A complex system for football player detection in broadcasted video

Sławomir Maćkowiak
Chair of Multimedia Telecommunication and Microelectronics,
Poznań University of Technology,
Polanka St. 3, 60-965 Poznań, Poland,
e-mail: smack@multimedia.edu.pl

Jacek Konieczny
Chair of Multimedia Telecommunication and Microelectronics,
Poznań University of Technology,
Polanka St. 3, 60-965 Poznań, Poland,
e-mail: jkonieczny@multimedia.edu.pl

Maciej Kurc
Chair of Multimedia Telecommunication and Microelectronics,
Poznań University of Technology,
Polanka St. 3, 60-965 Poznań, Poland,
e-mail: mkurc@multimedia.edu.pl

Przemysław Maćkowiak
Chair of Multimedia Telecommunication and Microelectronics,
Poznań University of Technology,
Polanka St. 3, 60-965 Poznań, Poland,
e-mail: przemyslaw.mackowiak@gmail.com

***Abstract*— In this paper a novel segmentation system for football player detection in broadcasted video is presented. The system is based on the combination of Histogram of Oriented Gradients (HOG) descriptors and linear Support Vector Machine (SVM) classification. Although recently HOG-based methods were successfully used for pedestrian detection, experimental results presented in this paper show that combination of HOG and SVM seems to be a promising technique for locating and segmenting players in broadcasted video. Proposed detection system is a complex solution incorporating a dominant color based segmentation technique of a football playfield, a 3D playfield modeling algorithm based on Hough transform and a dedicated algorithm for player tracking.**

Evaluation of the system is carried out using SD (720×576) and HD (1280×720) resolution test material. Additionally, performance of the proposed system is tested with different lighting conditions (including non-uniform pith lightning and multiple player shadows) and various camera positions.

I. INTRODUCTION

Due to large potential applicable demands, sports video analysis has been an active research topic in recent years. There have been a number of previous works for football video segmentation based on different techniques. One of the most interesting techniques are shape analysis-based approaches used to identify players and ball in the roughly extracted foreground [1,2] or techniques based on different classification schemes to segment football players and identify their teams [3,4,5]. However, most of existing approaches assume specific conditions such as fixed or multiple cameras, single moving object, and relatively static background. In football video broadcast, those strict conditions are not met. First, cameras used to capture sport events are not fixed and always move in order to follow the players. Secondly, the broadcasted video is a set of dynamically changed shots selected from multiple cameras according to broadcast director's instructions. Third, there are numerous players moving in various directions in the broadcasted video. Moreover, the background in sports video changes rapidly. These conditions make detection and tracking of players in broadcasted football video difficult. Therefore, future approaches should aggregate different techniques to detect and track the objects in football video.

Furthermore, few of these aforementioned approaches have the capability of online (means in real time or real time with a short period of delay) segmentation, which is important for many applications.

Based on observed characteristics of various broadcasted football games and analyses on difficulties of existed algorithms, we propose a novel approach which uses a dominant color based segmentation for football playfield detection, line detection algorithm based on the Hough transform to model the playfield and a combination of Histogram of Oriented Gradients (HOG) descriptors [15] with Support Vector Machine (SVM) classification [16] to detect players and player tracking system. The system is designed to detect location and orientation of the playfield as well as detect and track players on the playfield.

Original contribution of this paper is to apply low complexity techniques with significant potential, until now, used mostly for pedestrian detection. Therefore, the aim of this thesis is to explore the aptitude of the above methods and verify if proposed approach is sufficient for the purpose of segmentation of football video broadcast.

II. PREVIOUS WORK

In order to create a complex football video segmentation system several types of techniques need to be incorporated. Concerning the system presented in this thesis the following techniques were selected as the most important ones.

One of techniques used for dominant color detection is MPEG-7 dominant color descriptor (DCD), however, it operates on three dimensional color representation and its results are not illumination independent [6]. Approach [7] is based on Euclidean distance to trained dominant color in IHS color space. Ren et al. [8] presented an image block classification method based on color hue variance followed by hue value classification by trained Gaussian mixture model.

