Team Activity Recognition in Sports

Cem Direkoğlu and Noel E. O'Connor

CLARITY: Centre for Sensor Web Technologies*, Dublin City University, Ireland
{cem.direkoglu,noel.oconnor}@dcu.ie

Abstract. We introduce a novel approach for team activity recognition in sports. Given the positions of team players from a plan view of the playing field at any given time, we solve a particular Poisson equation to generate a smooth distribution defined on whole playground, termed the position distribution of the team. Computing the position distribution for each frame provides a sequence of distributions, which we process to extract motion features for team activity recognition. The motion features are obtained at each frame using frame differencing and optical flow. We investigate the use of the proposed motion descriptors with Support Vector Machines (SVM) classification, and evaluate on a publicly available European handball dataset. Results show that our approach can classify six different team activities and performs better than a method that extracts features from the explicitly defined positions. Our method is new and different from other trajectory-based methods. These methods extract activity features using the explicitly defined trajectories, where the players have specific positions at any given time, and ignore the rest of the playground. In our work, on the other hand, given the specific positions of the team players at a frame, we construct a position distribution for the team on the whole playground and process the sequence of position distribution images to extract motion features for activity recognition. Results show that our approach is effective.

1 Introduction

Analyzing complex and dynamic sport scenes for the purpose of team activity recognition is an important task in computer vision. Team activity recognition has a wide range of possible applications such as analysis of team tactic and statistics (i.e. especially useful for coaches and trainers), video annotation and browsing, automatic highlight identification, automatic camera control (useful for broadcasters) etc. Despite the fact that there is much research on vision-based activity analysis for individuals [1], group activity analysis remains a challenging problem. In group activity, there are usually many people located at different positions and moving in different individual directions making it difficult to find effective features for higher level analysis.

There are mainly two possible sources of sport videos: TV broadcasts and multiple video feeds from fixed cameras around the playing field. We first review

* This work is supported by Science Foundation Ireland under grant 07/CE/I114.

A. Fitzgibbon et al. (Eds.): ECCV 2012, Part VII, LNCS 7578, pp. 69–83, 2012.

group activity analysis techniques using broadcast videos, and then review techniques which investigate sport videos captured by fixed multi-camera systems.

1.1 Using the TV Broadcast

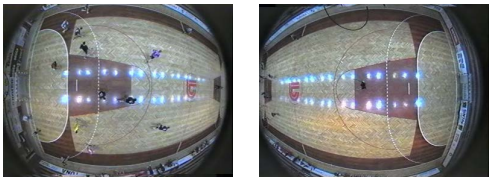
Kong et al. [2] use optical flow based features and the Latent-Dynamic Conditional Random Field model to recognize three different actions (i.e. left side attacking, stalemate and left side defending) in soccer videos. Later, Kong et al. [3] proposed an alternative approach to recognize the same activities in soccer videos. They use scale-invariant feature transform (SIFT) keypoint matches on two successive frames and a linear SVM to classify activities. Wei et al. [4] aims to discriminate group activities in broadcast videos targeting identification of football, basketball, tennis or badminton. They extract space-time interest points and use the probability summation framework for classification.

Despite the existence of such approaches, using a TV broadcast is not effective for group activity analysis, since the camera usually captures the region of interest (such as ball locations) and many players may not be in that region. Using broadcast cameras also suffers from inaccurate player localization because of occlusions, camera motion, etc.

1.2 Using Fixed Multiple Cameras

Most team activity analysis methods [5] [6] [7] [8] [9] [10] [11] use a fixed multi-camera system around the playing field to overcome the limitations of using broadcast data. The multi-camera system usually has a camera configuration to cover all locations on the playground and is therefore able to capture all players simultaneously. Player detection and tracking algorithms are employed in the videos to obtain the trajectories, and then these trajectories are transformed into the top-view of the playing field for more accurate analysis. In the activity analysis stage, features (e.g. position, speed and direction) are extracted using the explicitly defined trajectories and a model employed (e.g. Bayesian Net, Hidden Markov Models or SVM) to recognize the group activities such as different types of offense and defense. These models are summarized in the following paragraph.

