International Journal of Innovation and Scientific Research
ISSN 2351-8014 Vol. 23 No. 2 May 2016, pp. 233-242

http://www.ijisr.issr-journals.org/

Automatic Event Detection and Classification Based on Ball Trajectory in Broadcast Tennis Video Using SVM and HMM

M. Archana and M. Kalaiselvi Geetha

Department of Computer Science and Engineering,
Annamalai University,
Chidambaram, Tamil Nadu, India

ABSTRACT: An identifying event in sports video has many efforts of sports applications. In this paper, proposed a system for automatic detection of key events in Broadcast Tennis Video (BTV). The ultimate goal is to detect the events of complete tennis match. The detected tennis events are fault, rally and net approach, there are also other events in BTV, they all are considered as secondary one. To detect the events of tennis by analyzing the player`s position and ball tracking. The experiments done in different tennis tournament, which has the events (fault, rally and net approach), the some of the visual features are extracted from MHI (Motion History Image) and modelled by Support Vector Machines (SVM) and Hidden Markov Model (HMM) for recognizing tennis events. In result HMM gives a higher accuracy rate of 96.66% when compared to SVM rate of 86.42%.

KEYWORDS: Ball trajectory, event detection and classification, MHI (Motion History Image).

1 INTRODUCTION

In recent years, the significant increase of sports video is seen in internet and rapid increase in computer vision. Most of the researches are highlight detection, text recognition, scene identification and classification, structural tactics analysis, summarization of videos for broadcast tennis videos [1, 2]. In this paper, thoroughly investigate an automatic event detection and classification of broadcast tennis video [3]. According to the tennis rules and regulations, tennis events can be categorized into the following types:

FAULT:

Player 1 serves successfully, and player 2 fails to attend / return the ball, then considers as first serve fails, and the camera focus changes immediately out of the court view.

RALLY:

Player 1 serves successfully and also the player 2 returns successfully, it continued until any one of them fails to attend the ball.

NET APPROACH:

Player 1 serves successfully and the player 2 returns successfully [4]. One or both of them approach the net to stress his/her opponent.

Corresponding Author: M. Archana

2 RELATED WORK

[5] Have proposed a scheme called motion representation. A set of motion filters is used, based on these features classify using HMM the basketball video into 16 events and achieved 75% of accuracy. [6] Have addressed content based retrieval on broadcast videos. Different types of tennis strokes are recognized based on different feature using HMM. To recognize and classify the tennis strokes by using the knowledge based techniques.

[7] Have presented a technique to detect the tactic information in soccer video from the goal events. The goal events are detected using web-cast text and achieve best results on two tactic representation of players and ball multiple trajectories to discover the tactic.

[8] Have proposed a system to detect the events in baseball using scoring and last pitch. Event detection and event boundary detection is done using text recognition and processing speed of the ball respectively. The high caption recognition accuracy is achieved by the domain knowledge model, finally event boundary extraction is achieved by visual-based recognition.

3 PROPOSED APPROACH

The proposed approach is shown in fig.1. The given input video is divided into frames, then court – view frames are taken for consideration. The object player and the ball are identified using background subtraction method. After detecting the ball and player, followed event is detected using the contour.

