I-034

Home Video Summarization by Support Vector Machine

Nagul COOHAROJANANONE† **Kiyoharu AIZAWA**†

1. Introduction

Digital video (DV) cameras are receiving popular for personal usage in everyday life. The main application of a DV is to record events such as vacations, weddings, and graduation ceremonies. However, these recorded videos are easily summed up to many hours of the video data. To search for one particular record or play back the recorded video are a time consuming and boring task. Therefore, it is more desirable to edit or summarize the entire records for guide scanning. There has been an interest in developing efficient schemes for video summarization [1] [2] [3], however the algorithms were concentrated on low level features.

In this paper, we propose a home video summarization by personal interest using Support Vector Machine (SVM) algorithm. Our algorithm first extract the Representative frames (R-frames) from the video sequence. In order to extract the R-frames, we first employ shot duration, shot motion activity to calculate number of R-frames to be retrieved from each shot. We apply the adaptive sub-sampling to extract the R-frames from sequence according to the calculated number. From figure 1, Without user preference, the group of R-frames can be smoothed to produce the summarized video. For user preference, we apply the Support Vector Machine. By doing that, a group of R-frames are first marked as interesting and uninteresting. The selected frames are then trained. The SVM next classifies the video sequence into two classes (positive and negative). The group of the positive frames are either smoothed to become the summarized video or re-iterated until meet the satisfied result. In the experiment, the proposed algorithm gives the acceptable on the test home

In the paper, we first describe the summarization by representing the R-frames in the section 2. we next describe the summarization by the SVM algorithm in the section 3. The experiment results are shown in the section 4. The conclusion is explained in the section 5.

2. Summarization by R-frames

To extract the R-frames, we adopted the algorithm proposed in [5]. First, the boundary of the shot is detected by shot detection. In the shot detection, the difference between the current frame and the consecutive frame is calculated by the city-block distance measure. If the distance is greater than the given threshold, the shot boundary is declared. The duration of the shot can be calculated from the boundary. Motion activity in each shot is then calculated from the detected motion. Motion parameter can be retrieved from block matching algorithm or the motion information on the MPEG video.

The transient number of R-frames in each shot is determined by the duration. The weight parameter is calculated from the motion activity of the shot. R-frames are extracted from the shots by the adaptive sub-sampling. By the adaptive sub-sampling, the frames are retrieved where the frames contain the high frame difference. The extracted R-frames are then smoothed and formed the summarized video.

†The authors are with the Department of Electrical Engineering and Frontier Informatics, University of Tokyo

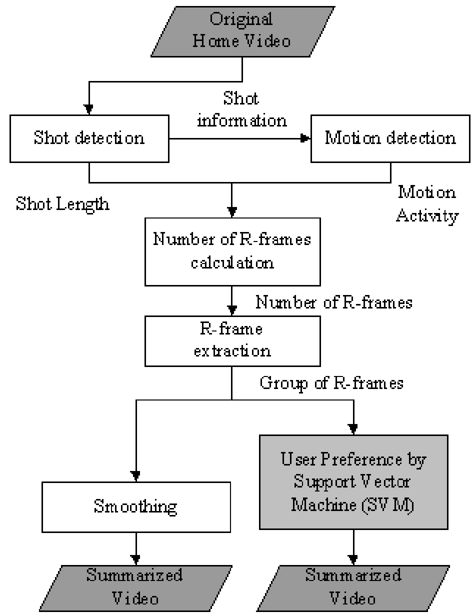

Figure 1: A Flowchart of the proposed algorithm

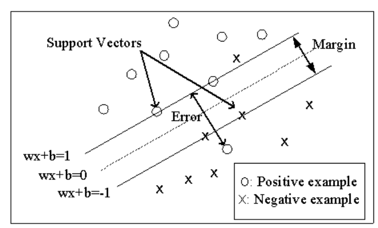

Figure 2: Basic idea of Support Vector Machine

3. Summarization by user preference

To summarize the video according to one interest, we make use the SVM algorithm. The SVM algorithm is designed to classify input vectors into classes based on a class of hyperplanes (Figure 2). Finally, and optimal hyperplane, defined as the one with the maximal margin of separation between the 2 classes, is constrained quadratic programming (QP) problem. The solutions of the QP problem consist of a subset of training patterns lying on the margin called Support Vectors, which carry all relevant information related to the classification. In cases where the input vectors are not well separable in the input space, kernel functions can be used to separate them nonlinearly in the input space.