Intille and Bobic [5] use Bayesian belief networks for probabilistically representing and recognizing multi-agent action from noisy trajectories in American football. Blunsden et al. [6] extract features from the trajectory data and classify different offense and defense types in European Handball using an SVM. Perse et al. [7] segment the play into three different phases (offence, defense and time out) in a basketball game using a mixture of Gaussians. Then a more detailed analysis is performed to define a semantic description of the observed activity. Perse et al. [8] also present another approach which uses petri nets (PNs) for the recognition and evaluation of team activities in basketball. Hervieu et al. [9] uses a hierarchical parallel semi-Markov model to represent and classify team activities in handball. Recently, Dao et al. [10] proposed a sequence of symbols which are derived from the distribution of players positions in a period of time to represent and recognize offensive types (e.g. side-attack, center attack) in soccer



(a) (b)

Fig. 1. Sample frames from the two fixed cameras for the European handball dataset

games. Li and Chellappa [11] also address the problem of recognizing offensive play strategies in American football using a probabilistic model.

2 Our Motivation and Contribution

In team activities, there is a group of people (the team) performing activities on the constrained playground. All of the existing trajectory-based methods analyse the specific positions (set of points) obtained by either vision-based tracking or GPS-based wearable sensors. There are two main drawbacks in these approaches. First, the position information is noisy. Second, and this is the most important drawback, they use only specific positions and ignore the rest of the playground. By its very nature, team activity takes place over the whole playground as the entire team reconfigures itself to either attack or defend. Thus we believe that a holistic approach is required rather than simply considering a collection of specific player locations.

In this paper, we propose an approach that analyses the entire playground. Given the team players positions from a plan view of the playing field at any given time, we solve a particular Poisson equation to generate a smooth distribution that we term the position distribution of the team. The position distribution is computed at each frame to form a sequence of distributions. Then, we process the sequence of position distributions to extract motion-information images for each frame, where the motion-information images are obtained using frame differencing and optical flow. Finally, we compute weighted moments (up to second order) of these images to represent motion features at each frame. The proposed motion features are experimented with Support Vector Machines (SVM) classification, and evaluated on a publicly available European handball dataset [12], using a similar multi-camera capture set-up to those reported previously, where sample frames from the handball dataset are shown in Figure 1. Results show that we can recognize six different team activities in the handball game, and we perform better than a method [6] that analyses the explicitly defined trajectories for recognition.

Our method is novel and different from other trajectory-based methods presented in section 1.2. These methods extract activity features using the explicitly defined trajectories, where the players have specific positions at any given time, and ignore the rest of the playground. In our work, on the other hand, given the specific positions of the team players at a frame, we construct a position

distribution for the team on the whole playground and process the sequence of position distribution images to extract motion features for activity recognition. The position distribution accounts the uncertainty of the positions and it is defined on the whole playground which can be considered as an intensity image. Representing the positions of the team players as an intensity image instead of a set of points at any given time, allows us to use frame differencing and optical flow, which are important techniques for image motion description. We extract motion features at each frame using the sequence of position distribution images instead of using the explicitly defined trajectories to represent activities.

In our preliminary work [13], we have verified that a particular Poisson equation can be used to determine the region of highest population, corresponding to the area with the highest density of the majority of players, and to estimate the region of intent, corresponding to the region towards which the team is moving as they press for territorial advancement. The approach proposed here significantly extends this early work to perform full classification of team activity. In this paper, we are not concerned about the region of intent or the region of highest population, and so consider the work reported here to be an independent piece of work.

3 Team Position Distribution Generation

We investigate the problem in the context of European handball, where the top-view of the handball field of play is shown in Figure 2(a) with the team player positions (a European handball team has 7 players). Given the positions of the team players at any time, we aim to generate a position distribution of the team defined on the whole playground. There are many possible probability distribution models (e.g. Gaussians, Laplace or Cauchy distribution), which can be centered on each player position and then summed up to generate a position distribution of the team. Since the activity is performed on the bounded playground and players have to be on the playground to be involved in the team-based activity, the position distribution must be zero outside the playground. This can be achieved by using the truncated versions (e.g. truncated Gaussians) of the probability distributions. However, all of the probability distribution models, which can be used to create a smooth distribution and account for uncertainty for the positions, are parameter dependent and the parameters need to be adjusted to optimize the performance of the team activity recognition. In our work, we choose to solve a particular Poisson equation to generate a position distribution since it has a unique and steady-state solution with respect to the given team player positions. The proposed Poisson equation is parameter free, and can model zero probability outside the playground without any truncation. The solution of the proposed Poisson equation only depends on the players positions.

