**Aalborg Universitet**

AALBORG UNIVERSITY
DENMARK

# Classification of sports types from tracklets

Gade, Rikke; Moeslund, Thomas B.

*Publication date:*
2014

*Document Version*
Accepted author manuscript, peer reviewed version

Link to publication from Aalborg University

*Citation for published version (APA):*
Gade, R., & Moeslund, T. B. (2014). *Classification of sports types from tracklets*. Abstract from KDD Workshop on Large-scale Sports Analytics, New York, United States.

**General rights**
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal -

**Take down policy**
If you believe that this document breaches copyright please contact us at vbn@aub.aau.dk providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from vbn.aau.dk on: April 10, 2024

# Classification of sports types from tracklets

Rikke Gade
Visual Analysis of People Lab
Aalborg University
Rendsburggade 14
Aalborg, Denmark
rg@create.aau.dk

Thomas B. Moeslund
Visual Analysis of People Lab
Aalborg University
Rendsburggade 14
Aalborg, Denmark
tbm@create.aau.dk

## ABSTRACT

Automatic analysis of video is important in order to process and exploit large amounts of data, e.g. for sports analysis. Classification of sports types is one of the first steps towards a fully automatic analysis of the activities performed at sports arenas. In this work we test the idea that sports types can be classified from features extracted from short trajectories of the players. From tracklets created by a Kalman filter tracker we extract four robust features; Total distance, lifespan, distance span and mean speed. For classification we use a quadratic discriminant analysis. In our experiments we use 30 2-minutes thermal video sequences from each of five different sports types. By applying a 10-fold cross validation we obtain a correct classification rate of 94.5 %.

## Keywords

Thermal imaging, Sports types, Classification, Tracking

## 1. INTRODUCTION

Manual analysis of video is very time consuming and expensive. Automating the analysis will enable a significantly higher amount of data to be processed and exploited for systematic analysis of, e.g., sports activities. The interest in sports analytics has grown significantly recently as governments, broadcasters, coaches, etc. see great potential in the data. In this work we focus on automatic recognition of sports types. For large amounts of video, this step will help separating the data into sequences of well-known sports types. Furthermore, for multi-purpose indoor arenas as well as outdoor fields, it can be of great interest to get a better knowledge of the use of the facilities, without having to perform manual annotation. We have previously proposed a method for activity recognition based on heatmaps produced from summed position data [3]. In this work we will try to estimate which type of sport is being performed based on motion features extracted from tracklets.

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 a fee.
*KDD Workshop on Large-Scale Sports Analytics* '14 New York, NY USA

Previous work on sports type recognition has often been based on the visual appearance of the court, such as court lines and dominant colour of the field [6, 11, 8]. The dominant colour has also been combined with motion features, such as camera/background motion [9, 10] or direction of motion vectors in image blocks [4]. In this work we will classify different sports types performed in the same indoor multi-purpose arena. The appearance of the court will therefore not be useful for classification. Furthermore, we use a static camera setup with thermal cameras, eliminating both camera motion features and any colour features. Thermal cameras are chosen in order to minimise the privacy issues of capturing video in public sports arenas.

Most relevant to this work then is mainly two papers. Lee and Hoff [7] detect players and use trajectory segments of three seconds from which they extract and test eight features based on speed, direction and path length. They find that two features maximises the classification accuracy. These features are average speed and the ratio of the overall distance to the path length. Using k-means clustering and decision tree classification, they achieve 94.2% accuracy. However, they test on only two sports types; Ultimate Frisbee and volleyball. Whether these two features will be sufficient to discriminate a larger set of sports types is therefore unknown. Gade and Moeslund [3] proposed sports type recognition based on classification of heatmaps produced from position data. The heatmaps are projected to a low-dimensional discriminative space using Fischers Linear Discriminant and new instances are classified as the nearest cluster. In this work five different sports types are classified with a precision of 90.8 %. Limitations of this work include the dependency on scale, direction and location on the field. To overcome these limitations, we will in this work extract local features, which are invariant to the position and direction of play. Based on trajectories (tracklets) from each player, motion features are extracted and used for classification.

