Sports Type Classification using Signature Heatmaps

Rikke Gade and Thomas B. Moeslund
Visual Analysis of People Lab
Aalborg University, Denmark
{rg, tbm}@create.aau.dk

Abstract

Automatic classification of activities in a sports arena is important in order to analyse and optimise the use of the arenas. In this work we classify five sports types based only on occupancy heatmaps produced from position data. Due to privacy issues we use thermal imaging for detecting people and then calculate their positions on the court using homography. Heatmaps are produced by summarising Gaussian distributions respresenting people over 10-minute periods. Before classification the heatmaps are projected to a low-dimensional discriminative space using the principle of Fisherfaces. Our result using two weeks of video are very promising with a correct classification of 90.76 %.

1. Introduction

Sport is an important part of the modern society. The amount of money invested in sport is huge, both from government, private and personal funding. A large part of those money is invested in the facilities every year for maintenance and new constructions. It is therefore of high interest to evaluate the use of the existing arenas in order to optimise the use of the facilities.

In this work we focus on recognising the activities observed in sports arenas. This information will be useful for both the evaluation made by the administrators of the arena as well as for the coach or manager of a sports team. Our goal is to recognise five common sports types observed in an indoor arena; badminton, basketball, indoor soccer, handball, and volleyball. To overcome any privacy issues we apply thermal imaging, which produces images where pixel values represent the observed temperature. Thereby it is possible to detect people without identification.

No previous work on recognising sports types has been based on thermal imaging. All existing work use visual cameras, and a few include audio as well. For features some works use edges that represent court lines and players. The sports categories can then be classified by edge directions, intensity, or ratio [6, 15]. Also based on the visual appearance of the court is a method that use the dominating colours of the image as features [9]. The dominant colour can also be combined with motion features [12, 11, 4] or combined with dominant gray level, cut rate and motion rate [10]. From the visual image SURF features [8] and autocorrelograms [13] can be extracted and used for classification. A combination of colour, edges, shape and texture has also been proposed by using six of the MPEG-7 descriptors [14]. One method is based only on location data and classifies sports categories by short trajectories [7].

After feature extraction the classification methods are based on well-known methods, such as k-means and Expectation Maximization for clustering, and decision trees, SVM, Hidden Markov models, Neural Network and Naive Bayesian for classification.

Most existing works are based on colour imaging, and many of them rely on the dominant colour of the fields as well as detection of court lines. These methods presume that each sports type is performed on a court designed mainly for one specific sport. In our work we aim to distinguish different sports types performed in the same arena, meaning that any information about the environment is not useful. Furthermore, due to privacy issues, we have chosen to use thermal imaging, which provides heat information only. Figure 1 shows an example of the thermal image, which is combined from three cameras in order to cover the full court.

Our hypothesis is that it is possible to classify five different sports types using a global approach based on position data only.

2. Approach

This work is based on occupancy heatmaps, which are summations of the registered positions of people over a given time span. It is believed that a heatmap is unique among a limited number of sports types.

Figure 2 shows examples of signature heatmaps, which are typical heatmaps for each sports type. Two heatmaps of miscellaneous activities are also shown. Each heatmap covers a 10-minute period.

978-0-7695-4990-3/13 $26.00 © 2013 IEEE
DOI 10.1109/CVPRW.2013.145

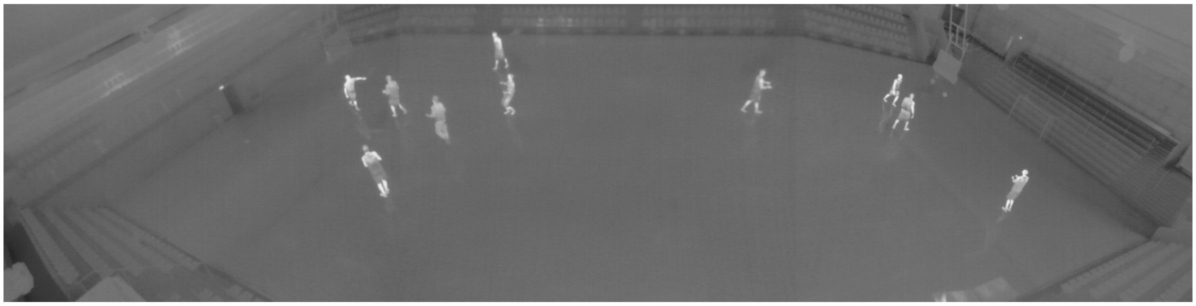


Figure 1. Example of an input image.