3.1 Background to the Poisson Equation

In mathematics, the Poisson equation is an elliptic type partial differential equation [14] which arises usually in electrostatics, heat conduction and gravitation. The general form of the Poisson equation, in two-dimensions, is given by,

(a) (b) (c)

Fig. 2. The Poisson equation is applied to generate the position distribution. (a) The top-view of the handball court with player locations. (b) The position distribution of the team. (c) The color mapped position distribution with level sets.

$$\nabla^2 I(\mathbf{x}) = -Q(\mathbf{x}) \tag{1}$$

where Q is a real-valued function of a space vector $\mathbf{x} = (x, y)$ and it is known as the source term, I is the solution which is also a real-valued function and ∇^2 is the spatial Laplacian operator. Given a source term $Q(\mathbf{x})$, we find a solution for $I(\mathbf{x})$ that satisfies the Poisson equation and the boundary conditions over a bounded region of interest. There are three general types of boundary conditions: Dirichlet, Neuman and Mixed. Here, we explain the Dirichlet condition, which is used in our algorithm. In the Dirichlet condition, the boundary values (solutions) are specified on the boundary. These values can be a function of space or can be constant. The Dirichlet condition is represented as $I(\mathbf{x}) = \Phi(\mathbf{x})$, where $\Phi(\mathbf{x})$ is the function that defines the solution at the boundary layer.

3.2 The Proposed Poisson Equation and Solution

The proposed Poisson equation and the resulting distribution (solution) are obtained based on the following considerations. The top-view image of the field of play is assumed to be a binary image where the player positions are one and the rest of the positions are zero at any time during the game. Although players are expected to be in the play area during the game, players sometimes can move a little outside for a variety of different reasons, such as to serve the ball, when the ball is out or in order to talk to the coach. Thus, we expand the binary image of the field of play to include the possibility that the players may move a little outside the lines. The binary image is defined to be the source term in the Poisson equation. The boundary condition is Dirichlet which has a specific solution, $I(\mathbf{x}) = 0$, at the boundaries of the expanded field of play. This means that there is no possibility for a player to be outside the region of interest. The proposed Poisson equation problem is,

$$\begin{aligned} \nabla^2 I(x, y) &= -\left(\sum_{i=1}^{N} \delta(x - x_i, y - y_i)\right) \\ I(x, y) &= 0, \quad \textit{boundary condition} \end{aligned} \tag{2}$$

where N is the number of players in the team and (x_i, y_i) is the position of player i. The source function is assumed to be a linear combination of

dirac-delta functions $\delta(.)$ in two dimensions. It is important to note that the proposed Poisson equation has a unique and steady-state solution at each frame. The solution is parameter free, and it only depens on the position of the players. Therefore, when players change their position from the previous frame to the current frame, the solution also changes in the current frame.

The numerical solution methods of the Poisson equation can be categorized as direct and iterative methods. In [15], Simchony et al. pointed out that direct methods are more efficient than multigrid-based iterative methods for solving the Poisson equation on a rectangular domain, since direct methods can be implemented using the Fast Fourier Transform (FFT). In our work, since the field of play is rectangular, we employ FFT-based direct methods to solve the proposed Poisson equation. The proposed equation has a Dirichlet boundary condition that needs discrete sine transforms (using FFT) to achieve an exact solution, where the detailed description of the solution method is given in [15]. The solution to the proposed equation forms peaks at the player positions. To smooth these peaks, we apply Gauss-Seidel iterations (5 iterations), as a post-processing stage, to relax the surface while maintaining the boundary condition ($I(\mathbf{x}) = 0$) outside the region of interest.

The resultant distribution provides the likelihood of a position to be occupied by players at any given time, and it is called the position distribution of the team. Figure 2(b) shows the position distribution for the given example and Figure 2(c) shows the same distribution with color mapping and with level sets. The resolution of the position distribution image is 220×120 in our experiments.

4 Motion-Information Images and Feature Extraction

Computing the position distribution for each frame provides a sequence of position distributions. We process the sequence of distribution images to generate motion-information images which can describe motion at each frame. The motion-information images are created using frame differencing and optical flow.

