IEEE/ACM Int'l. Conf. on Distributed Smart Cameras (ICDSC 09), September 2009

A Multi Camera System for Soccer Player Performance Evaluation

Marco Leo*, Tiziana D'Orazio* and Mohan Trivedi†
*Institute of Intelligent Systems for Automation
Bari, ITALY
Email: [leo,dorazio]@ba.issia.cnr.it
†Electrical and Computer Vision Engineering Laboratory
University of California
San Diego, CA, USA
Email: mtrivedi@soe.ucsd.edu

***Abstract*—This paper presents a multi-view approach to performance evaluation of soccer players by the analysis of the posture evolution. Some body-appearance features have been extracted and the most significant ones have been used to model the activity of the players involved in play. Continuous Hidden Markov Models have been used to model the temporal evolution of the body features in a multiple view decision making approach. Tests were carried out on different sequences of player activities extracted from matches played during the Italian "Serie A" Championship.**

I. Introduction

Some works have been presented in literature for tactical and team statistical analysis. Starting from the ball and player tracking information collected during the matches, these works try to extract tactical information for trainers ([2] and [3]), to evaluate player skills ([4],[5] and [6]) and to perform game analysis ([7] and [8]). These systems are not completely automatic but can greatly improve manual work during post processing analysis.

In [2] and [3] a system for showing goal events in a tactical mode to the coaches and sports professionals is described. After an initial phase in which goal events are detected by the analysis of web-casting text and broadcast video, tactical representations, known as "aggregate trajectory", are constructed based on multi-object trajectories using the analysis of temporal-spatial interactions among the players and ball. The acquisition of player trajectories in far-view shots is achieved by play-field detection using Gaussian mixture color models and Support Vector Classification on player candidates. A support vector regression particle filter keeps tracking the player in the frames. The distance covered by soccer players was measured in [4] with an automatic tracking system that is corrected manually when complex situations, such as when the player's trajectory changes during periods of occlusions, are not solved automatically. The segmentation and tracking phases for each game require respectively 6 and 4 hours of processing and in order to calculate the image-object transformation, before the games, 20 control points were established and measured directly onto the field.

In [5] a semi automatic system was developed to acquire player-possession for broadcast soccer video, its objective is to minimize the manual work. To acquire player possession the authors try to recognize the players touching the ball by assuming that they are those closest to it. Support Vector Machine methods are used to recognize the team of the player touching the ball. The view information and the player roles are used to produce the candidates for the player touching the ball. The selection among the possible candidates is done manually by the operator of the system.

Scout is a system presented in [7] for event game speed analysis and tracking. Background subtraction, connected component labelling, morphological filtering are used to segment and track moving objects. A vanishing point based method is proposed to map between the screen and physical coordinate systems. The system is designed to evaluate one player's performance at a time. In the case of complex player occlusions, more sophisticated tracking modules will be needed to avoid manual intervention. Analyzing the trajectories of moving objects, which consist of 22 players and a ball, in [6] much useful information is extracted in order to evaluate the performance of several players in a quantitative way. The proposed model is based on the trajectories of the players and the ball and their relationship. Several performance measures are introduced to analyze the performance concerning the interactions between the players of the same team and of their adversaries. The experiment was executed by collecting the trajectories of players and ball from a simulated soccer game. In [8] a model-based game analysis is carried out to map the real game process into an abstract representation obtained from features specifically designed for game analysis objectives. The system requires a real time positioning system that can continually track the position of the players' feet and the ball with an accuracy of a few centimeters. The authors propose the usage of *Cairos Technologies* (RFID-based technology) to collect the data and test their methodology to recognize passes, shots, and dribbles on Robot Cup Simulated soccer matches.

In this paper we propose a multi-view approach for player activity understanding. For the player performance analysis is not only important to measure the distance covered during the match by each player but also to evaluate how long the players have had an active role in the play. The aim of this

978-1-4244-4620-9/09/$25.00 ©2009 IEEE

work is to detect when a player is involved in the action either maintaining the ball control or interfering with opponent players who are conducting the play. The analysis of the body posture is fundamental in order to distinguish among different player activities. The proposed approach consists of two phases: an off-line learning phase and an on line testing phase. In the learning phase some sequences of player's involved activities are extracted, some body posture measures are evaluated and continuous Hidden Markov Models (HMM) are generated. In the test phase consecutive sliding windows of the player activities are provided to the HMM and the multi view behavior probability is evaluated. Tests were carried out on different sequences of player activities extracted from matches played during the Italian "Serie A" Championship.

The rest of this paper is organized as follows. Section II provides the system overview. In section III the body posture measures and their temporal statistical models are described; the multi view decision making procedure is detailed in section IV. Experimental results are reported in section V.

