

# Automated Service Scene Detection for Badminton Game Analysis Using CHLAC and MRA

Fumito Yoshikawa, Takumi Kobayashi, Kenji Watanabe, and Nobuyuki Otsu

**Abstract**—Extracting in-play scenes in sport videos is essential for quantitative analysis and effective video browsing of the sport activities. Game analysis of badminton as of the other racket sports requires detecting the start and end of each rally period in an automated manner. This paper describes an automatic serve scene detection method employing cubic higher-order local auto-correlation (CHLAC) and multiple regression analysis (MRA). CHLAC can extract features of postures and motions of multiple persons without segmenting and tracking each person by virtue of shift-invariance and additivity, and necessitate no prior knowledge. Then, the specific scenes, such as serve, are detected by linear regression (MRA) from the CHLAC features. To demonstrate the effectiveness of our method, the experiment was conducted on video sequences of five badminton matches captured by a single ceiling camera. The averaged precision and recall rates for the serve scene detection were 95.1% and 96.3%, respectively.

**Keywords**—Badminton, CHLAC, MRA, Video-based motion detection

## I. INTRODUCTION

DETECTING and segmenting in-play scenes in sport video sequences is necessary in various applications such as quantitative game and performance analysis and video browsing. In studies on video-based game and performance analysis of racket sports, much research efforts have been made to explore the relationships between predefined parameters and sports performance, and the methodologies for effective coaching (e.g. Pritchard et al. [1], Hughes et. al, [2]). Many of such the studies have necessitated manual operations in data processing. In more practical situations, however, quick feedback of the resultant quantitative data and/or videos is expected to be conducted for the relevant coaches and athletes. In order to reduce the laborious and time-consuming processes by human operators for more detailed analysis, speeding up and streamlining of the quantitative game analysis is required. Thus, first of all, automatic detection of the start and end of rally periods is essential for both the quantitative analysis and the effective video handling.

For the task to recognize the events in racket sports including

temporal segmentation of the in-play scenes, several different approaches have been developed. In most cases for broadcasted sport videos, several researchers have investigated the performance of deterministic and/or multimodal approaches by exploiting domain-specific rules and trajectories of the players and the ball (e.g. Sudhir et.al. [3], Han et. al. [4]). In our earlier work [5], we have examined the approach to infer each rally period from the motion trajectories of players and shuttlecock using a stationary ceiling camera, and confirmed the need to incorporate human motion recognition techniques for a more refined and higher-level sport video analysis.

Zhu et al. [6] develop a motion analysis-based approach to recognize the basic player action in broadcast tennis videos employing optical flow-based motion descriptors. For the similar task including serve detection, Zivkovic et al. [7] and Roh et al. [8] apply appearance-based approaches, and Han et al. [4] exploit features such as position and speed of the player. These methods require segmentation of the player regions, in which the segmentation error tends to affect final recognition. In addition, the conventional approaches include sequential and procedural processes require too special and tedious steps. These make it difficult to design adaptive and real-time systems.

In recent years, on the other hand, a scheme of adaptive vision system has been presented, which comprises two stages of feature extraction, namely, higher-order local auto-correlation (HLAC) or its extension CHLAC (Cubic HLAC) [9] and multivariate analysis [10], [11]. Concerning human motion and behaviour analysis, CHLAC approach has been successfully applied to motion recognition [9], unusual motion detection [12], [13] and motion segmentation [14].

In this paper, we apply CHLAC and multiple regression analysis (MRA) to detection of serve scenes in badminton in the simple statistical framework. In the experiment, we utilize the dataset which has been acquired by capturing elite athletes' matches by using a single ceiling camera. The experimental results demonstrate the effectiveness of the present method. The integration between the present approach and the object tracking-based approach proposed in our earlier work could contribute to more precise and depth analysis of racket sports and facilitate efficient utilization of video contents handled in sports institutes, training centers and gymnasiums.

