The State of the Art in Multiple Object Tracking Under Occlusion in Video Sequences

Pierre F. Gabriel, Jacques G. Verly, Justus H. Piater, and André Genon

{p.gabriel, jacques.verly, justus.piater, agenon}@ulg.ac.be
Department of Electrical Engineering and Computer Science,
University of Liège, Belgium.

Abstract

In this paper, we present a review of existing techniques and systems for tracking multiple occluding objects using one or more cameras. Following a formulation of the occlusion problem, we divide these techniques into two groups: merge-split (MS) approaches and straight-through (ST) approaches. Then, we consider tracking in ball game applications, with emphasis on soccer. Based on this assessment of the state of the art, we identify what appear to be the most promising approaches for tracking in general and for soccer in particular.

1. Introduction

There has been considerable research activity on the tracking of objects from video sequences over the last 20 years. This interest is motivated by numerous applications, such as surveillance, video conferencing, man-machine interfaces, and sports enhancement.

We consider the problem of simultaneously tracking one or more objects in one or more video sequences. In particular, we focus on the cases where two or more objects occlude each other, either partially or completely. Note that these objects can be rigid (e.g., cars) or deformable (e.g., persons). They can also be fixed (e.g., a column) or mobile, in which case they can be stationary or in motion.

To evaluate and compare the capabilities of the various video tracking systems developped by others for diverse applications, we have found it useful to develop formal notions of objects, groups of objects ("blobs") and occlusions. These concepts are presented in Section 2.

Based on this framework, we identify the primary signal processing components that any robust generic video tracking system should have. The first part of the discussion is centered on the use of a single camera. Then, we consider the added value of using multiple cameras. For each processing component, we list the main techniques that have been used by others. In each case, we make reference to the specific systems that use these techniques. This material is covered in Section 3 and 4.

As a concrete illustration, we consider, in Section 5, the particular application of video tracking in soccer and review the specialized techniques that have been proposed by others in this domain.

Finally, in Section 6, we identify the techniques that appear the most promising for dealing with occlusions in general and with soccer applications in particular.

2. Formulation of the occlusion problem

Tracking objects in crowded scenes necessarily leads to the problem of occlusion. Discussing this problem and comparing the capabilities of various, existing video tracking systems would be much easier if we could cast the occlusion problem in some formal framework. To the best of our knowledge, no such framework exists in the literature. As a result, we propose here a simple, first-cut framework that will allow us, first, to describe the problem of occlusion in generic terms and, then, to classify and to compare the existing tracking systems that deal with occlusions.

The primary entity is the "*blob*" (sometimes called target), which is defined as being a group of "objects". The exact nature of the objects is irrelevant; they can be persons, cars, columns, etc. Figure 1 shows the graphical representation of blobs and objects. It is important to note that a blob acts as a container that can have one or more objects. It is also important to understand that the things that are being detected, via image processing, and tracked, whether in the absence or in the presence of occlusions, are blobs, not objects. A blob could be recursively defined as being a group of blobs, instead of objects, but we do not feel this is necessary at this time.

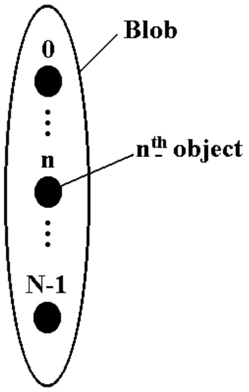

Figure 1. Representation of a blob containing N objects.

Objects and blobs are described by a series of attributes such as position, velocity and appearance. Blobs are also characterized by operations, such as **create**, **delete, merge**, and **split**. Create is used to create a new blob and to initiate a new track. Delete is used to remove a blob from future consideration. Active blobs are tracked at all times.

We assume that there exists a predicate Po that detects the fact that two or more blobs are occluding each other. The inner workings of Po are irrelevant and system specific. Once an occlusion is detected, there appear to be two useful, generic approaches for dealing with it.