II. System Overview

The experimental set up consists of six cameras placed on the two sides of the field, assuring that each area of the pitch is covered by two opposite cameras. The imaging solution uses DALSA Pantera SA 2M30 cameras, achieving a resolution of 1920x1080pixels (Full HD) at 25fps. The particular model uses a single CCD operating using a Bayer filter. The acquired images are transferred to the six processing nodes by fiber optic cables. Each node is equipped with two Xeon processors (Nocona series) running at 3.4Ghz with hyperthreading enabled. Each node features 2GB of RAM and uses 8 SCSI disks with an individual size of 73GB (configured in RAID0) to store the match, and another 120 GB SATA disk for the operating system and the software. The graphics subsystem is handled by a Geforce 7900GT card with 256MB.

Each node processes the image sequences separately, while a supervisor node collects all the data and fuses the data to evaluate the final behavior likelihood. In figure 1 we plot the system overview. Each node extracts moving regions by a background subtraction algorithm and detects the player's team by an unsupervised classification approach; than a tracking algorithm extracts the player tracks solving group situations by a splitting procedure that uses the knowledge of classes of involved players. Details of these two steps can be found in [10] and [9]. In this paper we consider only the body posture feature extraction step and the successive temporal modeling. The player tracks are analyzed during their permanence in the image. For each frame of the sequence different body posture features have been extracted. The player dimensions can vary according to their posture and position in the field, for this reason the features were normalized by using the blob's area or height. We represented the player's body by extending the feature vector introduced in [1] in order to make it more suitable for our application context. Then the most significant features were selected and provided to one continuous HMM. During the training phase four players were selected and several sequences, extracted from the phases which the players were involved in, were used to train the HMM. Then in the test phase different sequences of the same players but also of different players were used to estimate the behavior activity recognition. Images acquired by opposite cameras are processed by the corresponding nodes and then provided to a central supervisor that fuses the data and takes the final decision on the behavior activity likelihood.

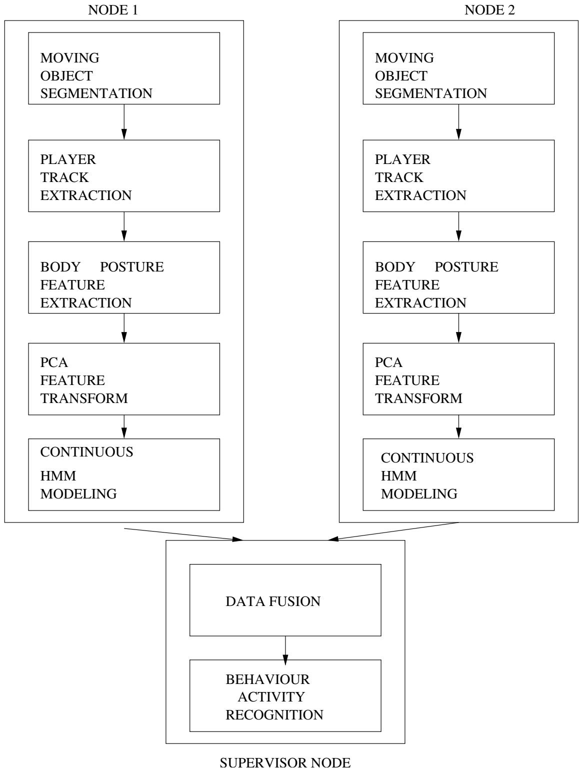

Fig. 1. System Overview

III. Player Activity representation

A. Body Posture Feature Extraction

Extracting the optimal set of features was subject of intensive work in the literature of recent years, since the selection of the most useful features, that have a high discriminative power, is fundamental to the success of the separation between classes. In particular, many works proposed a selection criterion based on metrics that maximize the interclass distance among posture classes. In our context we want to separate involved and not involved behaviors that are characterized not necessarily as variation of different postures, but as different evolutions of the same body posture. For this reason it is necessary not only to extract different body features that allow the discrimination among some player activities such as walking, standing, jumping, running, and so on, but also to evaluate their evolution to appreciate the variation of these activities and if the player is involved in the play. For example the running activity assumes different characteristics if the player is maintaining the ball control or he is just moving in the playing field. First of all we evaluated some body posture appearance features generating for the j player at frame k a feature vector

$$F_j^k = [O_j^k, W_j^k, L_j^k, R_j^k, Vx_j^k, Vy_j^k, H_j^k, U_j^k, \theta_j^k, B1_j^k, B2_j^k]$$

as follows. In this work we revised the F_j^k in [1] by adding more discriminating features for the considered application context. In particular, the considered features are the following:

