Football player detection in video broadcast

Abstract. The paper describes a novel segmentation system based on the combination of Histogram of Oriented Gradients (HOG) descriptors and linear Support Vector Machine (SVM) classification for football video. Recently, HOG methods were widely used for pedestrian detection. However, presented experimental results show that combination of HOG and SVM is very promising for locating and segmenting players. In proposed system a dominant color based segmentation for football playfield detection and a 3D playfield modeling based on Hough transform is introduced.

Experimental evaluation of the system is done for SD (720×576) and HD (1280×720) test sequences. Additionally, we test proposed system performance for different lighting conditions (non-uniform pith lightning, multiple player shadows) as well as for various positions of the cameras used for acquisition.

1 Introduction

Many approaches to football video segmentation and interpretation have been proposed in recent years. 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.

In this paper a novel approach to football broadcast video segmentation is proposed. We develop a hybrid segmentation system 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. The system is designed to detect location and orientation of the playfield as well as detect and track players on the playfield. Nevertheless the system is still under development and not all of the features are implemented.

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.

2 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.

3 System overview

To handle specific conditions in segmentation of the football video broadcast a dedicated system for football player detection was proposed (Fig. 1). The main operations in the system are: playfield detection, 3D playfield model fitting, object region recognition and object tracking module.

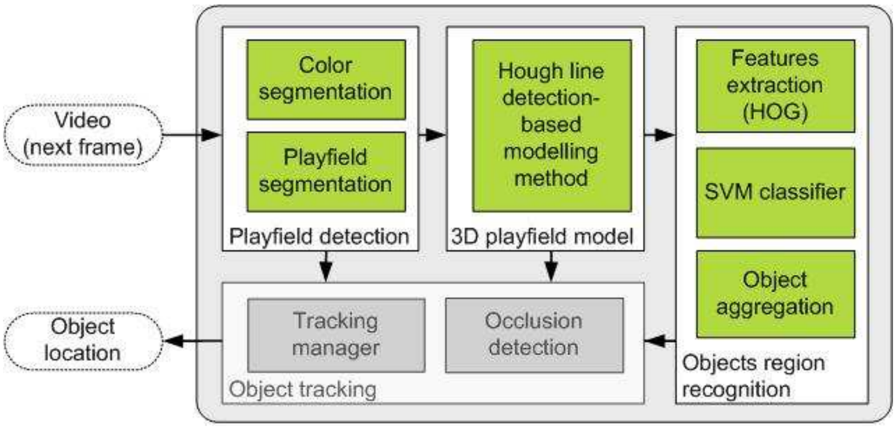

Fig. 1. Detection system overview.

3.1 Playfield detection

Information about playfield area position is crucial for further line detection, player detection and tracking. First, we assume that playfield is a homogenous region with relatively uniform color hue. Therefore the first step of playfield detection is vector quantization of color chrominance values. Quantized colors are then classified based on a-priori green color definition and their joint coverage area is considered as initial playfield mask. The playfield area is supposed to be the largest green area in the whole image. We perform labeling of all isolated areas and leave only the largest.

3.2 3D playfield model fitting

Playfield lines are detected using modified Hough transform [11] applied to binary line image. The first step is generation of image with playfield lines according to modified algorithm from [14] followed by thresholding and morphological thinning. Next, generated image is divided into blocks. For each block line parameters are estimated and used in voting procedure in modified Hough transform algorithm. Initial line candidates are obtained by searching Hough transform parameter space for local maxima. Next stage is candidate refinement using linear regression. Refinement stage is crucial for accuracy and false detection rejection. Candidates with too large regression error are rejected. Finally, candidates are aggregated based on their orientation and position. To achieve temporal consistency line candidates are estimated for each frame independently and then aggregated with lines from previous frames. We use predefined goal area model with appropriate dimension proportions as template and detected playfield lines as reference. For each frame every possible template position is considered. Each template line after transformation to playfield space is verified against playfield lines on input frame and a fitting error is computed. Playfield template position with smallest fitting error is considered as final solution.

3.3 Object region recognition

Our player detection module is based on HOG descriptor [15]. In order to represent player shape we use window size of 16×32 pixels. HOG Descriptor block size is 8 pixels, cell size equals 4 and block stride is also 4. Number of angle bins is set to 9. We use unsigned gradient and L2Hys [15] block normalization scheme. Detection is performed using multiple input image scales. Multiple scale approach ensures proper detection of players regardless of their distance to the camera. 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.