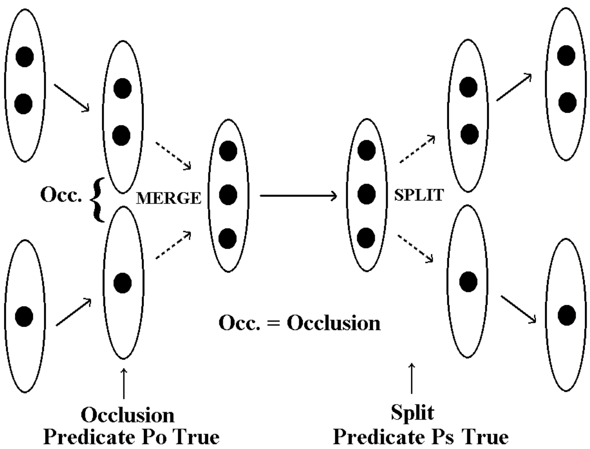

Figure 2. Merge-Split (MS) approach. Solid arrows correspond to blob motion and tracking.

We refer to the first approach as the *merge-split* (MS) approach (figure 2). As soon as blobs are declared to be occluding by Po, the system merges them into a single new blob. From that point on, the original objects are encapsulated into the new blob. The new blob is characterized by new attributes and is tracked as any of the active blobs in the system. We assume that there is a predicate Ps that is able to decide whether a blob containing at least two objects must be split or whether a new blob that has appeared next to a composite blob (with at least two objects) corresponds to one or more of the objects in the composite blob. The inner workings of Ps are also irrelevant and system specific. In either case, we must then properly split the composite blob into two other blobs.

We refer to the second approach as the *straight-through* (ST) approach (figure 3). Here, we simply continue to track the individual blobs (containing only one object) through the occlusion without attempting to merge them. The system relies on the same occlusion predicate Po used in the MS approach. Indeed, one generally needs to refine the tracking technique as soon as an occlusion is detected and as long as it persists.

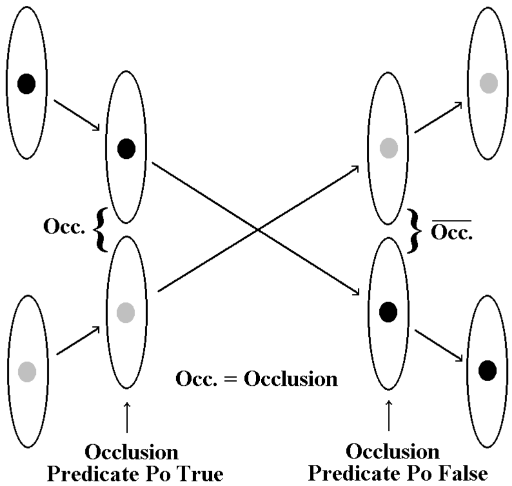

Figure 3. Straight-Through (ST) approach. Solid arrows correspond to blob motion and tracking.

3. Techniques based on single cameras

Some tracking systems do not deal with occlusion at all [2, 37]. Others minimize occlusions by placing the cameras so that they look down on the scene [4, 17, 8].

Most systems rely on the technique of *background subtraction*. One common approach is to build a simple statistical model for each of the pixels in the image frame. This model can be computed once and for all based on the first M frames or it can be continuously updated based on the last M frames. The model can then be used to segment the current frame into background and foreground regions: any pixel that does not fit the background model (e.g., for having a value too far from the mean) is assigned to the foreground. Some systems simply subtract consecutive frames.

A. Merge-split (MS) approach

In the MS approach, attributes of atomic blobs (i.e., containing a single object) are continuously updated until they come into an occlusion situation. At that point, the attribute of these objects are frozen and a composite blob is created containing these objects and their frozen attributes. The composite blob is tracked as any other and its own attributes are also continuously updated. When a split condition occurs, the problem is to identify the object that is splitting from the group. This is necessary if we want to have a continuous track of each object. For simplicity, consider that each object, e.g. A, belongs to an atomic blob; this blob is tracked until a merge occurs. If and when A emerges from this or another composite blob, we need to detect the fact that the splitting object is A. If we can successfully re-establish identities at each split, we can then track each object, such as A, during its entire existence in the video sequence.