- O_j^k: the number of foreground pixels divided by the area of the $j-th$ bounding box at frame K;
- W_j^k: Width of the bounding box divided by the height;
- L_j^k: Leftward span from the vertical major axis divided by the height;
- R_j^k: Rightward span from the vertical major axis divided by the height;
- Vx_j^k:Velocity of the $j-th$ bounding box along the x-axis of the image, divided by the height;
- Vy_j^k:Velocity of the $j-th$ bounding box along the y-axis of the image, divided by the height;
- H_j^k: Head ratio to height;
- U_j^k: Upper-body ratio to height;
- θ_j^k: Orientation of 2D Gaussian that represents the spatial distribution (x, y) of the $j-th$ foreground map with centroid $(\overline{X}, \overline{Y})$
- $B1_j^k$ width of the player's upper body divided by the height;
- $B2_j^k$ width of the player's legs divided by the height;

Naturally, the vector F_j^k, is not suited to be directly used as input to the classification step for two reasons: the high dimensional feature space makes it very difficult to build an efficient classifier (this problem is reported as *curse of dimensionality)* and, in addition, the measurements relative to considered features are not uncorrelated. A straightforward solution to these problems is to project the data onto low-dimensional subspaces obtained by dimensionality reduction transforms to extract the most significant and uncorrelated features.

B. PCA Body Posture Feature Transform

Principal component analysis (PCA) is a classical linear method that de-correlates the data and, at the same time, supplies a robust criteria to select the most significant features in the initial high dimensional set. PCA performs unsupervised dimensionality reduction by transforming a data set consisting of a large number of interrelated variables to a new set of uncorrelated variables, while retaining as much as possible the variations present in the original data set [11]. The key idea in PCA is that, by ordering the eigenvectors of the data covariance matrix according to the relative eigenvalues (largest first), one can create an ordered orthogonal basis with the first eigenvector having the direction of largest variance of the data. In this way, we can find directions in which the data set has the most significant amounts of energy and a new representation of the initial data (with minimum mean-square error) can be obtained by projecting them onto the most significant eigenvectors. The minimum number of significant components has to be experimentally evaluated by analyzing the relative eigenvalues: the percentage of data variance retained in each component is considered and, it is a common strategy, to select the first n components which preserve at least the 90% of the overall data variance. For example, in figure 2, the percentage of variance of the initial data preserved in the eigenvectors of the covariance matrix in the case of 100 measurements of the 11 aforesaid features in F_j^k is shown: the first three components represent almost the 95% of data variance; if also the fourth component is added the 99% of the variance of the initial data is retained. This preliminary test shows that, in the case of player posture measurements, it is possible to project initial data onto the first three or four components and to obtain a low-dimensional feature space that is a more suited input for a classifier. As a consequence, in the proposed framework, a new 4 dimensional feature vector

$$\tilde{F}_j^k = [C_{1_j}^k, C_{2_j}^k, C_{3_j}^k, C_{4_j}^k]$$

is built where the C_i^k are the first principal components obtained projecting vector F on the orthogonal basis consisting of the most significant eigenvectors of the covariance matrix of selected set of data measurements. The feature vector $\tilde{F}$ is then the input to the classification phase detailed in the following subsections.

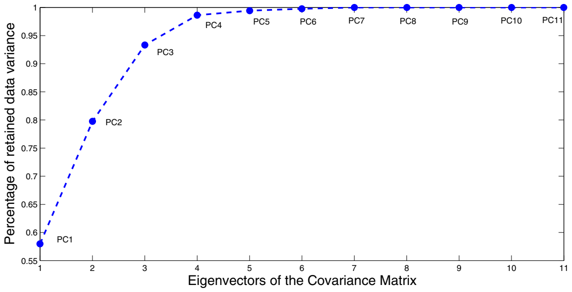

Fig. 2. Percentage of cumulative data variance preserved by the eigenvectors of the covariance matrix

C. Player Performance Evaluation in a Single View

In the recent years, a lot of approaches to recognizing human actions have been proposed. According to [12] these approaches can be categorized into three major classes: non-parametric, volumetric and parametric time series approaches. Nonparametric approaches typically extract a set of features from each frame of the video and then the features are then matched to a stored template. They have sufficient discriminating ability for several simple action classes such as sitting down, bending, crouching, and other aerobic postures but they lose discriminative power for complex activities due to overwriting of the motion history and hence are unreliable for matching [13]. Volumetric approaches consider a video as a 3-D volume of pixel intensities and extend standard image features such as scale-space extrema, spatial filter responses, etc., to the 3-D case. Unfortunately these approaches strongly depend on the the spatial and temporal activity execution, the body appearance, the noise, the possible occlusions and so on[14]. Parametric time-series approaches specifically impose a model on the temporal dynamics of the motion. The particular parameters for a class of actions are then estimated from training data. Examples of parametric approaches include Hidden Markov Models (HMMs), linear dynamical systems (LDSs), etc. One of the most popular statespace model is the Hidden Markov Model[15]. An HMM is a stochastic finite automaton, where each state generates (emits) an observation. Let X_t be the hidden state and Y_t the observation. Y_t might be a discrete symbol $Y_t \in 1...L$ or a feature-vector $Y_t \in \Re^L$. The parameters of the model are the initial state distribution

