Tracking multiple non-rigid objects in a cluttered scene

François Brémond and Monique Thonnat
INRIA Sophia Antipolis, projet Orion
2004 route des Lucioles BP93 - 06902 Sophia-Antipolis Cedex, France
Francois.Bremond@sophia inria.fr

Abstract

This paper presents a method to track multiple non-rigid objects in a cluttered scene. First, the particularities of non-rigid objects are shown. To cope with them a tracked target is defined as a moving region tracked individually or as a group of moving regions tracked globally. Then we explain how to compute the trajectory of a target and how to compute the correspondences between known targets and moving regions newly detected. In the case of an ambiguous correspondence a compound track is defined to freeze the associations between targets and moving regions until a more accurate information is available. The paper concludes by giving some results on the tracking algorithm implemented for a video-surveillance application.

Key words : Tracking, non-rigid objects, moving regions, cluttered scenes, image interpretation.

1 Introduction

In this paper we present a method to track multiple non-rigid objects in a cluttered scene. The class of applications we are interested in, is the automatic surveillance of indoor and outdoor partially structured scenes with a fixed monocular color camera. Given image sequences of a scene, an interpretation system has to analyze the behavior of real objects. In our case the real objects correspond either to humans or to vehicles. We have developed an interpretation system which is composed of three modules. First, from images the image processing module detects moving regions. Then the tracking

[0]In proc. of the *10th Scandinavian Conference on Image Analysis (SCIA)*, Lappeenranta (Finland), june 1997.

module associates these regions to track targets that correspond to real objects. Finally the scenario recognition module identifies targets as real objects and recognizes the scenarios relative to their behavior. In this paper we focus on the tracking module. The tracking problem is a central issue in scene interpretation because the lost of an object track implies the impossibility of analysing its behavior. The paper presents a tracking method, based on compound tracks that keep and suspend the track of ambiguous targets until their tracking becomes reliable enough. Our goal is to obtain a robust algorithm able to cope with real-time constraints. First, we show how to compute the correspondences between known targets and moving regions newly detected. Then we explain how to solve ambiguous correspondences and we conclude by providing examples on how this method works. Related works can be found in [2], [5], [6], [7].

2 Matching process

The tracking of moving regions is performed in a prediction-correspondence-update loop. At time *t*, the module predicts the next location of targets tracked at time *t-1* and tries to match these predictions with the moving regions detected at time *t*. This section explains how to compute these correspondences.

2.1 Non-rigid mobile objects and moving regions

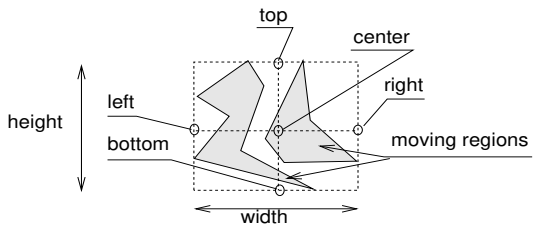

Figure 1: The five generic points (center, top, bottom, left, right) of a group of moving regions.

In our interpretation system the moving regions are detected by the image processing module. As the camera is fixed, this module subtracts the current image with the background image which is regularly updated to compensate for illumination change. Because of noise the tracking module has to face three kinds of problems : it has to fix detection problems (e.g. shadows, reflections, blinking lights), to take care of common tracking problems (e.g. occlusions,

merges of tracks, appearances of targets) and to face problems due to the non-rigid nature of objects (i.e. we cannot use a model of their shape). To face a part of these problems we use contextual information on the static environment of the scene [3]. We have divided the environment into zones, that provide us with clues on targets. For example, an *exit* zone indicates that targets inside this zone can leave the scene. So we call contextual information of a target the contextual information of the zone where it is located. To improve the tracking of non-rigid objects a solution as suggested in [6] and in [2] is to define a moving region as a set of primitives (e.g. corners, contour edges) and to track globally the moving region by combining all its primitives. So the lost of one primitive is corrected by the use of the other primitives. However in cases where images are really noisy, taking into account badly detected primitives can distort the global computation of the track. To less depend on detected primitives we propose to track generic points that represent a moving region (or more generally a group of moving regions). We have choosen five generic points as shown on figure 1. They are tracked separately and the one with the best match is used for the tracking of the global object. This method has two main interests. First, the tracking process is less sensitive to partial occlusion. If a point has a bad match, then it is not taken into account and gives a clue on the presence of a possible occlusion. Second, as generic points are not directly connected with reality they can be defined on any moving region, allowing us to track a moving region without having to know what it represents. Thus we separate the tracking module from the scenario recognition module in charge of recognizing real objects. So for us a target denotes any moving region tracked individually or any group of moving regions tracked globally. A target can either represent a noise (e.g. a shadow), a real object (e.g. a person), a part of a real object (e.g. the head of a person) or a group of real objects (e.g. a crowd).