The systems [32, 5] use only appearance features such as color, shape and texture to re-establish identity. Systems [18, 38] use both appearance and dynamic features. These last two systems use *Kalman filters* with first or second order blob motion model. These filters provide estimates of blobs locations in consecutive frames. As we re-establish identities, we can assign to the splitting object either its original speed or that of the blob it comes from. Brémond et al. [5] appear to be the only ones to track composite blobs. MacKenna et al. [32] track a person carrying an object picked up from the scene.

The system W4 [18] of Haritaoglu et al. uses gray-scale texture, silhouette shape information as well as a dynamic template. Piater et al. [38] detect occlusion by considering the interaction of adaptive gaussian regions of interest. Their technique allows them to avoid merging blobs that barely occlude each other. However, they do not address the issue of splitting and of re-establishing identity. MacKenna et al. [32] re-establish identity by using color histograms. The effectiveness of the method is illustrated in figure 4.

B. Straight-through (ST) approach

In the ST approach, we do not merge the occluding blobs so that the blobs always contain at most one object. In fact, if we only focus on this approach, we can simply talk about objects.

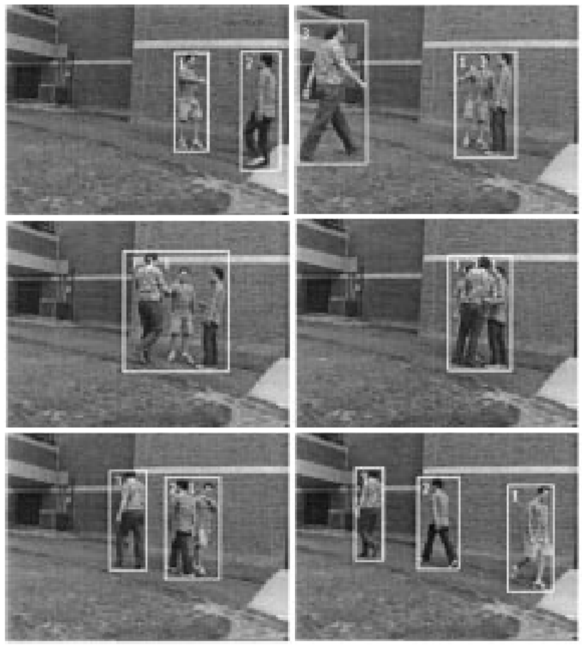

Figure 4. A sequence of people being tracked as they merge and split (taken from [32]).

In this approach, we must continue to track individual objects through each occlusion. In other words, we must be able, at all time, to classify any pixel in the vicinity of the occlusion region as belonging to exactly one of the occluding objects. Most systems rely on the appearance features of objects to classify these pixels [12, 43, 25, 19, 40]. One particularly useful feature is the relative depth between occluding objects. Relative depth (or depth for short) can be determined in various ways. Elgammal and Davis [12] determine the depth of people by evaluating different hypotheses regarding their spatial arrangement. As explained later, Senior et al. [43] order objects so that those that are assigned fewer “disputed” pixels are given greater depth. Of course, stereo from multiple cameras provides more direct depth information [3]. Below, we describe the significant feature of various systems.

Elgammal and Davis [12] build a model of each individual person prior to occlusion. The model consists of the color and spatial characteristics of the various significant parts of each person (e.g. head, torso, and legs). In the presence of occlusion, each pixel in the occluded group is assigned to a particular person’s part by a maximum-likelihood type algorithm. Given this assignment, the silhouette corresponding to each person is found immediately. The authors use ellipses to track people through each occlusion. These processing step are illustrated in figure 5.