In the remaining part of this paper, section 2 will describe the tracking algorithm used to produce tracklets, after which we choose the features to extract in section 3. In section 4 the classification approach is described, before the experiments and results are presented in section 5, and finally the conclusion is found is section 6.

## 2. TRACKING

To analyse the motion of people, we need their trajectories. Tracking multiple people through interactions, occlusion and complex motion is a problem no existing methods

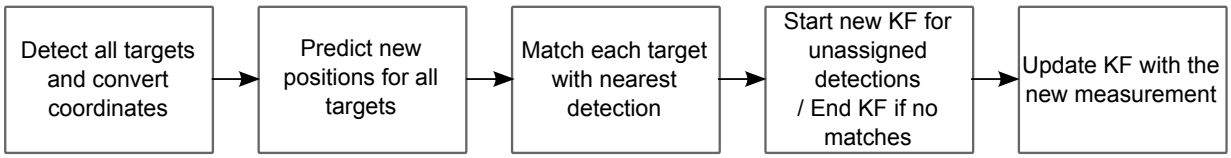


**Figure 1: Illustration of the tracking framework which is run for every frame.**

solve automatically yet. Instead, we here aim to obtain short but reliable trajectories (tracklets), from which we can estimate motion features.

In thermal images the appearance information of people is very sparse, as only temperature is measured. When people are observed from a distance of several metres, small differences in temperature patterns will not be visible, hence people will appear as grey blobs of similar temperature. A cropped input image is shown in figure 2(a). The similar appearance of people must be considered when choosing the tracking scheme.

We choose a classic approach based on the Kalman filter [5]. This method is one of the predict-match-update schemes, which predicts the next position of the object from the previous state (described by, e.g., position and velocity), then updates the estimate when a (probably noisy) measurement is obtained. Using Kalman filtering for multi-target tracking can be done by assigning a new Kalman filter for each new target, however, it implies some reasoning for assigning each detection to the right tracker. This is here determined by the shortest Euclidean distance within a given threshold. If a detection is not assigned to a tracker, a new Kalman filter is started. Likewise, if no detections are assigned to a tracker in $n$ consecutive frames, the Kalman filter is terminated. $n$ is experimentally set to 10 frames. Figure 1 illustrates the tracking process performed for each frame.

The first step illustrated in figure 1 is the detection of all targets in the frame. We use thermal imaging in an indoor arena, making it reasonable to assume that people appear warmer than a static background. The main step in our detection algorithm is therefore an automatic thresholding of the image. Figure 2 illustrates this step. In figure 2(a) and 2(b) people are nicely separated and easily segmented. However, occlusions can cause problems for detecting individual people, as shown in figure 2(c) and 2(d). To overcome some of these problems we try to detect these blobs containing more than one person and split them either horizontally or vertically. Further details on these procedures can be found in [2].

To be independent of the image perspective we transform the detected positions of people in the image into world coordinates before tracking. This is done by applying a homography matrix, calculated during initialisation.

Terminating the tracks with no possibility of re-identification later will naturally lead to more split trajectories. But as the identity of players has no role in this work, it is preferable to have short reliable tracklets instead of trying to resolve complex situations with a higher probability of false tracks.

Figure 3 shows examples of typical trajectories extracted from 2-minutes video sequences of each sports type. Each tracklet is assigned a random colour and are presented in world coordinates (top view of the court).

