Trajectory-based handball video understanding

Alexandre Hervieu
INRIA Rennes - Bretagne
Atlantique
Campus de Beaulieu
35042 Rennes Cedex, France
ahervieu@irisa.fr

Patrick Bouthemy
INRIA Rennes - Bretagne
Atlantique
Campus de Beaulieu
35042 Rennes Cedex, France
bouthemy@irisa.fr

Jean-Pierre Le Cadre
IRISA/CNRS
Campus de Beaulieu
35042 Rennes Cedex, France
lecadre@irisa.fr

ABSTRACT

This paper presents a content-based approach for understanding handball videos. Tracked players are characterized by their 2D trajectories in the court plane. The trajectories and their interactions are used to model visual semantics, *i.e.*, the observed activity phases. To this end, hierarchical parallel semi-Markov models (HPaSMMs) are computed in order to take into account the temporal causalities of object motions. Players motions are characterized using velocity informations while their interactions are described by the distances between trajectories. We have evaluated our method on real video sequences, and have favorably compared with another method, *i.e.*, hierarchical parallel hidden Markov models (HPaHMMs).

Categories and Subject Descriptors

I.4.8 [Computing Methodologies]: Image Processing and Computer Vision—*Scene Analysis*; I.5.4 [Computing Methodologies]: Pattern Recognition—*Applications*

Keywords

Sports videos, Computer vision, Semi-Markov models, Motion analysis, Pattern classification.

1. INTRODUCTION

Understanding human behavior and complex activities is of an increasing interest in the computer vision field. It is motivated by several applications in various domains such that video surveillance, sports video indexing... In these contexts, trajectories of mobile objects provided by tracking systems may be exploited. These trajectories form a high level description of dynamical contents in videos.

Several works have investigated the issue of exploiting mobile objects trajectories for video semantic analysis. Günsel et al. proposed a video indexing framework based on the analysis of “video objects” [2]. Their method relies on the dynamics of the tracked objects and the interactions between the corresponding trajectories. Based on these interactions,

other methods were developed to handle video content understanding [17, 5, 15, 16]. However, no practical extensive application of these methods has yet been reported.

Hence, in the sequel, we will concentrate on the analysis of video trajectories in a particular context: sports videos, and more specifically handball videos. Indeed, such activities are governed by rules and occur in a *a priori* known closed space, providing a favorable experimentation field for high-level video analysis methods [11].

Techniques relying on video trajectories have already been proposed for sports video activity analysis [4, 21, 1, 10]. Perše et al. proposed a basketball videos analysis method [19] that first segments tracked players trajectories into three different classes of basketball activities (offense, defense and timeout) [19]. It relies on Gaussian mixtures and an *EM* algorithm trained on manually labeled sequences. Then, using a partition of the court, a second step achieves a template-based complex activity recognition of the offense trajectory segments into three different classes.

We have also used players trajectories to reliably segment squash videos into “rally” and “passive” phases [3]. Our method involved a two-state semi-Markovian modeling of players’ trajectories characterized by invariant features in the image plane. In this paper, we extend our semi-Markovian modeling to express more complex activities and apply it to handball videos. First, a representation of the trajectories of players belonging to a given team is designed. It takes into account both dynamical and interaction information. Then, semi-Markovian models are developed to handle complex video interpretation including eight different activity classes. Audio information is also considered to help the interpretation and is integrated in the whole process.

In Section 2, we will describe the chosen representation that takes into account the dynamics of the players, the interactions between their corresponding trajectories and the specific role of the goal-keeper. This representation is then modeled using a eight-state hierarchical parallel semi-Markovian model (HPaSMM) which is used for handball video interpretation and presented in Section 3. This section also deals with the audio data processing and presents another method for comparison: the hierarchical parallel hidden Markov model (HPaHMM). Section 4 describes the processed data. Then, experiments and results are reported and discussed in Section 5. Finally, Section 6 gives information on computation time.

“Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or fee.

CIVR’09, July 8-10, 2009 Santorini, GR

Copyright © 2009 ACM 978-1-60558-480-5/09/07...\$5.00”

2. HANDBALL TEAM ACTIVITY FEATURES

This paper presents a method to retrieve phases of activities in handball videos. To this aim, motions (*i.e.*, the trajectories) of the players of a single team, in the court plane, are exploited. Input trajectories are presented in Section 4. Figure 3 shows the entire set of computed trajectories in the court plane.