4.1 Frame Differencing

The simplest way in which we can detect motion is by image differencing. Figure 3(a) shows the direction of movement of the team players from the current frame to the next frame (50 frames later), where the starting point of the arrow represents the position of the player at the current frame and the end point represents the position of the player at the next frame. We compute the position distribution for the team at the current and at the next frames. Since the team players move from the current positions to the next positions, they create higher position distribution values in the direction of movement. To detect motion with the direction, we apply change detection by simply subtracting the current distribution from the next distribution and keep the positive values while setting the negative values to zero, i.e. $(I(\text{x,y,n+m}) - I(\text{x,y,n})) > 0$, where $I(\text{x,y,n})$ represents the position distribution of the team at frame number n and m is the

(a) (b)

Fig. 3. Generating a motion-information image using frame differencing. (a) Team players movements. (b) The motion-information image.

number of frames between the current and the next frame. Frame differencing is applied with 50 frames (i.e. $m = 50$) of temporal extent in our experiments. Figure 3(b) shows the frame differencing whereby we keep the positive values and set the negative values to zero for the given example.

4.2 Optical Flow

Although frame differencing can provide some information about the movement, we cannot exactly see how the distribution points move. In order to describe the position changes at each frame, we compute optical flow vectors that can provide the displacement of the points with directions. We employ the classical Horn and Schunck (HS) method [16] for optical flow estimation. This is a differential approach which combines a data term that assumes constancy of some image property (e.g. brightness constancy, gradient magnitude constancy) with a spatial term that models how the flow is expected to vary across the image (e.g. smoothness constraint). An objective function combining these two terms is then optimized. In our experiments, we observed that using the gradient magnitude constancy assumption (i.e. $|\nabla I(\mathrm{x,y,n})| = |\nabla I(\mathrm{x+u,y+v,n+m})|$) instead of using the brightness constancy (i.e. $I(\mathrm{x,y,n}) = I(\mathrm{x+u,y+v,n+m})$) can estimate better optical flow, where u is the horizontal optical flow and v is the vertical optical flow. Therefore, in our work, we use the gradient magnitude constancy assumption together with the smoothness constraint to compute the optical flow on the playing field. The gradient of the position distribution is computed using the Sobel operator and the optical flow is computed from the current frame to the next frame with 8 frames of temporal extent. There are also two parameters that affect the solution of the HS method: a parameter that reflects the influence of the smoothness term is set to 0.1, and the number of iterations to achieve the solution is set to 200. Figure 4(a) shows the position distribution image and the estimated optical flow. For better illustration, Figure 4(b) shows the zoomed image from the red box in Figure 4(a). Note that this is a novel algorithm to compute the motion field on the top-view of the playground. Kim et al. [17] compute the motion field on the top-view of the playground by interpolating the player's motion vectors, where the player's motion vectors are generated using the specific positions of the players. On the other hand, in our algorithm, we use the specific positions to generate the position distributions, and then estimate the motion field using optical flow.

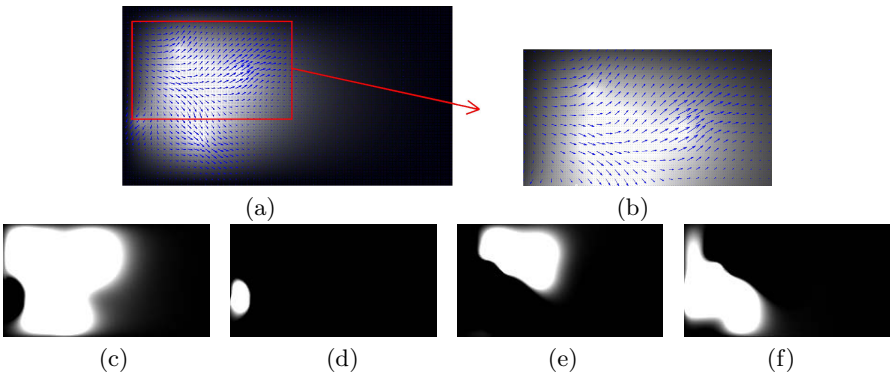


Fig. 4. Computing the directional speed images to represent the motion-information images. (a) The position distribution and the estimated optical flow. (b) The zoomed image from the red box in (a). (c) Directional speed image in the direction of positive x-axis, (d) negative x-axis, (e) positive y-axis and (f) negative y-axis.

The motion-information images, using the optical flow, are generated with the following considerations. The horizontal and vertical components (i.e. u and v) of the flow are two different scalar fields. Each of these components is half-wave rectified to generate four non-negative channels: u^+, u^-, v^+, v^-, so that $u = u^+ - u^-$ and $v = v^+ - v^-$. These channels, u^+, u^-, v^+ and v^-, represent directional speed images in the direction of positive x-axis, negative x-axis, positive y-axis and negative y-axis respectively. Note that the directional speed images have also been used in [18] for individual action recognition, but their usage for group activity recognition as proposed here is novel. The directional speed images are illustrated in Figure 4(c), 4(d), 4(e) and 4(f) for the given example.