2.2 Prediction of the target location

The predicted location of a target is computed thanks to its trajectory. In the literature probabilistic methods (e.g Kalman filter or conditional probabilities) are generally used to compute the motion of rigid objects [1], [6], [7]. However we think that these methods do not suit to the tracking of humans because they do not allow radical motion changes. To allow changes we suppose that if a target motion is constant, then it is locally linear, else a new motion has to be considered. Thus the **trajectory of a target** is defined as a set of line segments, the extremities of which are the locations of the target when its direction is changing. So the trajectory gathers an approximation of all past locations of the target. The **location of a target** is computed during the matching process between the target and a moving region detected in the

current frame. The location at time t of every generic point of the target is estimated by extending the target trajectory with the generic point location at time *t-1*. Then, every estimated location is matched with the location of the corresponding generic point of the moving region. The target generic point with the best match is chosen to compute the location of the target center thanks to the target height and width. Thus the trajectory of a target allow us to predict its location.

2.3 Ambiguity matrix

The matching process compares the predicted location of targets with the location of newly detected moving regions through the use of an ambiguity matrix, also called the association matrix. In this matrix the columns represent the targets and the rows represent the moving regions. The matrix elements measure the proximity between targets and moving regions which is defined as followed :

if intersection(moving region, target) $\neq \emptyset$
then proximity = 1 - $\frac{surface\ intersection}{surface\ union}$(moving region, target)
else proximity = 1 + $\frac{distance}{maximal\ size}$(moving region, target)

This proximity definition allows us to balance the distance between moving regions and targets through their size. So the ambiguity matrix gives the number of moving regions that are closed to the predicted location of each target. According to this number we define for a target four states :

- Visible : if the target corresponds to one moving region and the moving region has no other correspondence.
- Lost : if the target does not correspond to any moving region.
- Occluded : if the target does not correspond to any moving region, and if the contextual information of the target indicates the possibility of a static occlusion.
- Ambiguous : if the target corresponds to several moving regions or if the target corresponds to one moving region having other correspondences.

Thanks to the ambiguity matrix all the correspondences are computed. The remaining task of the tracking algorithm is first to update non ambiguous targets and second to solve the correspondences of the ambiguous targets.

3 Correspondence update

3.1 Update of non ambiguous correspondence

The update of a non ambiguous target depends on its state. If the target is visible, its trajectory is extended with the location of the moving region. If the

target has been lost since a short time or if it is occluded, we will try to retrieve its correspondence in the next frames. If the target has been lost for a time long enough and if the context indicates that the target is in an exit zone, then the track is finished and the target is eliminated. If the target has been lost for a time long enough and if it has no particular contextual information, then the target is supposed to be definitively lost and it is eliminated. When all targets are updated, new targets are initialized with the moving regions that have no correspondence with a known target. When all correspondences have been taken care, the targets are processed by the scenario recognition module to try to analyze whether or not targets are real objects. Using a priori knowledge on real objects a **belief coefficient** is established for each target in the framework of fuzzy sets [4]. Then, in a feedback stage the tracking module eliminates all the targets with a belief coefficient not high enough. This feedback stage provides the additional knowledge to eliminate the targets that correspond to detection errors. The method of the belief coefficient can be related to the method of the support of existence [7] which differs on the nature of the additional knowledge. The support of existence relies on tracking features (e.g. the target life time) whereas the belief coefficient relies on the behavior analysis of targets. As this coefficient belongs to the scenario recognition module, its working is not described in this paper.