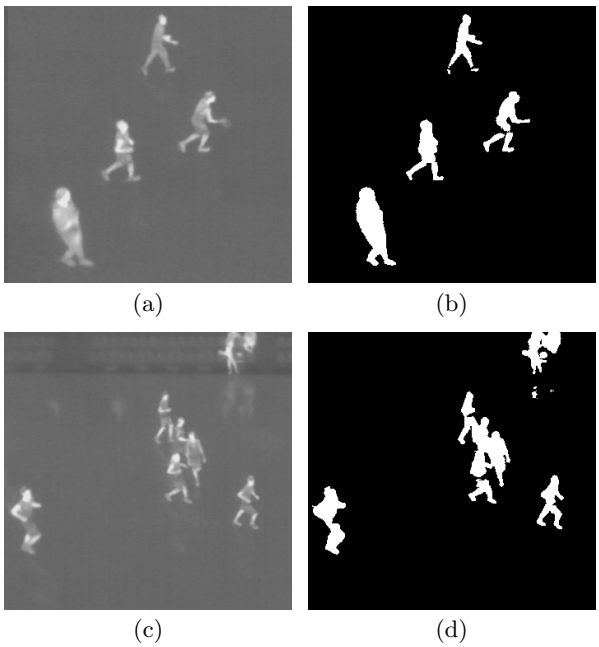


**Figure 2: Example of a thermal input images (a) and (c), and the corresponding binary image after automatic thresholding (b) and (d).**

## 3. FEATURES

From the trajectories we will extract features representing the typical type of motion for each sports type. We consider the following five types of features:

- **Speed:** Mean speed, acceleration, jerk
- **Direction:** Distribution of directions, change in direction
- **Distance:** Euclidean distance from start to end point, total distance travelled, largest distance span between two points
- **Motion:** Total motion per frame
- **Position:** Distance between people, area covered

As discussed in the introduction, we aim to find a few simple features, which should be invariant to the size and direction of the court, the position of the players according to the court and to the direction of play. The features must be robust to noisy detections and tracking errors as well. Acceleration, jerk, change in direction and euclidean distance from start to end point are all discarded because they are easily affected by tracking noise. The distribution

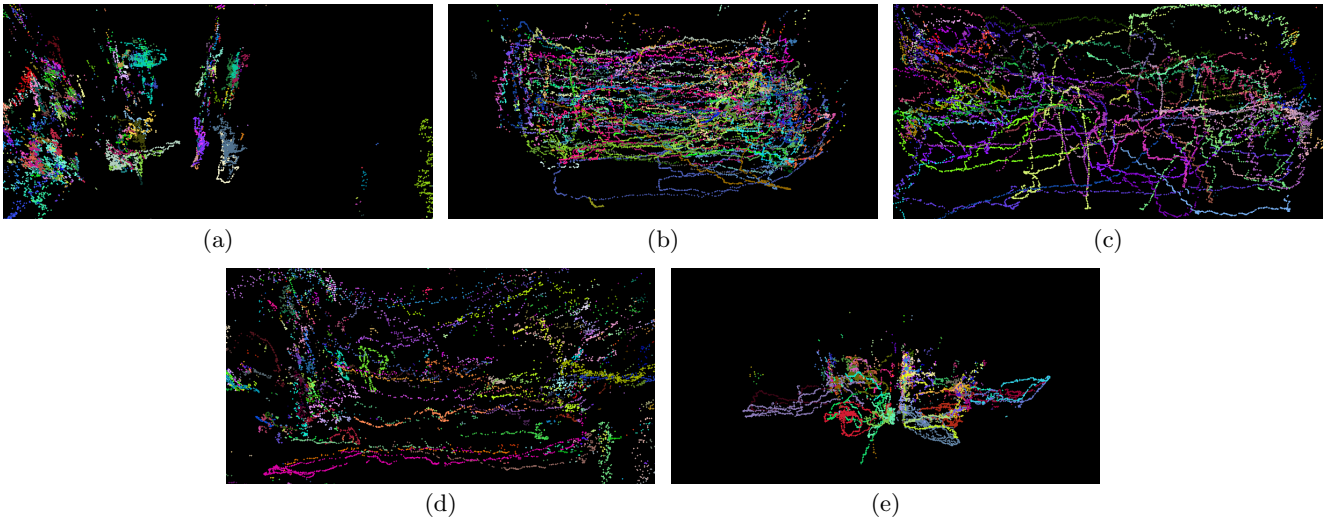

**Figure 3: Tracklets from a 2-minute period of (a) badminton, (b) basketball, (c) soccer, (d) handball, and (e) volleyball.**

of direction depends on the direction/rotation of play, and is therefore discarded. The motion and position features are discarded as they depend on number of people present on the court, as well as size of the play area. Hence, we end up with the following four features calculated for each tracklet:

