Estimating the Number of Soccer Players using Simulation-based Occlusion Handling

Huda, Noor Ul; Jensen, Kasper Halkjær; Gade, Rikke; Moeslund, Thomas B.

Published in:
2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)

DOI (link to publication from Publisher):
10.1109/CVPRW.2018.00236

Publication date:
2018

Document Version
Accepted author manuscript, peer reviewed version

Estimating the Number of Soccer Players using Simulation-based Occlusion Handling

Noor Ul Huda, Kasper H. Jensen, Rikke Gade and Thomas B. Moeslund
Visual Analysis of People Lab, Aalborg University
{nuh,khje,rg,tbm}@create.aau.dk

Abstract

Estimating the number of soccer players is crucial information for occupancy analysis and other monitoring activities in sports analysis. It depends on player detection in the field that should be independent of the environment and light conditions. Thermal cameras are therefore a better option over normal RGB cameras. Detection of non-occluded players is doable but precise estimation of number of the players in groups is hard to achieve. Here we propose a novel method for estimating number of the players in groups using computer graphics and virtual simulations. Occlusion conditions are first classified by using distinctive set of features trained by a bagged tree classifier. Estimation of the number of players is then performed by maximum likelihood of probability density based approach to further classify the occluded players. The results show that the implemented strategy is capable of providing precise results even during occlusion conditions.

1. Introduction

Soccer is the most popular sport around the world [34]. The application of soccer video analysis includes strategy understanding, player action recognition, occupancy analysis and many more. Estimation of number of players is the foundation of understanding soccer especially if we need to know the occupancy in a particular field. The occupancy analysis can be achieved by counting the number of players with respect to the time stamp over a large period [11]. Player detection and correct estimation of a number of players in the sports field is the basic step in every sports analysis. A number of solutions [36] have been proposed for soccer analysis. These solutions normally employ a large number of cameras or used broadcast videos for analysis. This makes the whole system very complex to deploy in local sports fields. Furthermore, the communal trust of large-scale, high-resolution camera systems in public fields is harder to come by due to the general privacy issues.

Precisely estimating the number of players is a challenging topic due to various factors. These factors consist of occlusion, motion blur, varying illumination, outdoor weather, changing player sizes and inconsistency in appearance of the players. Even though multi-camera solutions improves the precision by providing more information [8]. They also increase the hardware and complexity of the whole system. RGB cameras are also effective in many cases but they are challenged in varying illumination conditions.

Consequently, we propose a thermal camera based solution for estimating the number of players in groups and counting. Three thermal cameras are installed on a single pole. The setup is able to capture the complete soccer field from a distant location. One of the advantages of using three thermal cameras on the same location is its ease of installation. The other obvious advantage of the thermal camera over normal cameras is the privacy preservation, because thermal view makes people identification almost impossible. Contrary to this, it is hard to detect and estimate the correct number of players, specially when they are in groups. This is because thermal cameras provide less textual information about player appearance. The main contribution of this work is a simulation-based approach that efficiently deals with occlusion in the thermal camera view.

1.1. State of the art

1.1.1 Player Detection in Soccer Field

Various solution for the detection of players in soccer fields have been proposed. These methods include background subtraction and spectator region extractions [24, 30, 20, 19, 8, 5]. However, there may remain some noisy regions in the image (line segments and other discontinuities in the image). Yoon *et al.* [43] proposed a solution based on player regions separation by defining thresholds for size, compactness, ratio of vertical to horizontal length and color distribution. Haung *et al.* [18] presented a shape-based player detection method in order to remove noisy areas from connected components. Yao *et al.* [42] used the confidence map for segmentation of players in broadcast videos. The

(a) (b) (c) (d) (e)

Figure 1: (a) shows the top camera view of the soccer field with camera positions, (b) is the reference frame from left camera view, (c) shows the binary Image, (d) is the mask of the image and (e) is the final image we get after applying background mask and morphology.

confidence map is generated from the output of a Hough forest.

A Motion graphic feature based system is proposed by Liu *et al.* [25]. Their approach is based on motion analysis and action recognition of players using SVM and optical flow analysis in broadcast videos. Mahmoudi *et al.* [28] proposed another motion based system. Their system also utilized optical flow analysis with Lucas-Kanade algorithm for detection and tracking of players. A method based on Markov Chain Monte Carlo data association and Kalman filter is proposed by Liu *et al.* [27]. A combined appearance and motion model for evaluating player regions is proposed by Sermetcan *et al.* [1]. They used two camera system for detection and tracking players. Direkoglu *et al.* [7] proposed an 8 camera based system for detection of player in the field. They proposed a diffusion equation based solution to make the whole system invariant of color and rotation information.

Beetz *et al.* [2] utilized ontology models of game together with motion trajectories of players for detection. They suggested Blackwellized Resampling particle Filter for tracking of players. Intensity variance and color based segmentation is used for player segmentation in their work. Liu et al. [26] proposed a context-conditioned motion based tracking model. They work is based on the fact that the player response in an existing situation in only a limited number of ways. Yang *et al.* [41] proposed edge detection and threshold based player detection. They used broadcast videos for testing their algorithm. Heydari *et al.* [17] used Multilayer Perceptron Neural Networks for classification of players in broadcast videos. Gerke *et al.* [15] proposed color histogram and spatiograms for the detection of players. They enhanced their work using histogram based features for the identification of players [14]. They also tested their algorithm on broadcast videos. Most of the literature is either based on evaluation on broadcast videos or large camera setup is employed for player detection and tracking. This leads to the lack of more simple and robust approach for estimating number of players.