4.3 Feature Extraction

We use five motion-information images to describe motion at each frame, where one of them is obtained with frame differencing and the other four are obtained with optical flow. Frame differencing is applied with 50 frames of temporal extent while the optical flow is computed with 8 frames of temporal extent, so that frame differencing captures motion in a longer period of time while the optical flow captures motion in a shorter period of time. Our experiments show that describing the motion in this way performs better than other options.

Next, we compute weighted moments for each motion-information image to represent motion features at that frame. The discrete form of the equation is,

$$m_{pq} = \sum_{x} \sum_{y} w(\mathrm{x,y}) x^p y^q \Delta x \Delta y \tag{3}$$

Here, m_{pq} is the moment of order p and q, w(x,y) is the weight function, which we substitute for each motion-information image, $\Delta x = \Delta y = 1$ are spacing sizes of a pixel. We compute moments up to order $p + q = 2$, resulting in 6 moments per image and 30 moments in total to describe the motion at each frame.

Table 1. Team activities with their numbering

1. Slowly going into offense
2. Offense against set-up defense
3. Offense fast break
4. Fast returning into defense to prevent fast break
5. Slowly returning into defense
6. Basic defense

5 Classification Using the Motion Descriptors

We investigate the use of the proposed features with Support Vector Machine (SVM) classification. SVM is a powerful technique in classification. It maps each training data to higher dimensional space and constructs a separating hyperplane such that the distance between the hyperplane and a data point is maximized. Test data is then classified by the discriminant function. In our work, the test frame is classified using the 141 by 141 neighborhood frames (141 from past and 141 from future neighborhoods), which is determined experimentally. This means that the window size is 283 (including the test frame). Each of the frames in the window is labeled with the SVM classifier by using the one against all method. Then the most frequent class is selected to represent the activity of the test frame. In SVM, a Gaussian radial basis function kernel is used and the scaling factor is 2.4. The upper bound on the Lagrange parameters is 10. In addition, we use the sequential minimal optimization method to find the separating hyperplane since we have a large dataset and this method is computationally efficient.

6 Evaluation and Results

The proposed model is evaluated on the European handball which is usually an indoor game. In handball, there are seven players and it is played on a 40 by 20 meters court. The dataset for the handball game is from the publicly available CVBASE dataset [12]. The dataset consists of ten minutes of a handball game. The playground coordinates of the seven players of the same handball team are available throughout the sequence. The sequence consists of 14978 frames (approximately 10 minutes). These trajectories are extracted from two bird-eye view cameras, one above each part of the court plane, with semi-automatic tracking, where the details on trajectory extraction are given in [19]. The dataset providers obtained error estimates on players positions in the playground between 0.3 and 0.5 meters. There are mainly six different team activities in this dataset, where the starting and end times of the activities are also annotated. The definition of the six team activities with their numbering is given in Table 1. The length of each activity sequence ranges from 125 frames to 1475 frames. It should be noted that some of these activities can be split into more complex activity classes; however more information is required such as the ball trajectory or the trajectories of the opposing team to represent more complex activities, which is not provided in this dataset.

We evaluate our approach while comparing with a model, proposed by Blunsden et al. [6], which analyses the explicitly defined trajectories for team activity recognition. This method is designed to recognize the same activities in the same dataset, which we believe to be the best comparison we can perform given the current status of work in this area. They extract 5 features (i.e. positions, speed, directions) from each player trajectory, and then all the players features are concatenated to form 35 dimensional feature vector to represent the activity at each frame. A SVM classifier is then trained upon this data. They use the one against all method for classification. The test frame is classified using the 99 by 99 neighborhood frames that make the window size 199 (including the test frame). Each frame in the window is labeled with the SVM classifier and then the most frequent label represents the class of the test frame. A Gaussian kernel function is used and the scaling factor is 2.4. The upper bound on the Lagrange parameters is 10. The sequential minimal optimization method is used to find the separating hyperplane.

