Natural Video Browsing*

Cai-Zhi Zhu, Tao Mei
Department of EEIS
University of Science and Technology of China
Hefei 230027, P. R. China
zhucz@ustc.edu,meit@mail.ustc.edu.cn

Xian-Sheng Hua
Microsoft Research Asia
5F Sigma Center, 49 Zhichun Road
Beijing 100080, P. R. China
xshua@microsoft.com

ABSTRACT

In this demonstration, we show a novel system, Video Booklet, which enables nature personal video browsing and searching. Firstly representative thumbnails of video segments are selected and reshaped by a set of pre-trained personalized shape templates, and then printed out on a real booklet. When we want to watch the segment indicated by a certain thumbnail in the booklet, we are able to use camera phones or similar devices to capture the corresponding thumbnail, and send it to a computer via wireless network. Thereafter, the target thumbnail is accurately located by a Self-Trained Active Shape Model algorithm, and then the distortion of the captured image is corrected. Finally the Video Booklet system will automatically find the most similar thumbnail to the corrected one and begin to play the corresponding segment in the video library for us. Thereby, Video Booklet builds a seamless bridge between digital videos and analog albums.

Categories and Subject Descriptors: H.5.1 [Information Interfaces and Presentation]: Multimedia Information Systems – *video*

General Terms: Design, Experimentation.

Keywords: Video retrieval, Album, Active shape models.

1. INTRODUCTION

The quantity of multimedia data, in particular, home videos, is increasing dramatically in recent years with the popularity of digital camcorders. Unlike text data, which is much easier to be indexed and randomly accessed, it is well-known media data is difficult to be well indexed and efficiently accessed. It is time-consuming, as well as not convenient, for general users to search and browse their personal media data.

In this demonstration, we show a novel system, Video Booklet, which supports nature and efficient video library browsing and searching. As illustrated by Figure 1, the system consists of two sub-systems: Video Booklet Generation and Booklet-Based Video Browsing and Searching. In the first sub-system, videos in the library are segmented into scenes, shots and sub-shots, as well as the signature and a set of features are extracted from these segments [3] [2].

*This work was performed while Cai-Zhi Zhu and Tao Mei were visiting Microsoft Research Asia as research interns.

MM'05, November 6–11, 2005, Singapore.
ACM 1-59593-044-2/05/0011.

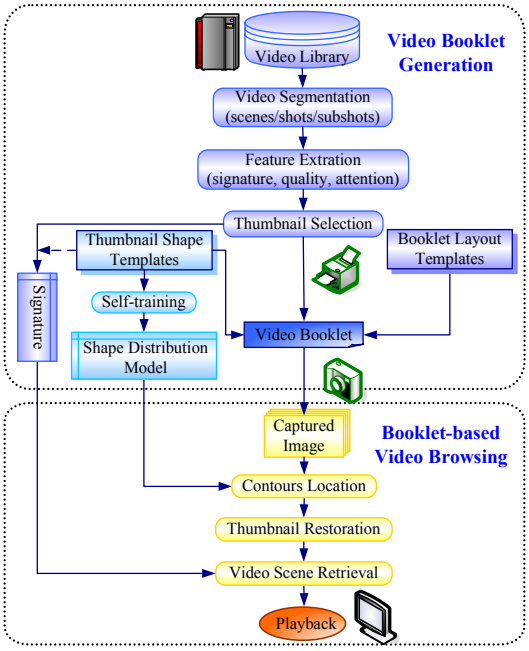

Figure 1: Flow chart of Video Booklet

Then based on these features, the temporal structure, and selected booklet layout template and thumbnail shape (contour) template, a video booklet is generated and printed out using a typical LaserJet printer (color or black and white) or developed into a real photo album. In the other subsystem, when the users want to browse their digital video library, they can firstly browse the corresponding booklets in a manner as browsing ordinary family albums. When they want to watch the segment indicated by a certain thumbnail in the booklet, they can use their camera phones to capture the corresponding thumbnail in a leisure manner. Then the captured image is sent to the computer or other device connected to the monitor via wireless network, and the Video Booklet system will automatically find the corresponding segment in the video library for the users and begin to play it. Video Booklet builds a bridge towards seamless communication between digital videos and analog albums.

2. BOOKLET GENERATION

Video booklet for a collection of video data is generated based on a thumbnail selection scheme, layout templates,

Figure 2: Shape template samples

and thumbnail shape templates. The thumbnail selection scheme helps choose the best set of thumbnails to be printed out on the album, which maximizes the signature differences, visual quality and representativeness [3].

Layout templates define the overall look-and-feel and the thumbnail layout of the to-be-printed booklet. With the layout templates, for the same selection of thumbnails, we can generate different sets of booklets in different forms. The layout templates are described by XML and can be shared among different users.