The characterization of handball team activity phases is based upon dynamics of the players and interactions between players. Three main ideas led the choice of the considered features:

- to concentrate interaction information within a limited number of features,
- to exploit the specific role of the goalkeeper,
- to take into account the global dynamics of fielders (*i.e.*, any handball player except the goal keeper).

To handle these three issues, five features are computed at each time instant t :

- the mean distance between the goalkeeper and the six fielders $d_{GF,t}$,
- the mean distance between the six fielders $d_{F,t}$,
- the min, mean and max values of distance between successive positions (*i.e.*, the distance covered between $t - 1$ et t) of the six fielders, denoted by $d_{intramin,t}$, $d_{intramean,t}$ and $d_{intramax,t}$.

Computed distances correspond to Euclidean distances between trajectories in the court plane. At time t , $d_{GF,t}$ and $d_{F,t}$ help specifying the interactions between players of a given team. The three last distances $d_{intramin,t}$, $d_{intramean,t}$ and $d_{intramax,t}$ give a global information, at every time instant t , on the dynamics of the fielders. Hence, characterizing team activity phases is achieved by five features:

$$d_{GF,t}, d_{F,t}, d_{intramin,t}, d_{intramean,t}, \text{ and } d_{intramax,t}.$$

The five features accounts both for the interactions between handball players and their dynamics. The values $d_{GF,t}$ are gathered in the vector D_{GF} :

$$D_{GF} = [d_{GF,1}, \dots, d_{GF,n-1}, d_{GF,n}],$$

where n is the size of the processed trajectories. The same holds for the other features.

To reduce computation time, a grouping procedure of the feature values is performed. For the five considered feature vectors D_{GF} , D_F , $D_{intramin}$, $D_{intramean}$ and $D_{intramax}$, groups of k_{group} consecutive values are formed. For each of these groups, the mean values are computed and used to construct five new feature vectors $\tilde{D}_{GF}$, $\tilde{D}_F$, $\tilde{D}_{intramin}$, $\tilde{D}_{intramean}$ and $\tilde{D}_{intramax}$ (having sizes k_{group} times smaller than D_{GF} , D_F , $D_{intramin}$, $D_{intramean}$ and $D_{intramax}$).

3. HPASMM MODELING OF HANDBALL ACTIVITY PHASES

To model activity phases observed in handball videos, we propose a hierarchical parallel semi-Markov model (HPaSMM). The proposed modeling is based upon a two-level hierarchy. In the lower layer, feature vectors are modeled using parallel hidden Markov models (PaHMMs, see [22]). These PaHMMs are used to characterize the states of the HPaSMM.

The states of the HPaSMM, denoted by S_i , compose the upper layer of the modeling. Each of these states correspond to one of the considered activity phases. Figure 1 contains an illustration of the developed modeling.

3.1 Activity modeling using semi-Markovian models

HPaSMM states S_i describe the activity phases. Eight phases of activity are considered, defined along with their numbering as follows:

- “slowly going into attack”: activity phase 1,
- “attack against set-up defense”: activity phase 2,
- “offense free-throw or timeout”: activity phase 3,
- “counterattack, fast break”: activity phase 4,
- “returning, preventing from fast break”: activity phase 5,
- “slowly returning in defense”: activity phase 6,
- “defense”: activity phase 7,
- “defense free-throw or timeout”: activity phase 8.

In the proposed HPaSMM, mixtures of Gaussian models (GMMs) are used to model the duration of the activity phases denoted by sd_i . These GMMs are fitted using “forward-backward” procedures. Initialization is performed using k-means algorithm. The set of parameters involved in the state duration models is denoted by ψ .

3.2 Feature modeling using PaHMMs

Features describing the dynamics of the trajectories and the interactions between trajectories $d_{GF,t}$, $d_{F,t}$, $d_{intramin,t}$, $d_{intramean,t}$, and $d_{intramax,t}$ are used to characterize the eight phases of activity S_i . They are denoted, for each activity phase S_i , as d_{GC}^i , d_C^i , $d_{intramin}^i$, $d_{intramean}^i$, and $d_{intramax}^i$. For each state of the upper level, the five corresponding feature vectors are modeled using a five-layer PaHMMs. The PaHMMs are here defined in a similar way than those presented in [22], where conditional probabilities B of observations are fitted by GMMs.