HOG descriptors are classified by linear SVM algorithm trained on player template database. Our player template database contains over 600 vertical frontal and vertical profile poses as positive examples and over 3000 negative vertical, non-player images. Positive templates were manually generated, negative examples were obtained manually and by bootstrapping procedure. We trained three SVM classifiers working in parallel in order to detect different poses of players: first one was trained on images with vertical frontal poses, second on vertical profile poses and the last on joint set of all vertical poses. All SVM classifiers were using the same negative sample set. If the SVM detector classifies analyzed image area as belonging to a player class, its location is marked using a rectangular area, called a bounding box.

Eventually, the considered object detection system uses a post-processing stage which aggregates resultant player bounding boxes from all SVM classifiers in order to increase the number of detected objects, decrease false detections rate and improve segmentation precision. 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 overcome this problem an additional merging operation is proposed, which 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 box aggregation (two overlapping boxes are merged into one resultant box).

3.4 Object tracking

Object tracking improves the system robustness in case of player occlusion and focus changes produced by a rapid camera movement or zoom. However, the tracking module is currently under intensive development and therefore is not used in presented revision of our system.

4 Evaluation of HOG+SVM player detection method

In this section a combination of HOG and SVM techniques is evaluated as a player detection method for our system. As the analyzed player detection algorithm should be independent from the input video resolution, performance of the system was evaluated on both SD (720×576) and HD (1280×720) test sequence resolution. Additionally, the created test set takes into consideration different lighting conditions (non-uniform playfield lighting, multiple player shadows etc.) as well as the different position of cameras used for acquisition. We selected 9 test sequences with football events and length of 25 to 50 frames for the evaluation. To perform the evaluation step the ground truth regions need to be selected manually for each frame in the test set.

The considered system is evaluated using the *precision* and *recall* performance metrics, defined as follows:

$$precision = TP/(TP+FP), \tag{1}$$

$$recall = TP/(TP+FN), \tag{2}$$

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:

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

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

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

$s_0(a,b)$ is called 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 is a threshold defining the degree of overlap required to determine two regions as overlapping. The set of ground truth regions G and detected regions D for a given frame are defined as: $G = \{g_1,\ldots,g_n\}$ and $D = \{d_1,\ldots,d_m\}$, with n – the number of ground truth regions and m – the number of detected regions in analyzed frame.

For the purpose of this thesis the analyzed system is evaluated with threshold values T equal 0.4 to 0.9 using three different Support Vector Machines (SVM), namely: vertical (V SVM), vertical frontal (VF SVM) and vertical profile (VP SVM) (see Section 3.3). Average results for threshold T=0.4 to 0.9 are illustrated in Fig. 2. Table 1 presents detailed evaluation results for threshold T=0.6.

Analysis of Table 1 show that selected test sequences provided diverse difficulty level for analyzed player detectors as the *precision* and *recall* metric values differ remarkably among the test set. The average *precision* values (Fig. 2) reached by the examined SVMs are between 0.19 to 0.80 depending on the threshold T parameter used for evaluation and the average *recall* metric values do not exceed 0.64 with a noticeable decrease for larger threshold T values. The impact of the threshold T on the evaluation results will be discussed in detail later in this section.

Table 1. Detection system evaluation results for threshold T=0.6 (P-*precision*, R-*recall*).

Test sequence	VF SVM		VP SVM		V SVM	
	P	R	P	R	P	R
1. fast camera pan, uniform lighting	0.92	0.60	0.92	0.66	0.92	0.67
2. non-uniform lighting, shadows, occlusion	0.48	0.40	0.65	0.63	0.57	0.60
3. occlusion, various player’s poses	0.56	0.46	0.54	0.44	0.49	0.44
4. non-uniform lighting, occlusion	0.68	0.41	0.73	0.51	0.73	0.50
5. non-uniform lighting, occlusion	0.68	0.53	0.82	0.62	0.75	0.61
6. interlaced, uniform lighting	0.97	0.99	0.93	0.99	0.87	0.98
7. motion blur, small figures, occlusion	0.65	0.39	0.71	0.39	0.72	0.47
8. motion blur, occlusion, uniform lighting	0.75	0.47	0.86	0.37	0.77	0.54
9. green uniforms, occlusion	0.90	0.59	0.86	0.62	0.82	0.65
Average	0.73	0.54	0.78	0.58	0.74	0.61

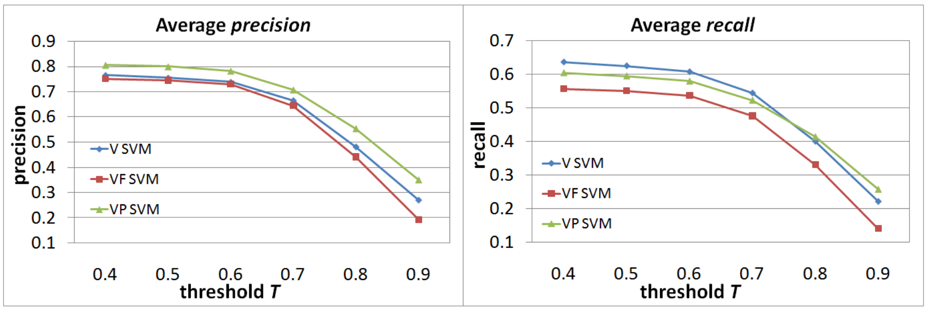

Fig. 2. Average values of *precision* and *recall* metrics for threshold parameter T=0.4 to 0.9.