1.1.2 Occlusion Handling

Occlusion is one of the major problems while dealing with sports videos. Khan *et al.* [31] proposed a color based segmentation method and Kalman filter for dealing with occlusion problems during tracking. Hayet *et al.* [16] performed detection based on point distribution model. They dealt with partial occlusion situations on specific video streams captured through multiple cameras with variable zoom and rotations. Iwase and Saito [20] proposed a solution based on 8 cameras for dealing with occlusion. Sabirin *et al.* [33] proposed free viewpoint based approach to cater occlusion while tracking. Kristoffersen *et al.* [23] perform people counting and occlusion handling by using stereo thermal camera setup in the street. They perform 3-D reconstruction and deal with occlusion based on clustering and tracking of the 3D point clouds. Manafifard *et al.* [29] suggested a detector that performs two-step blob detection (grass-based blob detection followed by an edge-based blob detection). They handled occlusion by a blob-guided PSO multi-player detection algorithm. Gade *et al.* [11] proposed a system based on identifying the occlusion in thermal cameras by defining thresholds on length and width of the blobs. They also enhance their work of player counting in indoor and outdoor sports arenas using constraint information of the stable periods by Graph search optimization [13]. In most of the literature, either the occlusion is handled in tracking of the players or complex system is employed to capture the videos at multiple angles. Choosing a threshold is also a compromise between false positives (spurious occlusions) and false negatives (missed occlusions) detections.

1.1.3 Thermal cameras

Automatic identification of human body includes both the visual and thermal information [4, 22, 37, 40, 32]. A comprehensive survey regarding thermal cameras and their application is performed by Gade *et al.* [12] Gade *et al.* [11, 13, 10] also proposed player counting and occupancy analysis by using thermal cameras in indoor sports arenas. Other work in the domain of thermal cameras for video

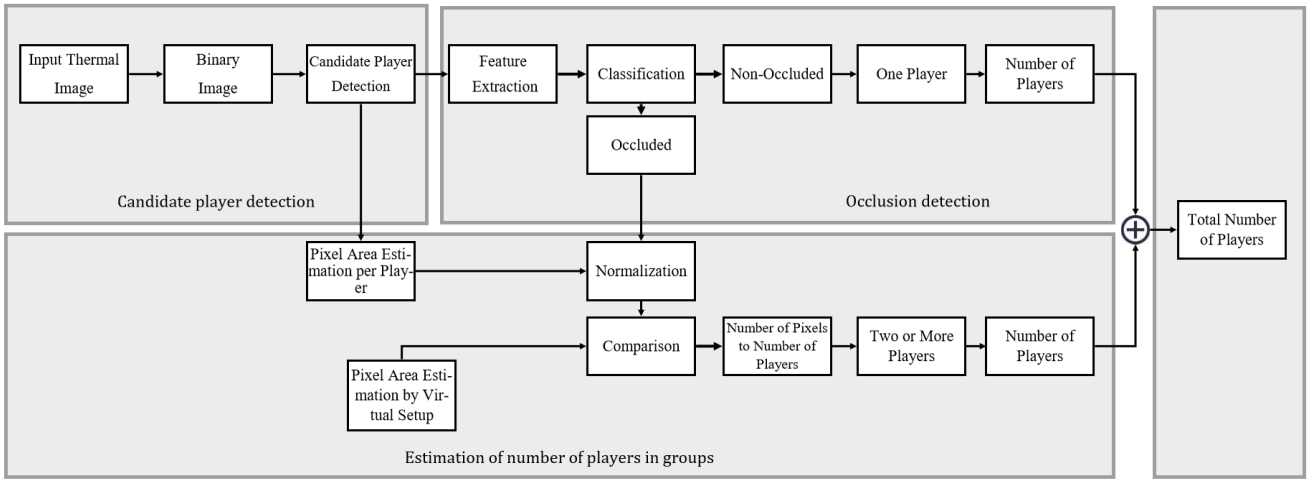


Figure 2: Flow diagram for the proposed system.

analysis and human detection is pedestrian detection and counting[23, 39]. Most of the work for utilizing thermal cameras in sports analysis is related to indoor sports arenas with a relevantly closed environment and small area of interest. Large outdoor sports fields are yet to be analyzed with thermal camera setups.

2. Proposed Method

Most vision systems for detection and counting of players in a sport field are either complex, in terms of number of cameras and controlled light environment, or they lack any supervised algorithm for the detection of player on a soccer field. In this work, we proposed a three staged supervised player detection and counting system i.e. candidate player detection, occlusion detection and estimation of number of players in each occluded group.

Player detection in soccer has always been a challenging task because of various factor i.e. weather (wind, rain, snow, clouds, etc.) and varying light conditions. Moreover, the shape, geometry and size of the player in the field vary with the angle and position of the camera.

To cope with all these issues here we propose a fixed three thermal camera-based solution that is independent of varying light and weather conditions and for the evaluation of our algorithm, we apply our approach to the outdoor field of soccer. The camera setup is shown in figure 1a. The proposed approach for estimating the number of players consist of the following steps. Given a video frame by the thermal camera, we compute a binary image. We have assumed that a player is any human on the field that could be a team player or a referee. From the resulting image a feature vector is extracted from each blob and afterwards, a bagged tree classifier is implemented to separate blobs into occluded and non-occluded players. Occluded players are then fed to a maximum likelihood of density estimation analysis to estimate the actual number of players in each occluded blob. The whole process, of occlusion detection and estimating number of players in the occluded blob, is illustrated in the figure 2 and described in details in the following.