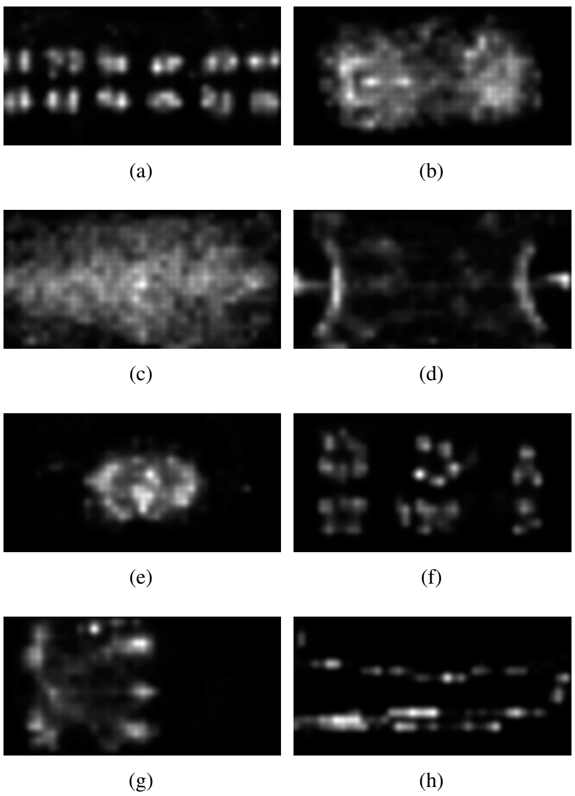


Figure 2. Signature heatmaps of (a) badminton, (b) basketball, (c) soccer, (d) handball, (e) volleyball, (f) volleyball (three courts), (g) miscellaneous, (h) miscellaneous.

The approach of this work is to detect individual people and use a homography to calculate their position at the court. A summation of the positions over time results in the occupancy heatmaps. These heatmaps will be classified after reducing the number of dimensions with PCA and Fisher's Linear Discriminant.

3. Detection

In recent work we have developed a method for counting people in sports arenas, based on thermal images [3]. We build upon this work by using the same detection method. For these indoor environments it is assumed that people are warmer than the surroundings, so that people can be segmented using an automatic threshold method. This method calculates the threshold value that maximises the sum of the entropy [5]. After binarising the image, ideally the people are now white and everything else in the image is black. There are, however, challenges to this assumption. This can be observations of non-human warm objects or reflections from people in the floor. These false detections must be reduced. Likewise, in order to detect individual people, partial occlusions must be handled. Solutions to these problems are described in [3]. Figure 3 shows the segmentation where detected people are marked with a green box.

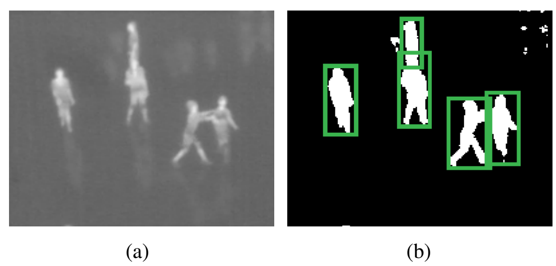


Figure 3. (a) is the input image and (b) shows the detected people (marked with a green box) after binarisation.

As spectators, coaches and other persons around the court are of no interest in this work, the image must be cropped to the border of the court before processing. Since each sports type has its own court dimensions, a single choice of border is not feasible. Handball and soccer are played on a 40×20 metres court, which is also the maximum court size in the observed arena. The volleyball court is 18×9 metres, plus a free zone around the court, which is minimum 3 metres wide, and the standard basketball court is 28×15 metres. Badminton is played on up to six courts of 13.4×6.1 metres. The court dimensions and layout in relation to each other are illustrated in figure 4. On the arena floor all court lines are drawn on top of each other, but here we have split it in two drawings for better visibility. Note that volleyball can be played on either three courts without free zones or on one court including the free zone.

