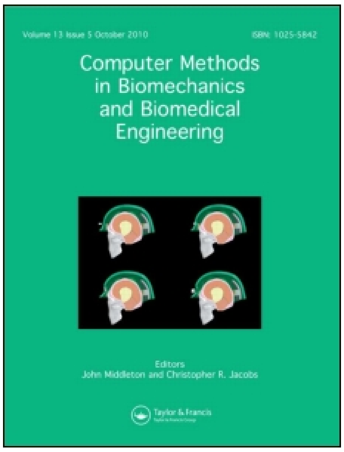

Computer Methods in Biomechanics and Biomedical Engineering
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713455284

Measuring handball players trajectories using an automatically trained boosting algorithm

Ricardo M. L. Barros[a]; Rafael P. Menezes[a]; Tiago G. Russomanno[a]; Milton S. Misuta[a]; Bruno C. Brandão[a]; Pascual J. Figueroa[a]; Neucimar J. Leite[a]; Siome K. Goldenstein[a]
[a] Department of Physical Education, University of Campinas, Campinas, Brazil

First published on: 14 December 2010

Computer Methods in Biomechanics and Biomedical Engineering
iFirst article, 2010, 1–11

Taylor & Francis
Taylor & Francis Group

Measuring handball players trajectories using an automatically trained boosting algorithm

Ricardo M.L. Barros*, Rafael P. Menezes, Tiago G. Russomanno, Milton S. Misuta, Bruno C. Brandão, Pascual J. Figueroa, Neucimar J. Leite and Siome K. Goldenstein

Department of Physical Education, University of Campinas, CX 6134, Campinas 13083-851, Brazil

(*Received 26 November 2009; final version received 14 May 2010*)

The aim of the present paper is to propose and evaluate an automatically trained cascaded boosting detector algorithm based on morphological segmentation for tracking handball players. The proposed method was able to detect correctly 84% of players when applied to the second period of that same game used for training and 74% when applied to a different game. Furthermore, the analysis of the automatic training using boosting detector revealed general results such as the training time initially increased with the number of figures used, but as more figures were added, the training time decreased. Automatic morphological segmentation has shown to be a fast and efficient method for selecting image regions for the boosting detector and allowed an improvement in the automatic tracking of handball players.

Keywords: tracking; sport; biomechanics

1. Introduction

In the past few years, both academic and industrial researchers have been investing considerable effort measuring the trajectories of players in team sports through video tracking. A special interest has been directed to the use of tracking in sports motion analysis because the kinematical analysis can provide useful information about player performance. For instance, the distance each player covers within the match period can lead to better planned routines and to the evaluation of player performance during a competition.

In sports-related applications, there are different constraints and goals in regard to mere academic researches. Achieving reasonable tracking is not enough. The user needs to use the application for long periods with accurate, robust and consistent results throughout different runs. Consider a handball game, which usually lasts 60 min, which accounts for 108,000 frames per camera. There is not a single automated solution that can track or automatically handle all complicated occlusion and serious clutter situations that occur in several key points during a match. Such problematic key points are often particularly interesting moments for the behaviour and movement analysis of players.

Some of the most widely used methods to obtain activity profiles and performances of team players are based on visual estimation of the distances covered during sporting matches, and most of the descriptions available were obtained using such approaches (Withers et al. 1982; Bangsbo et al. 1991; Mohr et al. 2003). The core task of these methods involves counting the number of strides taken during each discrete activity. This information is converted into distance, considering the length of an average stride for each type of movement (e.g. standing, walking, jogging and sprinting). Such methods are extremely time consuming and provide low spatial and temporal resolutions; moreover, most of them will not allow a simultaneous analyses of more than one player.

Global positioning system (GPS) or radio-frequency technology within sensor–transmitters attached to athletes has also been tried (Hennig and Briehle 2000). However, the results for using such approaches have only been noticed during simulating or training situations, due to restrictions imposed by the handball rules over the use of such devices in official competitions. In Edgecomb and Norton (2006), the measurement of distances covered by players, acquired from a global positioning and a computer-based tracking (CBT) system, relying on a human recorder that mechanically follows player movements on a calibrated miniaturised playing field, was compared during an Australian football game. A separate recorder operating a special computer was required for each tracked player. This recorder tracked such a player continuously from an elevated, midfield position in the stadium and recorded his movements with a conventional mouse or a commercially available pen tablet. According to the authors, distances measured by experienced researchers using the CBT system were as accurate as the results of the GPS technology.

Image processing and analysis have also been used for tracking, although the majority of these methods present only partial results. In Ohashi et al. (2002), only a single player was tracked per game. In Iwase and Saito (2004), all

*Corresponding author. Email: ricardo@fef.unicamp.br

ISSN 1025-5842 print/ISSN 1476-8259 online

DOI: 10.1080/10255842.2010.494602
http://www.informaworld.com

players were tracked, but only for short periods. In Toki and Sakurai (2005), all players were tracked for an entire game, but manually (frame-by-frame). The problem of tracking multiple players when occlusion and congestion are involved, especially when a scale must be considered, was addressed in Needham and Boyle (2001) but again, only partial results are reported.

Computer vision techniques have also been adopted in an attempt to develop efficient methods for tracking players during sports matches by Pers and Kovacic (2000) and Pers et al. (2002). Using CCD cameras with wide-angle lenses, mounted on the ceiling of a sports arena, these authors have developed, tested and optimised algorithms based on motion detection, template matching and colour-based tracking in their player's tracker.

The problem of labelling multi-targets was discussed in Nillius et al. (2006) and applied to soccer player tracking. The authors built a track graph describing the interactions between targets, explored the track graph and similarity measurements between the tracks to infer the most likely configuration of paths for all targets.

Okuma et al. (2004) suggested a vision system algorithm based on a boosted particle filter for multi-target detection and hockey player tracking. This algorithm combines the strengths of two successful algorithms: it mixes particle filters and the cascaded boosting algorithm from Viola and Jones (2001). According to the authors, the already learnt boosting detector enables the quick identification of players when entering the scene, whereas the particle filter tracking algorithm enables the tracking of individual players.