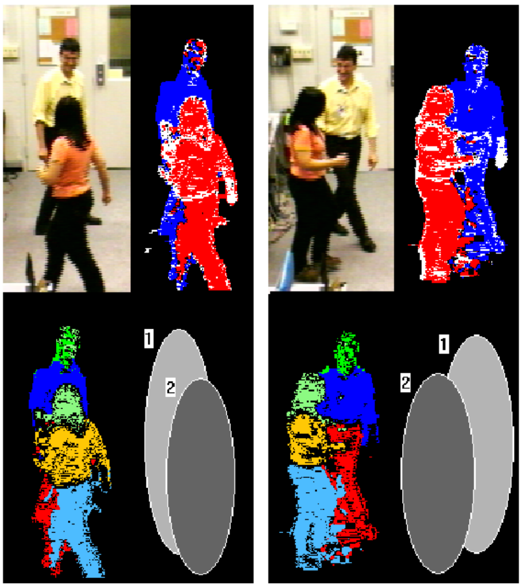

Figure 5. Left and right frames correspond to two different scenes. Subfigures in each are as follows. *Top-left*: original image, *bottom-left*: segmented regions for each person, *top-right*: segmented persons, and *bottom-right*: occlusion model (taken from [12]).

The approach of Senior et al. [43] is essentially as follows. The background is subtracted using a conventional statistical model. The result is a set of foreground objects. The models used for each object being tracked consists of a deterministic RGB color template and a registered probability mask. These two items are allowed to evolve over time. When two or more objects are detected to occlude each other, the occlusion is handled as follows. Since the templates and the masks of the N objects overlap, some image pixels will have nonzero probabilities in each mask. The key step is to assign these “disputed” pixels to one of the N objects. To do this, a simple maximum-likelihood classifier is used to assign the disputed pixels to one of the objects. These objects are ordered by distance based upon the number of disputed pixels assigned to them, a small number corresponding to a greater range. At this point, all disputed pixels can be re-assigned to each of the objects: each disputed pixel is simply assigned to the closest object. Figure 6 illustrates some important processing steps.

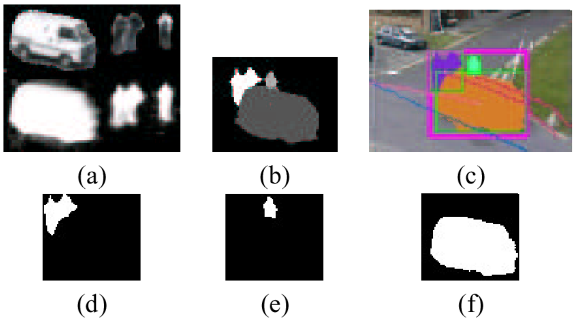

Figure 6. Example of occlusion resolution.
(a) Three appearance models, (b) selected foreground region, (d, e, f) pixels finally allocated to each object, and (c) resolved objects overlaid on original frame (taken from [43]).

Khan and Shah [25] present a system that tracks people in the presence of occlusion (see figure 7). First, they build a background model of the empty scene (consisting of the means and covariance of the color components for all pixels in the image). Second, they detect the first person entering the scene by looking for significant changes with respect to the background. Third, they detect and track new persons entering the scene. For every pixel, they define a vector [x,y,Y,U,V] corresponding to the spatial coordinates and color components. They assume that this vector has a gaussian distribution. The persons are segmented using the *Expectation Maximization* (EM) algorithm. Fourth, they use an a posteriori probability approach to track these persons from frame to frame.

Figure 7. Tracking two persons during occlusion (taken from [25]).

As for the MS approach, one can use a *Kalman filter* or an *extended Kalman filter* to estimate the positions and velocities of the moving objects [41, 11]. Such filters only require that the first and second order statistics of the states be known. Specifically, the form of these filters can be derived without assuming that the states have a Gaussian distribution. However, Gaussianity guarantees that the filter is optimal.