Authors are with the National Institute of Advanced Industrial Science and Technology, Central2, Tsukuba, Japan (corresponding author to provide phone: 029-861-5183; fax: 029-861-5046; e-mail: {Yoshikawa.Fumito, Takumi.Kobayashi, Kenji-Watanabe, Otsu.N} @aist.go.jp).

## II. METHOD OF SERVE SCENE DETECTION

### A. Input Image and Task

Badminton videos captured during competition and training usually contain both of the rally periods in question and the others: The in-play and break segments are alternatively concatenated. In practical environments, the videos are in most cases captured by cameras on the spectator stand behind the end line, but in some cases, by the ceiling camera in sports institutes, training centers and large-scaled gymnasiums hosting national and international sporting event as shown in Fig. 1 (upper).

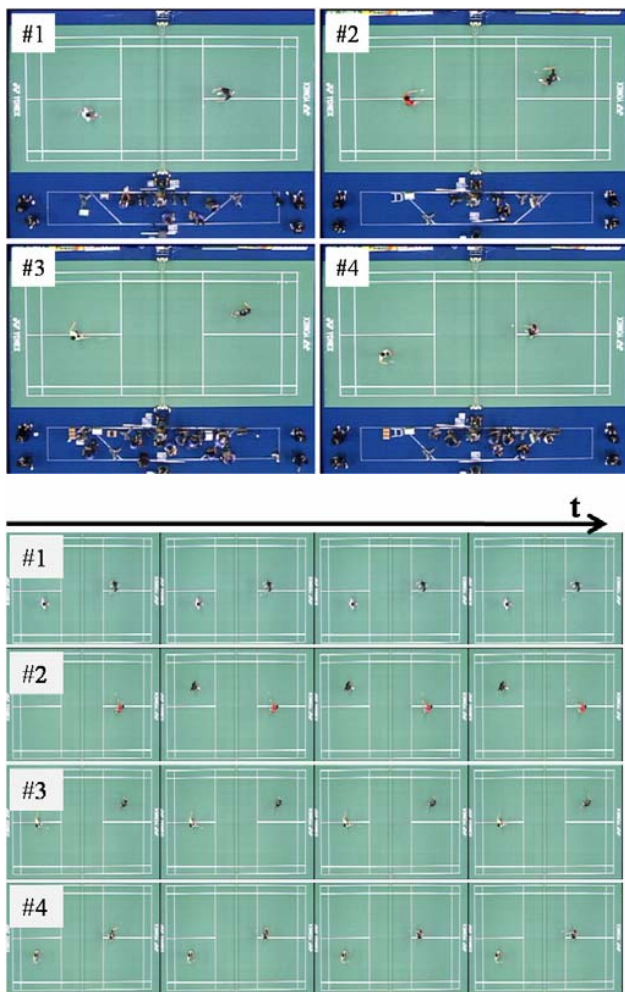


Fig. 1 Examples of video captured during matches: (upper) input images, (bottom) image sequences in serve motions

In order to clarify the task addressed in this paper, examples of input image sequences around serve motion corresponding to the occurrence of hitting shuttlecock are illustrated in Fig. 1 (bottom). The serve scenes consist of frames around a time point when the server hits the shuttlecock to deliver it forward the receiver. In our earlier work [5], the following domain-specific knowledge is employed for rally period detection;

- the serve is the first shot in each rally, and therefore the start point of each rally is determined as the point that the shuttlecock is delivered from the server's service court to the receiver's one while each player is detected and tracked in each service court during break.
- The shuttlecock is immobile for a moment after it lands, and accordingly the disappearance time of the shuttlecock can be a cue for detecting the end of each rally.

However the approach to infer each rally period from the motion trajectories of players and shuttlecock has some drawbacks due to difficulties in explicitly describing all of the possible contexts. In practice, the shuttlecock is not successfully detected.

The rest of this section describes the proposed approach to automatic detection of the serve scenes in racket sports.