$$\pi(i) = P(X_1 = i)$$

the transition model

$$A(i, j) = P(X_t = j|X_{t-1} = i)$$

and the observation model

$$P(Y_t|X_t)$$

with $\pi(\cdot)$ representing a multinomial distribution. The transition model is usually characterized by a conditional multinomial distribution: $A(i, j) = P(X_t = j|X_{t-1} = i)$, where A is a stochastic matrix (each row sums to one).

If the observations are discrete symbols, we can represent the observation model as a matrix:

$$B(i, k) = P(Y_t = k|X_t = i).$$

If the observations are vectors in $\Re^L$, it is common to represent $P(Y_t|X_t)$ as a Gaussian:

$$P(Y_t = y|X_t = i) = N(y, \mu_i, \Sigma_i)$$

where $N(y, \mu_i, \Sigma_i)$ is the Gaussian density with mean μ and covariance Σ evaluated at y:

$$N(y, \mu, \Sigma) = \frac{1}{(2\pi)^{\frac{L}{2}}|\Sigma|^{\frac{1}{2}}} \exp(-\frac{1}{2}(y - \mu)'\Sigma^{-1}(y - \mu))$$

A more flexible representation is a mixture of M Gaussians:

$$P(Y_t = y|X_t = i) = \sum_{m=1}^{M} P(M_t = m|X_t = i)N(y, \mu_{m,i}, \sigma_{m,i})$$

where M_t is a hidden variable that specifies which mixture component to use, and $P(M_t = m|Xt = i) = C(i; m)$ is the conditional prior weight of each mixture component.

In this paper an HMM model consisting of 3 hidden states and using MoG (Mixture of Gaussian) [17] as state output is used for each camera view to automatically distinguish between players involved or not involved in play. The input of the HMM is the 4 dimensional feature vector $\tilde{F}$ obtained after the application of the PCA feature selection approach described in the previous section. A set of properly selected sequences containing player involved in play is used to train the HMM. The training phase is performed by using a generalization of the EM algorithm for HMMs (often called Baum-Welch). After that, the trained HMM determines the likelihood that a player is involved or not in play: to do that, a sliding window (sized as the training sequences) is used to select a piece of the test sequence and then the selected subsequences are sequentially given as input to the HMM that computes the likelihood of observing a sequence relative to a player involved in play.

IV. Multi-view Player Involvement Evaluation

The player activity detection can be greatly improved if multiple views are considered. In the soccer context, especially when the players are involved in play, they move very close occluding each other in the camera's field of view. In these cases the opposite views can perceive different situations and their collaborative process increases the final likelihood of the activity recognition. Decision fusion approaches aim at combining the beliefs of the set of models used into a single, consensus belief. In this paper we tested two popular decision fusion approaches, i.e. the linear opinion pool and the logarithmic opinion pool. The linear opinion pool is a commonly used decision fusion technique that is convenient because of its simplicity [16]. The fusion output is evaluated as a weighted sum of the probabilities from each model.

$$P_{linear}(A) = \sum_{i=1}^{k} \alpha_i P_i(A) \tag{1}$$

where $P_{linear}(A)$ is the combined probability employing a set of models used for an event A; α_i is the weight given to the i-th model; $P_i(A)$ is the probability of the i-th model for the event A; and k is the number of models. The parameters α_i are generally chosen such that $0 \leq \alpha_i \leq 1$, and $\sum_i \alpha_i = 1$. The linear opinion pool is appealing in that the output is a probability distribution, and the weight α_i provides a rough measure for the contribution of the i-th model. However, the probability distribution of the combined output, namely, $P_{linear}(A)$, may be multimodal. An alternative to the linear opinion pool we tested also the log opinion pool. If the weights are constrained such that $0 \leq \alpha_i \leq 1$, and $\sum_i \alpha_i = 1$, then the log opinion pool also yields a probability distribution. However, as opposed to the linear opinion pool, the output distribution of the log opinion pool is typically unimodal. The log opinion pool consists of a weighted product of the model outputs

$$P_{linear}(A) = \prod_{i=1}^{k} P_i(A)^{\alpha_i} \tag{2}$$

Note that with this formulation, if any model assigns a probability of zero, then the combined probability is also zero. Hence, an individual model has the capability of a veto, whereas in the linear opinion pool, the zero probability is averaged out with other probabilities.