The set of trajectories extracted from the training videos is used to find the set of PaHMMs parameters denoted by ϕ . ϕ is composed of B , A' (state transition matrix) and π (initial state distribution) for each of the five considered features and for each activity phase.

3.3 Training procedures and Viterbi algorithm for activity phases retrieval

Suppose a upper level state sequence S containing R segments, *i.e.*, R successive phases of activity. Each of these activity phases is associated with a state S_i . Let q_r be the time index of the end-point of the r^{th} segment, the sequence of observations corresponding to segment r is given by $y_{(q_{r-1}+1, q_r]} = y_{q_{r-1}+1}, \dots, y_{q_r}$ such that $S_{q_{r-1}+1} = \dots = S_{q_r}$. A is the upper level HPaSMM state transition matrix, computed using transitions between activity phases in the training videos.

The entire set of parameters is finally given by $\theta = \{A, \phi, \psi\}$ and is estimated by a supervised learning stage. θ is then used to perform temporal activity phases recognition using a Viterbi algorithm.

The Viterbi algorithm gives the sequence of upper level HPaSMM states $\hat{S}$ maximizing the log-likelihood, *i.e.*, such that

Figure 1: HPaSMM modeling of handball composed of eight upper level states S_i . Each state corresponds to a given phase of activity. A phase of activity is modeled by a five-layer PaHMM (one layer for each feature vector) and by a GMM modeling state durations (sd_i is the state duration associated to state S_i).

$\hat{S} = \arg \max_S \log P(y, S|\theta)$. The likelihood $P(y, S|\theta)$ is defined, for an observation sequence y and a sequence of upper level state S , by:

$$\begin{aligned}
 P(y, S|\theta) &= \prod_{r=1}^R P(S_r|S_{r-1}) \\
 &\times \prod_{r=1}^R P(sd_i = q_r - q_{r-1}|\psi; S_{q_r}) \\
 &\times \prod_{r=1}^R P(y_{(q_{r-1}+1, q_r]}|\phi; S_{q_r}).
 \end{aligned}$$

3.4 Integrating audio information: detection of referee whistles

In order to facilitate the recognition of activity phases, we take into account audio information. Indeed, audio data contain important information to specify some phases as given by referee whistles.

Whistle instances in the audio stream are found using two softwares developed in our laboratory, Spro and Audioseg, which are available online [7, 6]. Spro first produces the description of the whistle request and of the processed audio stream contained in the handball video. This description is based upon cepstral coefficients of mel frequency [13] computed in time intervals defined using a sliding window. Then, Audioseg performs recognition of the request within the audio stream [14]. It compares the coefficients of each audio intervals using a Dynamic Time Warping procedure. This information is simply integrated in the HPaSMM method by considering that each whistle corresponds to an activity phase change. Hence, in the Viterbi decoding al-

gorithm, the hypothesis is made that as soon as a referee whistle is detected, the current activity phase is stopped and another phase begins. This gives a partition of the observed actions into successive segments Seg_k , where a segment is bounded by two referee whistles. Each segment is decoded separately by the Viterbi algorithm with the only hypothesis that the first activity phase of a segment Seg_{l+1} is different from the last activity phase found for the previous segment Seg_l .

3.5 Another method for comparison: hierarchical parallel hidden Markov models

Hierarchical parallel hidden Markov models (HPaHMMs) have a similar architecture as the HPaSMMs previously defined. In contrast, HPaHMMs do not model specifically the state duration. In a HPaHMM, the state duration follows a simple geometric law given by:

$$p(d_i) = a_{ii}^{d_i-1}(1 - a_{ii}),$$

where a_{ii} is the probability of staying in state S_i from time t to $t + 1$.

Hence, the training procedure of HPaHMM is very similar to the HPaSMMs one except for matrix A . Indeed, HPaHMMs training procedure computes A by searching the transitions between activity phases only at times q_r (the time index of the end-point of the r^{th} segment). For HPaHMMs training procedure, every time instant is considered to learn the activity-phases transition matrix A . To recognize phases of activity using a HPaHMM, a Viterbi algorithm is considered again.