The training process of the boosting detector algorithm involves the selection of a large number of image regions containing the objects of interest (players, referee, etc.) and another set with figures of non-players. For example, Okuma et al. (2004) used a set of 6000 figures of players in different poses and another set of 100 images containing non-player objects in order to train the algorithm for a hockey game. The authors implemented a program for the detection of low-intensity regions (i.e. hockey players) surrounded by high-intensity regions (i.e. rink surface). However, they warned that the data supplied by such a simple script are not ideal for training.

In two recent papers, we have proposed another successful approach for dealing with the automatic detection issue of soccer players in the analysis of video sequences. In the first (Figueroa et al. 2006a), we considered the problem of recovering background pixel information in order to segment and track objects in video images. The suggested solution involved a non-parametric morphological labelling operation, which considers lighting changes and the fact that a given scene may include both slow and fast motion. The segmentation of soccer players has been based on the difference between image sequences and the corresponding background representation recovered.

In the second paper (Figueroa et al. 2006b), we introduced a video system (DVideo) as a tracking solution for all players during the entire game. This technique utilised a representation based on the graph theory, with nodes corresponding to the blobs obtained by image segmentation and with edges, weighted using the information on blob trajectory in the image sequence, representing the distance between nodes. A novel approach for treating occlusions involving the split of segmented blobs based on morphological operators was presented, as well as backward and forward transversals of the graph to maximise the number of frames tracked automatically. The method automatically located players in 94% of the frames processed, with a relative error of only 1.4% of the covered extension. When the automatic tracking failed, a user interface was available to complete the trajectories manually. The results for the distances covered by soccer players and other kinematical variables obtained using this approach were reported in Barros et al. (2007).

The algorithms proposed in Figueroa et al. (2006a, 2006b) can automatically provide good accuracy in the segmentation of regions containing the candidate players (*blobs*). Furthermore, the use of graphs (Figueroa et al. 2004; Nillius et al. 2006) has proven very effective in solving players trajectory correspondence issues from one frame to the next. However, the application of morphological segmentation steps to all frames, especially when splitting the blobs using (forward traversal) direct and reverse (backward traversal) directions, makes the algorithm time consuming.

The present paper proposes and evaluates a new method for tracking handball players based on an automatically trained boosting algorithm. The method integrates the powerful algorithms proposed in Figueroa et al. (2006a, 2006b) with the advantages of the boosting detector (Okuma et al. 2004).

The text is organised in two main sections. In the first one, we theoretically describe the basics of the suggested method and its implementation. In Section 2, the experiments for evaluating the method are described. The following evaluations were conducted: (1) evaluation of measurement uncertainty; (2) evaluation of the boosting detector training; (3) evaluation of detection rates; (4) evaluation of automatically tracked trajectories and (5) analysis of the distances covered by handball players during official games using the suggested method.

2. Description of the proposed method

2.1 Data acquisition and camera calibration

One male and one female match of a Brazilian regional handball championship, for players under 21 years old,

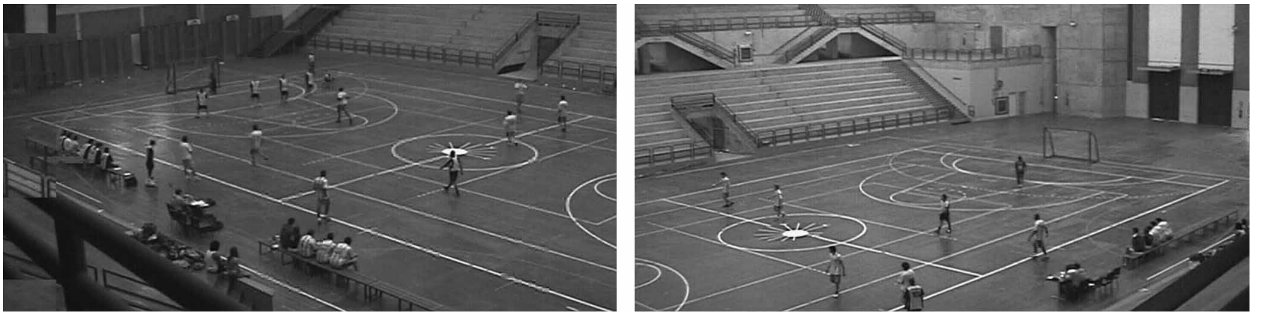

Figure 1. Synchronised cameras views used for tracking and measurement.

were filmed with two digital video cameras (JVC GR-DVL 9500), which were installed at elevated positions on the bleacher. Each camera covered little more than a half of the field as shown in Figure 1.

After recording, the video sequences were transferred to the PC and stored as an AVI file format (30 frames/s, with an image resolution of 720×480 and 24 bit colour resolution). The cameras were calibrated using the field's known dimensions. These parameters and the position of the players in the video sequences were used to reconstruct the 2D coordinates of each player using the direct linear transformation method, as described in Figueroa et al. (2003).

2.2 Training of the boosting detector algorithm

The traditional boosting detector algorithm, introduced by Viola and Jones (2001), uses a cascade of AdaBoost classifiers, with a very low false negative rate, a decreasing rate of false positives and an increasing complexity per cascade level, in order to decide if a given rectangular region of the image matches the object of interest or not.

Consider that for each coordinate (x, y) of the image, there is a rectangle $R_{w,h}(x, y)$ with (x, y) as its top-left pixel, width w and height h. At any given run of the boosting detector algorithm, w and h are fixed. Consider that each rectangle is initially marked as a positive example (i.e. it contains the object of interest). The first cascade level has the most simple (and fastest) classifier. It can be applied over the entire image very quickly and unmark examples that are clearly negative. As every classifier has a low false negative rate, most or all positive examples will remain marked. Likewise, as it is a simple classifier, many negative examples will remain marked. The next cascade level uses a more complex (and slower) classifier, but it will be applied over fewer examples. After the last cascade level is applied over the remaining marked examples, hopefully, only the positive examples will remain.

The boosting detector algorithm is usually applied many times over the image, using rectangles of different sizes (but with same aspect ratio), to detect objects of different scales. AdaBoost is a statistical technique to improve (hence the name *boost*) the classification power of weak classifiers, h_t, to build a strong classifier

$$H(x) = \operatorname{sign}\left(\sum_{t=1}^{T} \alpha_t h_t(x)\right),$$

where α_t is a weight associated to each weak classifier. A weak classifier is a simple learning algorithm with the classification ratio over 50% for the binary classification problem (decided between *is* or *is-not*). Decision tree algorithms such as C4.5 (Quinlan 1993) are popular choices for weak classifiers. Many times a simple stump classifier, i.e. a tree consisting of only the root node, is enough.