Most of line detection algorithms are based on Hough transform of binary line image [9] which can detect presence of a straight line structure and estimate its orientation and position. Some other approaches use modified Hough transforms like probabilistic Hough transform [10] or Block Hough transform [11] for computation speed improvements. Thuy et al. [12] proposeed Hough transform modification which allows line segment detection instead of straight line presence. On the other hand, random searching methods might be also used. Such methods [13] incorporate a random searching algorithm which selects two points and checks whether there is a line between them. Another issue is line image generation. Here, edge detection approaches and other gradient based techniques perform best [14].

Object detection is always based on extraction of some characteristic object features. Dalal et al. [15] introduced a HOG descriptor for the purpose of pedestrian detection and achieved good results.

Another important issue in object detection is object classification which separates objects belonging to different classes to distinguish requested objects from the others. One of the most commonly used object classifiers is SVM classifier which has been successfully applied to a wide range of pattern recognition and classification problems [16,17]. The advantages of SVM compared to other methods are: 1) better prediction on unseen test data, 2) a unique optimal solution for training problem, and 3) fewer parameters.

III. SYSTEM OVERVIEW

Specific conditions in segmentation of football video broadcast require an adequate approach therefore a dedicated system for football player detection was proposed (Fig. 1). The main components of the system are: playfield detector, playfield model fitter and object region recognition and object tracking module.

A. Playfield Detection

Accurate detection of a playfield area is very important for further segmentation process. It is assumed that playfield area is a homogenous region with relatively uniform color hue. Therefore the first step of playfield detection is a vector quantization of color chrominance values. After the quantization, colors are classified based on the a-priori green color definition. Joint coverage area of all colors considered as green is taken as initial playfield mask. The playfield area is supposed to be the largest green area in the whole image. Therefore all isolated areas are subjected to labeling algorithm and only the largest area combined from the input areas is selected as the playfield area mask.

B. Playfield Model Fitting

A model describing the playfield penalty field is helpful in case of player detection and also for event detection and classification. Fitting of the model requires information about position of some reference points in each frame. These points are obtained from playfield line crossings. Playfield lines are detected using modified Hough transform [11] applied to binary line image. First, playfield line image is generated according to [14]. Image is then thresholded and subjected to morphological thinning algorithm. In order to reject possible false detections of non-existing lines, binary line image is divided into blocks. For each block a set of line parameters is estimated independently and used in voting procedure in modified Hough transform algorithm. Initial line candidates are obtained by searching for local maxima in the Hough transform parameter space. To achieve greater accuracy, line candidates' parameters are refined via linear regression. Candidates with too large regression error are rejected. The final step is line candidate aggregation and tracking.

We use predefined penalty field model with appropriate dimensions proportion as a template. The template is matched to each possible position of a playfield based on the detected lines orientation. For each iteration a model fitting cost is computed and position with the smallest fitting cost is considered as a final solution.

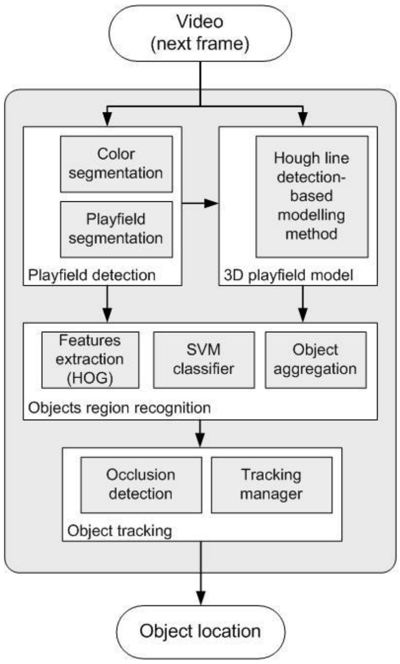

Fig. 1. Detection system overview.

C. Object Region Recognition

The player detection module is based on HOG descriptor [15]. Player shape is represented by a window size of 16×32 pixels. Consequently, HOG descriptor block size is 8 pixels, cell size is 4 pixels and block stride is also 4 pixels. Number of angle bins used is set to 9. Additionally, the unsigned gradient and L2Hys [15] block normalization scheme is used. In order to detect players of different size a multiple scale detector is applied. We assume that smallest player size must be at least 16×32 pixels and the largest can be at most half of the image height.