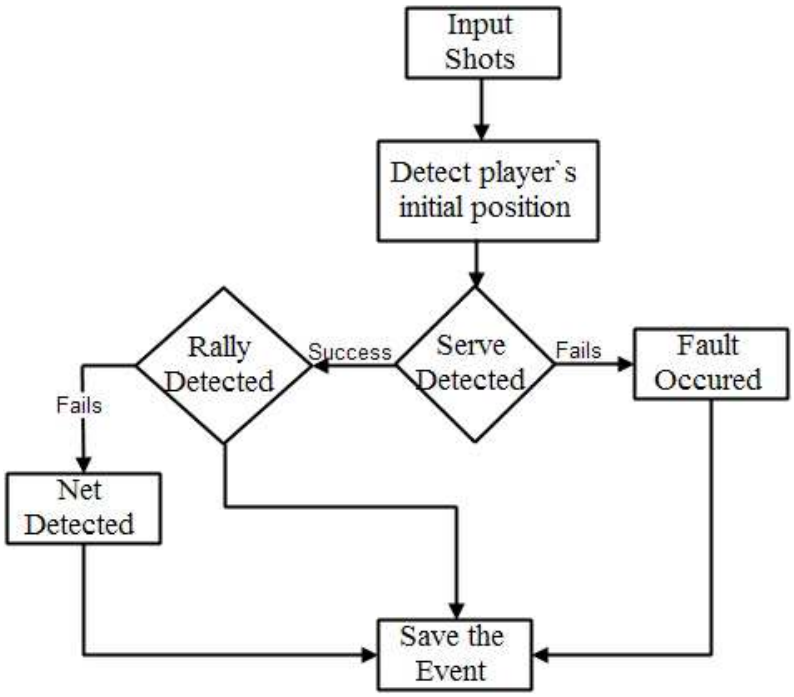

Fig. 1. Illustration of the proposed work

4 EVENT DETECTION

4.1 SMOOTHING THE IMAGE

Some noises are appearing very distinct and its pixel values are varied from its neighboring pixels, to eliminate the noises by changing its pixel value to the median of neighboring pixel values.

4.2 BACKGROUND MODEL

The ultimate aim of the background model is create a background in the Broadcast Tennis Video, which does not have a static background because of the moving cameras. The background model is created as,

$$B_G(x,y) = \frac{\sum_{K=1}^{n-1} I_t(x,y)}{F} \quad (1)$$

Where B_G (x, y) is the intensity of the background model, I(x, y) is the intensity of the pixel, t represents frame number and K represents number of the frames used to create a background model.

4.3 FRAME DIFFERENCE

Frame difference is to find the objects between the frames. Motion objects are extracted by pixel-wise differencing of successive frames. Motion information T_k or difference image is calculated by using Eq.

$$\boldsymbol{T_k} = \begin{cases} \mathbf{1}, & \boldsymbol{if\ D_k(i,j) > t} \\ \mathbf{0}, & \boldsymbol{otherwise}; \end{cases} \quad (2)$$

Where D_k is the difference image and t is threshold using Eq 3.

$$\boldsymbol{D_k(i,j) = |I_k(i,j) - I_{k+1}(i,j)|} \quad (3)$$

$$1 \le i \le w,\ 1 \le j \le h$$

4.4 CONTOUR DETECTION METHOD

To detect the ball contour and player initial position, present image as A and the next image as B are considered,