The data for the training algorithm are presented as $(x_1, y_1) \ldots (x_m, y_m)$, where $x_i \in X$ is an instance of the input (e.g. a numeric descriptor extracted from a rectangular region of the image) and $y_i \in Y$, $Y = \{-1.1\}$, specifies the class of the input (e.g. player or not player). There is a weight, $D_t(i)$, associated to each training datum (x_i, y_i) and a time t. The weight is used by the algorithm to emphasise inputs that are more difficult to classify. AdaBoost calls a weak classifier repeatedly in a series of rounds $(t = 1, \ldots, T)$. The basic AdaBoost algorithm (Schapire 2002) is as follows:

$$\text{Initialise } D_t(i) = \frac{1}{m}.$$

For $t = 1 \ldots T$

- Train the weak learner using the data and D_1, resulting in $h_t : X \rightarrow \Re$.
- Choose an $\alpha_t \in \Re$.
- Update the weights:

$$D_{t+1}(i) = \frac{D_t(i) \exp(-\alpha_t y_i h_t(x_i))}{Z_t},$$

where Z_t is a normalisation factor, so that $\sum_i D_{t+1}(i) = 1$.

The output is the strong classifier $H(x)$. For the binary classification problem, usually

$$a_t = \frac{1}{2}\ln\left(\frac{1-\varepsilon_t}{\varepsilon_t}\right),$$

where ε_t is the classification error. The bigger the T, the more complex the classifier becomes.

The boosting detector algorithm uses an integral representation of the image that allows quickly extracting useful numeric descriptors. The integral image representation takes each pixel as the sum of the intensities of the upper left pixels, i.e. $II(x,y) = \sum_{x'\leq x, y'\leq y} I(x',y')$, where $II(x,y)$ represents the integral image and $I(x',y')$ is the original image. The descriptors are calculated from integral image features as those of Figure 2. If S_j is the sum of the intensities of the pixels within a given rectangle, the value of the descriptor is $x_i = \sum_j^N B_j S_j$, where B_j is 1 if it is a white rectangle and -1 otherwise.

Using the integral image representation, the descriptor of Figure 3(a) can be calculated with as few as six additions or subtractions. Figure 3(c) shows the integral image feature applied over an image.

The boosting detector algorithm was trained to detect handball players on the court, using a cascaded classifier with 21 layers at most. All image regions extracted from the video frames were scaled to a resolution of 6×15 pixels. The bounding rectangle of each region is adjusted so that all regions have the same aspect ratio. The training sets were obtained using either manual or automatic selection of figures.

2.2.1 *Training boosting detector using manually selected figures*

The group of manually selected figures of handball players for training the boosting detector used up to 5000 figures. This was identified as the manual (M) training group and was subdivided into six subgroups, for evaluating the effects of the large number of positive figures in the training. These subgroups were designated as M100, M200, M500, M1000, M2000 and M5000, based on the number of figures used for training. Each figure was selected by an operator as a bounded rectangular box around the player. Non-handball-player sub-windows were also used to train the detector; these were generated from a set of 2869 image regions containing non-player objects (court and spectators). They were obtained by automatically extracting rectangles of different sizes, but with the same aspect ratio 2:5, from images of the court with no players in it. The number of regions is enough considering that the videos were shot on the same court. All figures were scaled to 6×15 pixels.

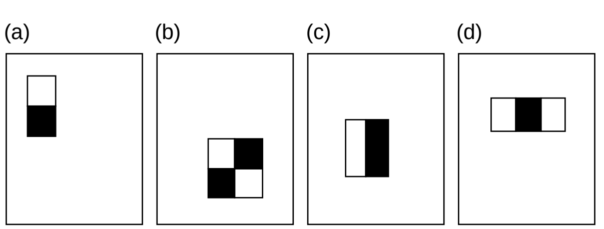

Figure 2. Integral image features (a)–(d).

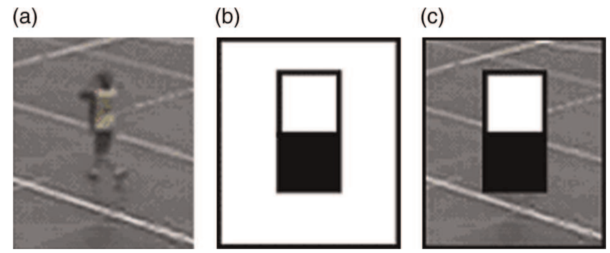

Figure 3. Integral image feature applied over an image. (a) A rectangular region of the image. (b) A typical feature, where the sum of the intensities of the pixels on the white rectangle is subtracted from the sum of the dark rectangle. (c) The region with the feature overlaid.

2.2.2 *Training boosting detector using automatically selected figures*

This second training group consisted of up to 15,000 figures, automatically chosen by morphological segmentation of blobs in a video sequence of a handball game. This group (A) was also divided into subgroups, with the eight automatically selected groups identified as A100, A200, A500, A1000, A2000, A5000, A10000 and A15000, also based on the number of figures used for training. The figures of the smaller groups were drawn from the A15000 group. The same non-handball-player sub-windows were used.

The morphological segmentation method used in the automatic selection of figures (blobs) in order to train the boosting detector, represented in Figure 6, consists of

(a) Loading the video sequence.
(b) Definition of image boundaries by an operator in the first sequence frame.
(c) Background extraction by applying a median filter to the pixels of a set of consecutive frames, thus, regularly updating the background image of those frames, with the intensity of each background pixel calculated as the mean of the most likely repeated intensities of such set of frames.
(d) Calculation of the difference between the image in the current frame and that corresponding to the extracted background.
(e) Image binarisation by thresholding.
(f) Morphological filtering (opening and closing) in order to eliminate noise. Labelling of the connected pixels and definition of the corresponding regions as blobs,