2.1. Player Detection

The thermal camera captures grey scale where a warm object, which is a player in our case, appears to be brighter than the surroundings and background. The first step in our algorithm is to detect and separate these objects from the image. Maximum entropy based thresholding that finds the threshold value based on the sum of entropies is employed for the segmentation of these light objects [21]. There may remain some blobs and noise outside the field because of intensity variations or spectators. Those blobs are removed from the image by a manually marked geometry based field mask. Blobs smaller than specified minimum area are discarded as they may belong to some noise and morphological closing is applied to join the other small blobs. The blobs are then labeled using a contour-finding algorithm [35].

2.2. Occlusion Detection

People standing beside or behind each other often merge into one big blob. These blobs come under the category of occlusion. One of the major contribution in this paper is occlusion handling by utilizing only blob information. Where the aim is to decide if a blob is one player or more than one (occluded player).

Images obtained from a thermal camera do not carry much textual or color information. Also, the players have different posture, shape and size that depending on the distance and orientation of the cameras. Moreover, the players on the border of the field appear too small to detect. Therefore, the normal state of the art feature extraction methods

fails. The only thing that can be utilized is the shape and orientation of the blobs.

2.2.1 Feature Vector Information

Players, whether occluded or non-occluded, appear larger near the camera. However, at the same pixel position, an occluded blob appears to be larger than a non-occluded blob. The non-occluded blob always has a vertically oriented shape, as players are always vertically oriented from the camera point of view. For an automated system to distinguish between occluded and non-occluded regions, a feature vector based on the blob information is formed for each candidate region. If a binary image I contains N potential candidate regions, then the set representation for an image I is $I = \{I_1, I_2....I_N\}$. Each candidate region is considered as a sample for classification and represented by a feature vector containing all M features, i.e. for a sample non-occluded blob I_i the feature vector is $I_i = \{f_1, f_2, f_3....F_M\}$, where $i = \{1, 2, 3, ...N\}$. The set of features utilized here are:

- **Connected point slope:** The connected points and branch points say $C(x, y)$ of the skeleton of the blobs are founded by using [9]. In the case of non-occluded blobs all the founded points, $C = \{C_1, C_2, C_3, ...C_k\}$, are connected in a single vertical symmetry. While the blob of an occluded group of players normally have two or more vertical symmetries. In these cases, slopes between each connected point provides a useful information to distinguish between occluded and non-occluded players' blobs and is computed according to the following equation,

$$slope_k = \frac{y_{k+1} - y_k}{x_{k+1} - x_k} \tag{1}$$

  Where k represents a point in the set of all potential connected points, K from $1 : k$, with $k = 1$ is the top branch point and $K = k$ is the bottom-most branch point. In the case of non-occluded player blobs the slopes would be greater between each connected point, whereas the in case of occlusion, where there are two or more vertical symmetries the slopes would be smaller.

- **Connected point distance:** This is the distance between the connected points in the skeleton of a blob. It can be observed in figure 3d that for some connected points, smaller slopes can occur in non-occluded cases as well. In that cases the distance between the connected points contain useful information. So, If two connected points with smaller slope have a larger distance between them, then they probably belong to an occluded blob otherwise not, as shown in the figure 3c.

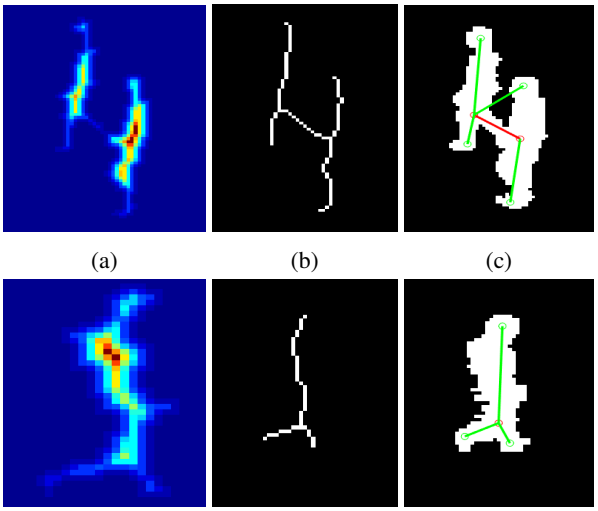


Figure 3: (a), (b) and (c) show the heat map, skeleton mask and connected points of an occluded region, (d), (e) and (f) show the heat map, skeleton mask and connected points of a non-occluded region. Note that red circles in (c) and (f) show the connected points whereas green circles and lines are the branch points.

- **Convex area:** It is the area of the convex hull of a blob. The convex area of occluded blobs appears to be larger than the convex area of non-occluded blob at the same position. But as we move away from the camera both occluded and non-occluded blobs appear to be small. So the blob area with respect to pixel distance from the camera is considered as a feature in our case. The pixel distance is calculated by equation 2.

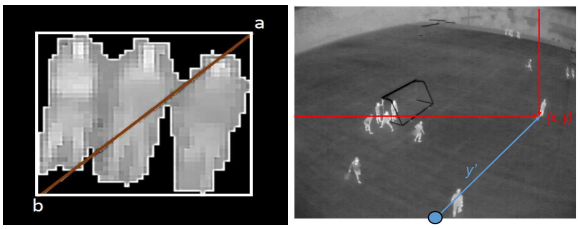
$$y' = \sqrt{|\ \frac{640}{2} - y\ |^2 + |\ 480 - x\ |^2} \tag{2}$$

  Here x and y are the pixel locations in image I (varying from 0 to 640 and 0 to 480 respectively). y' is the pixel distance with respect to the camera (see figure 4b) and 640 x 480 is the size of the image.