When the HOG descriptor is determined for analyzed area of the image a linear SVM trained on dedicated football player template database is used to classify it as a player or non-player class. Our player template database contains over

600 vertical frontal and vertical profile poses of the football players as positive examples and over 3000 negative non-player images. Templates were obtained by manual image segmentation for positive examples and by bootstrapping procedure for negative ones. In order to detect different poses of players three separate SVM classifiers are utilized. First one was trained to detect football players in vertical frontal poses, second is used for vertical profile poses and the last on joint set of all vertical poses. All SVM classifiers uses the same negative samples set. These three SVM detectors work in parallel and their results are aggregated to produce single set of detection results for our detection system. As we observed, a single player can cause multiple detections at a time and, hence, the resultant bounding boxes from single player detector overlap in many cases. In order to achieve a single bounding box surrounding the detected player an additional merging procedure is proposed. The procedure consists of the following stages: detection box filtering (boxes of not appropriate size or containing to many playfield pixels are rejected), overlap test for each pair of detected boxes (includes finding the biggest coherent area of non-playfield points in each box and testing if these areas overlap) and finally, box aggregation stage (two overlapping boxes are merged into one resultant box).

D. Object Tracking

Object tracking is applied to improve the system robustness in case of player occlusion and focus changes produced by a rapid camera movement or zooming. In proposed approach each object is described by the position and size of the detection box bounding this object. In each time instant, objects that are currently being tracked by the algorithm are compared with the boxes found by the SVM detector in the analyzed frame. For each object a similarity measure between the box associated with the object and a candidate box is evaluated using a dedicated cost function. The function incorporates both the overlap area and the size of the two boxes. After the cost evaluation, candidate box with the minimal cost is chosen to represent the object being tracked. Additionally, each object has a motion vector assigned to it. This motion vector is calculated based on the position of the object in previous frames and consequently, if the object cannot be matched with any box detected in the current frame, its position can be predicted using this motion vector.

IV. EVALUATION OF THE PLAYER DETECTION SYSTEM

In this section an evaluation of the proposed football player detection system is presented. In order to inspect the analyzed system in case of different input video resolution, performance of the system should be evaluated using both SD (720×576) and HD (1280×720) test sequence resolution. Additionally, prepared test set should consists of a video material with different lighting conditions, such as non-uniform playfield lighting, multiple player shadows etc., as well as different position of cameras used for the video acquisition. Taking the above conditions into consideration, 9 test sequences with football events and length of 25 to 50 frames were selected to form a test set for the system evaluation. For each test sequence the ground truth regions indicating football players were manually selected to create the desired output of the system.

Presented system is evaluated with the *precision* and *recall* performance metrics, defined as follows:

$$precision = \frac{TP}{TP + FP}, \tag{1}$$

$$recall = \frac{TP}{TP + FN}, \tag{2}$$

where TP is the set of true positives (correct detections), FP - the set of false positives (false detections) and FN - the set of false negatives (missed objects) defined as:

$$TP = \{r | r \in D: \exists g \in G: s_0(r,g) \geq T\}, \tag{3}$$

$$FP = \{r | r \in D: \forall g \in G: s_0(r,g) < T\}, \tag{4}$$

$$FN = \{r | r \in G: \forall g \in D: s_0(r,g) < T\}. \tag{5}$$

In the above equations: $s_0(a,b)$ indicates a degree of overlap between two regions a and b (i.e. bounding boxes of detected objects):

$$s_0(a\ b) = (a \cap b)/(a \cup b) \tag{6}$$

T determines a threshold which defines the degree of overlap required to consider two regions as overlapping the same area of the analyzed image. G is a set of ground truth regions and D is a set of detected regions for a given frame, defined as: $G = \{g_1,\ldots,g_n\}$ and $D = \{d_1,\ldots,d_m\}$, with n indicating the number of ground truth regions and m - the number of detected regions in analyzed frame.

For the purpose of evaluating the player detection system presented in this paper we use threshold T value equal 0.4. This value was chosen based on the interpretation of the T parameter: T=0.4 means that more than a half of the areas of a bounding box representing the detected object position and the ground truth box selected manually for this object overlap. This value is sufficient to evaluate how efficiently the objects are detected by the system. However, in case of evaluation of a system dedicated for the object segmentation and it's extraction from the background higher T values should be applied.