Isard and Blake developped the *condensation* algorithm [21] to track arbitrary contours in image sequences. This algorithm is an extension of *factored sampling* [16]. The condensation algorithm does not require second order statistics, but attempts to use estimates of the probability density function. In fact, it is a special case of *particle filtering*. Particle filters [30, 35, 36], such as the sequential Monte Carlo filter and sequential importance sampling, were developped to address nongaussianity and nonlinearity concerns.

MacCormick and Blake [31] use the condensation algorithm with *partitioned sampling* to perform robust contour tracking of multiple, interacting objects (see figure 8).

Figure 8. Occlusion reasoning using condensation algorithm with partitioned sampling (taken from [31]).

4. Techniques based on multiple cameras

Since the field of view (FOV) of video cameras is quite limited, there is a growing interest in using multiple cameras to increase the monitored area. However, multiple cameras can also help to track objects which are occluded from one or more viewing angles. Some system use static cameras [26, 24, 10, 7, 39] while others use automatically driven cameras [8, 45]. Stereo is sometimes used to track people [3, 9, 18].

In multiview monitoring, the data coming from the various cameras must be "fused" to handle occlusion. A Bayesian network can be used to perform this fusion such as in [10, 24, 7] (see figure 9). In these systems a Kalman filter is also used to refine the observed parameters.

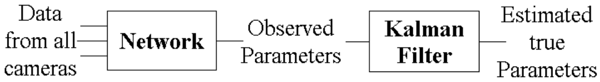

Figure 9. Data fusion in multiview monitoring.

Dockstader and Tekalp's [10] use the Bayesian net to iteratively resolve independence relationships and confidence levels, thereby producing the most likely vector of 3D state estimates for the tracked objects given the available data. They use the Kalman filter to update the 3D state estimates, thus enforcing temporal continuity.

The parameters used by Chang and al. [7] are epipolar line, landmarks, apparent color and height. When occlusion occurs in one camera, the objects' motion estimates provided by the Kalman filter are used to disambiguate the objects identities. In addition, the condensation algorithm is employed to track jointly the sets of features of each object that is being tracked. This is useful when a person is occluded in all video sequences.

5. Application to the soccer game

Over the past ten years, object tracking in video imagery has gained interest in the domain of sports to capture highlights and to gather statistics. Pingali and al. [39] use six cameras around a stadium to track the motion of a tennis ball to obtain its 3D trajectory. Intille and Bobick [20] use a *closed world* model to track players in american football for the purpose of video annotation.

Soccer is an important application for the tracking of multiple objects in the presence of frequent occlusions. The understanding of the movements of the players and the ball is essential for the analysis of matches, both for collecting statistics and for understanding tactics. References [53, 46, 50, 51] describe qualitative techniques. Yow et al. [53] propose a method for showing panoramic views of highlight scenes by mosaicking image sequences. Xie et al. [50] focus on the detection of two exclusive states of the game: play and break. These states are modelled with a *Hidden Markov Model*, and a maximum-likelihood approach is used to segment the game into the two states.

The detailed, automatic analysis of soccer games calls for quantitative methods that provide precise metric information. These methods must maintain precise track of all the players and the ball and give their precise position with respect to the landmarks of the field, such as side lines, half line, and center circle. Of course, the identity and team-association of each player must also be provided at all times. The three main issues are thus (1) identification of landmarks and precise calculation of their positions, (2) identification and tracking of the players and the ball and (3) precise calculation of their position and velocity vectors.

The methods used for coping with the occlusions of the players and the ball are those described in section 3. Iwase and Saito [23] use multiple cameras to overcome occlusions. The first to address all three points above is Seo et al. [44]. They track players by template matching and Kalman filtering (see figure 10). Occlusion reasoning is done by color histogram back-projection. Using a 2D-to-3D image-to-field transformation, the absolute position and the whole trajectory on the field model can be determined. The vertical color distributions of the templates are used to determine the team each player belongs to.