Figure 2: Three images corresponding to the same instant in three videos of the same handball action. The two images on the left have been extracted from two bird-eye view videos.

4. EXPERIMENTAL DATASET

To test the proposed method, we have exploited a set of trajectories of handball players belonging to a same team available in the database [8]. These trajectories were extracted from two bird-eye view cameras, one above each half of the court plane. A modified color-histogram-based condensation tracking procedure was used to extract the trajectories from the video sequences [12, 20]. Database providers supervised the tracking procedure by correcting errors that may have appeared during the process. An appropriate calibration step was introduced to map the image coordinates in the court plane and to compensate for the observed radial distortion. The providers obtained an error estimation on players coordinates in the court between 0.3 and 0.5 m [18]. 25 samples of trajectory data (one per frame) per second are available. A Gaussian smoothing was finally applied to reduce the tracking jitter while preserving the measurement accuracy.

Figure 3.1 shows three images corresponding to one given time instant of a handball match. Two images were extracted from videos acquired by bird-eye view cameras, the third one from a video shot by a handheld camera. Figure 3 contains the entire set of trajectories of seven players of the same handball team. They correspond to approximately ten minutes of a handball game (14664 images).

The database also provides an annotation of the activity (using different phases of activity). We used this annotation to construct the activity ground-truth based upon the eight activity phases defined in Subsection 3.1. It allows us to check that every whistle induces a change in the activity phases.

We now describe the experiments carried out to test the proposed HPaSMM and HPaHMM methods for trajectory-based handball activity recognition.

Let us first point out that the available set of trajectories is rather small for training HPaSMM and HPaHMM. Indeed, these models would require more trajectories to be completely and efficiently trained. Trajectories corresponding to several hours of videos of handball games would be necessary to correctly handle diversity of content.

In the following, the sets of trajectories respectively used for training and testing the proposed methods correspond to different periods of the ten minutes handball game. Hence, no overlapping data were used for the training and the testing procedures. However, these training and testing sets of trajectories correspond to sequences with showing the same team in a same game (due to the lack of available trajec-

ries), which is not ideal. It would be an add to test the proposed models on trajectories corresponding to other game and other teams.

Figure 3: Set of available trajectories of seven players belonging to the same handball team (each trajectory, corresponding to a different player, is plotted in a given color and is extracted from a ten-minutes video sequence).

In order to train the upper level state transition matrix A and also the GMMs modeling the state duration, activity observations from other video streams have been included. Indeed, trajectories are not required to train this subset of parameters so that manually recognizing activity phases in other handball videos provides additional information in term of transitions between activity phases and of durations of activity phases. They have been extracted from 2008 Beijing Olympic Games handball final. The videos are available online [9].

In the sequel, we will evaluate the performance of the method as the ratio of correctly classified images (with respect to the eight defined activity phases) and the total number of processed images. To this end, ground truth on the entire set of trajectories is exploited. All the reported results have been obtained with k_{group} value set at 8.

5. EXPERIMENTS

We first test the HPaSMM and HPaHMM methods without considering audio information. A first part of the handball players trajectories has been used for training the HPaSMM and HPaHMM models. It includes 6370 images, *i.e.*, more than 4 minutes of video. The second part of the trajectory set has been used for testing and comprises 8294 images, *i.e.*, little less than 6 minutes of video.

A rate of 76.1% correct recognition has been obtained using the HPaSMM method. Results are plotted in figure 4 which contains the ground truth and the obtained results. Using the HPaHMM method, a rate of 72.7% has been reached. It can be explained by the lack of trajectories for training. 4 minutes is not sufficient to efficiently train such models. Moreover, most of the errors occur for the offense and defense free-throw or timeout activity phases.

To alleviate these shortcomings, we now exploit audio information contained in the audio stream. We first present the results of referees whistle extraction. It corresponds to phase changes and then it helps detecting free-throw or timeout. We have conducted leave-one-out cross validation method, which enables to consider more training data.

Figure 4: Plot of the recognition results obtained by the HPaSMM method. The first part of the trajectory set (corresponding to 6370 images) is used for training while the second part (corresponding to 8294 images) is kept for recognition test. Ground truth is plotted in red while the recognition results are plotted in blue. Numbers on the vertical axis correspond to the numbering of activity phases described in Subsection 3.1. The horizontal axis denotes the numbering of the successive group of images ($k_{group} = 8$).