Table 1 presents detailed evaluation results of our detection system with the reference to each test sequence included in the test set.

Based on the analysis of Table 1 it can be concluded that selected test sequences provided diverse difficulty level for the player detection system as the *precision* and *recall* metric values differ among the test set. However, the average *precision* value reached by the system is 0.98 and the average *recall* metric values do not exceed 0.87.

TABLE 1.

DETECTION SYSTEM EVALUATION RESULTS FOR THRESHOLD T=0.4 (P-*PRECISION*, R-*RECALL*).

Test sequence	P	R
1. fast camera pan, uniform lighting	1.00	0.84
2. non-uniform lighting, shadows, occlusion	0.97	0.90
3. occlusion, various player's poses	1.00	0.88
4. non-uniform lighting, occlusion	0.97	0.94
5. non-uniform lighting, occlusion	1.00	0.87
6. interlaced, uniform lighting	1.00	1.00
7. motion blur, small figures, occlusion	0.96	0.80
8. motion blur, occlusion, uniform lighting	0.96	0.67
9. green uniforms, occlusion	0.96	0.96
Average	**0.98**	**0.87**

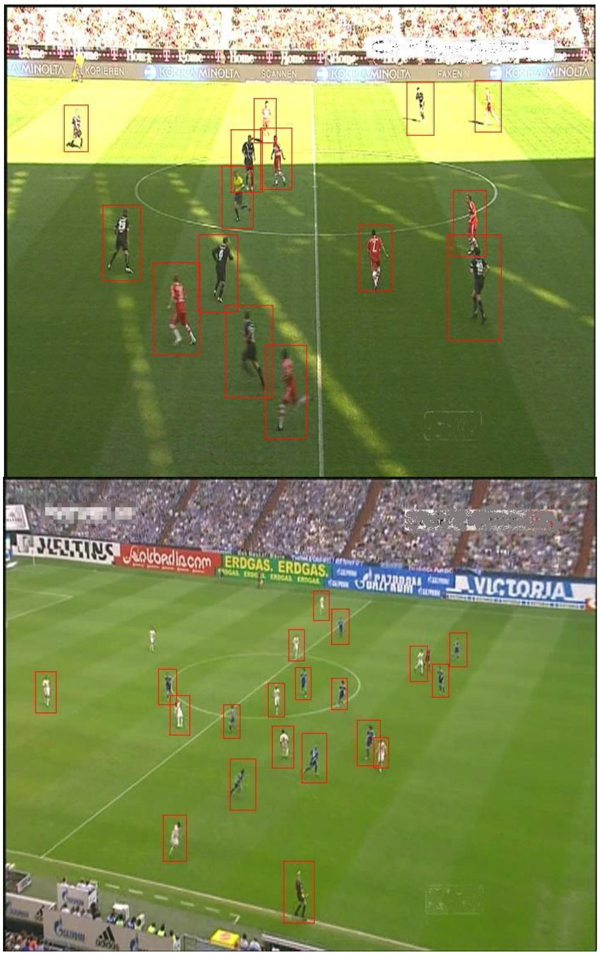

Fig. 2. Visualization of the player detection results.

Evaluation results presented above show clearly that number of false detections in the system is relatively small, resulting in *precision* metric values which exceed 0.95 in all analyzed cases. However, the number of missed objects changes remarkably among the test set - see the values of *recall* metric in Table 1. This is mainly caused by the occlusion between objects in the scene. Although a dedicated tracking algorithm was applied to the detection system to handle this problems, the algorithm is not robust in presence of additional temporal focus changes produced by a rapid camera movement or zoom.

Another possible solution to the problem of increasing the number of detected objects and, consequently, improving the *recall* metric is further development of the image database used to train the SVM detector.

In addition to the objective tests, performance of the proposed player detection system was also evaluated subjectively. Detection results in form of object bounding boxes were displayed on the original test sequences producing a clear illustration of the system performance. Exemplary results obtained for the analyzed test set are shown in Fig. 2.

