MANAGEMENT OF A MULTICAMERA TRACKING SYSTEM

C. Motamed and R. Lherbier
Laboratoire LASL EA 2600
Université du Littoral Côte d'Opale
Bat 2, 50 Rue F.Buisson
62228 Calais France

Keywords: Visual-surveillance, Multi-camera, Tracking, Sensor Management.

Abstract: This work is linked with the context of human activity monitoring and concerns the development of a multi-camera tracking system. Our strategy of sensors combination integrates the contextual suitability of each sensor with respect to the task. The suitability of a sensor, represented as a belief indicator, combines two main criteria. Firstly it is based on the notion of spatial "isolation" of the tracked object with respect to other object and secondly on the notion of "visibility" with respect to the sensor. A centralized filter combines the result of the local tracking estimations (sensor level) and then performs the track management. The main objective of the proposed architecture is to deal with the limitation of each local sensor with respect to the problem of visual occlusion.

1 INTRODUCTION

The development of vision-based system carrying out the monitoring of a global scene is an interesting field of investigation. Indeed, motivations are multiple and concern various domains as the monitoring and surveillance of significant protected sites, the control and estimation of flows (car parks, airports, ports, and motorways). Because of the fast evolution in the fields of data processing, communications and instrumentation, such applications become possible. Such systems are also helpful for automatic analysing of the sport activity.

This work concerns the development of a system for tracking of object from multiple fixed cameras. The tracking results based on the trajectories of objects permit to automatically estimate some specific individual or complex group behaviour. These tools allows a better understanding of athletes or team in order to propose an adapted training sequence or also to choose efficient tactical decisions with respect to a specific player or team. Such systems are useful for many sport activities as football, basketball, tennis etc. Other general utilities of these systems are the annotation for video database and the learning of human behaviour for the industry of simulation and digital games. Research challenges for development of these systems are multiples. From the detection and tracking points of view, difficulties concern: the detection artefact, the changing speed and direction of players and the problem of occlusion.

For wide area monitoring, single visual sensor is not generally sufficient. To maximize the capabilities and performance, it is often necessary to use a variety of sensor devices that complement each other. The approach presented in this paper concerns an approach of high level multi camera management. It is defined as a process that seeks to manage the usage of a suite of camera or measurement devices in a dynamic and uncertain environment, for to improve the performance of the perception.

The basic objective of sensor management may be the automatic tuning of internal parameter of a single sensor or a set of sensor with respect to a given task. A more general problems of multisensor management are, related to decisions about what sensors to use and for which purposes, as well as when, where and how to use them. This last side of high level management is linked with the concept of active perception strategy (Bajcsy, 1988).This strategy is particularly adapted where real-time

Motamed C. and Lherbier R. (2006).
MANAGEMENT OF A MULTICAMERA TRACKING SYSTEM.
In *Proceedings of the Third International Conference on Informatics in Control, Automation and Robotics*, pages 571-576
DOI: 10.5220/0001221105710576

performance is needed such as tracking, robot navigation, surveillance, visual inspection.

The active perception has been widely developed for designing the perception for mobile robotic. These systems and their adaptation capabilities offer a way to face those problems that were not solved adequately with conventional techniques. In fact real-time perception systems have theirs limitation in the computation of massive amount of input data with processing procedures in a reduced and fixed amount of time. This active strategy has the capacity to filter data and to focus the attention of the perception to relevant information or sensors and also can choose the best alternative by using the contextual information. Such proposed approach are clearly linked with the design of cognitive system which permits to combine knowledge and reasoning in order to develop smart and robust perception system.

In order to perform the high level management, it is essential to characterize each sensor with respect to its utility for the task. Generally the main characterization of the sensor is given by the measurement uncertainty. Many sensors have a variable uncertainty with respect to their functioning condition. And generally it is difficult to model the uncertainty globally. Another important characterization feature concerns the field of coverage of the sensor.

Tracking is an important task of computer vision with many applications in surveillance, scene monitoring, navigation, sport scene analysis and video database management. One of the most critical parts of a multi-object-tracking algorithm is the data association step. It has to deal with new objects, short or long term disappearing objects and occlusions.

