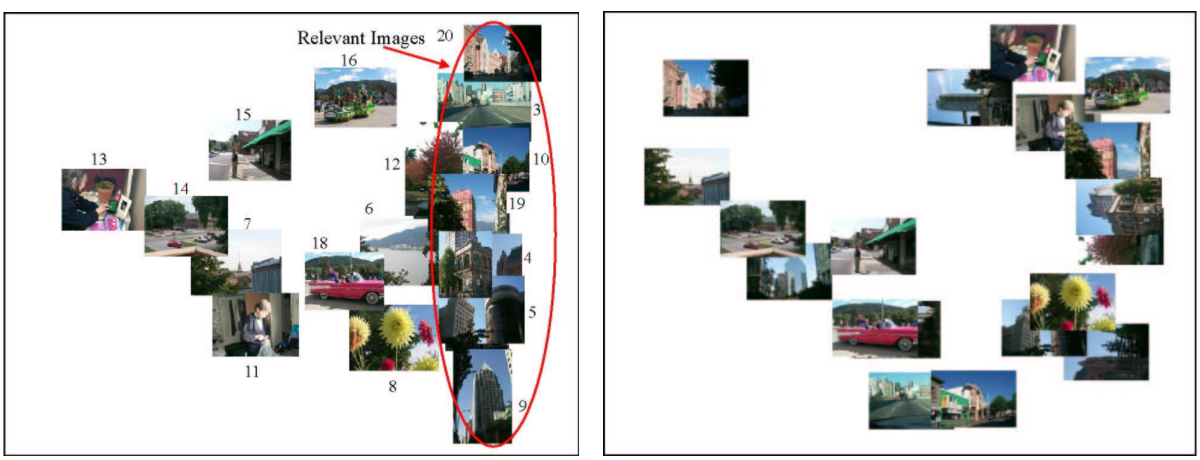


Figure 3. Top 20 retrieved images (ranked top to bottom and left to right; query image is shown first in the list)

Figure 3 shows the top 20 retrieved images given the top left image as the query. Figure 4(a) shows an example of PCA Splats for the top 20 retrieved images shown in Figure 3. The sizes of the images are determined by their visual similarity to the query. The higher the rank, the larger the size. There is a number close to each image in Figure 4(a) indicating its corresponding rank in Figure 3.



(a) PCA Splats

(b) Random layout

Figure 4. Different display/layout schemes

Clearly the relevant images are close to each other in the PCA Splats while these images are not consecutive to each other in the 1-D display in order of the decreasing similarities in Figure 3. PCA Splats also convey mutual distance information about all pair-wise similarities between images while the ranked 1-D display in Fig. 3 does not provide this perceptually informative. Naturally, neither does the 2-D random display in Figure 4(b).