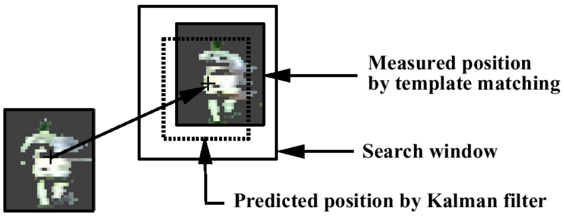

Figure 10. Tracking of player by Kalman filtering followed by template matching (taken from [44]).

Others have attempted to address some or all of the above issues [52, 49, 27]. Yoshinori et al. [52] track players by extracting shirt and pants regions. They cope with posture change and occlusion by considering the players' colors, positions, and velocities in the image. Kim et al. [28] analyze the 3D position of the ball by combining the 2D-to-3D image-to-field transformation with the physics-based constraint that a ball follows a parabolic trajectory.

Particle filtering was recently used to track soccer players. Needham and Boyle's [34] system fit each player being tracked with a model, and the sampling probability for the group of samples is calculated as a function of the fitness score of each player. Nummiaro et al. [35] suggest the integration of color distributions into particle filtering and show how these distributions can be adapted over time.

Multiple object tracking by particles filters tends to fail when two or more players come close to each other or overlap. The reason is that the filters' particles tend to move to regions of high posterior probability. Ok et al. [36] introduce an *occlusion alarm probability* (OAP) to solve this problem. This OAP is a form of the occlusion predicate Po introduced in Section 2. In fact, the OAP acts as a fuzzy predicate indicating the distance (in some statistical sense) between players.

6. Discussion and conclusion

The merge-split (MS) and straight-through (ST) approaches described in this paper are the main approaches to solve occlusion problems.

The main difficulty in MS approaches is to re-establish object identities following a split.

ST approaches do not suffer from this problem since object identities are maintained at all times. In regions-based ST approaches, the main difficulty is the assignment to a specific object of pixels that could belong to several objects ("disputed" pixels). In contour-based ST approaches, the main difficulty is the assignment to a specific object of some partial contours. Distance ordering of occluding objects appears to be a useful step in resolving occlusions: several ordering techniques are proposed in the literature.

Finally, let us indicate that there appears to be a tendency to move from Kalman filters to particle filters, to better handle nongaussian and nonlinear problems.

Acknowledgements : We thank all the authors who gave us permission to use their figures.

References

[1] J.K. Aggarwal and Q. Cai "Human Motion Analysis : A Review," in Proc. IEEE Nonrigid and Articulated Motion Workshop, 1997.

[2] A. Azarbayerjani and C. Wren, "Real-Time 3D Tracking of the Human Body," in Proc. of Image'com, 1996.

[3] D. Beymer and K. Konolige. "Real-time tracking of multiple people using continous detection," in Proc. International Conference on Computer Vision, 1999.

[4] A. F. Bobick et al. "The KidsRoom: A perceptually-based interactive and immersive story environment," in Teleoperators and Virtual Environment, 1999.