Moreover, Support Vector Machine is capable of manipulating large number of input vectors and large scale classification. It has also been shown to have outperformed other learning algorithm, especially in pattern recognition [4].

The SVM algorithm consists of two stages 1) Train-

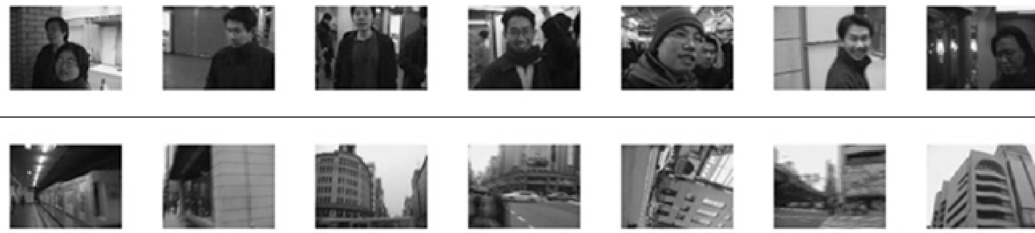

Figure 3: The above row shows the the extracted R-frames that are marked as interesting (positive examples), The below row shows the the extracted R-frames that are marked as uninteresting (negative examples)

Figure 4: First iteration: the query images are displayed with other positive images

ing and 2) Classifying. In order to avoid the training time consumption, only a group of R-frames are used in the training. User is asked to select the interesting and uninteresting frames equally from the extracted R-frames. The algorithm starts with training the selected data and continue to classify the video sequence. The SVM algorithm classifies the entire video sequence into two positive and negative group. Only the frames in the positive group are presented to the user.

In order to display the result to the user, we show the query images with the other positive images. Using the frame difference method will result in the query frames with the other most similar frames in video sequence. That might give a boring result. Instead of using frame difference, our algorithm display other the positive images that their level from the hyperplane are about the same level from the hyperplane of the query images. In this way, unsimilar image in the positive group can be viewed. Moreover, user can also re-iterate the SVM in the positive group until they meet the satisfied result.

4. Experiment

In the preliminary experiment, a test home video called "Lunch" (40 min) was summarized. For the summarization by R-frames, we chose to summarize to 25% summarization of original sequence, which gave the result of resulted in 259 R-frames. The 259 R-frames is later smoothed to compute the summarized video. With in the 259 R-frames, a user was asked to select the interesting and uninteresting frames (Figure 3). Seven frames that contained himself and his friends were selected as interesting frames and seven frames that did not contain him and his friends as uninteresting frames. The rest of the R-frames were set to Neural, The training method was done by transductive algorithm. In the experiment, 64 bins of RGB color histogram space were used as the vector feature. Right now, we are working on the SVM with the increased the vector features. The training time was approximately five minutes and the classifying time was approximately three minutes.

The first iteration result is shown in the Figure 4. In the result, the seven query images were display with other positive images that contain the distance from the hyper plane close to the query image. From our inspection, the image that look similar the negative examples were not be retrieved. These results can further be re-iterated for training and classifying upon the user's desire.

5. Conclusion

We proposed the home video summarization using shot characteristics and SVM. We applied the SVM to improve the summarized video according to user's interest. The experiment result gave the acceptable result. In order to increase the accuracy, we are now working on SVM by increasing the dimensional feature vectors.

References

[1] R. Lienhart, "Dynamic Video Summarization of Home Video", Proc.SPIE 3972: Storage and Retrieval for Media Databases, pp.378-389, 2000.

[2] B. Adams, C. Dora, S. Venkatesh, "Novel Approach Determining Tempo and Dramatic Story Sections in Motion Pictures", Proc. IEEE ICIP Sept. 2000.

[3] A. Divakaran, R. Radhakrishnan, K. Peker, "Motion Activity-Based Extraction of Key-Frames From Video Shots", Proc.IEEE ICIP Sept. 2002.

[4] Marti A. Hearst, "Support Vector Machines", IEEE Intelligent Systems, July/August 1998, pp. 18-28, 1998.

[5] N. Cooharojananone and K. Aizawa, "Home Video Summarization by Shot Characteristics", IEICE Technical Report, Vol.103 No.112, Jun. 2003.