Efficient Video Viewing System for Racquet Sports with Automatic Summarization Focusing on Rally Scenes

Shunya Kawamura[1,a] Tsukasa Fukusato[1,2] Tatsunori Hirai[1] Shigeo Morishima[2,3,b]

Waseda University[1] JST CREST[2] Waseda Research Institute for Science and Engineering[3]

[a]s_kawamura@ruri.waseda.jp, [b]shigeo@waseda.jp

1. Introduction

This paper presents a system for viewing important rallies efficiently in racquet sports videos. There are several long matches in a tournament; thus, watching racquet sports can be time-consuming. Liu et al. [2009] proposed a racquet sports video summarization system based on a supervised audio classification that generates the summary video composed of only rally shots. However, summarizing a video is time-consuming because video editors must label audio information every second in the first 30 minutes. To solve this problem, we propose a system to summarize racquet sports video automatically using audio and visual information. In addition, we propose an efficient viewing system that maintains understanding of games using fast-forwarding.

2. Rally Shot Detection

We perform unsupervised similar shot clustering using HSV color histograms. For example, clusters are rallies, zoom on players, the gallery, and so on. To select rally clusters, we focus on "camera work" and the "white line." We found that camera work was not significant in rally scenes, while the movement of the camera and subjects were significant in other scenes. Consequently, white lines are retained only in rally scenes to generate mean images from each shot. We convert mean images into binary images using color and edge features, and detect white lines using progressive probabilistic Hough transform.

Table 1 shows the performance for an approximately two-hour tennis video (clay court) and a comparison of Liu's supervised method with our unsupervised method. The results indicate that we can automatically detect rally shots and demonstrate high precision and recall rates.

Table 1: Comparison of the rate of rally shots detection

Method	Recall [%]	Precision [%]
Liu et al. 2009 (Supervised)	98.4	90.0
Our Method (Unsupervised)	98.4	90.0

Furthermore, we applied our method to other racquet sports video. The average recall rate for five videos was 99.1 [%], and the average precision rate was 88.9 [%]. This detection accuracy is sufficient to watch most rallies.

3. Evaluation of Rally Rank

To rank rallies, we extract three features from each rally shot. Each rally rank R_s for the s-th rally is calculated as follows:

$$R_s = \alpha L_s + \beta P_s + \gamma V_s + \delta \quad (1)$$

where L_s is the rally shot length, and P_s and V_s are the pitch and the volume in the most exciting period after the rally ends, respectively. The set of weighted coefficients (α, β, γ, and δ) are determined by multiple regression analysis of subjective values obtained by pre-experimentation. For example, our method for the tennis video showed that R_s values for first service faults were low. Hence, we remove insignificant rallies by adjusting a threshold.

4. Summarization and Fast-forward Viewing

A summary video composed of important rally shots is generated by high ranked rallies up to a specific time. We apply more efficient viewing by normal-speed playback for the last few hits (t seconds) for each rally shot and fast-forward ($\times\ n$ playback) for other times. This system is based on the following two tendencies of racquet sports video: replay scenes contain only a few hits in a rally because they are very important; we can understand rallies even if sound is inaudible during high-speed playback.

5. System Implementation

Table 2 shows the total time of the summary video for the two-hour tennis video (clay court). The summary video was generated without first service faults by adjusting the threshold (n = 2 or 3; t = 5). The results indicate that users can watch only important rallies in a short period of time. In addition, users can adjust the total summary time using the user interface (Figure 1).

Table 2: Examples of total watching time

	Total time [s]	Compression rate [%]
Input	6508	-
All rallies	1410	21.7
$n = 2, t = 5$	836	12.8
$n = 3, t = 5$	721	11.1

Figure 1: *User interface for racquet sports video. Bottom: time bar indicating shots. Right: viewing method settings (rally options, speed options).*

6. Results and Future Work

We have proposed an automatic racquet sports video summarization system. We achieved an efficient viewing experience based on fast-forwarding. By combining with detection methods for rally scenes and player movements, users will be able to watch the shots they wanted. In future, we aim to improve our system by allowing users to choose favorite scenes, such as replay scenes.

References

Chunxi, Liu. *et al*. 2009. A framework for flexible summarization of racquet sports video using multiple modalities, *Computer Vision and Image Understanding 113*(3), 415-424.

SIGGRAPH 2014, August 10 – 14, 2014, Vancouver, British Columbia, Canada.

ACM 978-1-4503-2958-3/14/08