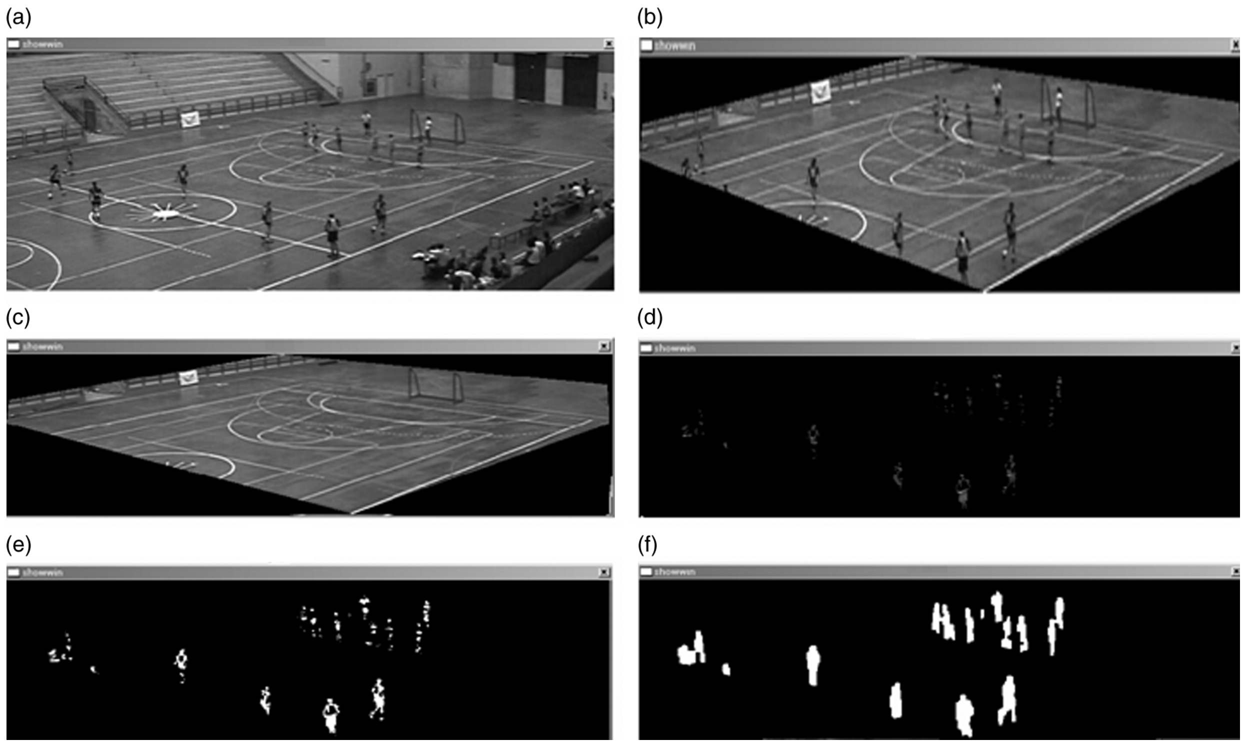

Figure 4. Steps of morphological segmentation. (a) Load the video sequence; (b) definition of image boundaries; (c) background extraction; (d) difference between the current frame and background; (e) image binarisation and (f) morphological filtering and labelling.

using watershed operator. This morphological operation defines a skeleton using zones of influence of the signal regional minima (the set of points with a given value m, such that its external boundary points are strictly greater than m). This skeleton constitutes the boundary of the zones of influence of the regional minima considered.

As it can be seen in Figure 4(f), the morphological segmentation provides a set of blobs, each containing at least one player. In order to avoid training the model with blobs containing more than one player, only those blobs smaller than a predefined size are used. Considering the fact that the figures required by the boosting detector algorithm have to be rectangles with a fixed aspect ratio, a correct bounding rectangle replaces the blob. Blobs are rejected if their scale does not match the scale expected for a handball player at its region on the image. This is done automatically using the homography between the image of the court and a virtual representation of the court with official dimensions.

Figure 5 shows the output of morphological segmentation and the input used in the boosting detector algorithm.

All training sessions involved figures selected during the first half-time of the game, whereas the detection and tracking procedures were evaluated during the second half.

Figure 5. On the left, automatic segmentation; on the right, automatically generated positive examples for boosting detector training.

2.3 Boosting detector

When running the boosting detector, we limit the multiple scale searches to regions where each scale is possible. As we have done during training, we used the homography between the image of the court and a virtual representation of the court with official dimensions in order for this to be automatically done.

2.4 Tracking using the DVideo system

In order to compare the performance of the automatically trained boosting detector with that of the original DVideo segmentation procedures, the results of the two methods were introduced into the DVideo system and submitted to identical tracking procedures, which can be summarised as follows:

(1) Manual frame-by-frame tracking (done by humans) assumed to provide true trajectories and used to compare the results of the next two conditions (manual tracking).
(2) Automatic tracking using the blobs detected by a boosting detector trained with 15,000 automatically selected Figures (A15000, boosting tracking).
(3) Automatic tracking using the original DVideo algorithm with morphological segmentation, followed by the split of blobs using graphs in forward and reverse directions (DVideo tracking).

Graphs were then developed from the set of blobs obtained during the detection (boosting tracking) and segmentation (DVideo tracking) steps, so that nodes represent blobs, whereas edges indicate the distance among these blobs. Therefore, each node of the graph stores the spatial information of a blob, whereas the edges convey the temporal information related to the dependence of blobs.

The tracking of each player is performed by minimal path searching through the graph. In addition to indicating the distance between blobs, the edges of the graph are weighted for information such as velocity, orientation and blob colour. In each step of the tracking method, the true 2D coordinates of the players on the court are reconstructed using calibration parameters and image coordinates.

Figure 6 presents a flowchart of the overall processing.

3. Results and discussion

3.1 Evaluation of measurement uncertainty

In order to estimate uncertainties in the determination of court positions, 22 points were measured manually, 10 times each, and the results were compared to the expected values. The mean 2D position of each point over the 10 repetitions was calculated, with the expected position obtained by direct measurement. Figure 7 shows the results of mean $\pm$ standard deviation of the distances between observed and expected positions in the different regions of the court.

The mean uncertainty involved in the determination of a position is 0.20 m. Considering that the field-of-view of each camera covers approximately half of the court (20×20 m), the relative uncertainty is less than 1%, and can be considered as acceptable.

3.2 Evaluation of boosting tracking

In this section, we have analysed the time spent in the manual and automatic selection of figures for the training of the boosting detector (selection time), as well as, the time spent in manual and automatic training (training time). These procedures were executed on a Pentium 4 PC (3.0 GHz, 2 GB RAM) running on Microsoft Windows XP.