6.1 Temporal Segmentation and Recognition

In our evaluation, the second half of the video is used for training (i.e. 7600 frames, 5 minutes and 4 seconds) and the first half is used for testing (i.e. 7328 frames, 4 minutes and 53 seconds). Both the first and second half include the six different team activities. In the first half, there are 1, 3, 3, 1, 2 and 3 instances and in the second half there are 3, 3, 2, 2, 2 and 4 instances for activity number 1, 2, 3, 4, 5 and 6 respectively. Since proper training is required for robust classification, we choose the second half for training purposes. The second half includes more activity samples than the first half, e.g. the activity number 1 is performed once in the first half and three times in the second half. In the training, there are at least two segments and at most four segments to represent an activity. On the other hand, in the testing, there are at least one segment and at most three segments to represent an activity. In addition, since we are testing the continuous sequence, there are also time-out segments which occur when the ball is out or when play is stopped. In handball, when it is time-out, teams keep moving and start to perform the next activity, e.g. if they are serving the ball, they move around to create space, on the other hand if the opponent team is serving the ball, they move around to prevent the pass. Therefore, each of the time-outs in the test sequence is defined to be the following activity in our experiments.

In continuous classification, we classify all individual frames. We evaluate our features with the SVM classification and all the details related to the classification are provided in section 5. The same evaluations are also conducted for the method proposed by Blunsden et al. [6] for comparison purpose. In evaluations, this model is denoted by FET+SVM, which means that Features are obtained using the Explicitly defined Trajectories and the classification is achieved using the Support Vector Machines. Figure 5(a) shows the temporal segmentation and recognition results obtained by the FET+SVM, while Figure 5(b) show the results obtained by the proposed features with SVM (Proposed features + SVM)

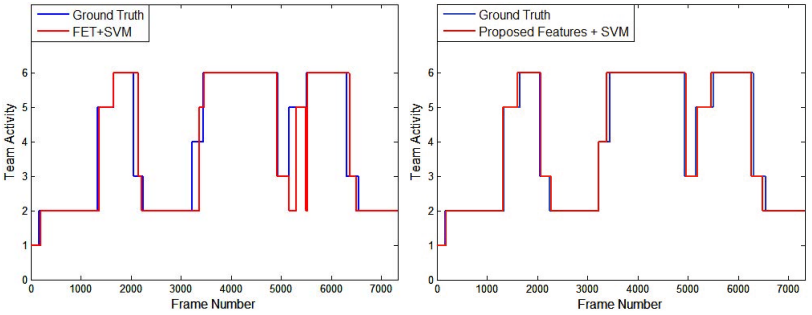


Fig. 5. Temporal segmentation and recognition of activities (a) FET+SVM (proposed by Blunsden et al. [6]) (b) Proposed Features with SVM

Table 2. Correct classification rates (CCR%) of the Proposed Features with SVM, and FET+SVM (total frames: 7328)

Methods	**FET+SVM [6]**	**Proposed Features+SVM**
CCR%	89.74%	94.61%

respectively. The blue graph represents the ground truth and the red graph represents the prediction. It is observed that the proposed features with SVM achieve better temporal segmentation and recognition than the FET+SVM. The FET+SVM cannot identify activity number 4 which is fast returning into defense, and confuses this with activity number 5 which is slowly returning into defense. The FET+SVM also confuses between the activity number 2 and 5, which is offense against set-up defense and slowly returning into defense respectively. There are also some errors when the activity switches in FET+SVM. The proposed features with SVM can recognize the six different activities and the errors occur when the activity switches.

As stated by [20], there are two basic units for scoring in the evaluation of activity recognition: frames and events. They are alternative to each other. Our evaluation is based on scoring the frames which is an acceptable validation method and which we believe puts us in line with best practice. We classify 7328 test frames in the evaluation, and Table 2 shows the correct classification rate (CCR%) for each method. The CCR% is computed as $CCR\% = (C_c/T_c) \times 100$, where C_c is the number of correct classification and T_c is the number of total classification. The FET+SVM achieves 89.74%, and the proposed features with SVM achieves 94.61% recognition rate. Results show that the proposed features with SVM performs around 4.9% better than the FET+SVM. Results show that the proposed features perform significantly better than the FET features [6] with the same classifier, i.e. SVM.

