Court-Aware Volleyball Video Summarization

Takahiro Itazuri
Waseda Univeristy
s132800732@fuji.waseda.jp

Tsukasa Fukusato
Waseda Univeristy
tsukasa@aoni.waseda.jp

Shugo Yamaguchi
Waseda University
wasedayshugo@suou.waseda.jp

Shigeo Morishima
Waseda Research Institute for Science and Engineering
shigeo@waseda.jp

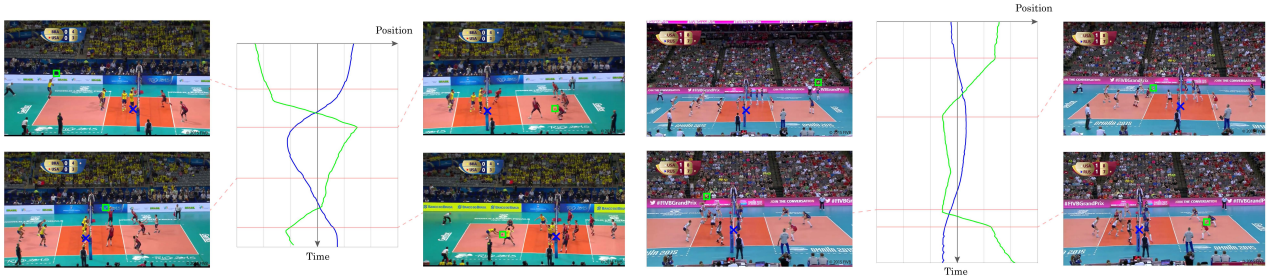

Figure 1: Examples of the transition of a ball and a court (x-component) in rally scenes. In the rally scenes, the global ball position (green) can be approximated by the center position of the court (blue) because the camera tends to follow the ball.

ABSTRACT

We propose a rally-rank evaluation based on the court transition information for volleyball video summarization considering the contents of the game. Our method uses the court transition information instead of non-robust visual features such as the position of a ball and players. Experimental results demonstrate the effectiveness that our method reflects viewers' preferences over previous methods.

CCS CONCEPTS

• **Computing methodologies → Video summarization**; *Scene understanding*; *Visual content-based indexing and retrieval*;

KEYWORDS

Sports video summarization, court-based rally-rank evaluation, rally scene detection.

ACM Reference format:
Takahiro Itazuri, Tsukasa Fukusato, Shugo Yamaguchi, and Shigeo Morishima. 2017. Court-Aware Volleyball Video Summarization. In *Proceedings of SIGGRAPH '17 Posters, Los Angeles, CA, USA, July 30 - August 03, 2017,* 2 pages.
https://doi.org/10.1145/3102163.3102204

SIGGRAPH '17 Posters, July 30 - August 03, 2017, Los Angeles, CA, USA

ACM ISBN 978-1-4503-5015-0/17/07…$15.00
https://doi.org/10.1145/3102163.3102204

1 INTRODUCTION

Watching sports videos is a time-consuming task because of the long duration of sports games. As a result, it is difficult for viewers who do not have enough time to watch a lot of games. To address this problem, many sports video summarization methods have been proposed.

Sports can be classified into two types: sports with fewer scoring points and sports with many scoring points. Previous methods for sports with fewer scoring points detect specific exciting events (e.g., home run in baseball and goal in soccer) and generate summary videos that consist of the detected events. On the other hand, previous methods [Kawamura et al. 2016; Liu et al. 2009] for sports with many scoring points (e.g., tennis, badminton and table tennis) detect play scenes and quantify the degree of excitement for each play scene by using "affective features" such as sound information and editorial information. Since volleyball is also one of sports with many scoring points, this approach is effective.

In order to generate summary videos that reflect viewers' preferences more, it is necessary to use semantic features representing the game content such as the position of a ball and players; however, such features have not been considered in previous methods. This is because the acquisition of them is still not robust and unreliable to generate the summary video. Instead, our method adopts the court transition information caused by camera operation, which can be acquired more robustly (Figure 1). We focus on a tendency that the camera follows the ball and assume that the global ball transition can be approximated by the court transition. To prove this assumption, we performed correlation analysis between the

Table 1: Experimental results (*P*: Precision, *R*: Recall, *F*: F-measure, *PCC*: Pearson correlation coefficient, *AF*: Affective features [Kawamura et al. 2016], *CF*: Court-based features, *TCR*: Time Compression Rate).