- **Diagonal Length of bounding box:** The diagonal distance of the bounding box (figure 4a) is the last feature to be used for classification. This distance would be larger for occluded blobs then non-occluded blobs.

2.2.2 Classification using Bragged Tree Classifier

Feature extraction is followed by the implementation of Bragged tree classifier [3] to distinguish between occlusion and non-occlusion blobs. The training dataset used



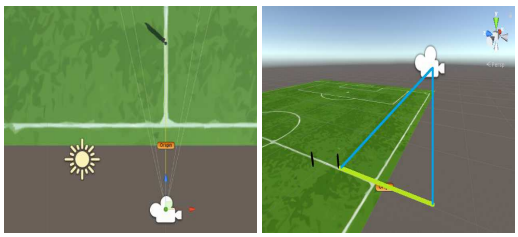
(a) (b)

Figure 4: (a) The diagonal length of a bounding box is measured between the corners 'a' and 'b', (b) shows the pixel distance calculation where the blue lines are the pixel distances and red lines are the original distances in the image plane.

for the classifier includes 1700 non-occluded and 120 occluded player blobs samples collected from over three different soccer videos. Evaluation of the proposed classifier was performed using k-folds cross validation with folds selected to be 5. Results are explained in section 3.2.

2.3. Estimating the number of players

Here we present a novel method for estimating the number of players in an occluded blob. Our method estimates the number of players in a blob by comparing the size of the detected blob with different likelihoods of learnt blob sizes created in virtual environment. A virtual setup of a football field with real world field coordinates, camera height, viewing angle and resolution is created using unity [38]. A human body is modeled as cylindrical blobs. The height, depth and width of the cylinders are taken as standard person height and width. Virtual player occlusion is created by considering the fact that occlusion can occur in a finite number of possible ways. All of these are simulated and a likelihood density is learned for each distance from the camera. Shadows are not considered in our work as the background surface is non-reflecting and no shadow occurs in case of thermal view. The virtual setup is illustrated in figure 5



(a) Top View of our setup (b) Side View of our setup

Figure 5: Virtual setup.

The process of simulating data for occlusion of two persons is

1. One static player is placed at minimum possible distance from the camera in the field.
2. The second player is shifted horizontally towards the first player from right to left in steps of 0.05 meter until they occlude in the 2D camera view.
3. The instant occlusion occurs the algorithm measures the combined pixel area of the blob.
4. The second player is shifted until the players are non-occluded in the camera. For each step the pixel area of the blob is measured.
5. The second person now shifts 0.05 meters above the previous position and steps 2-5 are repeated until no occlusion is present.
6. The first static person is then moved 1m further away from the initial position and the process (steps 2-6) is repeated for the entire field.

Since the size of a blob to a large degree is independent of the viewing direction, the steps above are only required for one particular viewing direction, and hence the size of a blob only depends on the distance from the camera. The process for three and four players follows a similar procedure except that it forms more combinations of static and moving players. The process could be repeated for higher number of players but the data set we are using is having a maximum of four occluded players.

2.3.1 Maximum likelihood based density estimation

The processes above result in 9978 possible occlusion combinations for two players, 12401 possible occlusion combination for three players and 33001 possible occlusion combination for four players. The size of each combination is normalized by the size of one simulated person at that particular distance. This results in a likelihood distribution that expresses the size as a function of the number of persons. This is illustrated in figure 6. After classification of occlusion, see figure 2, pixel distance with respect to camera is first calculated for each occluded blob (see equation 2). Then the blob is normalized with respect to the pixel area of a one-player blob, so the measure can be compared with the simulated data. In order to perform this normalization, an analytic function is learned from one-player blobs and their distances to the camera, see figure 7. This lead us to formulate a general relationship between pixel distance from the camera and average pixel area of a person that can be used for estimating the area of one person at every pixel location. The relationship is dependent on external parameters like camera height and tilt angle.