In this work, the α values in equations 1 and 2 are dynamically computed on the basis of a parameter measuring the quality of the feature in the vector $\tilde{F}$. As proposed in [1] each feature component of the vector $\tilde{F}$ has assigned a weight

$$W = (w_{c_1}, w_{c_2}, w_{c_3}, w_{c_4})$$

such that $|W| = 1$. The deviation of the appearance fidelity between consecutive frames for the i-th camera was then defined as:

$$D_F^i = \Omega(|W * (\tilde{F}_j^k - \tilde{F}_j^{k-1}|)$$

where Ω is a vectorial function that returns the largest element among its arguments. Therefore, the α values in equations 1 and 2 were computed as follows:

$$\alpha_i = (1 - \frac{D_F^i}{\Sigma_{j=1}^M D_F^i})$$

where M is the number of cameras having the considered player in their field of view at the considered time instant (in our experiments M=2). In this way the multi-view score takes into account the robustness of the feature extracted from each camera and then so the player's activity performance is more reliable than that obtained using a single view. Finally, in order to take a final decision and automatically label each frame of the test sequences containing a player involved in the play, a learned threshold is used. The threshold was set as the mean value of the log-likelihood output of the trained HMM when the training sequences are provided as input.

V. Experimental Results

The experiments were carried out on image sequences acquired during a real soccer game of the Italian "Serie A" championship. The HMM was trained using 4 pairs of sequences relative to 4 different players involved in play and acquired from opposite cameras. In particular the first pair of sequences was relative to a player receiving and then carrying the ball; the second pair was relative to a player shooting the ball, the third one to a player trying to challenge the ball carrier and, finally, the fourth pair was relative to a player quickly running forward to receive a forward pass. The training sequences were manually segmented and each of them was 50 frames long: actually, in the considered context, shorter training sequences did not provide adequate information to properly model the player activities and, on the other side, longer sequences require complex HMM architecture that usually are difficult to design and, in addition, often generate over-fitting problems.

After the manual segmentation, for each patch of the training sequences ($50 \times 8 = 400$ patches), the 11-dimensional feature vector F described in section III was built. These feature vectors were then used to generate the training data matrix A (each feature vector is a row, producing a matrix A sized 400×11). As described in section III, the eigenvectors of the matrix A were computed and the most significant directions in data were pointed out. The first four principal directions were then used to project initial training data and to represent them by a new 4-dimensional vector $\tilde{F}$ of uncorrelated features. Finally, the vectors $\tilde{F}$ were sequentially provided to the HMM in order to model the temporal evolution of the features corresponding to a player involved in play.

In figure 3 some patches extracted from the training sequences are reported. From left to right, players challenging the ball carrier, quickly running forward, receiving the ball from a teammate and shooting the ball are shown.

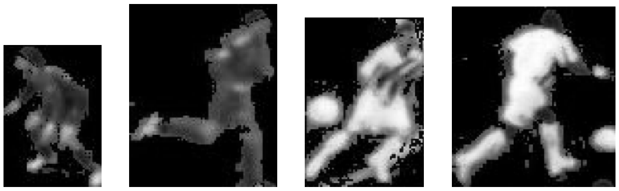

Fig. 3. Some patches extracted from the training sequences

The trained HMM was then tested on 10 pairs of sequences having different length. The 20 test sequences were preliminary observed by a human operator in order to build the ground truth of the player involvement in the play. In table I the ground truth generated by the human operator is reported.

TABLE I
Ground Truth relative to the 10 pairs of sequences used to test the proposed framework.

ID of the pair of Sequences	Cameras	Length	Frame intervals containing a player involved in play
1	3-4	1144	[65;158] [225;265] [530;600]
2	3-4	1150	[760;840]
3	1-2	965	[128;211] [554;721]
4	1-2	482	[327;351]
5	3-4	687	[441;580]
6	3-4	954	[125;231] [441;687]
7	5-6	447	[224;341]
8	1-2	258	[112;195]]
9	5-6	444	[224;341]
10	5-6	425	[124;376]

In figure 4 three pairs of patches relative to the same player acquired by opposite synchronized cameras are shown. It is possible to observe that the player appearances can strongly vary depending on the position of the player with respect the camera. This is particularly evident for the two patches on the right in figure 4 in which the same player appears with different sizes and strongly differs in his body configuration.

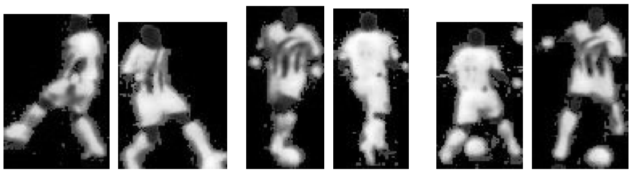

Fig. 4. Three pairs of patches extracted from opposite cameras and relative to the same test sequence

During experimental tests, first of all the likelihood of player involvement in the play from a single view was computed: for each sequence a sliding window (50 frames long) was used to select the portion of the sequence to be analyzed and then for each selected patch the 11-dimensional feature vector F described in section III was computed. The feature vector F was then linearly transformed by projecting it onto the orthogonal basis defined statistically analyzing the training data by the PCA method. The resulting 4-dimensional feature vector $\tilde{F}$ was finally given as input to the trained HMM in order to get the likelihood of observing, in the considered temporal window, a player who is involved in the play.

