FOOTBALL PLAYER TRACKING FROM MULTIPLE VIEWS

Using a Novel Background Segmentation Algorithm and Multiple Hypothesis Tracking

A. Koutsia, N. Grammalidis, K. Dimitropoulos
Informatics and Telematics Institute,Centre of Research and Technology Hellas
1st km Thermi-Panorama road, Thessaloniki, Greece

M. Karaman, L. Goldmann
Communication Systems Group, Technical University of Berlin, Einsteinufer 17, Berlin,Germany

Keywords: Multiple cameras, multi-target tracking, football game analysis, background segmentation.

Abstract: In this work, our aim is to develop an automated system which provides data useful for football game analysis. Information from multiple cameras is used to perform player detection, classification and tracking. A background segmentation approach, which operates with the invariant Gaussian colour model and uses temporal information, is used to achieve more accurate results. Information derived and matched from all cameras is then used to perform tracking, using an advanced Multiple Hypothesis Tracking algorithm.

1 INTRODUCTION

As football is one of the most popular sport activities, many research efforts have been devoted to analyse football games. Needham and Boyle (Needham, 2003) track the movements of football players using a single camera on an indoor court. A more recent approach, proposed in (Figueroa et al., 2004) performs player tracking using a graph representation of the moving objects in the scene. Another interesting approach using multiple cameras, was presented in (Xu et al., 2005) and was part of the INMOVE project.

The main contribution of this work is the use of a novel background segmentation approach and Multiple Hypothesis Tracking (MHT) to improve football player tracking. The resulting information can be used as input for an analysis tool for football games as well as for low bit-rate transmission and visualisation of the game.

The paper is organised as follows: the first section briefly describes the camera calibration technique used and is followed by a section on a novel background segmentation method. Next, the observation extraction and fusion steps are described, followed by a section on the tracking unit. Finally experimental results are presented and conclusions are drawn.

2 SYSTEM DESCRIPTION

2.1 Camera Calibration

In the proposed approach, multiple static cameras are used to capture images of a football game. To obtain the exact position of the targets in the real world, calibration of each camera is required, so that any point can be converted from image coordinates (measured in pixels from the top left corner of the image) to ground coordinates (horizontal and vertical distances in meters from the centre of the football pitch) and vice versa. A calibration technique based on point and line correspondences as well as a 3x3 homographic transformation was used (Dimitropoulos, 2005).

2.2 Background Subtraction-Based Segmentation

Segmentation of moving objects in videos in case of static or motion compensated camera is mostly

Koutsia A., Grammalidis N., Dimitropoulos K., Karaman M. and Goldmann L. (2007).
FOOTBALL PLAYER TRACKING FROM MULTIPLE VIEWS - Using a Novel Background Segmentation Algorithm and Multiple Hypothesis Tracking.
In *Proceedings of the Second International Conference on Computer Vision Theory and Applications - IU/MTSV*, pages 523-526

based on background subtraction. In (Karaman et al., 2005), the authors compared some selected state of the art background subtraction methods and proved that most of the proposed systems in literature are developed for specific environmental conditions and overcome explicitly or implicitly some of problems such as shadow, highlight, etc. Based on this investigation, they propose a method (Karaman et al., 2006) which combines the use of the Gaussian colour model (GCM) and temporal information. This method is extended and used for the purposes of this work.

Instead of the well-known colour spaces (RGB, YCbCr, HSV), the Gaussian colour model proposed by (Geusebroek et al., 2001) is used. It is based on the measurement of object reflectance in colour images and focuses on the description of colour invariants. Various invariant features are obtained from reflectance properties by its differentiations.

The proposed system includes two parts: pre-segmentation and post-classification. In the pre-segmentation part, a background reference model is generated using multiple frames without moving objects, calculating mean and standard deviation maps for each pixel. The absolute difference image of the input frame and the reference mean frame is binarised using unimodal thresholding (Rosin, 2001). Further, binary maps are obtained from image differencing results and standard deviation maps. Both binary results are combined with an AND operation. The results of each channel is further processed with an OR operation into a preliminary foreground mask.

The second part of the system eliminates the falsely detected regions. Therefore, an OR operation is applied to the last obtained final foreground mask and the mask from the successive frame (motion mask). Finally, an AND operator is applied to this motion segmented result and the preliminary foreground mask in order to restrict the extension of the moving objects to the motion region and achieves elimination of large noisy regions. Some morphological operations are also applied to reduce residual noise. Moreover, to consider temporal changes of the background, the background model is selectively updated.

If there are no frames without moving objects available, the background model estimation is more difficult. A median image is calculated based on multiple frames. Depending on the strength of movements, a variable step length is used for skipping frames considered for each intermediate median image. A final median image is generated by median filtering of these intermediate results and a clean background image without any moving objects can be obtained. Furthermore, for background modelling, multiple frames are required in order to derive statistical models for each pixel. For this, a simple background subtraction method is used based on the median image. For each current image, an over-segmented foreground mask is obtained by setting a very low threshold which after negation leads to an under-segmented background mask. This ensures that only very reliable background pixels are used for generating the final background model.