5.1 Integrating audio data: referees whistle detection

The method for detecting referees whistles detection supplied satisfying results. As shown on Figure 5, it correctly detects 29 referees whistles among 31. The request was elaborated from a whistle extracted in another audio stream. However, these results were obtained by manually choosing the optimal detection threshold (drawn in red in Figure 5). A completely automatic referees whistle detection would require a threshold selection procedure.

5.2 Integrating audio information for handball activity phases recognition

We can now use the referees whistles (29 referees whistles were correctly detected in the ten-minutes audio stream) in the HPaSMM and HPaHMM methods.

5.2.1 Leave-one-out cross validation method

Taking into account referees whistle was first considered to help recognition of activity phases. It also facilitates a leave-one-out cross validation (LOOCV) test method. The set of trajectory is “cut” into 30 segments (defined by the 29 detected whistles). Viterbi decoding of one segment may be

Figure 5: Plot of the referees whistle detection results obtained on a ten-minutes handball audio stream. The horizontal axis describes the time. The distance between the successive sliding windows describing the audio stream and the whistle request is plotted in blue, detection threshold is drawn in red. Hence, a whistle instance is detected each time the distance between the audio stream and the request is below the chosen threshold. The 29 detected referees whistles are indicated in green, while the two missed ones are in brown.

computed independently from the other segments, the only information to integrate is that the first activity phase of a segment Seg_{k+1} must be different of the last activity phase of the segment Seg_k . Hence, a LOOCV method has been exploited by decoding each segment independently while using the 29 other segments for training HPaSMM and HPaHMM models.

For each test segment, a set of 29 segments is used for training models. This procedure allows us to rely on more important training sets. Before processing a given segment, a set of about 9 minutes long can be used to train the models.

5.2.2 Results using audio information and LOOCV method for HPaSMM

Applying the LOOCV method to the 30 segments of video trajectories, a performance of 89.8% correct activity recognition was obtained. Figure 6 contains the results. Recognition errors occur when processing events unobserved in the other 29 training segments. It explains why segments corresponding to “7-meter throw” and “Jump ball at the court center” events (indicated by light blue boxes in Figure 6) are wrongly classified. These two events would require to extend the activity classes (by defining two new upper level states S_9 and S_{10}). However, these events are observed only once in the processed ten minutes video, so that it was not possible to learn them. With a greater amount of trajectories, these two events could have been modeled independently. This stresses the importance of a sufficient set of training data adapted to the proposed modeling.

Similarly, the event “Back in the court center” (after conceding a goal) is observed only twice in the ten-minutes video. This activity phase would also require its own upper state.

Figure 6: Plot of the recognition results obtained by the HPaSMM method using the LOOCV procedure. Ground truth is drawn in red while the recognition results are plotted in blue. Detected referees whistles are indicated by dotted lines. Events “7-meter throw” and “Jump ball at the court center” are located by light blue boxes. Numbers on the vertical axis correspond to activity phases numbering as described in Subsection 3.1. The horizontal axis denotes the numbering of the successive group of images ($k_{group} = 8$).

Figure 7: Plot of the recognition results obtained by the HPaSMM method using the LOOCV procedure, without considering the “7-meter throw” and “Jump ball at the court center” events. Ground truth is drawn in red while the recognition results are plotted in blue. Detected referees whistles are indicated by dotted lines. Numbers on the vertical axis correspond to activity phases numbering as described in Subsection 3.1. The horizontal axis denotes the numbering of the successive group of images ($k_{group} = 8$).

Again, it was not possible to efficiently model it so that we resorted to the activity phases “defense free-throw or timeout” and “slowly going into attack” for the ground truth, which depict similar motion content.

Figure 7 and Table 1 present recognition results obtained with the HPaSMM after discarding the five segments corresponding to events “7-meter throw” and “Jump ball at the court center” (segments indicated by light blue boxes in Figure 6). Figure 7 shows the recognition results obtained while Table 1 contains, for each activity phase, the recognition results compared to the ground truth.