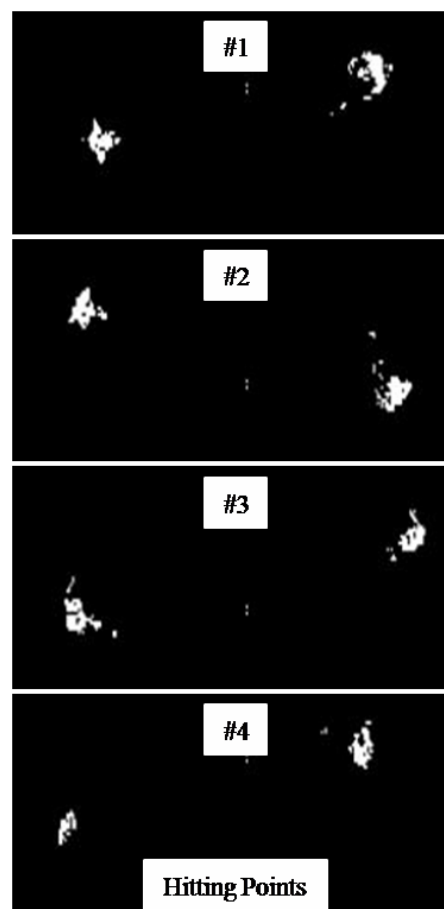


Fig. 2 Examples of the preprocessed image sequences in serve motions in badminton

### B. Preprocessing

In the preprocessing, we apply background subtraction which incorporates frame differencing and automatic thresholding [15]. This method is expressed in the following formula:

$$P_t(x, y) = \begin{cases} 1, & \text{if } \max(|I_t(x, y) - B(x, y)|, |I_t(x, y) - I_{t-1}(x, y)|) > \tau \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

where  $P_t$  is a label indicating whether objects exist or not at time  $t$ ,  $B$  is the background image and  $I_t$  is the video image frame at time  $t$ . These processes filter out both inherent noise and brightness information, which are irrelevant to both the posture and the motion of the players. Consequently, pixel values in each frame become either 1 (foreground) or 0 (background). The examples of the binary images are shown in Fig. 2.

### C. Motion Feature Extraction

In the stage of feature extraction, cubic higher-order local auto-correlation (CHLAC) [9] is employed. CHLAC enables simultaneous extraction of spatio-temporal features from the motion images. Let  $f(\mathbf{r})$  be three way data defined on the region (cubic data)  $D: X \times Y \times 3$  with  $\mathbf{r} = (x, y, t)^T$ , where  $X$  and  $Y$  are the width and height of image frame and  $T$  is the length of a time-window. Then, the  $N$ -th order auto-correlation can be defined as,

$$x_N(t; \mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_N) = \int_T \int_{X, Y} f(\mathbf{r}) f(\mathbf{r} + \mathbf{a}_1) f(\mathbf{r} + \mathbf{a}_2) \dots f(\mathbf{r} + \mathbf{a}_N) d\mathbf{r}. \quad (2)$$

CHLAC features possess important properties of shift invariance (rendering the method segmentation-free), additivity, and robustness to noise in data. Moreover, this method requires no prior knowledge or heuristics about objects. These favorable properties can benefit all aspects of approach to adaptively detect serve scenes including possible variability in terms of their appearances due to difference in server's and receiver's posture and kinematic profile.