This whole background subtraction step is applied to each camera view independently.

2.3 Observation Extraction

Using the masks of the moving objects and the calibration information it is possible to estimate the position of the targets in world (ground) coordinates. The uncertainty of each observation is also calculated as well as a classification based on shirt colour information.

Specifically, the following steps are used to estimate the field position of all players visible from each camera:

- A connected component labelling algorithm is applied to the binary foreground mask.
- Blob filtering is performed to discard moving blobs not corresponding to players.
- For each blob, a characteristic point is chosen. Since the ground plane assumption is used, a suitable point is the middle bottom of each blob.
- World (ground) coordinates of each player are calculated based on the calibration information.

Furthermore, using a version of (Borg et al, 2005), the observation uncertainty $\mathbf{R}$ is given by:

$$\mathbf{R}(x_w, y_w) = \mathbf{J}(x_c, y_c)\mathbf{\Lambda}\mathbf{J}(x_c, y_c)^T \qquad (1)$$

where $\mathbf{J}$ is the Jacobian matrix of the partial derivatives of the mapping functions between the image and the world (ground) coordinate systems and $\mathbf{\Lambda}$ is the measurement covariance at location (x, y) on the camera plane which is assumed to be a constant diagonal matrix.

The colours of the shirts define five different groups (two teams, two goal keepers and one for the referees). Similar to (Xu et al., 2004) by comparing a colour histogram for each observation and a model for each group, five probabilities are obtained which denote the classification of the observation.

Using a training set, a mean histogram is calculated offline for each class based on the half upper part of each blob (shirt). Histograms of $r' = r/(r+g+b)$ and $b' = b/(r+g+b)$ are used to achieve robustness against light intensity variations.

At runtime, a 2D histogram is calculated for each observation and is compared to each of the five class models using the histogram intersection method (Swain and Ballard, 1991). The resulting probabilities for each class model are then normalised to have a summation of one, forming a 5-value vector.

Finally, filtering is performed in order to avoid the use of erroneous observations. Filtering occurs when:

- The blob size or the height-to-width ratio cannot correspond to humans.
- Classification is uncertain, i.e. none of the probabilities is higher than a specific threshold.
- Real world position is out of the pitch.
- The observation is inaccurate, i.e. the maximum eigenvalue of $\mathbf{R}$ is higher than a specific threshold.

2.4 Observation Fusion

A target visible to multiple cameras may result in multiple observations due to inaccuracies in the choice of characteristic points or calibration errors. To group together these observations, a grid that separates the football pitch in cells is considered. Observations which belong to the same cell or to neighbouring cells are fused to a single observation. Fused observations are produced by averaging the parameters of the grouped observations.

The position $\mathbf{Z}$ and uncertainty matrix $\mathbf{R}$ of the fused observation are given by the equations (2):

$$\mathbf{R} = (\sum_{n=1}^{N} \mathbf{R}_n{}^{-1})^{-1}, \qquad \mathbf{Z} = \mathbf{R}\sum_{n=1}^{N} \mathbf{R}_n^{-1}\mathbf{Z}_n \qquad (2)$$

where $\mathbf{Z}_n$ and $\mathbf{R}_n$ are the position (in world coordinates) and uncertainty matrix of the n-th observation, respectively, in a group consisting of N observations.

To calculate the average classification vector, the uncertainty of each observation is used to specify a weight w_n that equals to 1, 2 or 3. The average classification vector is then calculated by (3). These parameters ($\mathbf{Z}$, $\mathbf{R}$, $\mathbf{c}$) of each fused observation comprise the input for the tracking unit.

$$\mathbf{c} = \frac{\sum_{n=1}^{N} w_n \mathbf{c}_n}{\sum_{n=1}^{N} w_n} \qquad (3)$$

2.5 Tracking Unit

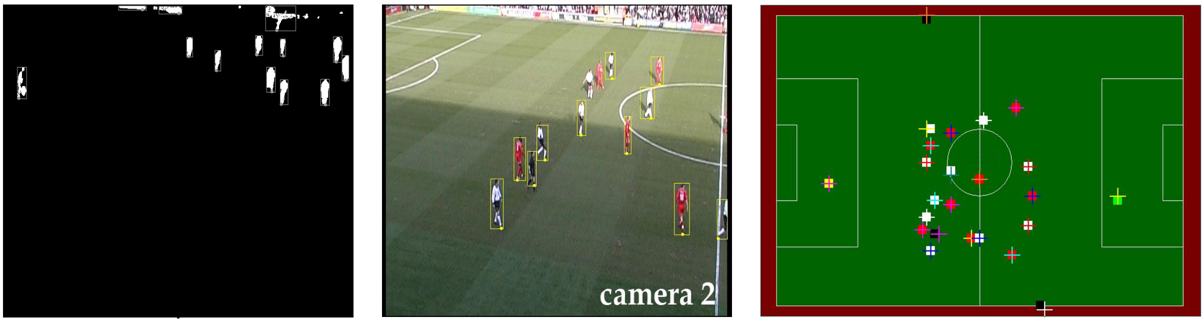
The next step is the generation and update of target tracks. The tracking unit is based on the Multiple Hypothesis Tracking (MHT) algorithm, which starts tentative tracks on all observations and uses subsequent data to determine which of these newly initiated tracks are valid. Specifically, the tracking unit was based on a fast implementation of the MHT algorithm by Cox et al. (Cox and Hingorani, 1996). A 2-D Kalman filter was used to track each target and additional gating computations are performed to discard observation – track pairs with incompatible classifications. Since only position measurements are available, a simple four-state (position and velocity along each axes) CV (constant velocity) target motion model in which the target acceleration is modelled as white noise provides satisfactory results.