3.2 Solving ambiguous correspondence

The main task of tracking algorithms is to solve the ambiguous correspondences between targets and moving regions. The way this task is achieved determines the robustness of tracking algorithms. According to the type of ambiguity we define four types of target track :

- Elementary : if a target is non ambiguous, its track is elementary.
- Merge1 : if a moving region corresponds to several ambiguous targets, a temporary target is initialized with the moving region and a track of type merge1 is created. The merge1 track gathers the temporary target and the ambiguous targets.
- Split : if several moving regions correspond to one ambiguous target, temporary targets are initialized for each moving region and a track of type split is created. The split track gathers the temporary targets and the ambiguous target.
- Merge2 : if several moving regions correspond to several ambiguous targets, temporary targets are initialized for each moving region and a track of type merge2 is created. The merge2 track gathers the temporary targets and the ambiguous targets.

We call **compound track** any track which type is not elementary. A compound track models an ambiguity. It tracks globally a set of temporary targets

while maintaining a set of ambiguous targets. Temporary targets represent the merging and newly detected moving regions, whereas ambiguous targets represent the historic of the temporary targets before the ambiguity has been met. The goal of the tracking algorithm is to associate the ambiguous targets with the temporary ones as soon as more accurate information is available. Meanwhile the compound track is used to freeze the associations.

The tracking algorithm is able to solve the ambiguity when a temporary target of the compound track moves away from the other temporary targets. In this case the tracking algorithm tries to find the ambiguous target that corresponds to the temporary target moving away. There are two mechanisms to find an ambiguous target according to the arrangement of ambiguous targets in the compound track. The first mechanism is used when ambiguous targets verify the *rigidity* condition. If ambiguous targets have the same parallel motion, then these targets are supposed to keep their spatial ordering inside the compound track. So the temporary target moving away is associated with the ambiguous target with the same ordering number. The second mechanism is used when ambiguous targets have distinct motions. In this case a local matrix of ambiguity is used to compute the correspondences between ambiguous and temporary targets. If these arrangements are not verified, the temporary target moving away is not associated to any ambiguous target. With these mechanisms an ambiguous target can be retrieved and the corresponding ambiguous correspondence can be solved without approximation.

When the compound track modeling an ambiguity is facing a second ambiguity, the tracking algorithm has to solve the first ambiguity to avoid an exponential explosion. So a second compound track has to be built to model the second ambiguity and the ambiguous targets of the first compound track have to be transformed into ambiguous targets of the second compound track. This transformation consists in retrieving the missing locations of the ambiguous targets during the time interval between the occurrence of the two ambiguities. In this goal the missing locations are approximated by the locations of the first compound track. The accuracy of the approximation depends on the type of the first compound track. With the split type there is no approximation because the split track contains only one ambiguous target. The merge1 type gives a better approximation than the merge2 type, because in a merge1 track the targets are closer to each other. Therefore all ambiguities are solved with more and less approximation.

The method of the compound tracks can be related to the *Multiple Hypothesis Tracking* method [5] and to the *beam search* method [7]. Unlike the MHT the compound track method does not keep all the past information on the correspondences but combines and approximates them. The compound track method differs also from the beam search method in that the compound track method does not track incoherent targets. It solves the ambiguity as late as

possible. Thus the compound track method keeps less information that the two other methods but approximates the computation. The comparative results depend on the frequency of the approximation.