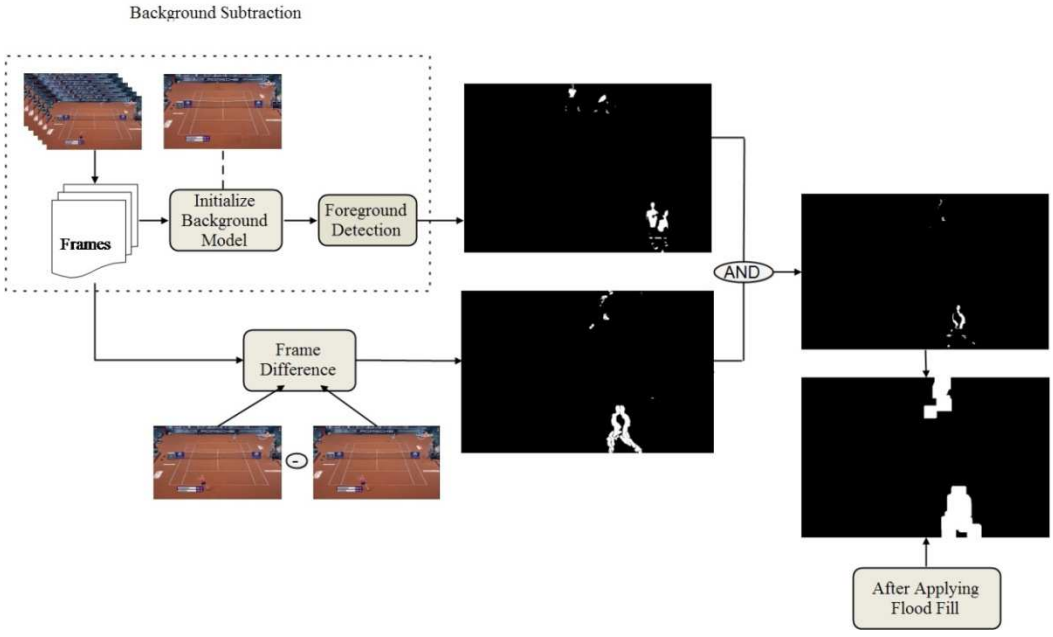

Fig. 2. Contour Detection method

- Frame difference of A and B is done and obtained C as result.
- The background model is created among the given frames and represent as D.
- The AND operation is done between C and D.
- Based on the size of the detected contour, the ball is identified and follows the motion of the ball.
- The biggest blob is splited, which is a part of the human body.

Apply flood fill techniques to complete the region as the human along with a tennis racket.

4.5 BALL TRACKING

The ball candidates are detected based on the aspect ratio, size, and compactness using the approach as described in section 4.4. To track the ball, the centroid is computed for the detected contour then follows the motion and finally, the ball is tracked [9].

4.6 PLAYER TRACKING

Player contour is detected using the approach as described in section 4.4. The largest blob is detected which belongs to parts of the human body. The other part of the human body is also detected based on the flood fill technique to reconstruct the player [10]. To track the player moving along with a camera to follow the action, divide the frame as upper and lower. The upper layer is a challenging task of tracking because of the upper player's size is too small and for lower layer is tracked using existing technique such as background subtraction.

5 TRACKING BASED EVENT DETECTION

5.1 FAULT EVENT

In the fault event, the player 1 serves successfully, where player 2 fails to attend the ball. In order to detect the event, the player initial position is founded while serve, the player`s contour height is too high compared to normal player contour, then based on that, the serve frames are detected [11]. There is any motion of ball contour, then follow the contour and track it. If a fault occurred there is no motion of the ball, hence the event is detected Fig. 3. Illustrates the fault occurred frames. The fault event is detected using the motion of the ball and player`s initial position only.

Fig. 3. Fault Occurred Frame

5.2 RALLY EVENT

In the rally event, the player 1 serves successfully and also player 2 returns successfully, to any one fails. In order to detect the event, the ball is identified using background subtraction method. Once the player serves, the ball is going to track and then follows the motion and hence the event is detected [12]. The rally event is detected using the motion of the ball and the number of hit counts, if the count is occurred, the event is happening Fig. 4. Illustrates rally sequence.

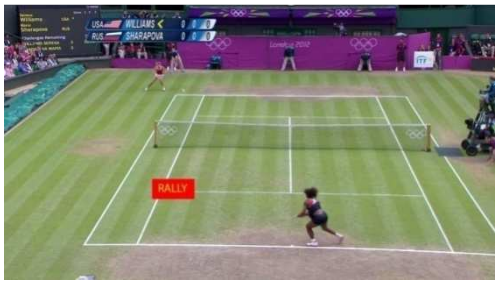

Fig. 4. Rally Frame

5.3 NET APPROACH

In the net approach event, the player 1 serves successfully and the player 2 also returns successfully. If any one of them through the ball to make the net stress, the event has occurred Fig. 5. Shows Net approach frames. In order to detect this event, along with ball tracking the net region is also monitored [13]. If any stress in net region by players, the event is happening.