[5] F. Brémond and M. Thonnat, "Tracking multiple nonrigid objects in video sequences," in IEEE Trans. on Circuits and Systems for Video Techniques, 1998.
[6] Q. Cai and J.K Aggarwal, "Tracking Human Motion Using Multiple Cameras," in Proc. International Conference on Pattern Recognition, 1996.
[7] T-H Chang, S. Gong, and E-J Ong, "Tracking multiple people under occlusion using multiple cameras," in Proc. 11th British Machine Vision Conference, 2000.
[8] R.T. Collins, A.J. Lipton, and T. Kanade, "A System for Video Surveillance and Monitoring," in Proc. 8th International Topical Meeting on Robotics and Remote Systems, 1999.
[9] T. Darell, G. Gordon, M. Harville, and J. Woodfill, "Tracking people with integrated stereo, color, and face detection," in Proc. Spring Symposium on Intelligent Environments, 1998.
[10] S.L. Dockstader and A.M. Tekalp, "Multiple camera fusion for multi-object tracking," in Proc. IEEE Workshop on Multi-Object Tracking, 2001.
[11] S. L. Dockstader and A. M. Tekalp, "Tracking multiple objects in the presence of articulated and occluded motion," in Proc. Workshop Human Motion, 2000.
[12] A. Elgammal and L. S. Davis, "Probabilistic framework for segmenting people under occlusion," in Proc. of IEEE 8th International Conference on Computer Vision, 2001.
[13] Th. Frank, M. Haag, H. Kollnig, and H.-H. Nagel, "Characterization of Occlusion Situations Occurring in Real-World Traffic Scenes," in Proc. 4th European Conference on Computer Vision, 1996.
[14] P. Gabriel, J. Verly, and A. Genon, "Video Sequence Analysis: Motion Detection and Object Tracking," in URSI Forum, 2002.
[15] D. Gavrila " The Visual Analysis of Human Movement : A Survey," in Computer Vision and Image Understanding, 1999.
[16] U. Grenander, Y. Chow, and D.M. Keenan "A Pattern Theoretical Study of Biological Shape", Springer-Verlag, New York, 1991.
[17] W.E.L. Grimson, C. Stauffer, R. Romano, and L. Lee, "Using adaptive tracking to classify and monitor activities in a site," in Computer Society Conference on Computer Vision and Pattern Recognition, 1998.
[18] I. Haritaoglu, "A Real Time System for Detection and Tracking of People and Recognizing Their Activities," Phd Proposal, University of Maryland, 1998.
[19] I. Haritaoglu, D. Harwood, and L. Davis, "Hydra : Multiple People Detection and Tracking Using Silhouettes," in Workshop of Video Surveillance, 1999.
[20] S. Intille, J. Davis, and A. Bobick. "Real-time closed-world tracking," in Proc. Computer Society Conference on Computer Vision and Pattern Recognition, 1997.
[21] M. Isard and A. Blake, "Condensation: conditional density propagation for visual tracking," in International Journal of Computer Vision, 1998.
[22] M. Isard and J. MacCormick, "BraMBLe: A Bayesian Multiple-Blob Tracker," in Proc. International Conference on Computer Vision, 2001.
[23] S. Iwase and H. Saito, "Tracking Soccer Player Using Multiple Cameras," in IPSJ SIGNotes Computer Vision and Image Media, 2000.
[24] V. Kettnaker and R. Zabih, "Bayesian MultiCamera Surveillance," in Computer Society Conference on Computer Vision and Pattern Recognition, 1999.
[25] S. Khan and M. Shah, "Tracking People in Presence of Occlusion," in Asian Conference on Computer Vision, 2000.
[26] S. Khan, O. Javed, Z. Rasheed, and M. Shah "Human Tracking in Multiple Cameras" in 8th IEEE International Conference on Computer Vision, 2001.
[27] H.W. Kim, "Automatic Soccer Game Analysis Using Vision Techniques," http://cafe.postech.ac.kr/research/vision/kimtw_1/kimm_research.html
[28] T. Kim, Y. Seo, and K.-S. Hong, "Physics-based 3D position analysis of a soccer ball from monocular image sequences," in International Conference on Computer Vision, 1998.
[29] D. Koller, J. Weber, and J. Malik, "Robust Multiple Car Tracking with Occlusion Reasoning," in European Conference on Computer Vision, 1994.
[30] P. Li, T. Zhang, and A.E.C. Pece, "Visual contour tracking based on particle filters," in Image and Vision Computing, 2003.
[31] J. MacCormick and A. Blake, "A Probabilistic Exclusion Principle for Tracking Multiple Objects," in International Journal of Computer Vision, 2000.
[32] S. McKenna, S. Jabri, Z. Duric, and H. Wechsler, "Tracking Groups of People," in Computer Vision and Image Understanding, 2000.
[33] T. B. Moeslund and E. Granum, "A survey of computer vision-based human motion capture," in Computer Vision and Image Understanding, 2001.
[34] C. Needham and R. Boyle, "Tracking multiple sports players through occlusion, congestion and scale," in British Machine Vision Conference, 2001.
[35] K. Nummiaro, E. Koller-Meier and L. Van Gool, "A Color-based Particle Filter," in First International Workshop on Generative-Model-Based Vision, 2002.
[36] H.W. Ok, Y. Seo, and K.S. Hong, "Multiple Soccer Players Tracking by Condensation with Occlusion Alarm Probability," in International Workshop on Statistical Methods for Vision Processing, 2002.
[37] T. Olson and F. Brill, "Moving Object Detection And Event Recognition Algorithms For Smart Cameras," in Proc. Image Understanding Workshop, 1997.
[38] J. H. Piater and J. L. Crowley, "Multi-modal tracking of interacting targets using Gaussian approximations," in Second IEEE International Workshop on Performance Evaluation of Tracking and Surveillance, 2001.
[39] G. Pingali, Y. Jean and A. Opalach, "Ball Tracking and Virtual Replays for Innovative Tennis Broadcasts," in 15th Int. Conference on Pattern Recognition, 2000.
[40] H.K. Roh and S.W. Lee, "Multiple People Tracking Using a Appearance Model Based on Temporal Color," in International Conference on Pattern Recognition, 2000.
[41] R. Rosales and S. Sclaroff, "Improved tracking of multiple humans with trajectory prediction and occlusion modeling," in Proc. IEEE Workshop on the Interpretation of Visual Motion, 1998.
[42] K. Sato and J. K. Aggarwal, "Recognizing Two-Person Interactions in Outdoor Image Sequences," in Proc. IEEE Workshop on Multi-Object Tracking, 2001.