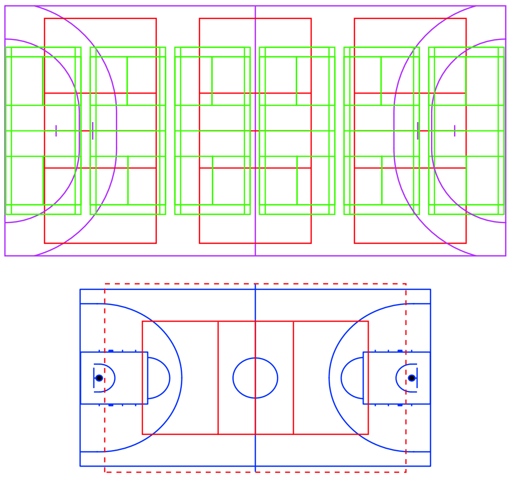


Figure 4. The outlines of the different courts illustrated. Purple: handball (and soccer), green: badminton, red: volleyball, blue: basketball. Drawn in two figures to increase the visibility.

During basketball and volleyball matches coaches and substitutes will be sitting within the dimensions of the handball court, and would be unwanted detections if we cropped only to the largest court dimensions. Considering the illustrated court dimensions it is therefore decided to operate with two different court sizes, 40×20 metres and 28×15 metres. In test cases it is of course not known which sport is performed and thereby not known which court size to choose. Instead both options will be tried out for all data. The classification process will be further described in section 4.

3.1. Occupancy heat maps

When a person is detected, the image coordinates of the bottom center of the bounding box are considered the position of the person, and will be converted to world coordinates using a homography. Since the input image is combined from three cameras, each observing either the left, middle or right part of the court, each camera needs one homography to calculate the transformation. This assumes that the cameras are perfectly rectified. During an initialisation we instead find the corresponding points in image and world coordinates for each five metres in both x- and y-direction. One homography is then calculated for each 5×5 m square.
To represent the physical area of a person, a standard person is represented by a 3-dimensional Gaussian distribution with a standard height of 1, corresponding to 1 person, and a radius corresponding to 1 metre for 95% of the volume. To take into account the uncertainty of the detections, the height of Gaussian distributions will be scaled by a certainty factor c. This depends on the ratio of white pixels (r) within a rectangle with a height corresponding to two metres, and a width being one third of the height. By tests it is found that most true detections have a ratio between 30 % and 50 %, while less than 1 % of the true detections lie below 20 % or above 70 %. We choose to discard all detections below 20 % and assign a value between 0.8 and 1 to all other detections:

$$c(i) = \begin{cases} 0, & \text{if } r < 20\% \\ 0.8, & \text{if } r > 60\% \\ 0.9, & \text{if } r < 30\% \parallel 50 < r < 60\% \\ 1, & \text{otherwise} \end{cases} \tag{1}$$

Figure 5 shows an example of the occupancy calculated for a single frame. Six people are detected with different certainty factors.

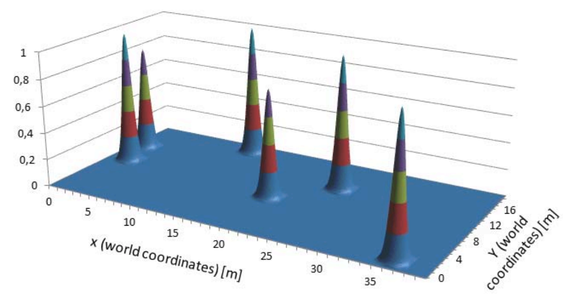
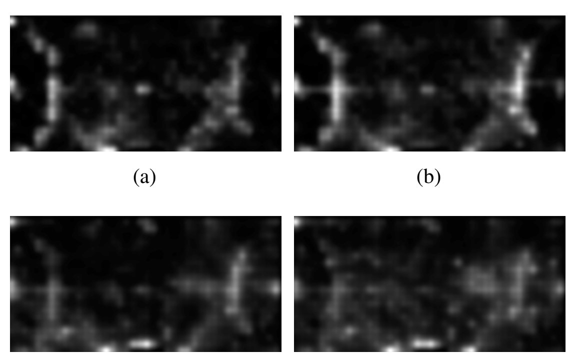


Figure 5. Occupancy for one single frame. Each person is represented as a Gaussian distribution.

The final occupancy heatmaps, as shown in figure 2, are constructed by adding up the Gaussians over time. The time span for each heatmap should be long enough to cover a representative section of the games and still short enough to avoid different activities to be mixed together. To decide on the time span, a comparison has been conducted between 5-, 10-, 20- and 30-minutes periods. An example of four heatmaps with the same end-time is shown in figure 6.