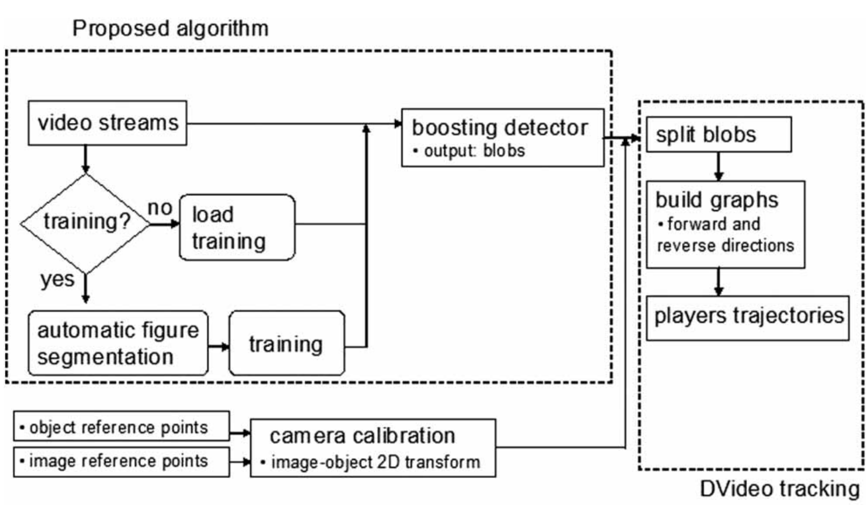

Figure 6. Flowchart of the overall processing.

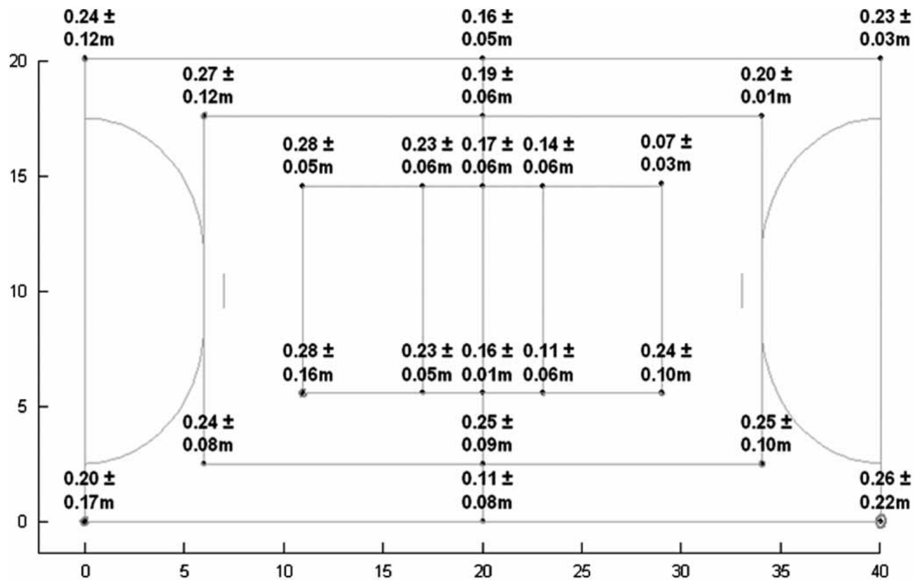

Figure 7. Representation of uncertainties to measure and reconstruct positions ($N = 10$) in 22 positions of the handball court. Mean ± standard deviation. Values are in metres.

3.2.1 *Evaluation of selection and training times*

Manual selection involved approximately 1250 figures per hour, or 0.35 figures per second. Of course, this process is extremely tedious and time consuming. The use of automatic selection of figures using the morphological segmentation algorithm suggested in this paper dealt with approximately 24 figures (blobs) per second. The training time required for using these figures is shown in Figure 8.

In both manual and automatic training times, the increase of figures (blobs) initially increased the training time, although later on the time decreased. The time required to train the boosting detector was maximal with the use of 1000 figures, for both manual and automatic training. One possible explanation for this is that, for our application, the cascaded boosting detector algorithm converges more rapidly when a larger number of figures are used for training. In general, the time required for training with manually selected figures is less than that with the automatically selected ones, due to the better quality of the manual selection. Interestingly enough, the time required for training the boosting detector algorithm with 15,000 automatically selected figures is almost the same as that with the use of 100 manually selected figures.

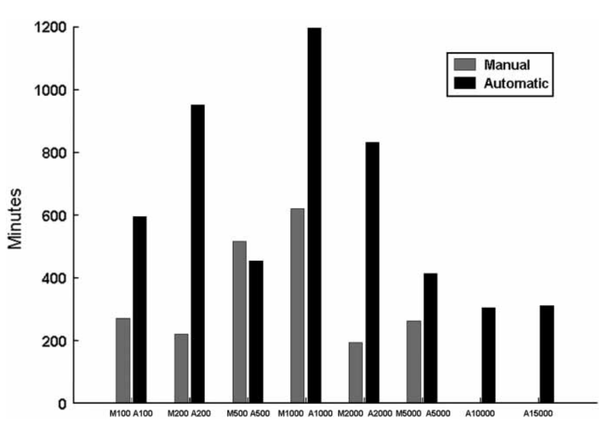

Figure 8. Manual (M) and automatic (A) training times of the boosting detector for handball players localisation. The number beside the letter M or A indicates how many figures were used for training.

3.2.2 *Evaluation of detection using manual and morphological segmentation training*

The effect of using manual and automatic training on the boosting detector and varying of the number of figures used in training can be analysed in Figure 9. It is important to emphasise that the algorithm was trained during the first period of a game, whereas detection was tested during the second period of that same game.

The larger the number of figures (blobs) used for training, the smaller the accumulated error in detection, regardless of training kind. The relationship between the number of images analysed and the accumulated error is approximately linear. This information can be useful in planning the number of figures that are necessary for the training process. Those periods in the sequence with non-linear behaviour between variables were analysed and revealed that game dynamics may have a slight effect on results. The transition phase, as players pass from defence to attack, as well as player agglomeration or simple pauses in the game may result an increase or decrease in the detection rate.