Qualitative approaches are a, alternative for motion correspondence (Rangarajan, 2001) (Sethi, 90) (Veenman, 2001). These approaches are less normative than conventional statistical methods used in tracking. The main advantage is their flexibility, because they permit an easy integration of several forms of a priori information and contextual information for constraining complex problem. Our approach belongs to the last category. In this work, vision sensors are distributed at different locations of the scene in a redundant and complementary manner (figure 1).

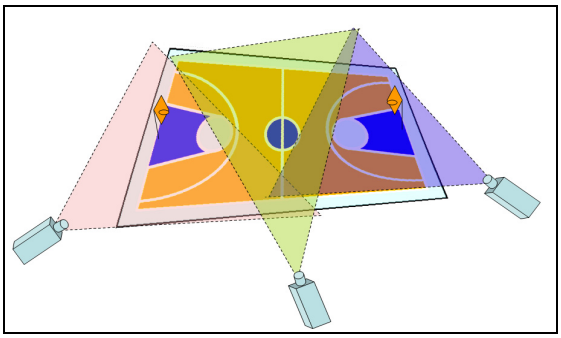

Figure 1: Distributed sensing architecture.

Our strategy of sensor combination integrates contextual information for representing the suitability of each sensor for the task. The suitability of a sensor combines two main criteria. Firstly it is based on the notion of spatial “isolation” with respect to other object and secondly on the notion of “visibility” with respect to the sensor.

Related works linked with the tracking of humans and vehicles with multiple visual sensors are numerous (Foresti,1995) (Utsumi, 1998) (Nakazawa, 1998) (Snidaro, 2004). The main contribution of this work is the manner of how the system controls the sensor information flow in order to improve the quality of the global tracking. The proposed system uses a hierarchical approach presented in the figure 2. A central level tracker uses results of local tracker associated with each sensor.

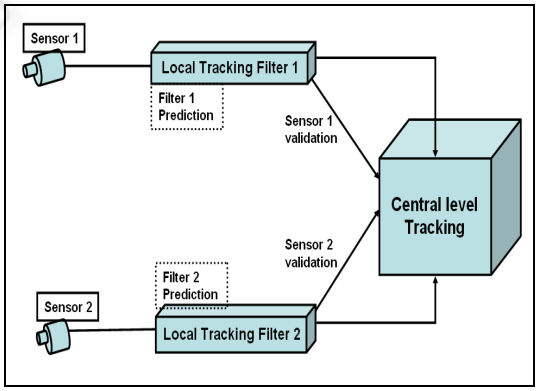

Figure 2: The global architecture of the tracking system.

The section 2 presents the local tracking unit. The Section 3 concerns the central level part of our tracking system. The Section 4 shows some experimental results by using a real multi-camera image sequence.

2 LOCAL VISUAL TRACKING

2.1 Detection and Tracking Strategy

The proposed local tracker basically uses a region-based approach. The tracking algorithms use cinematic and visual constraints for establishing correspondence. In our problem of object tracking for a visual-surveillance application, observations representing the detected regions are complex and are not corrupted by random noise only. They are also affected by detection errors, merging and splitting artefacts, which are difficult to model globally or statistically. In addition, the dynamic model of human activity may present significant variability according to the working context. For all these reasons, we have chosen a qualitative approach, which focuses mainly on data association quality rather than high precision object motion estimation. Assumptions underlying our approach concern several points: objects move smoothly from a frame to another, objects do not change rapidly their appearance colour; objects can interact with other objects or groups of objects. The Tracking algorithm uses globally the Nearest Neighbour (NN) strategy. However, in the presence of merging or splitting situations two specific procedures are launched in order to solve the association ambiguity.