To generate more impressive and personalized video booklet, thumbnail shape templates are supported, which enable users to reshape the selected thumbnails into flexible forms as illustrated by Figure 2.

As it is difficult to precisely locate the reshaped contours of the thumbnails in the captured degenerated images, especially when they are mixed with the background image, we adopted a method based on *Active Shape Models* (ASM) [1] to accomplish this goal. For typical ASM-based approaches, manual labeling of a large number of training data is generally required to learn the *Shape Distribution Models* (SDM) applied in the subsequent contour locating process. In this system, a *self-training* scheme for ASM is proposed, which provides a substitute for this typical labeling process and is able to obtain precise SDM for the booklet shape templates. With this self-training method, we are able to support a large number of shape templates in Video Booklet without any manual labeling. The system also supports personalized shape templates defined by users. To add a new template, users only need to draw a closed curve, and then the self-training scheme will automatically extract landmark points, generate training set, apply *Principal Component Analysis* (PCA), and obtain the SDM.

With the booklet shape templates, we can generate a set of reshaped/cropped thumbnails by applying these templates on the selected thumbnails.

3. BROWSING WITH VIDEO BOOKLET

In this section, we show how to accomplish nature and efficient video browsing/searching through the booklet generated in Section 2. The primary step is to locate the actual contour of the captured thumbnail, and then to restore the detected thumbnail area to its original shape by perspective transformation. Finally the thumbnail with the most similar signature in the video library is retrieved and the corresponding scene represented by the retrieved thumbnail is played back by the system.

In our previous work [3], a hybrid "bottom-up" approach is proposed to locate contours of captured thumbnails. This traditional approach detected local structures such as edges, assembled them into groups in an attempt to identify the objects of interest. Without a predefined model, this approach often fails to locate more complex shapes, such as thumbnails mixed with the background or badly degenerated.

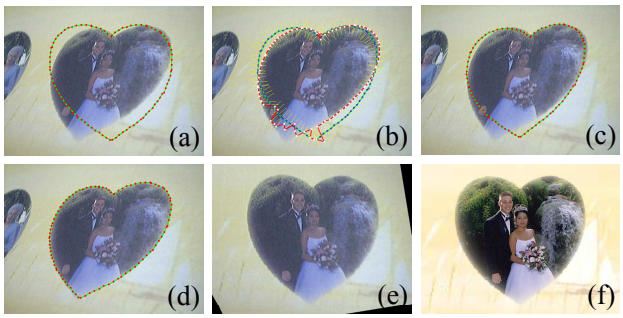

Figure 3: ASM searching process: (a) an initial approximation; (b) adjusted landmark points; (c) fitted shape template; (d) final located result; (e) restored thumbnail image; (f) original thumbnail before printing.

In this paper, a "top-down" approach is designed to tackle more general cases. Generally, the shape template of a printed thumbnail is either known or constrained to a predefined template set. Although the appearance of the shape contour may be blurred by noise and distorted due to change of viewing position, the contour prototype is often unchanged. So the most straightforward approach to locate the shape contour is to predefine an exhaustive shape template set, and then to select the best one matched to the test image while taking noise and distortion into account. Therefore, as aforementioned, the deformable models algorithm, ASM proposed by Cootes *et al.* [1], is adopted in our system.

Taking the advantage of SDM generated by the self-training scheme, ASM can be used both for boundary location and template classification. An example of ASM searching process is illustrated by Figure 3.

Finally, the signature of the restored image is extracted and applied to find the best match in the signature database. Thereafter, the most similar thumbnail in the video library is retrieved and the system will play the video from the beginning of the scene that the thumbnail represents.

4. CONCLUSION

In this demo, we have shown a novel system to browse and search personal video clips, named Video Booklet, which builds a seamless bridge between digital videos and analog albums. An ASM based method, including a novel self-training scheme, is adopted to enable generating impressive video booklet.

There are a number of extensions for Video Booklet. First, similarity search might be applied according to the gray or color characteristics. Thus users are able to find similar clips in the video library. Second, the system is also applicable for photo library. This also enables quick and convenient access of the photo collections.

5. REFERENCES

[1] T. F. Cootes, C. Taylor, D. Cooper, and J. Graham. Active shape models - their training and their applications. *Computer Vision and Image Understanding*, 61(1):38–59, Jan. 1995.

[2] X.-S. Hua, X. Chen, and H.-J. Zhang. Robust video signature based on ordinal measure. In *Proceedings of ICIP*, Singapore. October 24-27. 2004.

[3] X.-S. Hua, S. Li, and H.-J. Zhang. Video booklet. In *Proceedings of ICME*, Amsterdam, Netherlands. July 6-8. 2005.