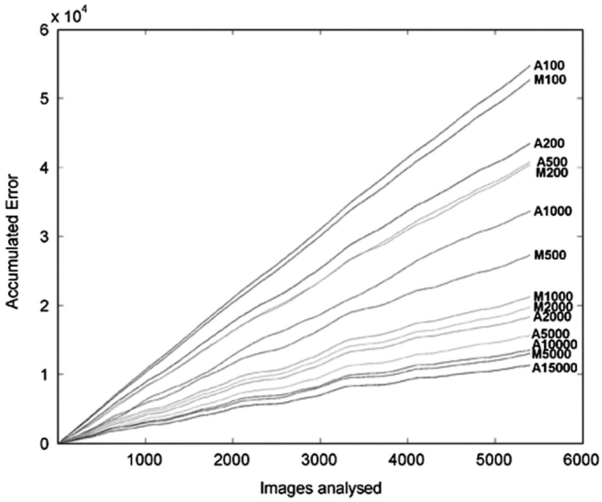

Figure 9. Accumulated error of the boosting detector in function of the number of images analysed for manual (M) and automatic (A) training, with different sizes of training set. Errors in pixels.

Figure 10 shows the percentage of successful detection as a function of the number of figures used for manual and automatic training.

In general, the manual training provides better results than that using automatic selection if the number of figures remained constant. Such better results were to be expected and can be explained by those situations in which the automatic selection of figures introduces false positive figures during training, a problem avoided by the manual selection. The training using automatic morphological segmentation has achieved the highest detection rate (82%) with the use of 15,000 figures.

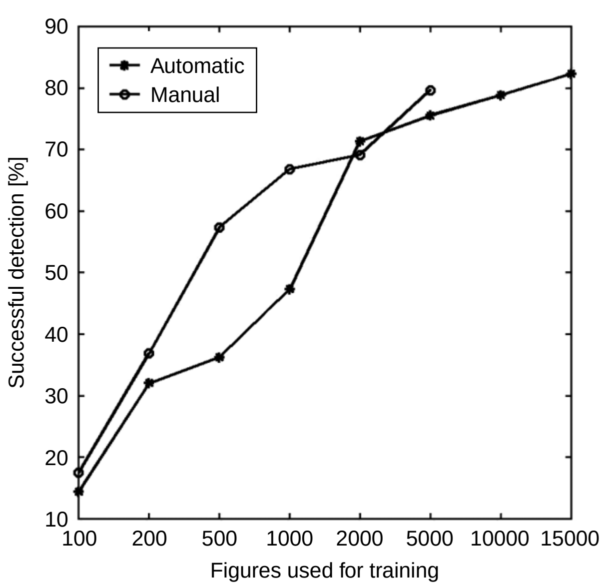

Figure 10. Percentage of successful detection for manual and automatic training.

3.3 *Evaluation of tracking*

Figure 11 shows the x-coordinate (main court direction) of the trajectory of a single player as a function of the frames analysed for 3 min, using manual tracking, boosting tracking and DVideo tracking. Figure 12 shows the y-coordinate for the same situations. Data were filtered using a third-order Butterworth digital filter.

In both figures, the curves present a similar shape revealing similarity of results when using different tracking conditions. Table 1 allows the quantitative comparison of the trajectories of four players tracked during the same 3 min of the game.

The mean percentage of the absolute difference from the manual measurements and that segmented by the DVideo system was 2.2% and the maximal percentage error reached was 7.8%. The booting tracking presented better results considering both the mean percentage differences (1.8%) and the maximal error (3.6%). It is, however, necessary to remember that the filter can affect the three sets of data differently; therefore, these results should be carefully analysed.

The percentage of frames automatically tracked was 67% with the DVideo tracking and 84% with the boosting tracking. This result can be considered a great advance regarding the manual and current automated methods applied to handball matches.

Better results were previously obtained when the same system was used to analyse soccer players, with mean of 94% with automatic tracking (Barros et al. 2007). The reduction observed here is probably related to the peculiarities of handball games. Further improvements can probably be achieved by positioning cameras in both sides of the court or at the ceiling, as suggested by Pers and Kovacic (2000).

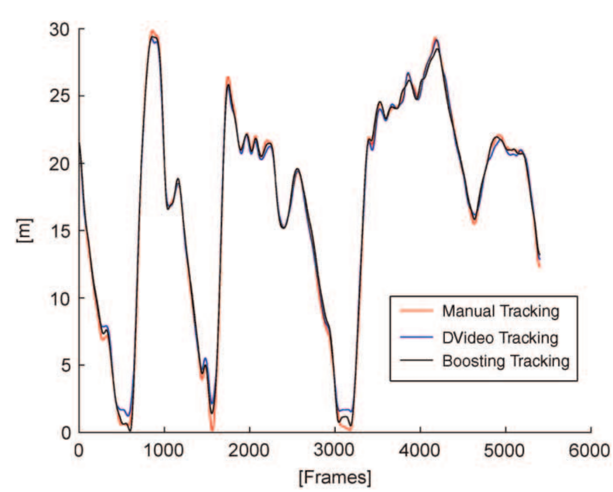

Figure 11. Motion of one player (x-coordinate) in function of frames analysed over 3 min by manual, boosting and DVideo tracking.

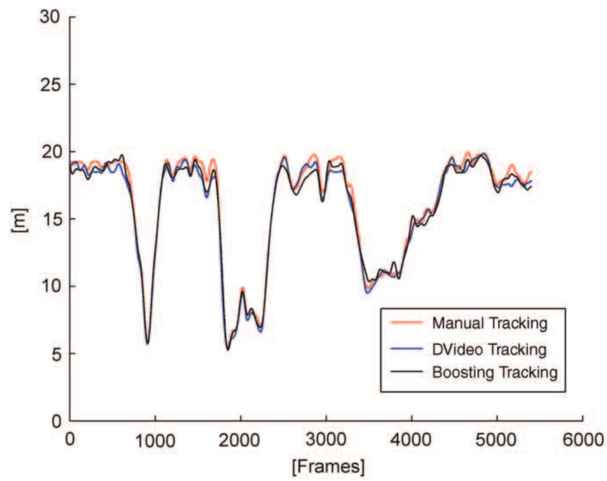

Figure 12. Motion of one player (y-coordinate) in function of the frames analysed over 3 min by manual, boosting and DVideo tracking.

3.4 Results of distances covered by handball players

Figure 13 presents the trajectories of six players over the first halt-time (30 min). Player 4 was replaced by player number 8 and, therefore, their trajectories were merged.

