Motion based Segmentation to improve Tracking of Non Rigid Objects

Elodie BICHOT, Laurent MASCARILLA and Pierre COURTELLEMONT

Laboratoire d’Informatique – Image – Interaction, University of La Rochelle
Av. Michel Crépeau, 17042 La Rochelle Cedex, FRANCE
{elodie.bichot, laurent.mascarilla, pierre.courtellemont}@univ-lr.fr

Abstract: In this paper, we focus on the problem of tracking non rigid objects in cluttered environment. We propose a tracking scheme in three steps. Firstly, CONDENSATION method enables us to track objects in presence of partial occlusions thanks to multiple hypotheses handled. Secondly, a motion segmentation is performed in the neighbourhood of the target to determine number and regions of motion clusters and to, finally, update target model if the confidence in the learned model is sufficient. Our main contributions concern definition of dissimilarity measure to compare geometric motion vectors and application of fuzzy relational clustering to motion segmentation problem. This motion segmentation embedded in a tracking procedure allows us to distinguish motion blobs and to disambiguate complex situations such as occlusions. Applications on fish tank sequences show the relevance of this technique.

Keywords: motion segmentation, tracking, Condensation, fuzzy relational clustering

1 Introduction

Tracking has a wide range of applications. It was largely applied to surveillance domain and today, new applications such as behaviour interpretation or motion study for learning appear. Our application consists in tracking fishes in order to improve their recognition. This work represents a part of the Aqu@thèque project which is an automatic fish recognition system in interactive live videos. Details are given in (Semani & al., 2002). In a few words, it is the design of an interactive interface showing in live a fish tank sequence and on which the user selects a fish to learn general information about its species. The recognition system is composed of three steps: a segmentation step to extract regions corresponding to fishes, a feature extraction step to characterize each region and a classification step to identify the fish of interest. Recognition system may not be able to acknowledge the selected fish in particular conditions. For example, if the fish is partially occluded by another fish, the recognition system considers the region defined by the two fishes and so, the recognition may fail. Hence, the tracking of the fish of interest during and after the occlusion is necessary to enable its recognition. Another example occurs when a fish undergoes a strong illumination for a short time because of lighting conditions of the fish tank. Recognition system may not identify it. A robust tracking keeps track of fish until the fish is subject to a normal lighting and the system recognition understands it.

The tracking application is a challenging task because of the problems induced. On the one hand, we have to choose some features to track and to cope with variations in appearance of fishes caused by changes in illumination or in orientation. These variations concern color, size and shape of objects. On the other hand, the complex motions of target and the background clutter due to illumination conditions and water motion are other difficulties that we have to tackle. Presence of occlusions may also disturb tracking algorithm. To overcome these problems, we suggest an association between motion segmentation and tracking: motion segmentation provides target model to the tracking module. Motion segmentation consists in estimating motion vectors and in grouping together similar vectors. For this purpose, we also propose an original application of fuzzy relational clustering which consists in determining medoids, i.e the best representative data among the input data set. It requires the definition of a dissimilarity measure between geometric motion vectors. This latter takes into account all information namely norm and direction provided by vectors and permits to cluster them. This segmentation allows us to distinguish the occluder and the occluded object.

Segmentation and tracking have already been associated and we describe this combination in section 2. We also introduce the drift problem of tracking. Then, we present our approach to tackle this problem in section 3. It includes Condensation method and an original motion segmentation technique relying on fuzzy relational clustering. Section 4 follows with some results and finally, we conclude in section 5.

2 Background

Segmentation and tracking are two main problems in computer vision. In this section, we propose a non exhaustive description of segmentation and tracking association. Tra-
ditionally, segmentation is used to group pixels with similar properties such as color or motion into blobs. Then, tracking procedure focuses on these blobs and estimates their location.

For example, Heisele performs in (Heisele, 2000) a color segmentation followed by a merging procedure on each color cluster with similar motion. A Kalman filter is integrated to predict color cluster locations in the next image which initializes clustering procedure. In other words, this approach consists in estimating color partition thanks to a prediction given by Kalman filter and in performing motion segmentation on this first partition by grouping color clusters with parallel trajectories. It preserves motion discontinuities which coincide with color region boundaries. It is applied to traffic scenes with cars and pedestrians.