Recognition results of Figure 7 were obtained when omitting the “7-meter throw” and “Jump ball at the court center” events, *i.e.*, by considering 87% of the whole ten-minutes video, that is more than 8 minutes 30 seconds. Using HPaSMM, a correct recognition rate of 92.2% has been obtained. Every activity phase has been correctly recovered, the 7.8% of errors corresponding to time-lags only. Most of these time-lags are one or two seconds long. The method thus supplies a quite satisfying interpretation of the observed video.

	Recognition rate	# images
Slowly going into attack	68.1	728
Attack against set-up defense	92.4	3776
Offense free-throw or timeout	87.1	1056
counterattack, fast break	95.6	1280
Returning, preventing from fast break	88.6	280
Slowly returning in defense	93.6	1112
Defense	100	3592
Defense free-throw or timeout	81	928

Table 1: Correct recognition rates (by comparing to ground truth) obtained by the HPaSMM method for the eight activity phases. Experiments were conducted with the LOOCV test method, without considering the “7-meter throw” and “Jump ball at the court center” events. The number of images corresponding to each activity phase is also indicated (ground truth).

5.3 Comparison with the HPaHMM method using audio information and LOOCV

The use of referee whistles in the HPaHMM method is done similarly as with the HPaSMM one. Hence, each segment is decoded separately and the LOOCV test method defined above is also considered when applying the HPaHMM method. When processing the set of trajectories of about 8 minutes and 30 seconds to test the HPaHMM method (still omitting the “7-meter throw” and “Jump ball at the court center” events), the correct recognition rate is 89.8%. It is lower than the 92.2% correct recognition rate obtained with the HPaSMM method.

For most of the 30 processed video segments, similar segmentations were obtained with the HPaHMM and the HPaSMM methods. However, for some segments, the HPaHMM method provided a less accurate segmentation than the HPaSMM method. This is illustrated in Figure 8 which presents the results obtained by both methods on the first segment of the processed handball video. Indeed, for that segment, correctness rate is 60% for HPaHMM whereas the HPaSMM method reaches a 89%. Let us recall that the recognition

rate is evaluated as the number of frames correctly classified.

These results show that modeling upper state durations is of interest since the HPaSMM method obtains better results than the HPaHMM one. First, this information enables to reduce time shifts between the ground truth and the extracted transitions between activity phases. Furthermore, as shown on Figure 8, this information also helps avoiding errors in the recognized activity phases that may occur with the HPaHMM method.

Figure 8: Plot of the recognition results for the first segment of the video. On the left, results obtained with the HPaSMM method. On the right, results obtained with the HPaHMM method. For both methods, ground truth is plotted in red and obtained results are plotted in blue. Numbers on the vertical axis correspond to activity phases numbering described in Subsection 3.1. The horizontal axis denotes the numbering of the successive group of images ($k_{group} = 8$).

6. COMPUTATION TIME

We now give information on the computation time corresponding to the conducted experiments. We used a Intel Pentium Centrino 1.86 GHz processor. k_{group} was set to 8 (see Section 2). We applied the LOOCV test procedure with a set of trajectories corresponding to 8 minutes and 30 seconds (still omitting the “7-meter throw” and “Jump ball at the court center” events).

The computation time required to train a HPaSMM model was about 3 seconds. Computation time for the recognition stage applied to one segment (about 35 seconds long) is about 10 seconds. Hence, the overall process took 2 minutes and 20 seconds. The HPaHMM computation time is a little lower than the HPaSMM one.

Referees whistles detection allows us to decode each segment independently. It is possible to decode a segment as soon as it ends. As soon as a whistle is blown, about ten seconds (depending on the size of the processed segment) are necessary to recognize the activity phases in the last segment. Handling audio information helps having a nearly real-time processing of the trajectories.

7. CONCLUSION

We have described an original HPaSMM framework and its application to processing trajectories extracted from videos. Trajectories of handball players belonging to a given team and reconstructed in the court plane are used for activity phases recognition. A specific representation taking into account both dynamics and interaction information has been proposed. The developed HPaSMM method has been tested on a set of trajectories corresponding to 10 minutes of a handball game and gave satisfying results. Extensions to

more complex model architectures could be handled with larger sets of trajectories. Application of this HPaSMM framework to other team sports may also easily be investigated.