The representation allows individual and collective analyses by the team staff about the tactical performances of players. For instance, the defensive and offensive team strategies can be identified by the team staff considering the region of the pitch that has been most frequently visited. The results in Table 1 showed that the tracking of player 7 was particularly difficult using the DVideo tracking. One possible explanation can be found looking at Figure 14, which shows that such a player has the pivot function in the team, therefore always playing in a region of great congestion, in the middle or behind the opposite defence.

Figure 14 presents the distances covered by outline players of both teams in the first and second half of the match. It is important to emphasise that this kind of simultaneous

Table 1. Distances covered by six players during 5 min of game calculated from the data obtained by manual, boosting and DVideo tracking. The percentage differences from DVideo and boosting tracking compared to manual tracking are indicated in brackets.

	Distances covered (m)		
Player number	Manual tracking	DVideo tracking	Boosting tracking
2	440.4	436.5 (−0.8%)	440.8 (0.1%)
3	447.3	454.7 (1.6%)	432.5 (−3.3%)
4	381.4	378.3 (−0.6%)	390.8 (2.1%)
5	325.5	318.3 (−1.6)	341.8 (3.6%)
6	367.2	370.6 (0.7)	367.0 (0.0%)
7	435.2	470.2 (7.8%)	443.2 (1.8%)

results for all outline players, during the whole game, obtained not in simulated but in actual competition condition, was not found within the literature. Furthermore, the distances covered provide useful information about the physical performance of players and it can be used for better planning the training process.

4. Conclusions

In this paper, we have evaluated and compared the use of a boosting detector, the publicly available OpenCV implementation, in the task of tracking, for a long period, handball players in official matches, with no controlled environment or lighting.

The mean uncertainty to determine absolute position in the handball court using the suggested vision system is 0.20 m and the relative uncertainties are less than 1% and can be considered acceptable.

The use of automatic selection of figures using the morphological segmentation algorithm proposed in this paper dealt with approximately 24 figures (blobs) per second as opposed to 0.3 figures per second provided by manual selection. The increment of figures in the training process did not cause always the increment in training time, as could be expected, because greater and better set of figures given for training will provoke a faster convergence of the cascade boosting detector algorithm. The time spent for automatically training the boosting detector algorithm with 15,000 figures is almost the same needed for training using 100 manually selected figures.

The larger the number of figures (blobs) used for training, smaller will be the accumulated error in detection, regardless of training kind, and the relationship between the number of images analysed and accumulated error is approximately linear. The highest successful detection rate (82%) was provided by the boosting detector trained with automatic morphological segmentation using 15,000 figures. The use of automatic morphological segmentation showed to be a fast and efficient method for training the boosting detector.

The percentage of frames automatically tracked was 67% with the DVideo tracking and 84% with the boosting tracking. The complication factor of the boosting algorithm is training and building the appropriate training set. To address this question, we used a boosting detector trained for the above situation to track a female handball match, recorded on a different day under distinct camera and illumination conditions with players using uniforms with different luminance and colour patterns. This experiment held an average detection rate of 74% (compared to a 5-min manually detailed sequence). As expected, the accuracy decreases, as the training is no longer specialised for the game conditions, nevertheless, this shows the incredible potential of the application of cross-training boosting detectors.

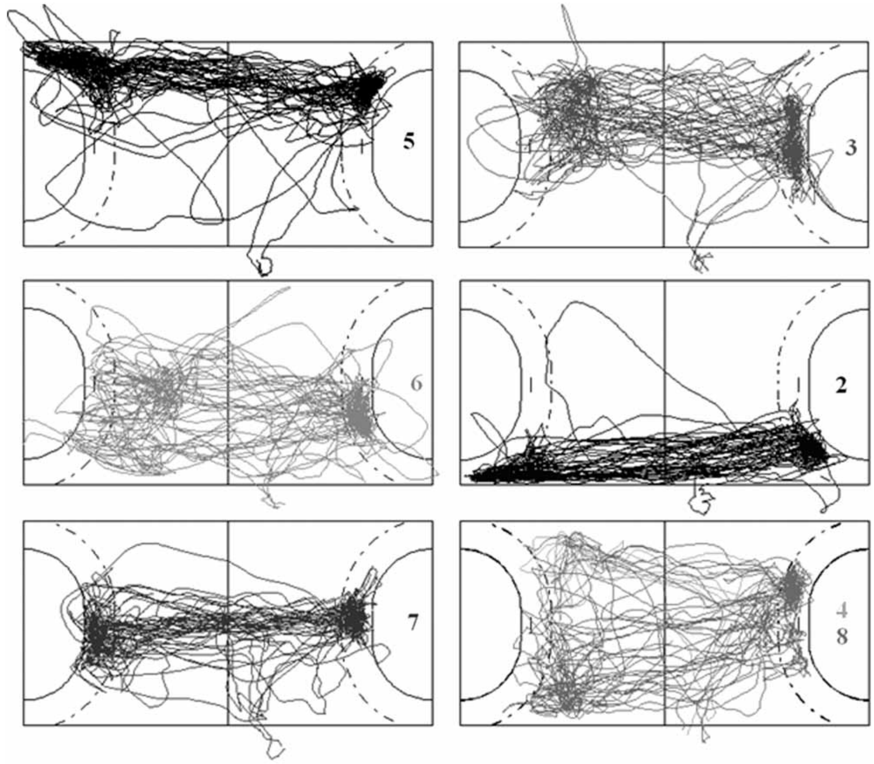

Figure 13. Trajectories of handball players in the first half-time (30 min). Defence on the right side.