[43] A.W. Senior, A. Hampapur, L. M. Brown, Y. Tian, S. Pankanti and R. M. Bolle, “Appearance Models for Occlusion Handling,” in 2nd International Workshop on Performance Evaluation of Tracking and Surveillance systems, 2001.
[44] Y. Seo, S. Choi, H. Kim, and K.S. Hong, "Where are the ball and players? Soccer game analysis with color-based tracking and image mosaick," in Proc. of the International Conference on Image Analysis and Processing, 1997.
[45] S. Stillman and T. Tanawongsuwan, “Tracking multiple people with multiple cameras,” in International Conference on Audio- and Video-based Biometric Person Authentication, 1999.
[46] T. Taki, J. Hasegawa, and T. Fukumura, “Development of motion analysis system for quantitative evaluation of teamwork in soccer games,” in Proc. of the International Conference on Image Processing, 1996.
[47] D. Tweed and A. Calway, “Tracking Many Objects Using Subordinated Condensation,” in Proc. of the British Machine Vision Conference, 2002.
[48] A. Utsumi, H. Mori, J. Ohya, and M. Yachida, “Multiple-Human Tracking Using Multiple Cameras,” in 3rd IEEE International Conference on Automatic Face and Gesture Recognition, 1998.
[49] O. Utsumi, K. Miura, I. Ide, S. Sakai, and H. Tanaka, "An object detection method for describing soccer games from video," in Proc. IEEE International Conference on Multimedia and Expo, 2002.
[50] L. Xie, S. Chang, A. Divakaran, and H. Sun, "Structure Analysis of Soccer Video with Hidden Markov Models", in IEEE International Conference on Acoustics, Speech and Signal Processing, 2002.
[51] P. Xu, L. Xie, S.F. Chang, A. Divakaran, A, Vetro, and H. Sun, "Algorithms and system for segmentation and structure analysis in soccer video," in Proc. IEEE International Conference on Multimedia and Expo, 2001.
[52] O. Yoshinori, J. Miura, and S. Yoshiaki, “Tracking Players and Estimation of the 3D Position of a Ball in Soccer Games,” in Proc. International Conference on Pattern Recognition, 2000.
[53] D. Yow, B. Yeo, M. Yeung, and G. Liu, "Analysis and Presentation of Soccer Highlights from Digital Video," in Proc. Asian Conference on Computer Vision, 1995.