Evaluation results presented above show clearly that image database used for training of analyzed SVM detectors need to be extended in order to increase the *recall* value especially. Despite insufficient size of the database there is a number of further improvements which should also help in reducing the missed objects rate for presented detection system. First, we observe that the playfield detection algorithm often classifies green and white player uniforms as a playfield which is an obvious mistake. Next, size of the resultant bounding boxes from the aggregation algorithm should be determined more accurately, as the playfield model fitting can provide additional information about expected size of a player depending on his or her location in analyzed image. This information, together with application of an additional tracking algorithm should also be useful for improving the *recall* metric of presented system in case of player occlusion and temporary focus changes produced by a rapid camera movement or zoom.

Another important improvement is increasing the number of parallel player detectors used in the system and aggregating their results to produce a single output. This step includes both increasing the number of SVM classifiers to detect e.g. horizontal player poses as well as application of other detection methods based on object color and shape. Above improvement is planned in future release of the system.

We now turn to the discussion on the impact of the threshold T parameter on the object detection system evaluation. Taking into consideration the evaluation result

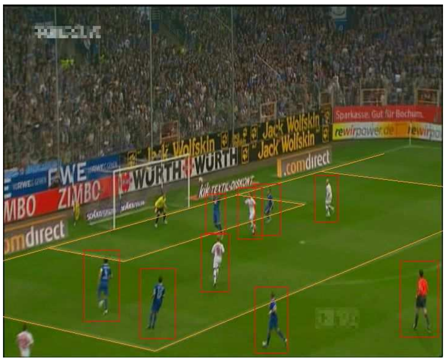

Fig. 3. Player detection and 3D playfield fitting results.

presented in Table 1, we believe that in case of analyzed system, threshold values T larger than 0.6 are too restrictive. Our research show that applied evaluation metrics turn out to be very sensitive for bounding boxes inaccuracies, especially for small objects detection. In case the object size is 16×32 points, decreasing the size of the bounding box only by 1 pixel in each dimension, which seems to be rather slight inaccuracy, decreases the $s_0(a.b)$ (eq. 6) by about 0.1. This issue influences the results even more if we consider the fact that the ground truth bounding boxes usually do not retain a fixed aspect ratio and the detection system produces bounding boxes with one constant aspect ratio, which is our case. Large bounding box inaccuracies may result only by the above fact and not the detector mistakes. On the above analysis it can be concluded that the most reasonable approach would be adaptation of the threshold T parameter depending on the size of detected object, e.g. in range of 0.4-0.8. Moreover, in case of presented evaluation results, we can observe a clear saturation of both precision and recall metric values for T=0.6 (Fig. 2), which turns to be the most reliable result confirmed also by the subjective detection results evaluation (Fig. 3).

5 Conclusion

In the paper, a novel approach to football broadcast video segmentation is proposed based on a combination of several techniques used to detect players and playfield: dominant color based segmentation, 3D playfield modeling based on Hough transform, HOG descriptor detection and SVM classification which show potential robustness in case of great inconstancy of weather, lighting and quality of video sequences. A dedicated test set and ground-truth database were created to perform objective evaluation of the player detection algorithm applied in our segmentation system and to explore the aptitude of selected methods. Results show that there are some works deserving further research in proposed approach. Image database used for training of analyzed SVM detectors need to be extended as well as the number of parallel detectors used in the system. Our future work will also focus on dealing with 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)
10. 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)
11. 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)
12. Thuy, T., N., Xuan, D., P., Jae, W., J.: An improvement of the Standard Hough Transform to detect line segments, ICIT (2008)
13. Jiang, G., Ke, X., Du, S., Chen, J.: A straight line detection based on randomized method, ICSP (2008)
14. Li, Q., Zhang, L., You, J., Zhang, D., Bhattacharya, P.: Dark line detection with line width extraction, ICIP (2008)
15. Dalal N., Triggs B., Histograms of oriented gradients for human detection, Computer Vision and Pattern Recognition 1, 886-893 (2005)
16. 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)
17. Paisitkriangkrai S., Shen, C. Zhang, J.: Performance evaluation of local features in human classification and detection, IET Computer Vision 2, 236-246 (2008)