Visualization of Multi-Video Summaries

Yingbo Li, Feng Wang and Bernard Merialdo
EURECOM
BP 193, 06904 Sophia Antipolis, France
E-mail: {Yingbo.Li, Feng.Wang, Bernard.Merialdo}@eurecom.fr

Abstract

In this paper, we describe two visualization approaches for multi-video summaries. Visualization of video summary is important for discovering the relations among the frames of different videos. And it is useful in evaluating the video summary and comparing the summaries from different algorithms. In one of our approaches, we exploit a circle space to visualize the relations between multi-video frames. While in the second approach, we illustrate a summarization method based on hierarchical clustering and present its corresponding visualization. Both presentations are used in actual experiments for producing summaries from a set of videos.

1 Introduction

Digital video documents are widely used nowadays. Many technologies are developed by researchers to classify, analyze and summarize the videos. Video summarization [3] [5] [6] is a popular topic in recent years. Some research has focused on multi-document [4] [7], but only little work has been proposed for multi-video [8]. In particular, how to visualize multi-video summaries has not been perfectly resolved yet. In this paper, we investigate some visualization approaches for multi-video summaries. Two visualization methods are proposed. One uses a circle space to demonstrate the relations between multi-video frames, and the other one shows the dynamic construction of a multi-video summary based on hierarchical clustering [1] [2] [9]. We implement our algorithms on set of videos corresponding to various news items from the website "Wikio.com". This site gathers news articles from various sources, so that different descriptions and comments about the same event can be browsed at once.

In the rest of paper, we first present the principles of multi-video visualization based on circle space, and then present an illustration of the result. In Section 3, we explain the algorithm for multi-video summarization based on hierarchical clustering, which is followed by an experimental result. Finally, we finish the paper with the conclusion.

2 Multi-Video Visualization

We explain the principle of circle visualization of multi-video frames in Section 2.1, and the experimental result in Section 2.2.

2.1 Principle of Multi-Video Visualization

Assume that we have a group of videos, $\{V_1 \ldots V_m \ldots V_n\}$. For all the videos, we can compute the similarity value between every video and every frame of this group. It is described as following:

$$SIM(f, V_m) = \max_{k \in V_m}\{SIM(f, k)\} \tag{1}$$

where f is a frame from video group $\{V_1 \ldots V_m \ldots V_n\}$, and V_m a video in this group. So the similarity value between frame f and video V_m, $SIM(f, V_m)$, is considered as the maximum similarity value between frame f and every frame k of video V_m. In the proposed algorithm, cosine value between the HSV histogram of frame f and HSV histogram of video frame k is used as the similarity value, $SIM(f, k)$.

$$SIM(f, k) = \cos(X, Y) = \frac{X \cdot Y}{\|X\|\|Y\|} \tag{3}$$

where X (resp. Y) is the HSV histogram of frame f (resp. k). A circle is used to visualize the relations of videos and frames. The videos are averagely placed at regular intervals on the boundary of the circle. The frames from different videos are displayed as points inside the circle, at a position which is related to their similarities with the videos. The coordinates of frame points in the circle are computed as:

$$P_f = \sum_{m=1}^{n} P_m \cdot SIM(f, V_m) \tag{6}$$

where P_f is the position of the frame f, P_m is the position of video m, and $SIM(f, V_m)$ is defined in Equation (1). With this formula, the more similar a frame is to a video, the closer its location inside the circle is to the location of the corresponding video.

2.2 Experimental Example: Final Frame of Demo Video

The final frame of demo video is displayed as Figure 1. Points with the same color indicate frames from the same video. The comparison of the color point location allows to quickly have a look at the similarities between videos. Points that are close to the center of the circle correspond to frames that belong to the general visual ambiance of the video set. On the contrary, points which are near the boundary will correspond to frames that are specific to one particular video.

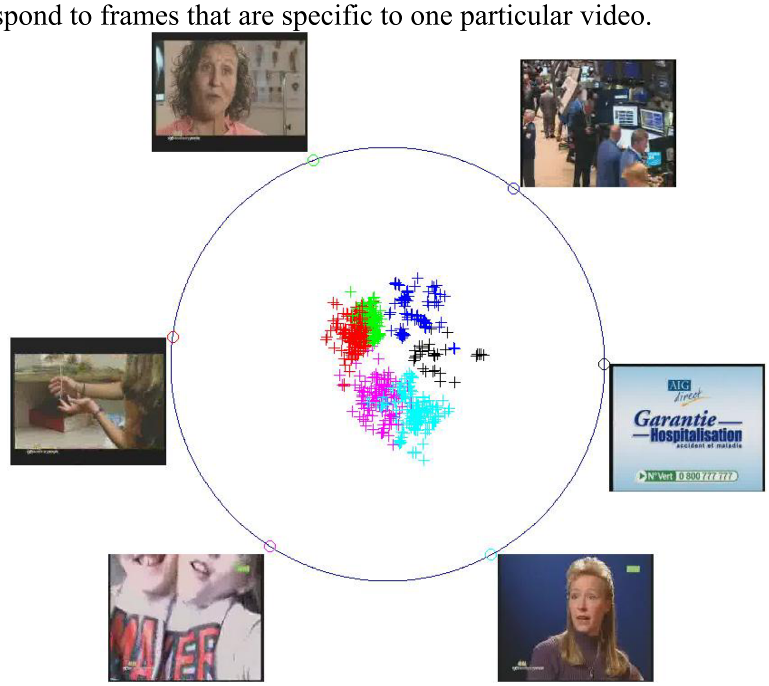


Figure 1: Example of Multi-Video Visualization

3 Multi-Video Summarization

We now explain the principle of multi-video summarization based on hierarchical clustering in Section 3.1, and the illustration of experimental result is in Section 3.2. This is based on our earlier work on single video summarization [1].