***Lifespan [frames]*** is measured in number of frames before the tracklet is terminated. This feature implicitly represents the complexity of the sequence; the lifespan of each tracklet will be shorter when the scene is highly occluded:

$$ls = n_{end} - n_{start} \tag{1}$$

where $n$ is the frame number.

***Total distance [m]*** represents the total distance travelled, measured as the sum of frame-to-frame distances in world coordinates:

$$td = \sum_{i=0}^{ls-1} d(i, i+1) \tag{2}$$

where $d$ is the Euclidean distance function.

***Distance span [m]*** is measured as the maximum distance between any two points of the trajectory. This feature is a measure of how far the player move around at the court:

$$ds = \max(d(i,j)), \quad 0 < i < ls,\ 0 < j < ls \tag{3}$$

***Mean speed [m/s]*** is measured as a mean value of the speed between each observation:

$$ms = \frac{td \cdot n_{seq}}{ls \cdot t} \tag{4}$$

where $t$ is the duration of the video sequence in seconds, and $n_{seq}$ the duration of the sequence in number of frames.

For each video sequence used in the classification, we will use the mean value for each feature and combine the features with equal weighting. We test all combinations of the features described above, from using a single feature to using all four. We find that the best results are obtained when using all four features, indicating that none of them are redundant or misleading.

## 4. CLASSIFICATION

For the classification task we choose to use a supervised learning method. We provide labelled training data and aim to find a function that best discriminates the different classes. For this purpose we apply discriminant analysis with both a linear (LDA) and a quadratic discriminant function (QDA). The simpler linear function LDA estimates the planes in the n-dimensional space that best discriminates the data classes [1]:

$$g_l(\mathbf{x}) = w_0 + \sum_{i=1}^{n} w_i x_i \tag{5}$$

where the coefficients $w_i$ are the components of the weight vector $\mathbf{w}$ and $n$ is the number of dimensions of the space. The quadratic function estimates an hyperquadric surface:

$$g_q(\mathbf{x}) = w_0 + \sum_{i=1}^{n} w_i x_i + \sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij} x_i x_j \tag{6}$$

The best choice of discriminant function depends on the nature of the data, and we will test both linear and quadratic functions.

Each of the five sports types is considered a class. In the classification phase, each new sample is assigned to the class with smallest misclassification cost.

In this work we do not consider undefined activities, such as warm up and exercises, as the number and variety of these activities might be unlimited, thus not representable in a single class.

## 5. EXPERIMENTS

For the experiments we use sports types which can be easily defined and thereby unambiguously annotated. From

| Truth \ Classified as | Badminton | Basketball | Soccer | Handball | Volleyball |
|---|---|---|---|---|---|
| **Badminton** | **29** | 0 | 0 | 1 | 0 |
| **Basketball** | 0 | **27** | 2 | 0 | 1 |
| **Soccer** | 0 | 0 | **29** | 0 | 1 |
| **Handball** | 0 | 0 | 1 | **27** | 0 |
| **Volleyball** | 0 | 2 | 0 | 0 | **26** |

**Table 1: Classification results for 146 video sequences used for tests in a 10-fold cross validation.**

recordings made in two similar indoor multi-purpose arenas we have five well-defined sports types available: Badminton, basketball, handball, soccer, and volleyball. We use 60 minutes of video recordings from each of the five sports types and divide them into 2-minutes sequences to get a total of 150 video sequences. The experiments are run as 10-fold cross validation; using one 10th of the data for test and the remaining part for training, then repeating the process 10 times with a new data subset for test each time.