In figure 5 the likelihood values (in logarithmic representation) relative to the sequences of the first pair in table I are shown (the likelihoods of the two opposite cameras 3 and 4 are reported). The lowest logarithmic likelihood values correspond to the time intervals in which the player was most probably involved in play. The differences between the two graphs demonstrate how the same player is viewed differently by the opposite cameras. To compare these behaviors we plotted in figure 6 the two likelihood curves superimposed with the ground truth. The red line is relative to the likelihood values of the camera 3 and the blue line is relative to the likelihood values of the camera 4. The three areas delimited by vertical dot lines correspond to the frame intervals which the human operator indicated as involved in the play (see first row in table I).

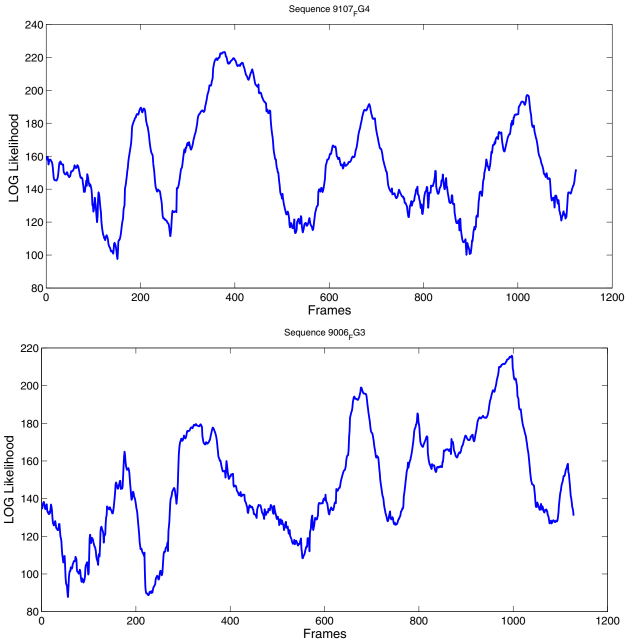

Fig. 5. The likelihood values computed for the two sequences of the first row in table I. On the top the results obtained by the camera 3, on the button the results obtained by the camera 4.

Then, as described in section IV, the likelihood scores coming from opposite views and relative to the same player were combined in order to get a multi-view likelihood score. Two different strategies were tested, i.e. the log and linear opinion pool described in section IV.

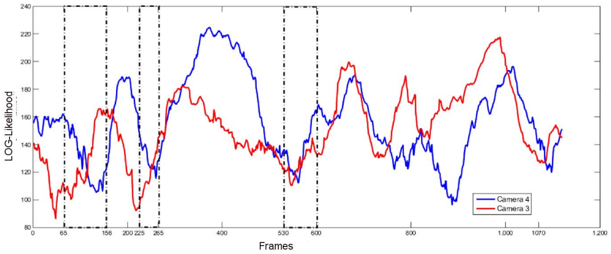

Fig. 6. The two curves of figure 5 superimposed with the ground truth intervals. The red and blue curves are relative to the cameras 3 and 4 respectively.

In figure 7 the deviation D_F of the appearance fidelity between consecutive frames for the sequence 1 acquired by camera 4 is reported. The peaks in this plot represent the points in which features are not reliable and the the α values have to give a greater weight to the opposite view. Figure 8 reports the multi-view likelihood score computed by the log and linear opinion pool formulas. These graphs are quite similar, anyway in the following tables some different behaviors between the two functions can be appreciated.

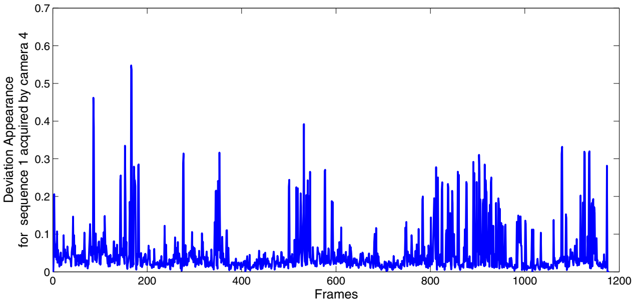

Fig. 7. Deviation of the appearance fidelity between consecutive frames for the sequence 1 acquired by camera 4