3.1 Multi-Video Summarization Based on Hierarchical Clustering

We subsample the frames from the video group, $\{V_1 \ldots V_m \ldots V_n\}$ at the rate of one per second, and cluster this set using a hierarchical clustering. Initially each frame is a cluster. At each step of the clustering, the number of clusters is reduced by one through merging the closest two clusters. The hierarchy is completely built until only all frames are in the same cluster, as illustrated in Figure 2. The distance between two frames is the Euclidean distance of HSV histogram, like Equation (4). The distances between clusters are the average distances of all the possible frame pairs in the clusters, like Equation (5).

$$DIS(f,s) = \sqrt{\sum_{i=1}^{j}(X_i - Y_i)^2} \tag{4}$$

where *f* and *s* are frames from video group. *X* is HSV histogram of frame *f*, and *Y* is HSV histogram of frame *s*. The argument *j* is the total number of tonal variations of histogram.

$$DIS_{cluster} = \frac{\sum(DIS_{each\ pair})}{number\ of\ pairs} \tag{5}$$

If we want to construct a summary of N seconds, we just select the hierarchy level which has N clusters and concatenate representative video segments for every cluster of this level.

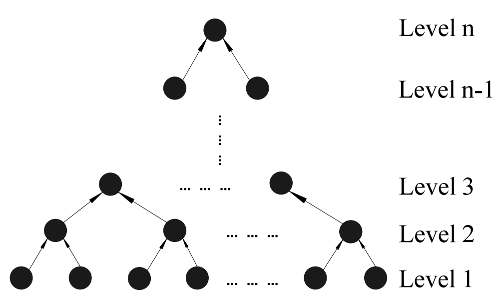
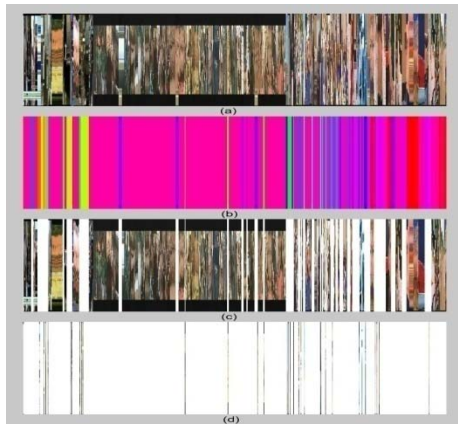


Figure 2: Clustering Hierarchy

3.2 Experimental Example

Figure 3 is a snapshot of the demonstration which illustrates the procedure of selecting the relevant frames for the summary. This figure is explained as following: Figure 3(a) in the first row is a trace image showing the concatenation of all the videos (the middle column of pixels is used to represent each frame of the concatenation of the videos). Figure 3(b) shows the result of the clustering of the frames at the selected level, where each color represents one cluster. The rest of the demonstration is dynamic and shows the selection process of the frames in the summary: the most important frames left in Figure 3(c) are chosen and inserted in the summary in Figure 3(d), while the corresponding cluster is removed from Figure 3(c).



(a)
(b)
(c)
(d)

Figure 3: Example of Multi-Video Summarization

4 Conclusion

In this paper, we have explained the principles of two visualization techniques for representing and constructing multi-video summaries. At the same time, two figures are used to briefly illustrate the experimental results. Multi-video visualization shows the relations of frames of similar videos. And multi-video summarization exploits clustering to obtain the summaries of videos, and directly displays the video summary. Both technologies could visualize multi-video summary, and the experimental results are able to be utilized to intuitively analyze the video summary. In the future work, we are interested in analyzing the relation of video frames by visualization approaches.

References

[1] E. Dumont and B. Merialdo, "Sequence Alignment for Redundancy Removal in Video Rushes summarization", *International Multimedia Conference Proceedings of the 2nd ACM TRECVid Video Summarization Workshop*, pp 55-99, 2008.

[2] E. Dumont and B. Merialdo, "Rushes Video Parsing using Video Sequence alignment", *15th International Multimedia Modeling Conference*, January 2009.

[3] Itheri Yahiaoui, Bernard Merialdo, and Benoit Huet, "Automatic video summarization", *Interactive Video, Algorithms and Technologies by Hammoud*, Riad (Ed.), 2006, XVI, pp 27-41.

[4] Ying Li, "Chapter 5: Movie Content Analysis, Indexing and Skimming via Multimodal Information", *Video Mining, Encyclopedia of Database Systems*, Kluwer Academic Publishers, 2003.

[5] Arthur G.Money, "Video summarisation: A conceptual framework and survey of the state of the art", *Journal of Visual Communication and Image Representation (19)*, February 2008, pp 121-143.

[6] Michael G.Christel, "Evaluation and user studies with respect to video summarization and browsing", *Multimedia Content Analysis, Management, and Retrieval*, *SPIE AND IS&T*, 2006.

[7] Dingding Wang, "Multi-document summarization via sentence-level semantic analysis and symmetric matrix factorization", *Special Interest Group on Information Retrieval*, 2008.

[8] Howard D.Wactlar, "Multi-Document Summarization and Visualization in the Informedia Digital Video Library", *New Information Technology Conference*, 2001.

[9] Krishnamachari, S. and Philips Res., "Image browsing using hierarchical clustering", *Computers and Communications, Proceedings. IEEE International Symposium on*, 1999.