The detection of moving objects is performed by comparing the three RGB components:

$$\begin{aligned} &if \quad \left(\max_{c=R,G,B} \left| I_c^k(P) - R_c^{k-1}(P) \right| > \omega \right) \\ &\rightarrow D^k(P) = 1 \qquad (1) \\ &else \rightarrow D^k(P) = 0 \end{aligned}$$

D^k represents the detection decision (1: moving object, 0: background). The parameters ω is the inter-frame detection threshold and the stability elementary step. The parameter ω is adjusted to twice the median absolute deviation of image noise estimated over a preliminary training sequence (ω= 20 with 8bits/color channel). The raw motion detection result generally presents many artefacts. Two kinds of cleaning procedures are applied. Firstly, we use standard morphological operations of erosion and dilatation for reducing noise in the foreground. Then, too small uninteresting image regions are removed. In our algorithm this size threshold is defined globally in an empirical manner for the entire image sequence.

In order to perform the tracking step, a Kalman filter is attached to each detected object (Position state, and speed as a hidden state), and the information obtained by the filter, is used to predict the position of the object on the next frame. For each detected region, three visual features are computed: centre of gravity, bounding box, and colour histogram. Colour histogram in particular, has many advantages for tracking non-rigid objects as it is robust to partial occlusion, is rotation and scale invariant and is computed efficiently. The histogram is reduced to 32 bins per colour channel in order to reduce the information quantity. However, in the context of multi-player sports based on teams, the colour information will mainly discriminate teams and in a lower manner individual players of the same team. However in the context of the basketball, or foot ball the mutual physical interactions mainly occur between players of opposite teams.

The matching is based on the spatial proximity of regions and their visual compatibilities. The algorithm evaluates explicitly the quality of each association. This information is summarized by a specific belief indicator called “consistency” indicators. The indicator is recursively updated and stored during the tracking step.

The consistency indicator of tracked objects firstly permits effective new objects to be validated after consecutive observations. Secondly, it permits to tolerate some temporary loss of the objects having a reliable track. It reacts as a robust filter at the object level. The indicator increases when no significant variation in the object features (colour and bounding box size and the speed vector) is perceived. Otherwise or in extreme situations when the target is lost, indicator is decreased. The dissimilarity between the object colour histograms is performed by the Bhattacharya distance.

The updating process of the consistency indicator of each tracked object is controlled in terms of fixed time delay. The track termination is decided for lost objects after a defined period of consecutive zero value consistency indicator.

The maintenance of the identity of each object is performed only when the object is present in front of the camera view. The track of the object is also terminated when it goes out of the field.

2.2 Merging and Splitting Procedure

An important difficulty of visual tracking is when objects come to merge from the camera viewpoint. This merging situation is immediately detected by testing the intersection of predicted object bounding box and the size variation of the newly detected region. When a merging situation is detected, firstly a notion of a temporary group of objects is defined in order to track the global region containing visually merged objects. Each group is considered as a new specific entity to track with its indicators. In addition to the temporary group tracking, the algorithm attempts to maintain the track of each individual object inside the global group region. The estimation of position of individual objects during the merging situation is based on their appearance model. We have chosen the mean shift algorithm. This algorithm has been adopted as an efficient technique for appearance-based blob tracking (Comaniciu, 2000). The mean-shift algorithm is a nonparametric statistical method for seeking the nearest mode of a point sample distribution (Cheng, 1995). The dissimilarity between the object model and the object candidates is performed by the Bhattacharya distance. In the normal mode, the algorithm stores continuously the latest sub-image of each object obtained from the motion detection stage. It permits, at the beginning of the merging situation, to get ready the initial object model useful for the mean shift algorithm.

In some situation, an object may be hidden entirely or partially by another one, so that the object can not be located by its appearance. In order to maintain a tracking continuity in all situations for each merged object, the notion of the artificial observation is introduced. It is positioned for each object, on an estimated location P* which is obtained by a weighted combination of the best appearance position Pa and the global group position Pg.

$$P^{*} = P^{a}.w + (1-w).P^{g} \quad (2)$$

The weight w is a normalised distance between the object model and the best location candidate obtained from the appearance approach. In this strategy, when the object cannot be identified by the appearance approach, the group position is favoured.