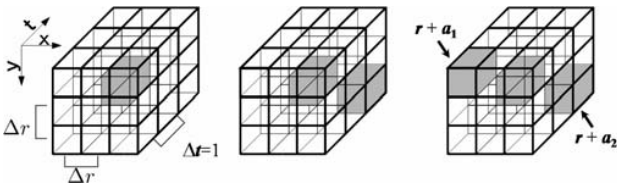


Fig. 3 Examples of mask patterns: (left)  $N=0$ ; (middle),  $N=1$ ,  $\mathbf{a}_1 = (\Delta r, \Delta r, 1)^T$ ; (right),  $N=2$ ,  $\mathbf{a}_1 = (-\Delta r, -\Delta r, -1)^T$ ,  $\mathbf{a}_2 = (\Delta r, \Delta r, 1)^T$

### D. Linear Regression

In the training phase, effective features for the serve scene detection are extracted from the given training example. The pairs of the motion feature vector  $x_i$  and the teacher signal  $c_i$  at time  $i$  are given. We apply multiple regression analysis (MRA), which determines the optimal linear coefficients  $\mathbf{a}$ , to estimate  $c$  from  $x$ :  $c \approx \mathbf{a}'x + b = \hat{\mathbf{a}}'x$ , where  $b$  is constant,  $\hat{\mathbf{a}} = [a', b]$ ,  $\hat{x} = [x', 1]$ . In this study, the supervised training signals are binary, assigning 1 at times during the serve motion, and otherwise assigning 0. The segment over a given time-window around the frame corresponding to the occurrence of hitting the shuttlecock is determined as that of the serve motion. Given the

motion feature  $\mathbf{x}$ , the existence of the target motion segment can be estimated by  $c = \mathbf{a}'x + b$ . In the method, the target scene are finally identified by detecting the local peak along the time axis and thresholding it after applying moving average to the estimated values over a time-window  $T$ .

## III. EXPERIMENTS

The proposed method was applied to automatic detection of serve scenes from the image sequences of actual badminton matches. The dataset used in this experiment comprises video sequences of five matches performed by top-level athletes; three matches of men's singles, two matches of women's singles. The dataset had been acquired by using a single ceiling camera in a gymnasium. These data were captured at 30 frames per second (fps) and 320 x 240 pixels (QVGA).

CHLAC features are obtained by using all mask patterns of mask patterns of  $\Delta r = 1, 3, 5$ . The time-window  $T$  for smoothing the estimation results is 25 for the serving motion detection. In non-maximum suppression for peak detection, number of neighbors is set to 3. To evaluate the performance, we calculated recall (R) and precision (P) rates for each match. In this evaluation, the detected point was regarded as correct if it was within each time duration of the target serving motion which was strictly determined by hand as ground truth. We define precision and recall as follows;

$$\text{Precision} = \text{detected\_correct\_peaks} / \text{detected\_peaks},$$

$$\text{Recall} = \text{detected\_correct\_peaks} / \text{correct\_points}.$$

The eight serves in each badminton match are used as the supervised training data. The frames in non-training section are used as the test data. Table I shows the experimental results of serve detection by the proposed method. The reason for incorrect detection in certain cases is that the current version of our method does not account for variation of serves such as forehand-long, forehand-short, backhand-short. The proposed method could facilitate efficient utilization of video contents as indicated in Table I. The prototype system that we developed is shown in Fig. 4.

TABLE I  
EXPERIMENTAL RESULTS FOR THE SERVE DETECTION

	Serves	Recall(%)	Precision(%)
Match #1	57	100	98.3
Match #2	55	91.7	90.2
Match #3	56	94.9	93.3
Match #4	56	94.9	96.6
Match #5	68	100	97.1

## IV. CONCLUDING REMARK

We have presented a novel approach to automatic detection of serve scenes in actual badminton matches. The present method consists of feature extraction by CHLAC and prediction by MRA, and yields favorable detection performances. Future work includes integrating our earlier and present approaches and incorporating human motion recognition techniques for a more refined and higher-level sport video analysis, such as analysis and characterization of

each player's motion, skill and tactics.

We applied this method to other sport motion, such as the start and the end point detection of weightlifting motion, and obtained the similar results, which shows the validity and generality of our method [16], and thus the proposed method will also be application to the other sports.



Fig. 4 Snapshot of our prototype viewer using indexes obtained by our rally period segmentation method

#### ACKNOWLEDGMENT

We thank Mr. Shigemitsu Imai at Nippon Badminton Association for his cooperation on data gathering.