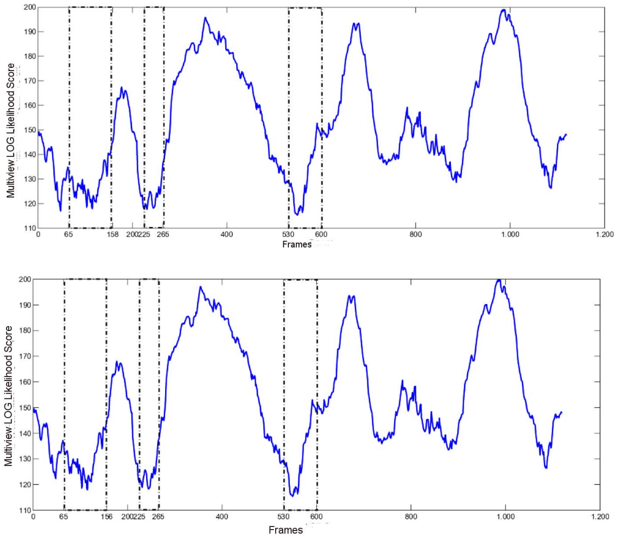

Fig. 8. Log opinion pool (top) and linear opinion pool (bottom) scores for the first pair of sequences in table I.

In order to take the final decision about the player involvement we used $th = 123.0459$ as a decision threshold, i.e. the mean value of the likelihood values obtained providing the 8 training sequences as input to the trained HMM. In figure 9 we report the output values of the trained HMM relative to the eight training sequences (red point) and the corresponding threshold value (red line) computed as their mean value.

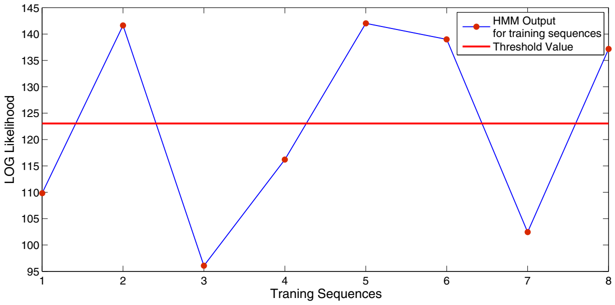

Fig. 9. The threshold value selected for the decision making process about player involvement in play

The system performances for all the test sequences were evaluated by using the following error measure E:

$$E = \frac{F_N + F_P}{F}$$

where F_N and F_P are the number of false negative and false positive labels and F is the total number of frames in the considered sequence. In table II the results obtained on 10 pairs of test sequences using log pool score are reported whereas in table III the corresponding results obtained using linear pool score are reported.

TABLE II
THE RESULTS OBTAINED ON 10 TEST SEQUENCES USING LOG POOL SCORE.

ID of the pair of Sequences	**Number of Frames**	F_N	F_P	P
1	1144	25	33	0.05
2	1150	41	28	0.06
3	965	55	33	0.091
4	482	44	11	0.114
5	687	11	12	0.033
6	954	66	18	0.088
7	447	41	11	0.116
8	258	35	49	0.325
9	444	19	15	0.076
10	425	11	20	0.072
Overall	**6956**	**348**	**230**	**0.083**

TABLE III
THE RESULTS OBTAINED ON 10 TEST SEQUENCES USING LINEAR POOL SCORE.

ID of the pair of Sequences	**Number of Frames**	F_N	F_P	P
1	1144	28	24	0.045
2	1150	37	26	0.054
3	965	46	38	0.087
4	482	29	19	0.099
5	687	18	18	0.052
6	954	38	14	0.054
7	447	32	13	0.100
8	258	31	25	0.217
9	444	22	18	0.090
10	425	18	22	0.094
Overall	**6956**	**299**	**217**	**0.074**

Tables II and III demonstrate that the proposed approach can discriminate in many cases if the players are involved or not in the play. In addition it is possible to conclude that the choice of pool score strategies did not alter the experimental results, even if, in general the linear pool score lightly outperformed the log pool score in many cases. Finally it is possible to observe that, generally, wrong labeling occurred in presence of misleading body configurations: for example, in the test sequence number 7, most of the false positive occurrences were generated because the player raised his arms (most probably to draw a teammate's attention). In that case the system wrongly recognized the player as involved in the play whereas the human operator did not. As clearly visible in figure 10 the player body configuration of the player with the arms raised (on the left) is very similar to that of a player shooting the ball and included in the training sequences (on the right).

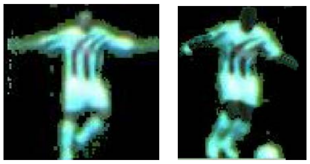

Fig. 10. Two very similar patches however containing a player not involved (on the left) and involved (on the right).)