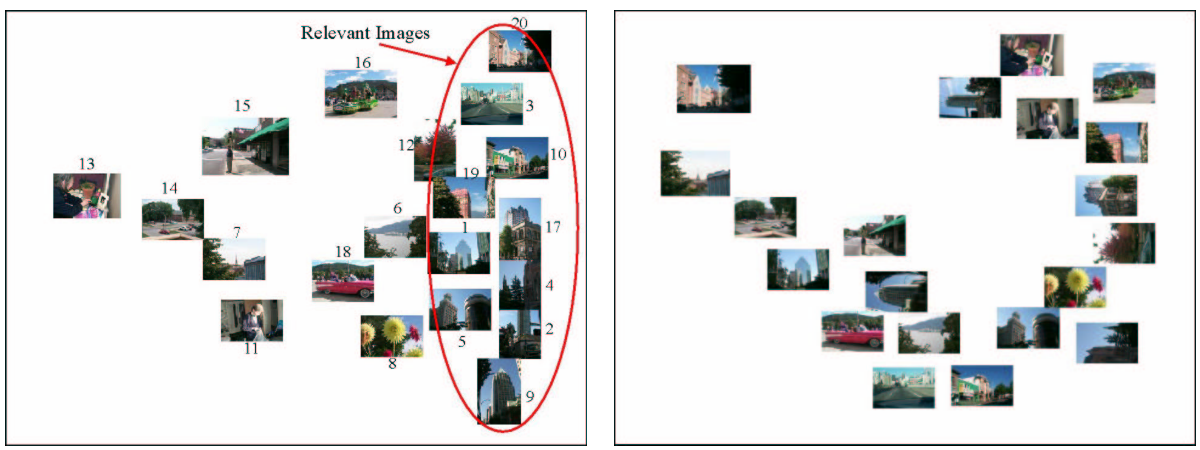
3.2 Layout Optimization

However, one drawback of PCA Splats is that some images are partially or totally overlapped which makes it difficult to view them at the same time. This overlap will be even worse when the number of retrieved images becomes larger, e.g. larger than 50. To solve the overlapping problem an optimization technique is proposed in [10].

Given the sets of the retrieved images and their corresponding sizes and positions, the layout optimizer tries to find a solution that places the images at the appropriate positions while deviating as little as possible from their initial PCA Splats positions. 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.

The total cost function is a linear combination of the individual cost associated with each of the competing factors mentioned above. The optimization process is to minimize the total cost by adjusting the size and positions of the images until a local minimum cost is reached. The images will be redisplayed based on the optimized sizes and positions. Interested readers should refer to [10] for details.

Figure 5 (a) shows the optimized PCA Splats for Fig. 4(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. The number close to each image indicates their rank of visual similarity to the query image. An improved visualization can also be achieved when this technique is applied to the random display of Fig. 4(b) simply because the overlapping is minimized as shown in Figure 5(b).



(a) Optimized PCA Splats (b) Optimized random display

Figure 5. Optimized displays

4. User Modeling

We propose a novel scheme to generate or mimic user layouts of personal photos. The layout can be on the table-top display, part of the PDH device (Figure 1). Given information from the layout, i.e., positions and mutual distances between images, a novel feature weight estimation scheme, noted as α-estimation is proposed based on a non-negative least-squares optimal solution. The algorithm will mimic the original (user) layout using the learned feature weights. These features can be visual features, audio features, semantic features or any combinations of them.

Figure 6(a) shows an example of a layout. It is a PCA Splats of the images with their high dimensional features weighted by some (unknown) α. Figure 6(b) shows the corresponding computer-generated layout. It is also a