During merging, the history of each tracked object included in a group (positions and indicators) is continuously updated by means of its artificial observation. This approach of compound observation is close to the PDAF algorithm (Bar-Shalom 1988) which combines the influence of multiple candidates in the validation gate. During the merging situation the consistency indicator of an object is updated only by taking into account the speed stability, estimated thanks to its artificial observation. If during tracking a sole object joins a group, the initial group with updated objects is maintained. The global procedure during the merging situation may be relatively time-consuming. So the decision of group creation is robustly validated once it has been predicted by using the proximity of consistent tracked objects. It permits to tolerate efficiently some fugitive false-detections. On the other hand, this last strategy cannot initialise a group if one of the merged objects is newly created and has a low consistency indicator. In this last case one object is only considered and no group is created.

As in the merging situation, the splitting has to be detected immediately. The Splitting situation is detected once a new object is detected close to a temporary group region. In order to reduce the influence of some detection errors when a real sole object or a group is split by the motion segmentation errors during a short delay, the decision of splitting is provided only when it is confirmed during a fixed time delay (1 second in this implementation).

When the algorithm detects split regions associated to a known group, a specific procedure focuses its attention toward the identity of objects. After splitting, the group updates its individual object. When the group is reduced to a known sole object, the group entity is destroyed.

A visual comparison between objects before merging and after splitting permits to affect the best object identity for each region. When a region considered as a sole object splits, separated detected regions are associated with new objects and inherit the history of the initial object (previous tracks position and consistency indicators).

3 CENTRAL LEVEL TRACKING

From a multi-sensor organization, we have chosen the hierarchical approach. A centralized filter combines the results of the local tracking filters (sensor level) and then performs the track management (Fig. 2). The sensor validation indicator is computed at the predictive location of

the tracked object obtained by the sensor level estimator.

A planar homography transformation HS is performed in order to bring each local estimation from sensor 's' into the same reference plan R0. The position in image concerns the position of player on the ground plane obtained by the lower segment of the object bounding box. The sensor validation information permits to guide the tracking process by a context driven strategy. It controls the combination of local trackers through the central level tracking. The validation is represented by a confidence value for each current tracked object and is based on dynamic contextual information with respect to each sensor. The main objective is to favor the best views with respect to dynamic occlusion of each tracked object. The sensor validation indicator represents a notion of suitability of a sensor for the tracking of a defined object. This validation indicator clearly favor un-occluded tracked object with respect to the camera views.

For each object with respect to each sensor 's', the indicator (λs) is set to its maximum=1 if the tracked object is known as sole object from the local sensor. When the object goes out of the field of the sensor the indicator takes the zeros value. In presence of occlusion, highlighted by the notion of group managed by the local sensor 's', the indicator takes the degree of visibility : ws estimated during the merging procedure of local tracker (section2.1).

The central level tracker based on a Kalman filter uses the result of local trackers (positions P*s and indicators) as its observation. Before filtering, it performs a combination of these observations weighted by their validation indicators in order to generate a fused position Pglobal.

$$P_{global} = \frac{1}{\sum_s \lambda_s} \cdot \sum_s H^S(P_s^*).\lambda_s \qquad (3)$$

4 EXPERIMENTATIONS

We present in this section some experimental results of our multi-camera tracking system. The experimentation is based on a real competition of basket-ball recorded from two distant wide angle views during 20 min. The application permits in particular to estimate the total length of the whole trajectory covered by each team.

The figure 3 illustrates the complementarity of the two views and shows an instantaneous result of local tracking and also the notion of group with respect to each view.

We have study the performance of our multi-camera tracking system with respect to our evaluation video sequence. The evaluation is focused on the data association step with respect to the notion of confusion of identity of players. The comparison is performed between the result of local trackers and the global tracker. In our system the local trackers do not maintain a global identity during the whole sequence. Each object (player) is initialised when it enters in the field of the camera. So, for the local tracker, we count only local errors of association occurring during the presence of the object in the field of the camera. The result of this evaluation from of this experimentation is satisfactory (Table 1). It shows in particular that the majority of occlusion containing the same objects does not occur simultaneously under the two distributed views. Figure 3 illustrates a case of such situation when the same player belongs to a group from each view.