VI. Conclusion

This paper presented a multi-view approach for performance evaluation of soccer players by the analysis of the posture evolution. Some body-appearance features have been extracted and transformed in an uncorrelated vectorial space defined by the Principal Component Analysis. The most significant components have been used to model the player activity during involved or not involved situations in the play. Continuous Hidden Markov Models were used to model the temporal evolution of the body features in a multiple views decision making approach. The temporal model was finally used to automatically recognize if a player was involved or not actively in the play in 10 long test sequences. Experimental tests carried out during Italian "Serie A" matches, demonstrated the reliability of the proposed approach. Future works will deal with a more detailed automatic player performance analysis performed by introducing multiple HMMs to model different activities as running, kicking, receiving the ball, and so on.

References

[1] S.Park, M. Trivedi, *Understanding human interactions with track and body synergies (TBS) captured from multiple views*, Comput. Vis. Image Underst., vol. 111 (1), 2008, pp. 2–20

[2] Guangyu Zhu, Qingming Huang, Changsheng Xu, Yong Rui, Shuqiang Jiang, Wen Gao, Hongxun Yao, *Trajectory based event tactics analysis in broadcast sports video*, MULTIMEDIA '07: Proceedings of the 15th international conference on Multimedia, 24-29 September, 2007, pp. 58-67, Augsburg, Germany

[3] Guangyu Zhu, Changsheng Xu, Yi Zhang, Qingming Huang, Hanqing Lu, *Event Tactic Analysis Based on Player and Ball Trajectory in Broadcast Video*, Conference on Image and Video Retrieval (CIVR), Niagara Falls, Canada, 7-9 July, 2008

[4] Barros, R.M.L. and Misuta, M.S. and Menezes, R. P. and Figueroa, P.J. and Moura, F.A. and Cunha, S. A. and Anido , R. and Leite, N.J. , *Analysis of the distances covered by first division Brazilian soccer players obtained with an automatic tracking method*, Journal of Sports Science and Medicine, vol. 6, n. 2, June, 2007, pp. 233-242

[5] X. Yu, T.Sen Hay, X. Yan, E. Chng, *A player possession acquisition system for broadcast soccer video*, Proceedings of the IEEE International Conference on Multimedia & Expo, Amsterdam,The Netherlands, 6-8 July, 2005

[6] C. kang, J. Hwang, N. K. Li, *Trajectory Analysis for Soccer Players*, Proceedings of the 6th IEEE Int. Conference on Data Mining - Workshops (ICDMW'06), Hong Kong, China, 18 December,2006

[7] P.S. Tsai, T. Meijome, P.G. Austin, *Scout: a game speed analysis and tracking system*, 2007,Machine vision and Application, vol. 18, n. 5, October, pp. 289-299

[8] M. Beetz, B. Kirchlechner, M. Lames, *Computerized real-time analysis of football games*, Pervasive Computing, IEEE, num. 3, pp. 33-39, vol. 4, July-September, 2005

[9] P.Spagnolo, T.D'Orazio, M.Leo, N.Mosca, M.Nitti, *A Background Modelling Algorithm based on Energy Evaluation*, VISAPP 2006: Proceedings of the International Conference on Computer Vision Theory and Applications,2006

[10] T.D'Orazio, M.Leo, P.Spagnolo, P.L.Mazzeo, N.Mosca, M.Nitti, *A Visual Tracking Algorithm for Real Time People Detection*, Wiamis2007: Proceedings of the International Workshop on Image Analysis for Multimedia Interactive Services, 2007

[11] I. T. Jolliffe, *Principal component analysis,second edition*, Springer Serires in Statistics,2002

[12] P. Turaga, R. Chellappa, V.S. Subrahmanian, O. Udrea, *Machine recognition of Human Activities: A Survey*, IEEE Transactions on Circuits and Systems for Video Technologies, vol.18, No. 11, November 2008

[13] A.F. Bobick, J. Davis, *The Recognition of Human Movements using Temporal Templates*, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 23, No.3, pp.257-267, 2001

[14] J. C. Niebles, H.Wang, and L. F. Fei, *Unsupervised learning of human action categories using spatial-temporal words*, Proc. British Mach. Vis. Conf., pp. 12491258, 2006

[15] L. R. Rabiner, *A tutorial on hidden Markov models and selected applications in speech recognition*, Proc. IEEE, vol. 77, no. 2, pp. 257286, Feb. 1989

[16] I. Bloch, *Information combination operators for data fusion: a comparative review with classification*, IEEE Trans. Syst. Man Cybern.Part A: Syst. Humans 26 (1996), pp. 5267

[17] J. Bilmes, *A Gentle Tutorial on the EM Algorithm and its Application to Parameter Estimation for Gaussian Mixture and Hidden Markov Models*, Technical Report, University of Berkeley, ICSI-TR-97-021, 1997.