Video	Rally Scene Detection						Correlation Analysis	Rally-Rank Evaluation			Video Summarization		
	Ours			Kawamura et al.			*PCC*	Adjusted R-squared			Input[s]	Rally[s]	*TCR*[%]
	P	*R*	*F*	*P*	*R*	*F*		Ours	*AF*	*CF*			
1	1.000	0.863	0.926	0.948	0.877	0.811	−0.825	0.935	0.759	0.707	6275	876	14.0
2	0.876	0.966	0.919	0.942	0.830	0.882	−0.802	0.735	0.430	0.685	6743	1516	22.5
3	1.000	0.903	0.949	0.957	0.903	0.929	−0.836	0.702	0.569	0.612	4823	1096	22.7
4	0.994	0.855	0.919	0.927	0.825	0.873	−0.921	0.896	0.708	0.697	9452	1587	16.8
5	1.000	0.698	0.822	0.515	0.269	0.353	−0.718	0.976	0.868	0.903	9446	1364	14.4
Total	**0.974**	**0.857**	**0.907**	0.858	0.740	0.770	**−0.820**	**0.544**	0.474	0.483	-	-	**18.1**

ball position and the court position. Experimental results demonstrate the effectiveness that our method reflects viewers' content preferences over previous methods.

2 RALLY SCENE DETECTION

Some rally scene detection techniques in racket sports videos have been proposed recently. Kawamura et al. [Kawamura et al. 2016] perform automatically rally scene detection based on the assumption that rally scenes are usually obtained with a fixed camera. They perform Liu's unsupervised shot clustering based on HSV histogram [Liu et al. 2009] and determine clusters containing shots where the white lines of the court can be detected in an average image as rally scenes. However, rally scenes in volleyball videos contain camera operation such as panning, so it remains problematic to apply Kawamura's method to volleyball videos. In contrast, we assume that "the court is taken by a specific direction in rally scenes" extended Kawamura's assumption. Based on this assumption, we apply court detection based hierarchical rally scene detection approach; cluster-level and shot-level. That is, we follow the steps: (1) determining rally clusters with the court from the clusters and (2) determining rally shots with the court from the rally clusters.

Experimental results indicate that our detection method provides higher accuracy than the previous method (Table 1). This is because our method performs court detection instead of white-line detection and selects the rally scenes from rally clusters. On the other hand, recall values are smaller than precision value since our method cannot detect rally scenes taken by a different direction from camera direction constraint.

3 CORRELATION ANALYSIS

We perform correlation analysis for verifying our assumption that the ball position (x-component) can be approximated by the court position (x-component of the court center) on total 5341 frames (10 rally scenes from each of 5 rally videos). In correlation analysis, we calculate Pearson correlation coefficient (*PCC*) between the court position and the ball position obtained manually. As a result, the absolute value of *PCC* is more than 0.7 in all videos, which proves that there is strong correlation (Table 1).

4 VIDEO SUMMARIZATION

We evaluate a rally-rank for each rally scene and generate a summary video containing only the rally scenes with higher rally-rank than a threshold. The threshold is determined automatically so that the summary video becomes within viewer-specified time length. Given rally length L_r, pitch P_r, volume V_r, total court movement distance D_r, court average speed S_r, and court maximum speed M_r, each rally-rank I_r for the r-th rally is calculated as follows:

$$I_r = \alpha L_r + \beta P_r + \gamma V_r + \delta D_r + \epsilon S_r + \zeta M_r + \eta \quad (1)$$

where $\alpha, \beta, \gamma, \delta, \epsilon, \zeta, \eta$ are weight coefficients determined by multiple regression analysis of the subjective experiment values. The subjective experiments was performed for 10 rally scenes from each of 5 volleyball videos and we asked 45 subjects to rank the interest of the rally scene at 7-point Likert scale.

To verify that our rally-rank evaluation can reflect viewers' interest, we calculated adjusted R-squared between average subjective values and evaluated values (Table 1). The results show that all adjusted R-squared values in our method is higher than those in Kawamura's method. Therefore, our rally-rank evaluation demonstrates the effectiveness of satisfying viewers' interest. In our rally-rank evaluation, the rally scene with more rally counts tends to be evaluated with a higher rank. Time compression rates of summary videos containing all rally scenes against original input videos indicate that viewers can watch all rally scenes within a short period of time.

5 CONCLUSION

We have proposed a court-based rally-rank evaluation for volleyball video summarization. By combining the affective features and the court-based features as semantic features, our method indicate the effectiveness of reflecting viewers' preferences. In the future, we aim to improve our system by considering the game flow and temporal connectivity.

ACKNOWLEDGMENTS

This work was supported in part by JST ACCEL (grant JPMJAC1602).

REFERENCES

Shunya Kawamura, Tsukasa Fukusato, Tatsunori Hirai, and Shigeo Morishima. Rsviewer: An efficient video viewer for racquet sports focusing on rally scenes. In *International Conference on Information Visualization Theory and Applications*, volume 2, pages 247–254, 2016.

Chunxi Liu, Qingming Huang, Shuqiang Jiang, Liyuan Xing, Qixiang Ye, and Wen Gao. A framework for flexible summarization of racquet sports video using multiple modalities. *Computer Vision and Image Understanding*, 113(3):415–424, 2009.