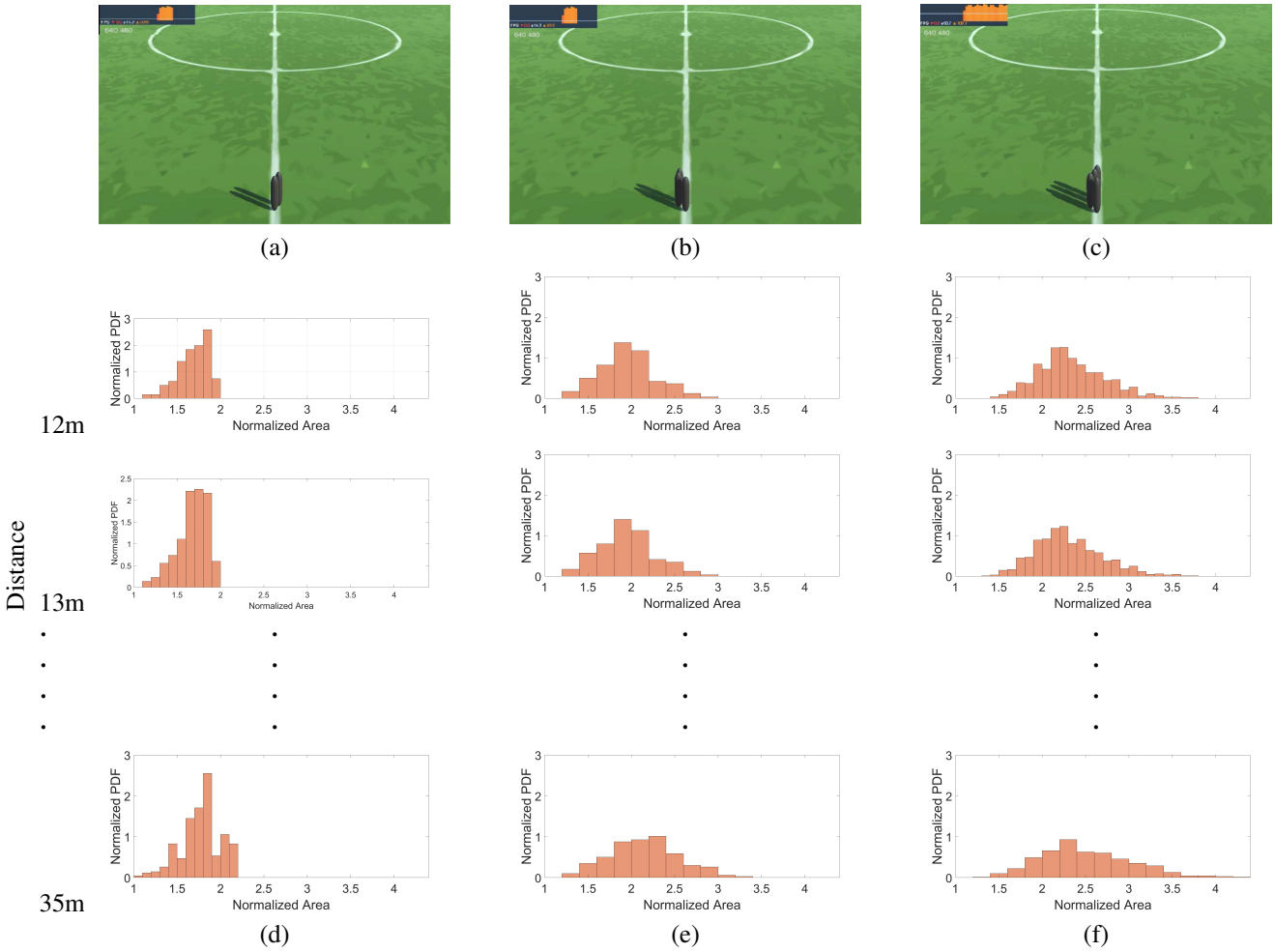


Figure 6: (a), (b) and (c) are the examples of virtual projections of two, three and four occluded players, respectively. (d), (e) and (f) are the normalized probability densities with respect to distance.

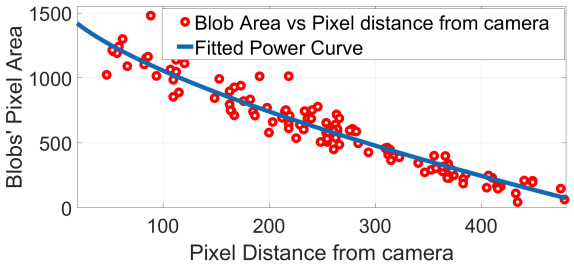


Figure 7: Relationship between blob area and camera distance.

Normalized area with its pixel distance is compared with the virtually generated likelihood distributions. The one that matches best determine the actual number of players in a particular blob. In case, the pixel distance does not match with any of the learn likelihood densities, the nearest neighbor distance is considered in that case.

3. Experiments

3.1. Data and Setup

As there exist no publicly available soccer data captured with thermal cameras, we have used our own captured data for evaluation purposes. The video is captured with an AXIS Q1922 LWIR sensor with 57 degrees of horizontal Field of View (FOV) and a resolution of 640x480. The camera is placed on a pole that is 4.6 meter away from the field at a height of 10.5 and tilt angle of 27°. The data that are used for the testing contain 5 minutes of video with 8990 frames containing 71443 players. The ground truth is marked by manually counting the number of players in each frame.

3.2. Results

In this paper, we present the results for the left view camera. Here the evaluation of our features with Bagged tree classification model. This evaluation includes the comparison of our features with state of art human detection histogram of oriented gradients (HOG) features [6] which are still used in many state of the art player detection algorithms [1, 15]. [7, 29, 15] have also performed this comparison analysis. HOG is trained on grey scale images for classification of occluded and non-occluded blobs. Figure 8 shows the comparison of two features in terms of ROC. Other comparison measures for evaluation of our features are presented in Tables 1.

Clearly our proposed features gather with classifier per-

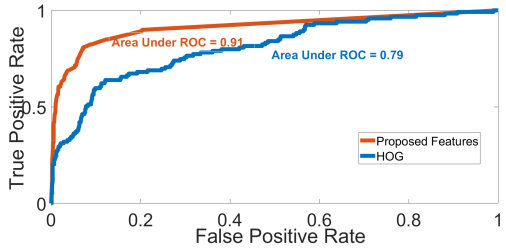


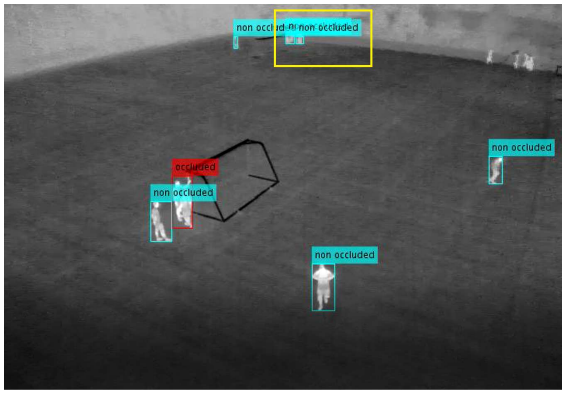
Figure 8: Receiver Operating Curve for the [6] and proposed method.