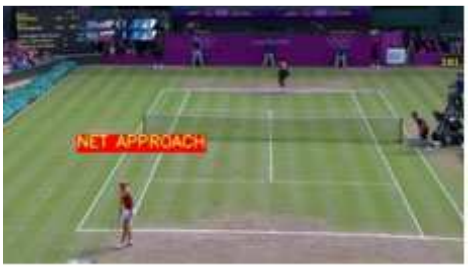

Fig. 5. Net Approach Frame

6 FEATURE EXTRACTION

6.1 MOTION HISTORY IMAGE (MHI)

In order to capture motion feature, motion information is considered as foreground object motion and background motion. The MHI is used for human motion recognition and analysis. It is a cumulative gray scale image which represents the motion information [14]. It holds a history of temporal variations at each pixel location, which then decay over time. MHI variants and representation are presented in the figure. In MHI image the motion stream or sequence of motion is incorporated by using each pixel intensity. MHI is obtained by using:

$$H_\tau(x,y,t) = \begin{cases} \tau, \ if \ \emptyset\ (x,y,t) = 1 \\ \max(0, H_\tau\ (x,y,t-1) - \delta, \ otherwise; \end{cases} \quad (4)$$

Here (x, y) and t show the position and time, ϕ(x, y, t) signals the presence of motion in the current video image, the duration τ decides the temporal extent of the movement, δ is the decay parameter.

The update function ϕ(x, y, t) for every new video frame analyzed in the sequence and is defined by,

$$\emptyset(x,y,t) = \begin{cases} 1, D(x,y,t) > \in; \\ 0, otherwise\ ; \end{cases}$$

Some possible image processing techniques for defining this update function ϕ(x, y, t) are background subtraction, image differencing and optical flow. The MHI in this work is generated from a binarized image, obtained from frame subtraction, using a threshold ∈ and D (x, y, t) are computed. In Fig. 6. Shows the MHI samples.

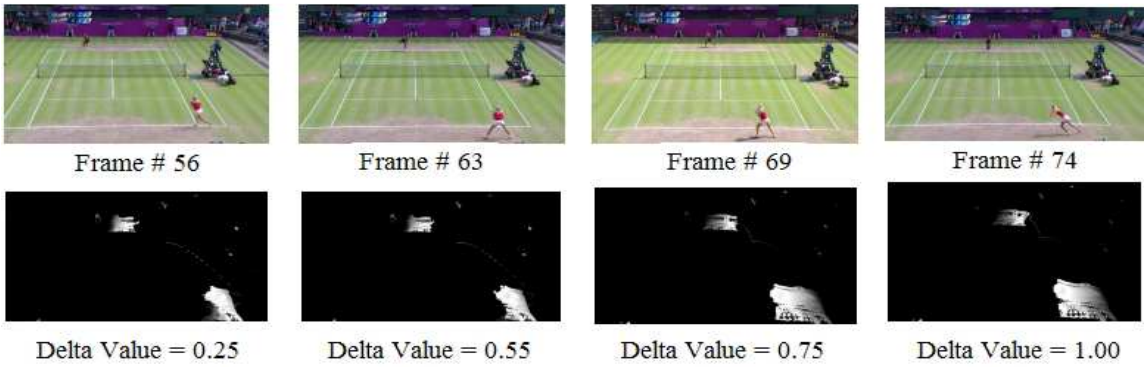

Fig. 6. MHI Sample Images

6.2 HISTOGRAMS OF ORIENTED GRADIENTS (HOG)

In order to detect the player pose in MHI image, the Histograms of Oriented Gradients (HOG) are considered. Player pose detection in images is a challenging task because of the various ways of appearance and the wide range of poses [15]. To solve this robust feature is needed, in which allows the player form cleanly in addition to background under different illuminations. For this HOG descriptor performs excellent compared to other exciting feature such as SIFT descriptors.

A typical method is to apply a one-dimensional discrete differential mask as the horizontal orientation (D_k = [-1 0 1]) and vertical orientation ($D_Y = [-101]^T$).