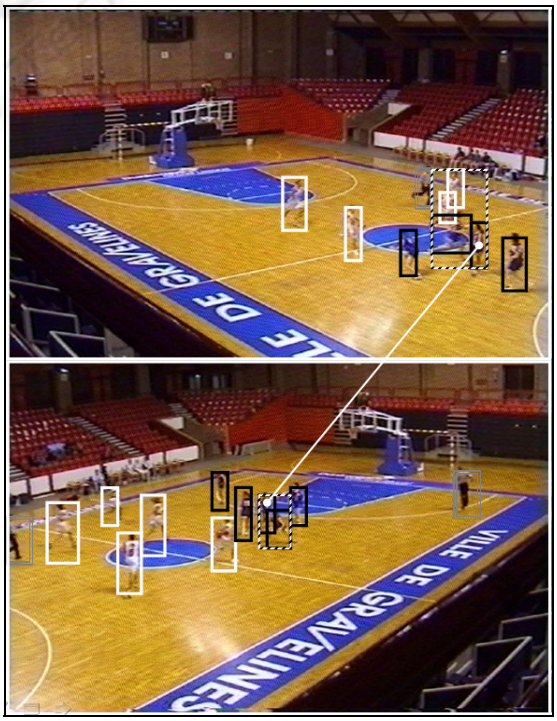

Figure 3: An illustration of local tracking from two views.

Table 1: Performance evaluation with respect to the identity confusion.

Tracking level	Local tracking Left camera	Local tracking Right camera	Central level tracking
Number of object identity confusion	32	26	14
Number of object in group	132	115	--
Number of objects simultaneously appearing in a group from 2 views	31	31	--

5 CONCLUSION

The development of a tracking system from multiple fixed cameras is helpful for many applications in the domain of surveillance or monitoring. Our strategy of sensor combination integrates contextual information for representing the suitability of each sensor. The main advantage of our qualitative motion correspondence and sensor selection is its capacity to cope naturally with known modelled tracking ambiguities. At the local level the tracking takes into account the regions merging, splitting, and target missing. At the central level the tracking solves a good part of occlusions. The use of the sensor validation from local tracker permits in an active manner, to focus the central level, on appropriate information.

ACKNOWLEDGEMENTS

This work is supported by the Nord/.Pas de Calais, French Regional Council and FEDER (European funding for regional development) in the context of the RaVioLi Project.

REFERENCES

Bajcsy R. 1988, Active Perception, *Proceedings of the IEEE,* 76(8):996-1005.

Bar-Shalom Y. & Forthmann T.E, 1988, *Tracking and data association. Academic Press*

Cheng Y. 1995, Mean Shift, Mode Seeking, and Clustering. *IEEE Trans. Pattern Analysis and Machine Intelligence, Vol 17(8*), pp.790-799.

Comaniciu D., Ramesh V, and Meer P. 2000, Real-TimeTracking of Non-Rigid Objects using Mean Shift. *IEEE Computer Vision and Pattern Recognition, Vol II,* , pp.142-149.

Foresti G.L., Murino V., Regazzoni C., and. Vernazza G. 1995,A multilevel fusion approach to object identification in outdoor road scenes. *International Journal of Pattern Recognition and Artificial Intelligence, 9(1):23–65,*.

Nakazawa A., Kato H., and Inokuchi S.. 1998 Human tracking using distributed vision systems. *In Proceedings of the 14th ICPR, pages 593–596,*.

Rangarajan K. and Shah M., 1991. Establishing Motion Correspondence. *CVGIP: Image Understanding*, 54 (1), pp. 56-73.

Sethi I.K. and Jain. R., 1990. Finding trajectories of feature points in a monocular image sequence. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 9 (1), 56-73.

Snidaro L., Luca Foresti G., Niu R., Varshney P. K., 2004. Sensor fusion for video surveillance, *International Conference on Information Fusion 2004, pp. 739-746, Sweden.*

Utsumi A., Mori H., Ohya J., and M. Yachida. 1998, Multiple view-based tracking of multiple humans. *In Proceedings of the 14th ICPR, pages 597–601.*

Veenman C.J. and Reinders M.J.T. and E.Backer 2001. Resolving Motion Correspondence for Densely Moving Points. *IEEE Transactions on PAMI* vol23, pp54-72.