	Acc	P %	R%	AUC	TT
HOG+SVM [6]	94.3	62.1	77.3	0.79	142.0sec
Proposes	96.1	77.3	84.1	0.91	1.5sec

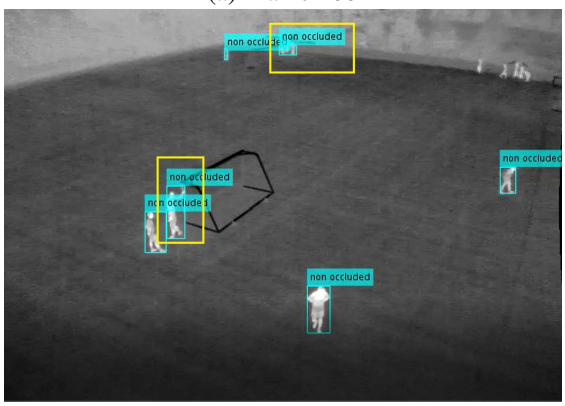
Table 1: Comparison of the methods for occlusion detection with [6] in terms of accuracy (Acc), precision (P), Recall (R), area under receiver operating curve (AUC) and training time (TT).

formed better than [6]. 100% accuracy can not be achieved due to more false negative cases because of the factors that include fully occluded cases and occlusions in farther blobs interpreted as non-occlusions. This is because the blobs appear to be too small to detect and classify, see figure 9.

Next we evaluate our method for estimating the number of players compared with the ground truth. The results are is presented in figure 10. Figure 11 shows some qualitative results after estimation. The results presented demonstrate the precision of the proposed method. Another important observation from figure 10 is that the proposed approach yields better accuracies in former frames. The reason is that most of the players appear at the farther boundary of the field in the last part of the test data. This makes the occlusion detection more challenging and hence estimation become more uncertain.



(a) Frame 400



(b) Frame 480

Figure 9: Non-Occluded vs Occluded, yellow boxes show some false negative cases.

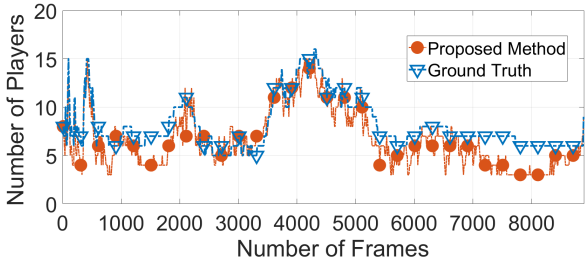
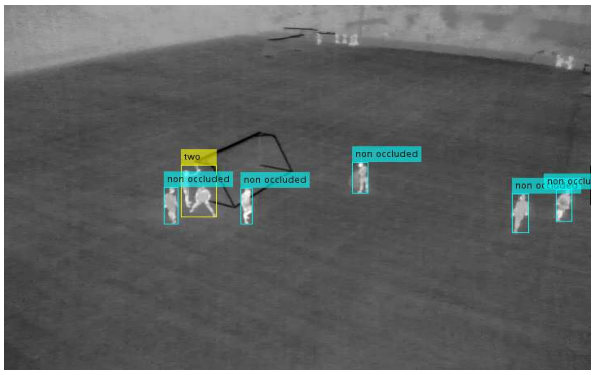


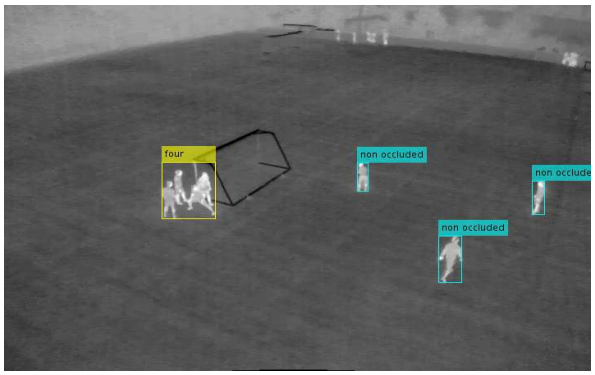
Figure 10: Orange line represents the estimated number of people and the blue line shows the ground truth.

For the quantitative evaluation, a comparative analysis of our proposed algorithm with previously proposed algorithms is presented in table 2.

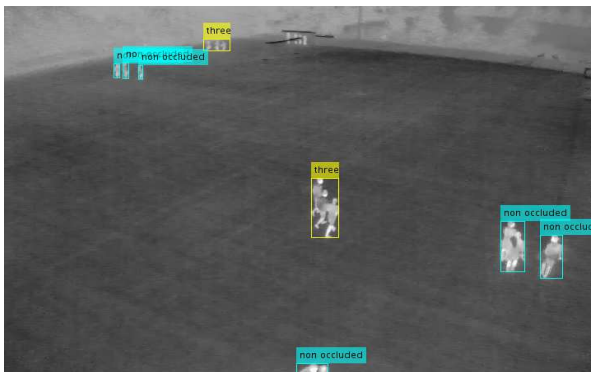
Our system clearly outperforms in terms of precision. It should be noted that everyone has used their own local datasets and hence no direct comparison is possible. [27, 17, 29, 14] have used broadcast videos for the detection of players. We are performing estimation of number



(a) Frame = 50



(b) Frame = 350



(c) Frame = 4000

Figure 11: Results after estimation.

of players rather than just detection. Also we are working on non-commercial videos and not utilizing any temporal information since this is not desirable in ordinary setups where bandwidth and on-site processing power can be problematic. [11] have used thermal cameras for evaluation of their counting algorithm but tested their methodology in indoor sports arenas having closed environment and relatively small area of interest. We are testing our algorithm in a large outdoor soccer field. Better accuracy can probably be achieved by including boundary information and temporal data like in [11].