Convolution mask of horizontal orientation: $I_X = H(x, y, t) * D_X \infty$ (5)

Convolution mask of vertical orientation: $I_Y = H(\mathrm{x}, \mathrm{y}, \mathrm{t}) * \mathrm{D}_\mathrm{Y}$ (6)

Size of gradient: $|G| = \sqrt{I_X{}^2 + I_Y{}^2}$ (7)

Orientation of gradient: $\theta = \arctan \frac{I_Y}{I_X}$ (8)

Signed gradient: $\alpha_{Signed} = \begin{cases} \alpha & \alpha \geq 0 \\ \alpha + 360 & \alpha < 0 \end{cases}$ (9)

Unsigned gradient: $\alpha_{Signed} = \begin{cases} \alpha & \alpha \geq 0 \\ \alpha + 180 & \alpha < 0 \end{cases}$ (10)

When the image is created, convolution masks of the horizontal and vertical orientations (Equations (5) and (6), respectively) are applied to the image, and the orientation and gradient size are calculated. Second, histograms of the divided cells are calculated. Each pixel value in the cell is calculated as the orientation of the gradient through an advanced gradient calculation. These values are spread on orientation histogram bands, which are set as the number of bins. Cells are comprised of rectangular shapes in the image [16]. As an expression of the gradient, the histogram bands are evenly distributed from 0 to 360 degrees (Equation (9)) or from 0 to 180 degrees (Equation (10)). MHI image are received as the input and the HOG feature is then created. This process is shown in Fig.7. In addition, in our system, the number of cells is three (3× 3) and the number of bins is nine. In therefore create a 81-dimensional vector and also varied the cells as (4 x4) and (5x5) and examined.

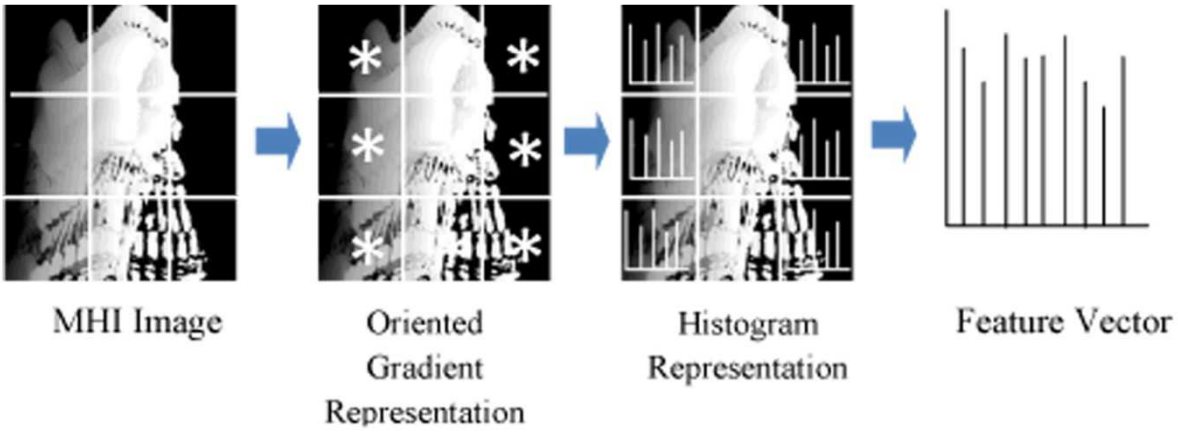

Fig. 7. Extraction process of MHI-HOG feature vector.

7 EXPERIMENT RESULT

For experiments, different BTV from broadcast channels from Australian Opens, Wimbledon Opens, French Opens and US Opens are recorded and are shown in Fig. 8. The sequences of tennis videos are red, clay, artificial and grass courts. The dataset consists of 73 clips for single (10 to 35 secs) at 25 fps, as shown in Table. 1.

Table 1. Dataset used for experiments

Datasets	Single(min)	No. of plays
Australian Opens	115	96
Wimbledon Opens	102	94
French Opens	93	92
Us Opens	90	80

Fig. 8 Example video frames

7.1 RESULTS WITH DISCUSSION

The HOG feature is extracted from MHI image, by varying the cell size of 3x3, 4x4, 5x5 and 6x6, the feature dimensions are 91, 125, and 225,324 respectively. The dimensions are reduced by applying Prinicipal Component Analysis (PCA) to 36, 31, 16 and 36 where the size of the cell is from 3x3, 4x4, 5x5 and 6x6 respectively. The response of these dimensions the accuracy is measured and analyzed in Fig.9. Based the accuracy finally consider the 5x5 cell size it gives the highest accuracy when compared to others is illustrated in Fig. 9.