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

Figure 6. Heatmaps with same end-time and different time span: (a) 5 minutes, (b) 10 minutes, (c) 20 minutes, (d) 30 minutes.

The comparison in figure 6 illustrates the situation where a handball team start with exercises and warm-up, before playing a short handball match. The end-time for each heatmap is the same. The 30-minute period (figure 6(d)) is too long, the warm-up and game is mixed together such that no activity is recognisable. Between the 5-, 10- and 20-minute periods the 10-minute period (figure 6(b)) shows the most clear pattern. The same is observed in other comparisons, therefore it is chosen to let each heatmap cover 10 minutes. We will shift the starting time 5 minutes each time, so that the periods overlap and the resolution of classifications will be 5 minutes.

4. Classification

We wish to classify the sport based on the heatmaps only. These are images with a resolution of 200×400 pixels, so each heatmap can be considered a sample in a 80,000-dimensional space. Principal Component Analysis (PCA) is a well-known method for dimension reduction, but since it uses non-labeled data and seeks the dimensions with largest variance between all samples, there is a risk that the differences between classes are not conserved. Fischer's Linear Discriminant (FLD) seeks the directions that are efficient for discrimination between classes [2]. However, using FLD introduces the small sample size problem: In order to have a non-singular within-class scatter matrix (S_W) it is necessary to have more samples than dimensions. As we have a 80,000-dimensional space, it is not realistic to have a sample size of $n >$ 80,000. In order to solve this problem we will adapt the idea of Fisherfaces for face recognition [1]: First, project the sample vectors onto the PCA space of r dimensions, with $r \leq rank(S_W)$ and then compute the Fisherimage in this PCA space.

4.1. Dimension reduction

PCA is performed by pooling all training samples and calculating the directions with largest variance. The PCA will only have as many non-zero eigenvalues as the number of samples minus one, which will be significantly less than the original 80,000 dimensions. We choose to reduce the space to the 20 dimensions with largest eigenvalues, as these eigenvalues represent a significant part of the total variance. All heatmaps are projected to the new 20-dimensional space before further processing.

4.2. Fischer's Linear Discriminant

Using Fisher's Linear Discriminant the optimal projection of the data is found such that the ratio of the between-class scatter S_B and the within-class scatter S_W is maximised:

$$W_{opt} = arg \max_W \frac{|W^T S_B W|}{|W^T S_W W|} \tag{2}$$

where W_{opt} is a matrix with orthonormal coloumns, consisting of the set of generalised eigenvectors of S_B and S_W corresponding to the m largest eigenvalues. There are at most $c - 1$ non-zero generalised eigenvalues, where c is the number of classes.

The between-class scatter matrix S_B and the within-class scatter matrix S_W are defined as:

$$S_B = \sum_{i=1}^{c} N_i(\mu_i - \mu)(\mu_i - \mu)^T \tag{3}$$

$$S_W = \sum_{i=1}^{c} \sum_{x_k \in X_I} N_i(\mu_i - \mu)(\mu_i - \mu)^T \tag{4}$$

where μ_i is the mean image of class X_i, and N_i is the number of samples in class X_i [1].

4.3. Final classification

The training data is projected to the new space found by FLD, and the mean coordinate for each class is calculated. Each test sample to classify is projected to the same space, and the nearest class is found using the Euclidean distance. We use video from a public sports arena, which includes a lot of undefined activities. Besides the five traditional sports types we therefore define a category of miscellaneous activities. This can include everything from specific exercises and warm-up, to cleaning the floor and an empty arena. This category will be trained as a class in the same way as each sports type. Since miscellaneous contains very different heatmaps, it could be argued that this class will end up as a mean image of all other classes. However, by treating it as a class like the other sports types, the FLD will take care of projecting the data to a space that, as far as possible, discriminates the classes.

As described in section 3 we will use two different court dimensions for tests. The final classification for each time span should therefore choose between the classification of these two heatmaps. If they agree on the same class, the final classification is simply that class. If one heatmap is classified as a sports type, while the other is classified as miscellaneous, the final classification will choose the regular sports type. Are both heatmaps classified as sports types, the sample with shortest distance to the class mean will decide the classification.