For classification we test both linear and quadratic discriminant functions as described in section 4. The quadratic function fits the data best and obtain a correct classification rate of 94.5 %, while the linear discriminant function has a correct classification rate of 90.4 %. Table 1 shows the classification result of the 146 video sequences used for tests during ten iterations, using the quadratic discriminant function.

Of the 146 sequences, 138 are correctly classified and only 8 sequences are wrongly classified, giving a total correct classification rate of 94,5 %. The errors are distributed with 1-3 wrongly classified sequences for each sports type.

Comparable work from [3] obtained a correct classification rate of 90.8 % using the same five sports types, plus a miscellaneous class.

The Kalman tracking algorithm, including detection of people, is implemented in C# and runs real-time with 30 ms per frame. The 10-fold classification is implemented in Matlab and takes only 33 ms in total. Both are tested on an Intel Core i7-3770K CPU 3.5 GHz with 8 GB RAM.

# 6. CONCLUSION

In this paper we introduced a new idea for sports type classification. Based on tracklets found by a Kalman filter we extract four simple, but robust, features. These are used for classification with a quadratic discriminant analysis. Using a total of 150 video sequences from five different sports types in a 10-fold cross validation we obtained a classification rate of 94.5 %. The result is better than what was previously obtained in [3], while this new approach is also more general; it doesn't depend on the position of the players or direction of play.

Due to privacy issues, we used thermal imaging only. However, the classification approach presented is applicable for other image modalities. Only the detection step should be substituted with a different method, which could be a HOG detector or another general person detector.

The proposed method is independent of the type of arena and it is expected that it could easily be extended to outdoor arenas as well. With the current set-up where the entire arena is monitored from a far-view, the level of details available for each person is limited. In a future perspective higher resolution imaging devices is expected to be available, enabling a more fine-grained analysis of individual people, such as pose and motion of each body part.

# 7. ACKNOWLEDGMENTS

This project is funded by *Nordea-fonden* and *Lokale- og Anlægsfonden*, Denmark. We would also like to thank Aalborg Municipality for support and for providing access to their sports arenas.

# 8. REFERENCES

[1] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification*. Wiley-Interscience, 2nd edition, 2001.

[2] R. Gade, A. Jørgensen, and T. Moeslund. Long-term occupancy analysis using graph-based optimisation in thermal imagery. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2013.

[3] R. Gade and T. Moeslund. Sports type classification using signature heatmaps. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, June 2013.

[4] X. Gibert, H. Li, and D. Doermann. Sports video classification using HMMS. In *International Conference on Multimedia and Expo (ICME)*, 2003.

[5] R. E. Kalman. A new approach to linear filtering and prediction problems. *Transactions of the ASME–Journal of Basic Engineering*, 82(Series D):35–45, 1960.

[6] C. Krishna Mohan and B. Yegnanarayana. Classification of sport videos using edge-based features and autoassociative neural network models. *Signal, Image and Video Processing*, 4:61–73, 2010.

[7] J. Y. Lee and W. Hoff. Activity identification utilizing data mining techniques. In *IEEE Workshop on Motion and Video Computing (WMVC)*, Feb 2007.

[8] P. Mutchima and P. Sanguansat. TF-RNF: A novel term weighting scheme for sports video classification. In *IEEE International Conference on Signal Processing, Communication and Computing (ICSPCC)*, 2012.

[9] D.-H. Wang, Q. Tian, S. Gao, and W.-K. Sung. News sports video shot classification with sports play field and motion features. In *International Conference on Image Processing (ICIP)*, 2004.

[10] J. Wang, C. Xu, and E. Chng. Automatic sports video genre classification using Pseudo-2D-HMM. In *18th International Conference on Pattern Recognition (ICPR)*, 2006.

[11] Y. Yuan and C. Wan. The application of edge feature in automatic sports genre classification. In *IEEE Conference on Cybernetics and Intelligent Systems*, 2004.