#### REFERENCES

- [1] Pritchard, S., Hughes, M., and Evans, S., "Rule changes in elite badminton," In: Mike Hughes and I. M. Franks (eds.) *pass.com*. Cardiff: CPA, UWIC. pp. 213 – 221, 2001
- [2] Hughes, M. Hughes, M.T., and Behan, H., "The Evolution of Computerised Notational Analysis through the Example of Racket Sports," In *International Journal of Sports Science and Engineering*, Vol. 1, No. 1, pp.3 -28., 2007
- [3] Sudhir, G., John C. M. Lee, Anil K.Jain, "Automatic Classification of Tennis Video for High-level Content-based Retrieval," In: *IEEE International Workshop on Content-Based Access of Image and Video Databases*, pp.81-90, 1998
- [4] Han, J., Farin, D., P. H. N. de With, Lao W., "Real-Time Video Content Analysis Tool for Consumer Media Storage System," In: *IEEE Transactions on Consumer electronics*, Vol. 52, No.3 pp. 870-878, 2006
- [5] Yoshikawa, F., Raytchev, B., Kobayashi, T., Otsu, N., "Automated Indexing System for Badminton Game Analysis," In: *13<sup>th</sup> World Congress of the International Association for Sports Information*, 2009
- [6] Zhu, G., Xu, C, Huang, Q, Gao, W, Xing L., "Player Action Recognition in Broadcast Tennis Video with Applications to Semantic Analysis of Sports Game," In: *the 14th annual ACM international conference on Multimedia*, 2006
- [7] Zivkovic, Z., van der Heijden, F., Petkovic, M., Jonker, W., "Image Segmentation and Feature Extraction for Recognizing Strokes in Tennis Game Videos," In: *Proceedings 7th Annual Conference on the Advanced School for Computing and Imaging*, 2001
- [8] Roh, M.C., Christmas, B., Kittler, J., Lee, S.W., "Gesture Spotting for Low-resolution Sports Video Annotation," *Pattern Recognition* 41, pp. 1124-1137, 2008
- [9] Kobayashi, T., Otsu, N., "Three-way Auto Correlation Approach to Motion. Recognition," *Pattern Recognition Letter*, Vol. 30, No. 3, pp. 212- 221, 2009
- [10] Otsu, N., Kurita, T., "A New Scheme for Practical Flexible and Intelligent Vision System," In: *IAPR Workshop on Computer Vision*, 1998
- [11] Otsu, N., "CHLAC Approach to Flexible and Intelligent Vision Systems," In: *ECSIS Symposium on Bio-inspired, Learning, and Intelligent Systems for Security*, 2008
- [12] Iwata, K., Satoh, Y., Sakaue K., Kobayashi, T., Otsu, N., "Development of Software for Real-time Unusual Motions Detection by Using CHLAC," In: *ECSIS Symposium on Bio-inspired, Learning, and Intelligent Systems for Security*, 2008
- [13] Nanri, T., Otsu N., "Unsupervised Abnormality Detection in Video Surveillance," In: *IAPR Conference on Machine Vision Applications*, 2005
- [14] Kobayashi, T., Yoshikawa, F., Otsu, N., "Motion Image Segmentation Using Global Criteria and DP," In: *International Conference on Automatic Face and Gesture Recognition*, 2008
- [15] Otsu, N., "A threshold selection method from gray-level histogram," In: *IEEE Transactions on System Man Cybernetics*, Vol. SMC-9, No. 1, pp. 62-66, 1979
- [16] Yoshikawa, F., Kobayashi, T., Watanabe, K., Shirai, K., Otsu, N., "Start and End Point Detection of Weightlifting Motion Using CHLAC and MRA," In: *1st International Workshop on Bio-inspired Human-Machine Interfaces and Healthcare Applications*, 44-50, 2010