Table 3 illustrates the precision and recall results, for each activity class, obtained using each method. Here, the precision for a class is defined as $P\% = (P_c/P_t) \times 100$, where P_c is the number of frames correctly predicted as belonging

Table 3. Precision and recall of the Proposed Features with SVM, and FET+SVM for each activity

Activity Number	# of Frames	FET+SVM [6]		Proposed Features+SVM	
		Precision(P%)	Recall(R%)	Precision(P%)	Recall(R%)
1.	164	82.41%	100.0%	88.17%	100.0%
2.	2914	86.48%	98.79%	97.98%	98.28%
3.	675	96.96%	61.48%	84.76%	84.88%
4.	225	0.0%	0.0%	97.38%	66.22%
5.	675	83.92%	70.37%	99.46%	82.07%
6.	2675	94.19%	98.80%	92.79%	98.28%

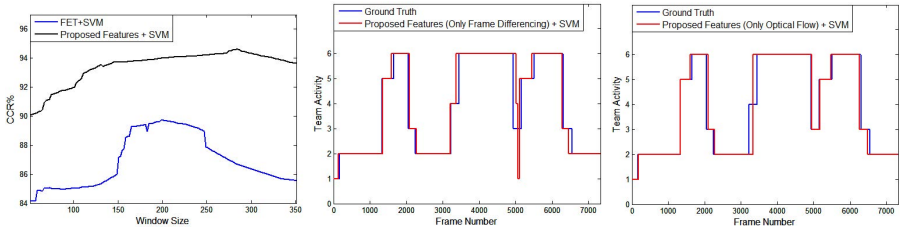
to that class and P_t is the total number of frames predicted as belonging to that class. The Recall for a class is defined as $R\% = (R_c/R_t) \times 100$, where R_c is the number of frames correctly predicted and R_t is the total number of frames that actually belong to that class. In this table, both the precision and recall must be high for a method to show that it can handle the activity switches and provide sufficient discrimination. There is only one activity, i.e. Activity 6, in Table 3 where the FET+SVM [6] has slightly better precision and better recall than the Proposed features+SVM. In general, the proposed features+SVM has better performance than the FET+SVM [6]. The main problem of the FET+SVM method is that it cannot discriminate the activity number 4 and it is sensitive to activity switches. On the other hand the proposed features with SVM can discriminate all activities, and can handle activity switches better than the FET+SVM [6]. Figure 7 shows sample frames with the automatically recognized activities by the proposed features with SVM.

6.2 The Effect of Window Size

We present the effect of differing window size in the classification performance (CCR%). Figure 6(a) show the CCR% for the proposed features with SVM and for the FET+SVM model [6]. The window size ranges from the 51 to 351 in our evaluation. It is observed that the proposed feature with SVM performs better than the FET+SVM model at each window size. The optimal window size for the proposed features with SVM is 283. For the FET+SVM model, it is 199.

6.3 The Effect of Motion-Information Images

We present the influence of motion-information images and report what the temporal segmentation and the classification results would be if only frame differencing or only optical flow was used. Figure 6(b) shows the result obtained by using only frame differencing (one motion-information image). Figure 6(c) shows the result using only optical flow (four motion-information image), and Figure 5(b) illustrates the result using the combination of them (five motion-information image). Only frame differencing achieves 90.96%, only optical flow



(a) (b) (c)

Fig. 6. The effect of window size and the effect of motion information images. (a) The classification performances (CCR%) with differing window size. (b) Temporal segmentation and recognition using only frame differencing, (c) using only optical flow.

Table 4. The computation time for each stage of the methods

Stages	FET+SVM [6]	Proposed Features+SVM
Feature extraction in whole video	1.48 seconds	13702 seconds
Training all activities in the second half	19.89 seconds	25.25 seconds
Classifying all activities in the first half	0.31 seconds	0.29 seconds

(a) (b)

Fig. 7. Team activity is automatically recognized by the proposed features with SVM

achieves 92.82% and the combination achieves 94.61%. Results indicate that using the combination improves the CCR% and the discrimination.

6.4 Computational Efficiency

The computational time for each stage of the methods are given in Table 4. Results are obtained using Matlab 7 on a Windows 7 Operating System with Intel Core i3-870, 2.93GHz and 8MB RAM. It is observed that the FET+SVM method is more efficient than proposed method with SVM especially in feature extraction. Although, the proposed features with SVM is computationally less efficient in feature extraction, it has significantly better classification accuracy in comparison to FET+SVM.