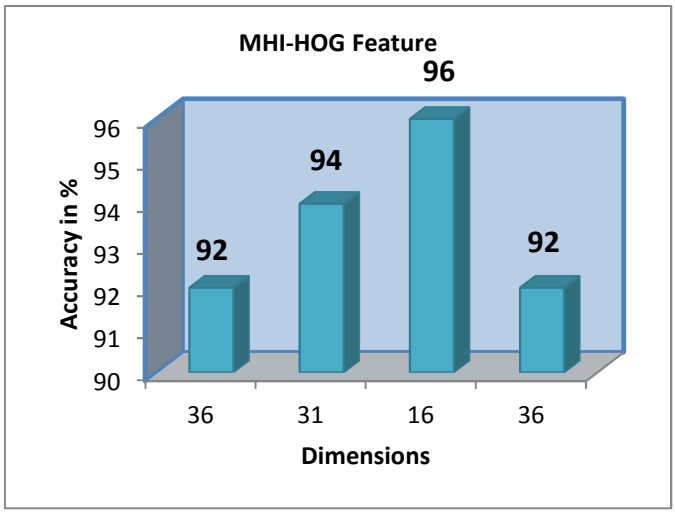

Fig. 9. HOG performance

8 PERFORMANCE MEASURE

Performance of classification is measured by using the matrices like Precision, Recall and F-measure. The accuracy can be computed by, the correctly recognized samples are true positives (TP), the incorrectly recognized samples are false positives (FP), correctly recognized which are not belong to the class are true negatives (TN), incorrectly recognized either assigned to the class are false negatives (FN). Recall gives how good an event is identified correctly. Specificity gives a measure of how good a method is identifying negative activity correctly. Precision is a measure of exactness and F-measure is the harmonic mean of Precision and Recall. Finally, Accuracy (A) shows the overall correctness of the event recognition. The statistical measures of Sensitivity (Recall), Specificity, Precision, F-measure and Accuracy is defined as

$$Precision = \frac{No.of\ TP}{No.of\ TP + No.of\ FP}$$

$$Recall = \frac{No.of\ TP}{No.of\ TP + No.of\ FN}$$

$$\mathrm{F_{measure}} = \frac{2 * \mathrm{precision} * \mathrm{recall}}{\mathrm{precision} + \mathrm{recall}}$$

$$Accuracy = \frac{TP + TN}{TN + FP + TP + FN}$$

Table 2 Performance measure of various tennis events

Classifier	Events	Precision	Recall	F-measure
SVM	Fault/Double fault	73.2	62.0	66.6
	Net Approach	68.1	64.8	65.2
	Baseline Rally	78.2	68.2	72.4
HMM	Fault/Double fault	94.9	99.1	96.9
	Net Approach	92.0	90.3	91.1
	Baseline Rally	93.2	88.8	90.9

Table 3 Confusion matrix for HMM in %

Events	Fault/Double fault	Net Approach	Baseline Rally
Fault/Double fault	**96.4**	4.5	2.1
Net Approach	6.3	**95.0**	1.7
Baseline Rally	2.6	4.2	**96.2**

Table 4 Confusion matrix for SVM in %

Events	Fault/Double fault	Net Approach	Baseline Rally
Fault/Double fault	**73.2**	19.5	7.3
Net Approach	18.6	**68.1**	13.3
Baseline Rally	4.6	17.2	**78.2**

9 CONCLUSION

In this proposed method for event detection and classification in Broadcast Tennis Video (BTV) using HOG feature from MHI. The events such as fault, rally and net approach are detected using the proposed method. Then from that detected event the MHI is generated which represents a motion and position of the player and HOG feature is extracted. This feature evaluated using SVM, gives an accuracy around 86% only because of conflict between the fault and net approach events. To achieve better results, go for HMM gives accuracy around 96% also performs well and also achieved good recognition results for all events**.**

REFERENCES

[1] Xu, G., Ma, Y.F., Zhang, H.J. and Yang, S, "A HMM based semantic analysis framework for sports game event detection," In *Image Processing, International Conference on ICIP Proceedings,* Vol. 1, pp. 1-25, 2003.