5. Experiments

For testing the classification approach, two continuous weeks of data has been captured. Capturing from 7am to 11pm this is a total of 224 hours of recordings, of which people are present in the arena in 142 hours and 82 hours are empty. Video from the first week is used for training data and the second week is used for test. Many undefined activities are observed during a day, from warm-up and exercises, to more passive activities, such as transitions between teams, "team meetings", cleaning, etc. Only well-known sports types performed like in matches will be used

for classification. Exercises related to a specific sport, such as practising shots at goal, will not be considered a specific sports type, but will be counted as miscellaneous activity. We do, however, allow variety in the play, such as different number of players and different number of courts in use for badminton and volleyball.
The sports types that are observed during both weeks and will be used in this work are badminton, basketball, indoor soccer, handball, and volleyball. As shown in figure 4 two different layouts of volleyball courts are observed, one which fit three volleyball courts playing in the same direction as badminton (denoted volleyball-3) and the other version with only one court in the middle of the arena, playing in the opposite direction (denoted volleyball-1). These will be treated as two different classes, both referring to volleyball. This results in seven classes to classify, including miscellaneous.

The heatmaps used for training and test of each sports type are samples that are manually labelled to be a regular performed sport. The miscellaneous heatmaps are chosen as samples that represent the various kind of random activities that takes place in the arena. The number of heatmaps used for each class is listed in table 1.

Category	Training heatmaps	Test heatmaps
Badminton	35	19
Basketball	16	12
Soccer	20	22
Handball	18	15
Volleyball-1	33	13
Volleyball-3	15	8
Misc.	163	30
Total	300	119

Table 1. Data set used for training and test.

Of the very large number of available heatmaps from miscellaneous activities, we choose 30 different samples for testing. This is the main reason that the number of heatmaps used for test are lower than the number of training heatmaps.
In order to test the system under real conditions, which will be continuous video sequences of several hours, we do also perform a test on video captured on one day continuously from 7am to 11pm. This video contains recordings of volleyball, handball and soccer, as well as miscellaneous activities. The training data described in table 1 is used again for this test.

5.1. Results

Table 2 shows the result for the first test with data sampled from one week. The ground truth is compared with the classification. This results in an overall true positive rate of 90.76 %. This result is very satisfying, considering that we classify seven classes based only on position data.
A low number of seven heatmaps are wrongly classified as miscellaneous instead of the correct sports type. Four of them are from videos where only one of the three volleyball courts are used, and this situation is not represented in training data. The error could therefore be reduced by capturing more training data. Three heatmaps representing soccer are misclassified as volleyball played on the centre court. Inspecting these images, there are some similarity between the sports, depending on how they are performed. At last, one miscellaneous image is classified as handball. This is a situation where the handball team practise their play towards one goal. It is therefore very close to a real handball video, but manually labelled as miscellaneous.
The result of classifying one full day is illustrated in figure 7. The green periods illustrate volleyball matches. Before these matches there is a warm-up period, where short periods of exercises are confused with basketball and volleyball played on the three courts. The last case is obvious, because some of their warm-up exercises include practising volleyball shots in the same direction as volleyball is normally played using the three courts. This test do like the previous test show that soccer can be misclassified as volleyball. The overall result is very promising, showing that of the total of 191 heatmaps that are produced and classified for the full day, 94,24 % are correctly classified. This result shows that our approach works very satisfying for the challenging situation of a full day's video, even with a better result than the first test.

5.2. Comparison with related work

A comparison of our results with the reported results in related work is listed in table 3. It should be noted that each work has its own data set, making it hard to compare the results directly. All related works use normal visual cameras, where we use thermal cameras. In addition to that most work use video from different courts for each sports type, where we use video from the same indoor arena. Our result is comparable with the related work using an equal number of sports types. It is also seen that we test on a large amount of data compared to other works.

Reference	Sports types	Video length	Result
Gibert et. al [4]	4	220 min.	93 %
Mohan and Yegn. [6]	5	5 h. 30 min.	94.4 %
Lee and Hoff [7]	2	Approx. 1 hour	94.2 %
Li et. al [8]	14	114 hours	88.8 %
Mutch. and Sang. [9]	20	200 min.	96.65 %
Sigari et. al [10]	7	(104 video clips)	78.8 %
Wang et. al [11]	4	(173 test clips)	88 %
Wang et. al [12]	3	16 hours	100 %
Watcha. et. al [13]	7	233 min.	91.1 %
Xu et. al [14]	4	1200 frames	N/A
Yuan and Wan [15]	5	N/A	97.1 %
Our work	5	54 hours	90.76 %

Table 3. Data set used for training and test.