8. REFERENCES

- [1] N. Anjum and A. Cavallaro. Multifeature object trajectory clustering for video analysis. *IEEE Trans. on Circuit and Systems for Video Technology*, , Special issue on event analysis in video, 18(11):1555–1564, November 2008.
- [2] B. Günsel, A. M. Tekalp, and P. J. L. van Beek. Content-based access to video objects: temporal segmentation, visual summarization, and feature extraction. *Signal Processing*, 66(2):261–280, April 1998.
- [3] A. Hervieu, P. Bouthemy, and J.-P. L. Cadre. Activity-based temporal segmentation for videos of interacting objects using invariant trajectory features. In *IEEE International Conference on Image Processing*, San Diego, US, October 2008.
- [4] A. Hervieu, J.-P. L. Cadre, and P. Bouthemy. A statistical video content recognition method using invariant features on object trajectories. *IEEE Trans. on Circuit and Systems for Video Technology*, , Special issue on event analysis in videos, 18(11):1533–1543, November 2008.
- [5] S. Hongeng, R. Nevatia, and F. Bremond. Large-scale event detection using semi-hidden markov models. In *IEEE International Conference on Computer Vision*, Nice, France, October 2003.
- [6] <http://gforge.inria.fr/projects/audioseg>.
- [7] <http://gforge.inria.fr/projects/spro>.
- [8] <http://vision.fe.uni-lj.si/cvbase06/download/dataset.html>.
- [9] <http://www.dailymotion.com/fr>.
- [10] C. R. Jung, L. Hennemann, and S. R. Musse. Event detection using trajectory clustering and 4-d histograms. *IEEE Trans. on Circuit and Systems for Video Technology*, , Special issue on event analysis in videos, 18(11):1565–1575, November 2008.
- [11] A. Kokaram, N. Rea, R. Dahyot, M. Tekalp, P. Bouthemy, P. Gros, and I. Sezan. Browsing sports video (trends in sports-related indexing and retrieval work). *IEEE Signal Processing Magazine*, 23(2):47–58, Mars 2006.
- [12] M. Kristan, J. Perš, M. Perše, M. Bon, and S. Kovačič. Multiple interacting targets tracking with application to team sports. In *International Symposium on Image and Signal Processing and Analysis*, Zagreb, Croatia, September 2005.
- [13] P. Mermelstein. Distance measures for speech recognition, psychological and instrumental. In *Workshop on Pattern Recognition and Artificial Intelligence*, Hyannis, US, June 1976.
- [14] A. Muscariello, G. Gravier, and F. Bimbot. Variability tolerant audio motif discovery. In *International Conference on Multimedia Modeling*, Sophia-Antipolis, France, January 2009.
- [15] P. Natarajan and R. Nevatia. Coupled hidden semi-markov models for activity recognition. In *IEEE International Joint Conference on Artificial Intelligence*, Hyderabad, India, January 2007.
- [16] P. Natarajan and R. Nevatia. Hierarchical multi-channel hidden semi-markov models. In *IEEE Workshop on Motion and Video Computing*, Austin, US, February 2007.
- [17] N. M. Oliver, B. Rosario, and A. P. Pentland. A bayesian computer vision system for modeling human interactions. *IEEE Transactions Pattern Analysis and Machine Intelligence*, 22(8):831–843, August 2000.
- [18] J. Perš and S. Kovačič. Tracking people in sport: making use of partially controled environment. In *International Conference on Computer Analysis of Images and Patterns*, Warsaw, Poland, September 2001.
- [19] M. Perše, M. Kristan, S. Kovačič, G. Vučkovič, and J. Perš. A trajectory-based analysis of coordinated team activity in a basketball game. *Computer Vision and Image Understanding*, available online 28 March 2008.
- [20] M. Perše, J. Perš, M. Kristan, G. Vučkovič, and S. Kovačič. Physics-based modeling of human motion using kalman filter and collision avoidance algorithm. In *International Symposium on Image and Signal Processing and Analysis*, Zagreb, Croatia, September 2005.
- [21] C. Piciarelli, C. Micheloni, and G. L. Foresti. Trajectory-based anomalous event detection. *IEEE Trans. on Circuit and Systems for Video Technology*, , Special issue on event analysis in videos, 18(11):1544–1554, November 2008.
- [22] C. Vogler and D. Metaxas. Parallel hidden markov models for american sign language recognition. In *International Conference on Computer Vision*, Kerkyra, Greece, September 1999.