Method	Acc%	P%	R %
Liu *et al.* [27]	–	88.6-92.3	88.8-92.1
Heydari *et al.* [17]	96.5	–	–
Manafifard *et al.* [29]	–	93	91
Gade *et al.* [11]	95.5	–	–
Gerke *et al.* [14]	–	83-90	66-78
Proposed Method	**81.4**	**97.8**	**78.8**

Table 2: Comparison of our proposed method against previous techniques in terms of Accuracy(Acc), Precision(P) and Recall(R).

4. Conclusion and Discussion

This paper proposed an automated system for precise counting of players using thermal cameras. A detailed feature vector for each candidate region is formed by using the shape and geometry of the blobs. We used Bagged tree classifier for detection of occlusion. In order to further classify the number of players in occlusion, we proposed a simulation based method. 8990 frames are used for evaluation of the proposed technique for detection of occlusion and estimation of number of players. No ideal conditions are assumed, so it is critical to know that the datasets that we have used contain all types of variations with respect to posture and position of players. The results showed that our proposed method for estimating number of players achieved a high precision, which makes our system suitable for counting precise number of players in groups. Our proposed system is not dependent on light and weather conditions, which make our system more practical for local non-commercial sports analysis.

The mapping from visual appearance of occluding people to the number of individuals could be based on other features than the number of pixels as done in this work. Given sufficient training data more sophisticated hand-crafted features or automatically extracted features via a deep learning approach could probably work. But such approaches are likely to require large amounts of annotated data to generalize to arbitrary setups. And since our work is to be applied in many different setups focus has been on a simple feature and an easy training procedure. In fact, for a new setup we need only to input the external camera parameters to the virtual simulation and re-render figure 7 and then learn the size of a 1-person blob as a function of distance to the camera in a particular setup. This makes our approach easy to adapt. However, as more fields are analyzed annotated data are automatically collected and future work therefore includes an investigation into the use of deep learning for learning a general mapping from blobs (or bounding boxes) to the number of people.

References

[1] S. Baysal and P. Duygulu. Sentioscope: A soccer player tracking system using model field. *IEEE Transactions on Circuits and Systems for Video Technology*, 26(7):1350–1362, 2016. 2, 7

[2] M. Beetz, S. Gedikli, J. Bandouch, B. Kirchlechner, N. V. Hoyningen-Huene, and A. C. Perzylo. Visually tracking football games based on tv broadcasts. *20th international joint conference on Artificial intelligence*, 2007. 2

[3] L. Breiman. Bagging predictors. *Machine Learning, Springer*, 24(2):123–140, 1996. 4

[4] C. Chen, R. Jafari, and N. Kehtarnavaz. A survey of depth and inertial sensor fusion for human action recognition. *Multimedia Tools and Applications, Springer*, 76(3):4405–4425, 2017. 2

[5] K. Choi and Y. Seo. Automatic initialization for 3d soccer player tracking. *Pattern Recognition Letters*, 32(9):1274–1282, 2011. 1

[6] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, 2005. 7

[7] C. Direkoglu, M. Sah, and N. E. O'Connor. Player detection in field sports. *Machine Vision and Applications, Springer*, 29:187–206, 2018. 2, 7

[8] T. D'Orazio, M. Leo, P. Spagnolo, P. L. Mazzeo, N. Mosca, M. Nitti, and A. Distante. An investigation into the feasibility of real-time soccer offside detection from a multiple camera system. *IEEE Transactions on Circuits and Systems for Video Technology*, 19(12):1804–1818, 2009. 1

[9] J. Feldman and M. Singh. Bayesian estimation of the shape skeleton. *Proceeding of National Academy of Sciences of the USA*, 2006. 4

[10] R. Gade, A. Jørgensen, and T. B. Moeslund. Occupancy analysis of sports arenas using thermal imaging. *International Conference on Computer Vision Theory and Applications*, 2012. 2

[11] R. Gade, A. Jørgensen, and T. B. Moeslund. Long-term occupancy analysis using graph-based optimisation in thermal imagery. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2013. 1, 2, 8

[12] R. Gade and T. B. Moeslund. Thermal cameras and applications: a survey. *Machine Vision and Applications, Springer*, 25(1):145–262, 2014. 2

[13] R. Gade and T. B. Moeslund. Constrained multi-target tracking for team sports activities. *IPSJ Transactions on Computer Vision and Applications, Springer*, 10(2), 2018. 2

[14] S. Gerke and K. Mller. Identifying soccer players using spatial constellation features. *ACM KDD Workshop on Large-Scale Sports Analytics*, 2015. 2, 7, 8

[15] S. Gerke, S. Singh, A. Linnemann, and P. Ndjiki-Nya. Unsupervised color classifier training for soccer player detection. *Visual Communications and Image Processing (VCIP)*, 2013. 2, 7

[16] J. B. Hayet, T. Mathes, J. Czyz, J. Piater, J. Verly, and B. Macq. A modular multi-camera framework for team sports tracking. *Advanced Video and Signal Based Surveillance, IEEE*, 2015. 2

[17] M. Heydari and A. M. E. Moghadam. An mlp-based player detection and tracking in broadcast soccer video. *International Conference on Robotics and Artificial Intelligence (ICRAI)*, 2012. 2, 7, 8