4 Results

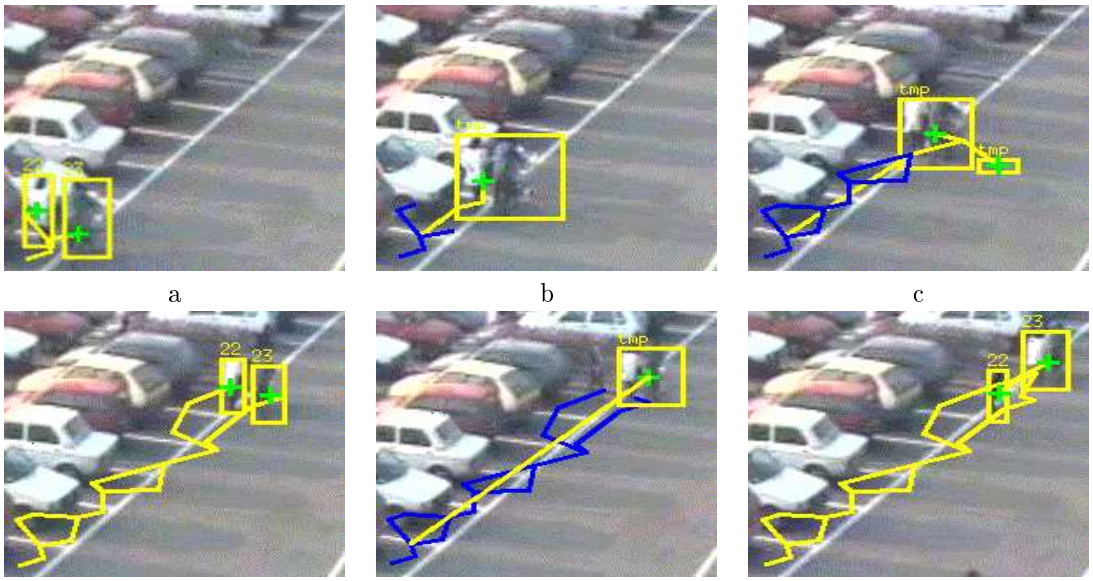

Figure 2: This figure shows two persons being tracked. On frames a and d they are tracked as individual, whereas on the other frames they are tracked as a group. The figure shows also the trajectory of the group and of each person.

In the framework of a video-surveillance application we have developed a system implementing the tracking algorithm described in this paper. Only the part using the contextual information has not yet be linked to the global system. We replace it temporary by providing directly the system with the suitable information. To validate our system we use image sequences taken in the framework of the european Esprit PASSWORDS project. For a color image frame at 512x512 resolution the average time of the global processing is 2.5 seconds on a Sun Spark station SS-10. The success of the tracking algorithm depends on the quality of the image sequence. With little noise the tracking algorithm is able to track correctly real objects. For example figure 2 shows two persons (numbered 22 and 23) walking together splitting (frame a) and merging

(frame b) along their way. Despite the shadows that add noise (frame c) the two persons are still tracked (frame d). Frame e shows that thanks to compound tracks the tracks of the group and of each person are not lost. However on frame f the target (numbered 22) is mislabeled because a third person is going out from cars (he was not detected on frame e) and simultaneously is merging with the two first persons.

5 Conclusion

This article presents a tracking method, called compound tracks, the goals of which are to not lose any target and to work in real-time. This method consists in freezing the association of ambiguous targets through compound tracks until more accurate information is available.

The results are satisfactory besides that in some cases targets are still lost. There are many ways to improve the different stages of the tracking algorithm. But as the main source of errors is the detection of moving regions we think that these improvements have to be done in the same time on image processing. Our future works will consist in the addition of contextual information to try to fix the errors of moving region detection.

References

[1] Y. Bar-Shalom and T. Fortmann. *Tracking and data association.* Academic press, London, 1988.

[2] A. Baumberg and D. Hogg. An adaptive eigenshape model. In *British Machine Vision Conference*, Birmingham, September 1995.

[3] F. Brémond and M. Thonnat. A context representation for surveillance systems. In *Proc. of Workshop on Conceptual Descriptions from Images at the European Conference on Computer Vision (ECCV)*, Cambridge, April 1996.

[4] F. Brémond and M. Thonnat. Interprétation de séquences d'images et incertitude. In *Proc. of the Rencontres sur la Logique Floue et ses Applications (LFA)*, Nancy, December 1996.

[5] I. Cox and S. Hingorani. An efficient implementation of reid's Multiple Hypothesis Tracking algorithm and its evaluation for the purpose of visual tracking. In *IEEE Transactions on pattern analysis and machine intelligence*, volume 18, September 1996.

[6] F. Meyer and P. Bouthemy. Region-based tracking in an image sequence. In *Proc. of European Conference on Computer Vision*, pages 476–484, May 1992.

[7] Z. Zhang. Token tracking in a cluttered scene. Research report 2072, I.N.R.I.A., Sophia Antipolis, October 1993.