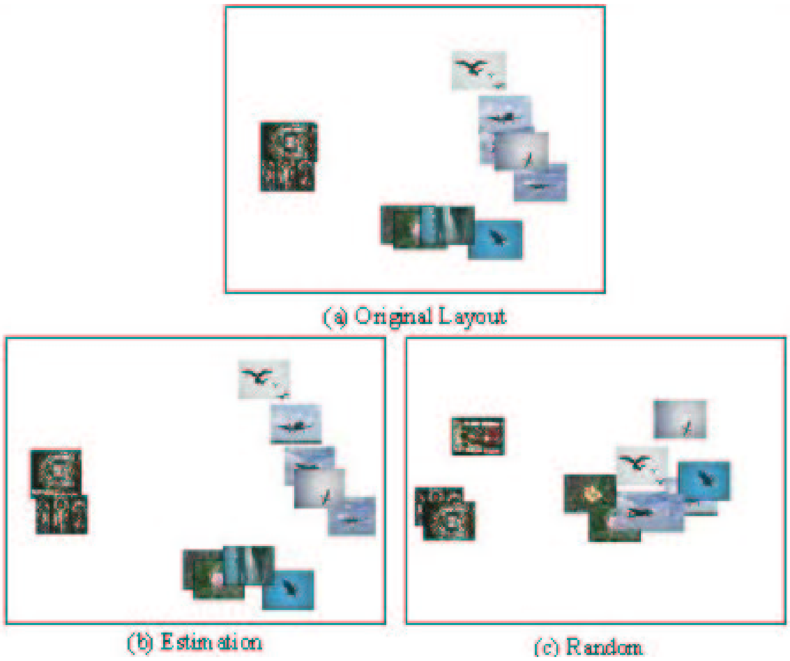
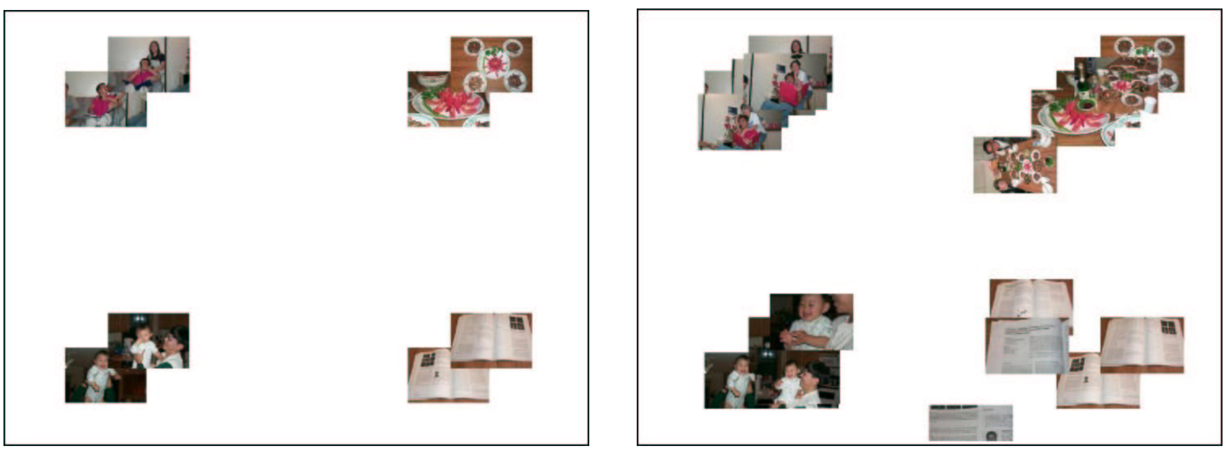


Figure 6. (a) An example user layout, (b) Computer-generated layout based on the learned feature weights, and (c) randomly generated layout (for comparison).



(a) a user-guided layout: 8 photos (4 classes and 2 photos from each class)

(b) computer layout for larger set of photos

Figure 7: User-modeling for automatic layout

PCA Splats of the images but with their high dimensional feature vectors weighted by the estimated α, which is determined solely from the configuration of Fig. 6(a). For comparison, Figure 6(c) shows the PCA Splat of the same images with their high dimensional feature vectors weighted by some random number (arbitrary α).

Figure 7(a) shows an example of user-guided layout. Assume that the user is describing her family story to a friend. In order not to disrupt the conversation flow, she only lays out a few photos from her personal photo collections and expects the computer to generate a similar and consistent layout for a larger set of images from the same collection. Figure 7(b) shows the computer-generated layout based on the learned feature weights from the configuration of Fig. 7(a). The computer-generated layout is achieved using the α-estimation scheme and post-linear, e.g., affine transform or non-linear transformations. Only the 37 visual features (9 color moments [7], 10 wavelet moments [8] and 18 water-filling features [9]) were used for this PCA Splat. Clearly the computer-generated layout is similar to the user layout with the visually similar images positioned at locations.

We have conducted a preliminary user study that has shown the superior performance of the proposed α-estimation over the random weighting as a control (or sanity check). Table 1 shows the user test results. There are two tests. The first one is to test whether the original feature weighting will generate a better layout than the control (i.e., random weightings) which is denoted as "α-matters." The second is to test whether the estimated feature weights will generate a better layout than random weightings, and is denoted as α-estimation matters. 12 naïve users were instructed in the basic rules for the user test and asked to perform either the α-matters and α-estimation-matters tests. The users were given the following instructions: (1) Both absolute and relative positions of images matter (2) In general, similar images, like cars, tigers, should cluster. (3) The relative positions of clusters matter. Each user performed a single forced-choice test only.

