Personalization of Video Content Features and its Application

Kok-Meng ONG† and Wataru KAMEYAMA†
† Waseda University

Abstract In this paper, we propose to extract personalized video features based on the viewer’s interest. The viewer’s interest toward the video content is modeled in real time by monitoring their pupil size, gazing point and heart rate. The experimental results show that the video summary generated using weighted video features achieves higher precision and recall ratios for video with high narration nature if compared to video summary generated by video features without personalized weight prediction.

1. Introduction

Video suggestion, retrieval and abstraction systems which rely on the video features that are extracted based on signal or image processing techniques are common [1]. However, because human preference differs from one person to another, the systems which are constructed based on the common video features might not be applicable for everyone. Therefore, in this paper, we propose to extract personalized video features based on the viewer’s bio-signals.

2. Personalization of Video Features

Our proposal to extract personalized video features is shown in Fig. 1. The video content is analyzed on shot basis. Two important stages of our proposal are as follow:

2.1. Extraction of Human Interest

In order to take human perception into account, instead of calculating the shot importance directly from the video features, the Human Interest, $|H|$ is extracted. This $|H|$ is modeled based on the viewer’s bio-signals. Therefore, it is effective in reflecting the viewer’s perception towards the video content. $|H|$ is generated based on the fusion of the following components using Attention Fusion Function [2]:

1. Cumulative Pupil Response, CPR
2. Frequency Component of pupil, FC
3. Gini index of gaze, Gini
4. Low Frequency of heart rate, LF

5. Ratio of High Frequency to Low Frequency of heart rate, HF/LF
6. Average heart rate, AvgHR,
7. Gradient of heart rate, GradHR

2.2. Prediction of Video Feature’s Weight

Conventional signal based approach calculates the shot importance $(0 \leq |A| \leq 1)$ based on the video feature vector, $A$ extracted using the low level signal:

$$|A| = \frac{1}{M} \sum_{m=1}^{M} A_m$$ (1)
Where $M$ is the total number of video feature components. However, this does not necessarily satisfy the preference of each individual. Therefore, $|H|$ is used to predict individual preference.

First, the correlation coefficient, $\rho_m$ between $|H|$ and the corresponding video feature component, $A_m$ is calculated. Then, in order to find out the shot importance based on weighted video feature, equation (1) is modified to:

$$|A| = \sum_{m=1}^{M} w_m A_m$$

(2)

Where the contribution of each video feature component is weighted by its corresponding weight factor, $w_m$ which is learned based on the correlation coefficient, $\rho_m$ as follow:

$$w_m = \frac{\rho_m}{\sum_{m=1}^{M} \rho_m}$$

(3)

3. Application in Video Summary

In order to verify our proposal, an experiment was carried out on 5 subjects, on a short film test video [3]. The video feature components in $A$ are extracted based on Arousal Model proposed by Hanjalic et. al. [4], where $M = 3$ is the total number of video feature components, which are corresponding to (1) Motion, (2) Rhythm and (3) Sound.

Video summary is generated by selecting 30% shots with highest importance calculated based on Arousal Model using equation (1), and Weighted Arousal Model using equation (2). These selections are matched to the ground truth shots, which are picked by the subjects, in order to calculate the precision, recall and F-measure.

4. Result

The average results for all subjects are shown in Fig. 2. From Fig. 2, it is observed that results for video summary based on Weighted Arousal Model are higher if compared to Arousal Model without weight prediction.

5. Conclusion and Future Works

In this paper, a personalized video features extraction method is proposed. By personalizing the video features, better result could be obtained for the application in video summary. This is achieved by the prediction of the weight for video features by finding the correlation between Human Interest and the video feature components. By applying the weight to the original video features, improvement in precision and recall is achieved for shot selection in video summary.

Our future works include prediction of Human Interest from the video features, and extension of our proposal to more video types.

Acknowledgement

This research is conducted jointly with VIS Research Institute [5].

References