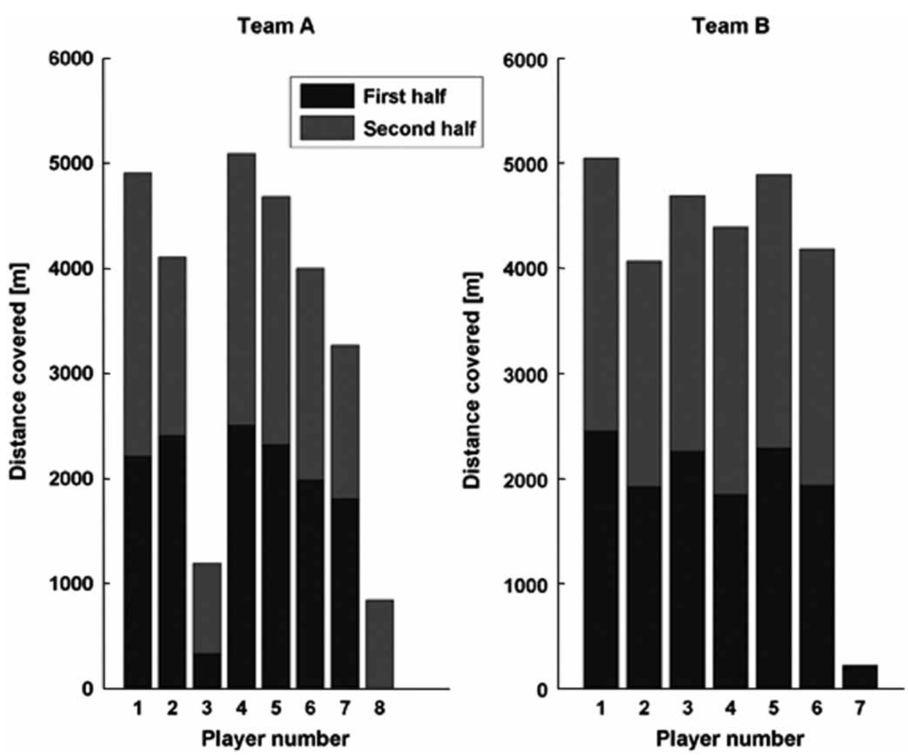

Figure 14. Distances covered by players of teams A and B in the first and second half-times of a handball game.

Although these results can be considered a great advance compared to manual methods, further developments are required due to the large number of frames to be manually measured or corrected.

Official sports events pose a series of difficulties for the use of computer vision. Most of the time, there are severe restrictions on where you can place your cameras. Additionally, there is no control over the colours or patterns of uniforms and equipment used by athletes. In the examples we used in this paper, it is very difficult for a human to distinguish between two different teams. The strength against variation in the video quality is also something not entirely explored within the literature, and deserves a lot of further work.

Additionally, little is described on the computer vision literature regarding the qualitative nature of the algorithms failures, and how to proceed in a real system when they eventually occur. In sports analysis, this means human intervention, and not all errors are treated equally. New methodologies to measure accuracy and false negatives have to be developed in order to take this into account.

References

Bangsbo J, Norregaard L, Thorso F. 1991. Activity profile of competition soccer. Can J Sport Sci. 16(2):110–116.

Barros RML, Misuta MS, Menezes RP, Figueroa PJ, Moura FA, Cunha SA, Anido R, Leite NJ. 2007. Analysis of the distances covered by first division Brazilian soccer players obtained with an automatic tracking method. J Sports Sci Med. 6(2):233–242.

Edgecomb SJ, Norton KI. 2006. Comparison of global positioning and computer-based tracking systems for measuring player movement distance during Australian Football. J Sci Med Sport. 9(1–2):25–32.

Figueroa PJ, Leite NJ, Barros RM. 2003. A flexible software for tracking of markers used in human motion analysis. Comput Methods Programs Biomed. 72(2):155–165.

Figueroa P, Leite N, Barros RML, Cohen I, Medioni G. 2004. Tracking soccer players using the graph representation. In: Proceedings of the 17th International Conference on Pattern Recognition (ICPR'04), IEEE Computer Society, Los Alamitos, CA, USA, Vol. 4. p. 787–790.

Figueroa PJ, Leite NJ, Barros RML. 2006a. Background recovering in outdoor image sequences: an example of soccer players segmentation. Image Vision Comput. 24(4): 363–374.

Figueroa PJ, Leite NJ, Barros RML. 2006b. Tracking soccer players aiming their kinematical motion analysis. Comput Vis Image Underst. 101(2):122–135.

Hennig E, Briehle R. 2000. Game analysis by GPS satellite tracking of soccer players. In: XI Congress of the Canadian society for biomechanics, Montreal. p. 44.

Iwase S, Saito H. 2004. Parallel tracking of all soccer players by integrating detected positions in multiple view images. In: Proceedings of the 17th international conference on pattern recognition (ICPR'04), Cambridge, UK, Vol. 4. p. 751–754.

Mohr M, Krustrup P, Bangsbo J. 2003. Match performance of high-standard soccer players with special reference to development of fatigue. J Sports Sci. 21(7):519–528.

Needham CJ, Boyle RD. 2001. Tracking multiple sports players through occlusion, congestion and scale. In: Proceedings of the British machine vision conference, Manchester, UK. p. 93–102.

Nillius P, Sullivan J, Carlsson S. 2006. Multi-target tracking – linking identities using Bayesian network inference. IEEE computer society conference on computer vision and pattern recognition, 2006; FORTH, Greece. p. 2187–2194.

Ohashi J, Miyagi O, Nagahama H, Ogushi T, Ohashi K. 2002. Application of an analysis system evaluating intermittent activity during a Soccer match. London and New York: Routledge.

Okuma K, Taleghani A, de Freitas N, Little JJ, Lowe DG. 2004. A boosted particle filter: multitarget detection and tracking. In: European conference on computer vision, 2004; May, Prague. p. 28–39.

Pers J, Kovacic S. 2000. A system for tracking players in sports games by computer vision. Electrotech Rev. Ljubljana, Slovenija 67(5):281–288.

Pers J, Bon M, Kovacic S, Sibila M, Dezman B. 2002. Observation and analysis of large-scale human motion. Hum Mov Sci. 21(2):295–311.

Quinlan JR. 1993. C4.5: programs for machine learning. San Mateo, CA: Morgan Kaufmann Publishers, Inc.

Schapire RE. 2002. The boosting approach to machine learning. An overview. MSRI Workshop on Nonlinear Estimation and Classification.

Toki S, Sakurai S. 2005. Quantitative match analysis of Soccer games with two dimensional DLT procedures. In: XXth congress of international society of biomechanics, Cleveland. p. 911.

Viola P, Jones MJ. 2001. Rapid object detection using a boosted cascade of simple features. In: IEEE conference on computer vision and pattern recognition, IEEE Computer Society, Los Alamitos, CA, USA, Vol. 1. p. 511.

Withers RT, Maricic Z, Wasilewski S, Kelly L. 1982. Match analyses of Australian professional soccer players. J Hum Movement Stud. 8:159–176.