V. Conclusions

In the paper, a novel segmentation system for football player detection in broadcasted video is proposed. The proposed system is a complex solution which incorporates several techniques which are used to detect players and playfield: dominant color based segmentation, 3D playfield modeling based on Hough transform, HOG descriptor detection and SVM classification. These methods were selected based on their potential robustness in case of great inconstancy of weather, lighting and quality of the input video sequences. The performance of the system was evaluated with a dedicated test set and a ground-truth database was created to perform objective evaluation of the player detection algorithm. Results show that proposed solution seems to achieve high objective and subjective notes in terms of precise location of the detected objects, however, the number of missed objects still needs to be decreased. Consequently, there are some works deserving further research in the proposed approach. Image database used for training of the SVM detectors needs to be extended and our future work will also focus on developing more advanced algorithms to handle the occlusion issue for player detection and tracking.

References

[1] Haiping, S., Lim, J. h., Tian, Q., Kankanhalli, M. S.: Semantic labeling of soccer video, Proceedings of IEEE Pacific-Rim Conference on Multimedia ICICS-PCM, 1787-1791 (2003)

[2] Huang, Y., Llach, J., Bhagavathy, S.: Players and Ball Detection in Soccer Videos Based on Color Segmentation and Shape Analysis, Lecture Notes in Computer Science, Volume 4577/2007, 416-425 (2007)

[3] Nuñez, J. R., Facon, J., Brito Junior, A. d. S.: Soccer Video Segmentation: referee and player detection, 15th International Conference on Systems, Signals and Image Processing, 2008. IWSSIP 2008, 279 – 282 (2008)

[4] Vandenbroucke, N., Ludovic, M., Postaire, J-G.: Color image segmentation by pixel classification in an adapted hybrid color space. Application to soccer image analysis, Computer Vision and Image Understanding 90, 190–216 (2003)

[5] Guangyu, Z., Changsheng, X., Qingming, H., Wen, G., Automatic multi-player detection and tracking in broadcast sports video using support vector machine and particle filter, Int. Conf. Multimedia & Expo, 1629-1632 (2006)

[6] Hong, S., Yueshu, W., Wencheng, C., Jinxia, Z.: Image Retrieval Based on MPEG-7 Dominant Color Descriptor, ICYCS, 753-757 (2008)

[7] Ying, L., Guizhong, L., Xueming, Q.: Ball and Field Line Detection for Placed Kick Refinement, GCIS, vol 4, 404-407 (2009)

[8] Ren, R., Jose, J.,M.: Football Video Segmentation Based on Video Production Strategy, Lecture Notes in Computer Science, 3408, 433-446

[9] Candamo, J., Kasturi, R., Goldgof, D., Sarkar, S.: Detection of Thin Lines using Low-Quality Video from Low-Altitude Aircraft in Urban Settings, Aerospace and Electronic Systems, IEEE Transactions on, vol.45, no.3, pp.937-949, July (2009) Guo, S., Y., Kong, Y.,G., Tang, Q., Zhang, F.: Probabilistic Hough transform for line detection utilizing surround suppression, International Conference on Machine Learning and Cybernetics (2008)

[10] Yu, X., Lai, H.C., Liu, S.X.F., Leong, H.W.: A gridding Hough transform for detecting the straight lines in sports video. ICME (2005)

[11] Thuy, T., N., Xuan, D., P., Jae, W., J.: An improvement of the Standard Hough Transform to detect line segments, ICIT (2008)

[12] Jiang, G., Ke, X., Du, S., Chen, J.: A straight line detection based on randomized method, ICSP (2008)

[13] Li, Q., Zhang, L., You, J., Zhang, D., Bhattacharya, P.: Dark line detection with line width extraction, ICIP (2008)

[14] Dalal N., Triggs B., Histograms of oriented gradients for human detection, Computer Vision and Pattern Recognition 1, 886-893 (2005)

[15] Yu-Ting Ch., Chu-Song Ch.: Fast Human Detection Using a Novel Boosted Cascading Structure With Meta Stages, IEEE Transactions on Image Processing 17, 1452-1464 (2008)

[16] Paisitkriangkrai S., Shen, C. Zhang, J.: Performance evaluation of local features in human classification and detection, IET Computer Vision 2, 236-246 (2008)