[2] Almajai, I., Kittler, J., de Campos, T., Christmas, W., Yan, F., Windridge, D. and Khan, A, "Ball event recognition using HMM for automatic tennis annotation," *17th IEEE International Conference on in Image Processing (ICIP),* pp. 1509-1512, 2010.

[3] Johnson, D.O. and Agah, A, "Recognition of Marker-less human actions in videos using hidden Markov models," *In Proceedings of the ICAI, pp*. 95-100, 2011.

[4] Jiang, Y.G., Bhattacharya, S., Chang, S.F. and Shah, M, "High-level event recognition in unconstrained videos," *International Journal of Multimedia Information Retrieval*, Vol. *2*, no. 2, pp.73-101, 2013.

[5] Tien, M.C., Wang, Y.T., Chou, C.W., Hsieh, K.Y., Chu, W.T. and Wu, J.L, "Event detection in tennis matches based on video data mining," *IEEE International Conference In Multimedia and Expo,* pp. 1477-1480, 2008.

[6] Xu, G., Ma, Y.F., Zhang, H.J. and Yang, S, "Motion based event recognition using HMM. In *Pattern Recognition," 2002. 16th International Conference on IEEE Proceedings.* (Vol. 2, pp. 831-834). 2002.

[7] Petkovic, M., Jonker, W. and Zivkovic, Z, "Recognizing Strokes in Tennis Videos using Hidden Markov Models," *In VIIP,* pp. 512-516, 2001.

[8] Kijak, E., Gravier, G., Gros, P., Oisel, L. and Bimbot, E, "HMM based structuring of tennis videos using visual and audio cues," *International Conference on ICME'03. Proceedings in Multimedia and Expo,* Vol. 3, pp. 300-309, 2003.

[9] Gravier, G., Demarty, C.H., Baghdadi, S. and Gros, P, "Classification-oriented structure learning in Bayesian networks for multimodal event detection in videos," *Multimedia tools and applications*, Vol. *70, no.* 3, pp.1421-1437, 2014.

[10] Vis, J.K., Kosters, W.A. and Terroba, A, "Tennis patterns: player, match and beyond," *In 22nd Benelux Conference on Artificial Intelligence (BNAIC 2010), Luxembourg,* pp. 25-26, 2010.

[11] Zhu, G., Xu, C., Huang, Q., Gao, W. and Xing, L, "Player action recognition in broadcast tennis video with applications to semantic analysis of sports game," *International conference on Multimedia In Proceedings of the 14th annual ACM,* pp. 431-440, 2006.

[12] Zelnik-Manor, L. and Irani, M, "Event-based analysis of video," *IEEE Computer Society Conference In Computer Vision and Pattern Recognition, Proceedings of the CVPR,* Vol. 2, pp. 120-123, 2001.

[13] Zhu, G., Xu, C., Gao, W. and Huang, Q, "Action recognition in broadcast tennis video using optical flow and support vector machine," *In Computer Vision in Human-Computer Interaction Springer Berlin Heidelberg*, (pp. 89-98), 2006.

[14] Chang, C.K., Fang, M.Y., Kuo, C.M. and Yang, N.C, "Event detection for broadcast tennis videos based on trajectory analysis," *2nd International Conference in Consumer Electronics, Communications and Networks (CECNet),* pp. 1800-1803, 2012.

[15] Xu, C., Zhang, Y.F., Zhu, G., Rui, Y., Lu, H. and Huang, Q, "Using webcast text for semantic event detection in broadcast sports video," *IEEE Transactions on Multimedia, Vol. 10*, no. 7, pp.1342-1355, 2008.

[16] Kapela, R., Świetlicka, A., Rybarczyk, A. and Kolanowski, K, "Real-time event classification in field sport videos," Signal *Processing: Image Communication*, Vol. *35*, pp.35-45, 2015.