7 Conclusions and Future Work

We have presented an approach for team activity recognition in sports. Given the positions of team players from a plan view of the playing field at any given time,

we solve a particular Poisson equation to generate a position distribution for the team. Computing the position distribution for each frame provides a sequence of distributions, which we process to extract motion features at each frame. Then the motion features are used to classify team activities. Results show that the proposed approach is effective, and performs better than a method (FET+SVM) that extracts features from the explicitly defined trajectories. Currently, we are working on field hockey datasets and our preliminary results indicate that it is possible to use the proposed approach in this domain as well. In future, we will present our results on other sporting domains.

References

1. Aggarwal, J.K., Ryoo, M.S.: Human Activity Analysis: A Review. ACM Computing Surveys 43(3), 16 (2011)
2. Kong, Y., Zhang, X., Wei, Q., Hu, W., Jia, Y.: Group action recognition in soccer videos. In: Proc. ICPR, pp. 1–4 (2008)
3. Kong, Y., Hu, W., Zhang, X., Wang, H., Jia, Y.: Learning Group Activity in Soccer Videos from Local Motion. In: Zha, H., Taniguchi, R.-I., Maybank, S. (eds.) ACCV 2009, Part I. LNCS, vol. 5994, pp. 103–112. Springer, Heidelberg (2010)
4. Wei, Q., Zhang, X., Kong, Y., Hu, W., Ling, H.: Group Action Recognition Using Space-Time Interest Points. In: Bebis, G., Boyle, R., Parvin, B., Koracin, D., Kuno, Y., Wang, J., Pajarola, R., Lindstrom, P., Hinkenjann, A., Encarnação, M.L., Silva, C.T., Coming, D. (eds.) ISVC 2009, Part II. LNCS, vol. 5876, pp. 757–766. Springer, Heidelberg (2009)
5. Intille, S.S., Bobick, A.F.: Recognizing planned, multiperson action. CVIU 81(3), 414–445 (2001)
6. Blunsden, S., Fisher, R.B., Andrade, E.L.: Recognition of coordinated multi agent activities, the individual vs the group. In: ECCV Workshop on Computer Vision Based Analysis in Sport Environments (CVBASE), pp. 61–70 (2006)
7. Perse, M., Kristan, M., Kovacic, S., Vuckovic, G., Pers, J.: A trajectory-based analysis of coordinated team activity in a basketball game. CVIU 113(5), 612–621 (2009)
8. Perse, M., Kristan, M., Pers, J., Music, G., Vuckovic, G., Kovacic, S.: Analysis of multi-agent activity using petri nets. Pattern Recognition 43(4), 1491–1501 (2010)
9. Hervieu, A., Bouthemy, P., Cadre, J.P.L.: Trajectory-based handball video understanding. In: International Conference on Image and Video Retrieval, vol. 43, pp. 1–8 (2009)
10. Dao, M.S., Masui, K., Babaguchi, N.: Event tactic analysis in sports video using spatio-temporal pattern. In: Proc. ICIP, pp. 1497–1500 (2010)
11. Li, R., Chellappa, R.: Recognizing offensive strategies from football videos. In: Proc. ICIP, pp. 4585–4588 (2010)
12. CVBASE 2006 dataset, in workshop on computer vision based analysis in sport environments (2006), `http://vision.fe.uni-lj.si/cvbase06/downloads.html`
13. Direkoglu, C., O'Connor, N.E.: Team behavior analysis in sports using the Poisson equation. In: Proc. ICIP (2012)
14. Braun, M.: Differential equations and their applications. Springer (1993)
15. Simchony, T., Chellappa, R., Shao, M.: Direct analytical methods for solving Poisson equations in computer vision problems. T-PAMI 12(5), 435–446 (1990)

16. Horn, B.K.P., Schunck, B.G.: Determining optical flow. Artificial Intelligence 17, 185–203 (1981)
17. Kim, K., Grundmann, M., Shamir, A., Matthews, I., Hodgins, J., Essa, I.: Motion fields to predict play evolution in dynamic sport scenes. In: CVPR, pp. 840–847 (2010)
18. Efros, A.A., Berg, A.C., Berg, E.C., Mori, G., Malik, J.: Recognizing action at a distance. In: Proc. ICCV, pp. 726–733 (2003)
19. Janez, M.K., Kovacic, S.: Multiple interacting targets tracking with application to team sports. In: International Symposium on Image and Signal Processing and Analysis, pp. 322–327 (2005)
20. Ward, J.A., Lukowicz, P., Gellersen, H.W.: Performance metrics for activity recognition. ACM Trans. on Intelligent Systems and Technology 2(1), 6, 1990 (2011)