The average percentage of the vote rate for the original weight in α-matters test is 84.6% and user's consistency rate is 91.7%. The average percentage of the vote rate for the estimated weights in α-estimation-matters test is 95.8% and the user's consistency rate is 96.7% - indicated that the estimated weights are indeed favored over the random weights. (50% would indicate no difference or preference for estimation vs. random).

α-matters	Vote rate for Original weights	Vote rate for Random weights	User Consistency
User 1	82.5%	17.5%	90%
User 2	77.5%	22.5%	90%
User 3	85%	15%	90%
User 4	92.5%	7.5%	100%
User 5	80%	20%	80%
User 6	90%	10%	100%
Average	84.6%	15.4%	91.7%

(a) α-matters test

α-estimate matters	Vote rate for Original weights	Vote rate for Random Weights	User Consistency
User 1	90%	10%	100%
User 2	97.5%	2.5%	90%
User 3	97.5%	2.5%	90%
User 4	95%	5%	100%
User 5	97.5%	2.5%	100%
User 6	97.5%	2.5%	100%
Average	95.8%	4.2%	96.7

(b) α-estimation-matters test

Table 1. User study results.

5. Discussion

The PDH project is at its initial stage. We have just begun our work in both the user interface design and photo visualization and layout algorithms. Many interesting questions still remain as our future research in the area of content-based information visualization and retrieval. The next task is to carry out an extended user-modeling study by having our system learn the feature weights from various sample layouts provided by the user. We have already developed a framework to incorporate visual features with semantic labels for both retrieval and layout. Incorporation of relevance feedback in our framework seems very intuitive and is currently being explored. Another challenging area is automatic "summarization" and display of large image collections. Since summarization is implicitly defined by user preference, α-estimation for user-modeling will play a key role in this and other high-level tasks where context is defined by the user.

Acknowledgement

This work was partially supported by Mitsubishi Electric Research Laboratory (MERL), Cambridge, MA and NSF Grant CDA 96-24396, EIA-99-75019.

References

1. Kang, H., Shneiderman, B., "Visualization Methods for Personal Photo Libraries: Browsing and Searching in the PhotoFinder", in *Proceedings of IEEE International Conference on Multimedia and Expo* (ICME2000), New York City, New York.
2. Balabanovic, M., Chu, L., Wolff, G., "Storytelling with Digital Photographs", *in the Proceedings of CHI 2000,* (April, 2000. The Hague, The Netherlands).
3. Leibe, B., Starner, T., Ribarsky, W., Wartell, Z., Krum, D., Singletary, B., "The Perceptive Workbench: Toward Spontaneous and Natural Interaction in Semi Immersive Virtual Environments", in *Proceedings of IEEE Virtual Reality 2000* (New Brunswick, NJ, March 2000).
4. Y. Rubner, "Perceptual metrics for image database navigation", Ph.D. dissertation, Stanford University, 1999.
5. W. S. Torgeson, *Theory and methods of scaling*, John Wiley and Sons, New York, NY, 1958.
6. Jolliffe, I.T., *Principal Component Analysis*, Springer-Verlag, New-York, 1986.
7. M. Stricker and M. Orengo, "Similarity of Color Images", *Proc. SPIE Storage and Retrieval for Image and Video Databases*, 1995.
8. 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.
9. 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.
10. Q. Tian, B. Moghaddam, Thomas S. Huang, "Display Optimization for Image Browsing", will appear in *2nd International Workshop on Multimedia Databases and Image Communications (MDIC'01)*, Sep. 17-18, 2001, Amalfi, Italy.