3 RESULTS AND CONCLUSIONS

The proposed system was applied to 8 synchronised image sequences of 2500 frames each, captured for the INMOVE project. A sample of the background masks produced using the novel background segmentation algorithm is shown in Figure 1(a). Figure 1(b) shows a sample frame where the extracted observations are marked with bounding boxes. The small circles at the bottom of each box are the chosen characteristic points.

After processing the information obtained for each observation, the system produces a representation of the players' positions on the pitch (Figure 1(c)) which gets updated for each frame. The squares show the position of the fused observation along with their classification. The crosses (tracks) have different colours based on the track ID. The experimental results showed that the position of each player is quite precise and that the class information is accurately extracted from the image sequences. Furthermore, examination of the tracking sequences showed that, in most cases, the track IDs are successfully maintained through time.

The aim of this paper was to implement a multiple camera system which performs football player



(a) (b) (c)

Figure 1: A sample background segmentation mask, observations generated for camera 2 and output of the tracking system.

tracking. The information obtained can then be usable for game analysis purposes and football viewing through low bit rate channels (e.g. flash animations or mobile phones). Football players' positions and class (shirt colour) were successfully estimated and tracked. The summarised information of the typically twenty five tracks is only a few floating point numbers per track, thus possible to be transmitted via low bit rate channels.

To improve the usability of the proposed system some additional functionalities need to be implemented, e.g. ball detection and tracking. Also a football game analysis tool can be built to use the information provided by this system according to coaching needs.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge Dr. John Seeley, Dr. James Orwell, and Fulham Football Club for their parts in making the football datasets available for experimental analysis. This work was supported by the European Commission under the FP6 IST Network of Excellence: "3DTV-Integrated Three-Dimensional Television - Capture, Transmission, and Display" (contract FP6-511568).

REFERENCES

Borg, M., Thirde, D., Ferryman, J., Fusier, F., Valentin, V. , Brémond, F., Thonnat, M., Aguilera J., Kampel, M, 2005. *Visual Surveillance for Aircraft Activity Monitoring. The Second Joint IEEE International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance (VS-PETS 2005) in Beijing, China.*

K. Dimitropoulos, N. Grammalidis, D. Simitopoulos, N. Pavlidou and M. Strintzis, 2005. Aircraft Detection and Tracking using Intelligent Cameras, *IEEE International Conference on Image Processing (ICIP 2005), Genova, Italy, pp. 594-597.*

Cox, I.J., Hingorani, S.L., 1996. An Efficient Implementation of Reid's Multiple Hypothesis Tracking Algorithm and its Evaluation for the Purpose of Visual Tracking. *IEEE Transactions on pattern analysis and machine intelligence, Vol. 18, pp. 138-150.*

Figueroa, P., Leite, N., Barros, R.M.L., 2004. Tracking Soccer Players Using the Graph Representation. In *Proc. of the 17th Int. Conf. On Pattern Recognition.*

Geusebroek, J. M., van den Boomgaard, R., Smeulders, A. W. M., Geerts, H. , 2001. Color invariance. IEEE *Trans. Pattern Anal. Machine Intell. 23(12), pp. 1338–1350.*

Karaman, M., Goldmann, L., Yu, D., Sikora, T., 2005. Comparison of Static Background Segmentation Methods. *Proceedings of Visual Communications and Image Processing (VCIP).*

Karaman, M., Goldmann, L., Sikora, T., 2006. A New Segmentation Approach Using Gaussian Color Model And Temporal Presentation, *IS&T/SPIE Electronic Imaging (EI).*

Needham, C.J, 2003. *Tracking and Modelling of Team Game Interactions*. Ph.D. Thesis, University of Leeds.

Rosin, P. L., 2001. Unimodal thresholding. *Pattern Recognition 34(11), pp. 2083–2096.*

Swain, M. J., Ballard, D. H., 1991. Color indexing. *Int. J. Comput. Vision, 7:11-32.*

Xu, M., Orwell, J., Jones, G., 2004. Tracking football players with multiple cameras. *ICIP 2004 International Conference.*

Xu, M. Orwell, J. Lowey, L. Thirde, D, 2005. Architecture and algorithms for tracking football players with multiple cameras. *IEE Proceedings - Vision, Image, and Signal Processing Volume 152, Issue 2, p. 232-241.*