[18] Y. Huang, J. Llach, and S. Bhagavathy. Players and ball detection in soccer videos based on color segmentation and shape analysis. *Multimedia Content Analysis and Mining*, 4577:414–425, 2007. 1

[19] N. Inamoto and H. Saito. Virtual viewpoint replay for a soccer match by view interpolation from multiple cameras. *IEEE Trans Multimedia*, 9(6):1155–1166, 2007. 1

[20] S. Iwase and H. Saito. Tracking soccer players based on homography among multiple views. *Visual Commun Image Proc*, 5150:283–292, 2003. 1, 2

[21] J.N.Kapur, P.K.Sahoo, and A.K.C.Wong. A new method for gray-level picture thresholding using the entropy of the histogram. *Computer Vision, Graphics, and Image Processing, Springer*, 29(3):273–285, 1985. 3

[22] T. Ko. A survey on behavior analysis in video surveillance for homeland security applications. *37th IEEE Applied Imagery Pattern Recognition Workshop*, 2008. 2

[23] M. S. Kristoffersen, J. V. Dueholm, R. Gade, and T. B. Moeslund. Pedestrian counting with occlusion handling using stereo thermal camera. *SENSORS*, 62(16), 2016. 2, 3

[24] M. A. M. Laborda, E. F. T. Moreno, J. M. del Rincn, and J. E. H. Jaraba. Real-time gpu color-based segmentation of football players. *Multimedia Tools and Applications*, 73(3):1617–1642, 2012. 1

[25] G. Liu, D. Zhang, , and H. Li. Research on action recognition of player in broadcast sports video. *International Journal of Multimedia and Ubiquitous Engineering*, 9(10):297–306, 2014. 2

[26] J. Liu, P. Carr, R. T. Collins, and Y. Liu. Tracking sports players with context-conditioned motion models. In T. B. Moeslund, G. Thomas, and A. Hilton, editors, *Computer Vision in Sports*, chapter 6, pages 133–132. Springer, 2014. 2

[27] J. Liu, X. Tong, T. W. W. Li, Y. Zhang, and H. Wang. Automatic player detection, labeling and tracking in broadcast soccer video. *Pattern Recognition Letters, Elsevier Science Inc*, 30(2):103–113, 2009. 2, 7, 8

[28] S. Mahmoudi, M. Kierzynka, P. Manneback, and K. Kurowski. Real-time motion tracking using optical flow on multiple gpus. *Bulletin of the Polish academy of sciences, Technical Sciences*, 62(1), 2014. 2

[29] M. Manafifard, H. Ebadi, and H. A. Moghaddam. Multi-player detection in soccer broadcast videos using a blob-guided particle swarm optimization method. *Multimedia Tools and Applications, Springer*, 76(10):12251–12280, 2016. 2, 7, 8

[30] R. Martn and J. M. Martnez. A semi-supervised system for players detection and tracking in multi-camera soccer videos. *Multimedia Tools and Applications*, 73(3):1617–1642, 2014. 1

[31] I. K. MM. Khan, TW. Awan and Y. Soh. Tracking occulded objects using kalman filter and color information. *International Journal of Computer Theory and Engineering*, 6(5), 2014. 2

[32] T. B. Moeslund, A. Hilton, V. Krüger, and L. Sigal. *Visual Analysis of Humans: Looking at People*. Springer, Germany, 2011. 2

[33] H. Sabirin, Q. Yao, K. Nonaka, H. Sankoh, and S. Naito. Semi-automatic generation of free viewpoint video contents for sport events: Toward real-time delivery of immersive experience. *IEEE MultiMedia*, (99), 2018. 2

[34] A. statistical based analysis of world's most popular sports. Biggest Global Sports. http://www.biggestglobalsports.com/worlds-biggest-sports/4580873435. [Online; accessed 03-March-2018]. 1

[35] S. Suzuki. Topological structural analysis of digitized binary images by border. *Computer Vision, Graphics, and Image Processing, Springer*, 30(1):32–46, 1985. 3

[36] G. Thomas, R. Gade, T. B.Moeslund, P. Carr, and A. Hilton. Computer vision for sports: Current applications and research topics. *Computer Vision and Image Understanding, Science Direct*, 159:3–18, 2017. 1

[37] P. Turaga, R. Chellappa, V. S. Subrahmanian, and O. Udrea. Machine recognition of human activities: A survey. *IEEE Transactions on Circuits and Systems for Video Technology*, 18(11):1473–1488, 2008. 2

[38] Unity Technologies. Unity. https://unity3d.com/unity. 5

[39] W. Wang, J. Zhang, and C. Shen. Improved human detection and classification in thermal images. *IEEE International Conference on Image Processing*, 2010. 3

[40] W. Wei and A. Yunxiao. Vision-based human motion recognition: A survey. *Second International Conference on Intelligent Networks and Intelligent Systems, IEEE*, 2009. 2

[41] Y. Yang and D. Li. Robust player detection and tracking in broadcast soccer video based on enhanced particle filter. *Journal of Visual Communication and Image Representation Elsevier*, 46:81–94, 2017. 2

[42] A. Yao, D. Uebersax, and J. G. V. Gool. Tracking people in broadcast sports. *Joint Pattern Recognition Symposium, Springer*, 6376:151–161, 2010. 1

[43] H.-S. Yoon, Y. lae J. Bae, and Y. kyu Yang. A soccer image sequence mosaicking and analysis method using line and advertisement board detection. *ETRI*, 24(6):443–454, 2002. 1