Truth \ Classified as	Badminton	Basketball	Soccer	Handball	Volleyball-1	Volleyball-3	Misc.
Badminton	**17**	0	0	0	0	0	2
Basketball	0	**12**	0	0	0	0	0
Soccer	0	0	**18**	0	3	0	1
Handball	0	0	0	**15**	0	0	0
Volleyball-1	0	0	0	0	**13**	0	0
Volleyball-3	0	0	0	0	0	**4**	4
Misc.	0	0	0	1	0	0	**29**

Table 2. Classification result for data samples from one week. The number of heatmaps classified in each category.

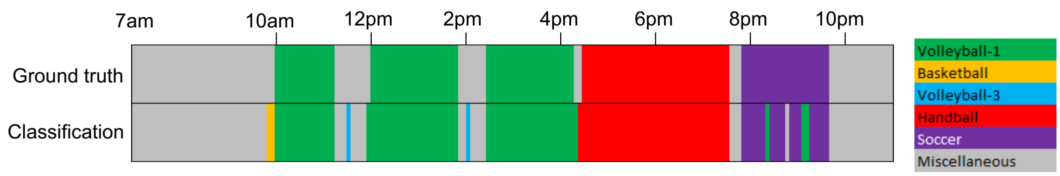


Figure 7. Comparison of ground truth and classification of video from one full day.

6. Conclusion

In this work we showed that sports types can be classified based only on the position data of people. Heatmaps are produced by summarising the position data over 10-minute periods. These heatmaps are projected to a low-dimensional space using PCA and Fischer's Linear Discriminant. Our result is an overall recognition rate for five sports types of 90.76 %. This is a very promising result, considering that our work is the first to use thermal imaging for sports classification. Furthermore, we use video from the same indoor arena, meaning that no information about the arena can be used in the classification.

For this work we have concentrated on sport played in match-like situations. Problems could rise if trying to classify a video of sport played in the opposite direction of usual, e.g. on half the court, or if trying to classify exercises related to one sports type. To overcome these limitations future work will investigate the possibility of including local features. These could be clues from short trajectories, such as speed and path length and straightness to overcome these limitations.

References

[1] P. Belhumeur, J. Hespanha, and D. Kriegman. Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection. *PAMI*, 19(7):711 –720, jul 1997. 4

[2] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification*. Wiley-Interscience, 2nd edition, 2001. 4

[3] R. Gade, A. Jørgensen, and T. B. Moeslund. Long-term occupancy analysis using graph-based optimisation in thermal imagery. In *CVPR*, 2013. 2

[4] X. Gibert, H. Li, and D. Doermann. Sports video classification using HMMS. In *International Conference on Multimedia and Expo (ICME)*, 2003. 1, 5

[5] J. Kapur, P. Sahoo, and A. Wong. A new method for gray-level picture thresholding using the entropy of the histogram. *Computer Vision, Graphics, and Image Processing*, 29(3):273 – 285, 1985. 2

[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. 1, 5

[7] J. Y. Lee and W. Hoff. Activity identification utilizing data mining techniques. In *IEEE Workshop on Motion and Video Computing (WMVC)*, 2007. 1, 5

[8] L. Li, N. Zhang, L.-Y. Duan, Q. Huang, J. Du, and L. Guan. Automatic sports genre categorization and view-type classification over large-scale dataset. In *17th ACM international conference on Multimedia (MM)*, 2009. 1, 5

[9] 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. 1, 5

[10] M. Sigari, S. Sureshjani, and H. Soltanian-Zadeh. Sport video classification using an ensemble classifier. In *7th Iranian Machine Vision and Image Processing (MVIP)*, 2011. 1, 5

[11] 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. 1, 5

[12] 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. 1, 5

[13] N. Watcharapinchai, S. Aramvith, S. Siddhichai, and S. Marukatat. A discriminant approach to sports video classification. In *International Symposium on Communications and Information Technologies (ISCIT)*, 2007. 1, 5

[14] M. Xu, M. Park, S. Luo, and J. Jin. Comparison analysis on supervised learning based solutions for sports video categorization. In *IEEE 10th Workshop on Multimedia Signal Processing*, 2008. 1, 5

[15] Y. Yuan and C. Wan. The application of edge feature in automatic sports genre classification. In *IEEE Conference on Cybernetics and Intelligent Systems*, 2004. 1, 5