Another example is the approach proposed by Khan and Shah (Khan & Shah, 2000). They use E.M algorithm to segment detected people in regions of similar colors. Then, they track these regions by performing a frame to frame matching technique. In this case, segmentation is the initialisation step of the tracking which operates on color clusters. This method is applied to the tracking of people which is a difficult task because of non rigidity of targets and presence of occlusions. Moreover, the tracking scheme presented in this paper that is, detection of moving regions and frame to frame matching of these regions, is frequently encountered in the literature as a tracking method (Lipton & al. ., 1998) (Park & Aggarwal, 2002).

Segmentation and tracking have also been combined in (Raja & al. ., 1998) in order to distinguish between background and foreground entities. Background and foreground objects are modeled by their colors by means of Gaussians mixture whose parameters are determined by E.M algorithm. These models allow pixels to be labeled as background or foreground objects. Then, tracking is performed by looking for object model in a searching area. This method processes in real time but relies only on color of objects and hence, can not cope with occlusions.

Hence, traditionally, segmentation is made to detect object of interest to initialize tracking process. We propose a segmentation to understand complex situations such as occlusions and to enable the update of target model only from target region in order to avoid drift of tracking described in (Matthews & al. ., 2003) (Nguyen & al. ., 2001).

This event is the distraction of tracking from target to focus on other objects or on background. It occurs when tracking relies on appearance model of target and when the update of this model is made naively. Many solutions have already been proposed to tackle this problem. In (Nguyen & al. ., 2001), Nguyen and al. use Kalman filter to provide an optimal and adaptive model of target, which allows the estimate of target position in the scene by a matching procedure. Occlusions are detected in order to not update target model when they occur. Nguyen and Smeluders integrate photometric features in the appearance model making the method robust to changes in illumination (Nguyen & Smeluders, 2002). In (Matthews & al. ., 2003), Matthews and al. suggest to update target model every frame but correct the drift errors in aligning current model on a reference learned at the first time. Thus, they are able to track cars in outdoor sequence characterized by strong changes in illumination. At last, McKenna, Raja and Gong use observed log-likelihood measurements to detect occlusions and not update target color model in these erroneous images (McKenna & al. ., 1998). They reinitialize this model when the occlusion is at end, i.e when likelihood measurements is more important.

We rely on the idea that segmentation is used to initialize tracking to suggest a local motion segmentation providing target appearance model to the tracking procedure.

3 Motion Segmentation to improve Tracking

3.1 Overview

We propose a cooperation between segmentation and tracking to achieve a robust tracking of non rigid objects whose appearance changes with illumination and orientation. We suggest the use of a segmentation map to provide target model to the tracking procedure. Our approach is composed of three modules: a tracking module, a segmentation module and an update module.

The aim of tracking module is to estimate target’s state and particularly target’s location in the current frame. The input of this module is the appearance model of the target obtained from the update module at a previous image. Then, a motion segmentation performed at the neighborhood of the target determines number of motion clusters and spatial support of each cluster. If number of clusters indicates presence of background and target, we can update target model from the corresponding support. If number of clusters shows presence of additional motion regions, it informs on occurrence of occlusions and we will not update target model. We provide the chosen model to the tracking module for the next iteration. The proposed framework is summarized in figure 1.

This framework supposes at time $t + 1$ target appearance does not differ too much from the appearance at time $t$. This cooperation between segmentation and tracking is necessary. Indeed, if target model is not updated, tracking would fail after only a few iterations because target appearance evolves along time according to illumination or orientation and target model does not reflect these changes. Systematic adaptation of tracker is dangerous since the tracker could focus on other objects or on background. Our cooperation tracking-segmentation will prevent tracker to adapt on false targets and to focus on them by distinguishing target pixels from others. It will be a new solution to the drift problem. After this overview, we will now describe tracking and segmentation module separately.


3.2 Tracking module

Tracking is performed by particle filtering and precisely the Condensation method introduced in computer vision by Isard and Blake (Isard & Blake, 1998). Condensation approach evaluates posterior probability distribution on configuration object conditionally to all measurements. It consists in propagating conditional probability density on object state through time. Each density is approximated by a weighted sum of Dirac functions centred on each particles. Particle represents a hypothesized state for object and associated weight the probability of this state. The set of weighted particles is propagated in three steps to explore configuration space: prediction, correction and resampling. Prediction relies on dynamic model to spread particles in state space. Then, correction assigns a weight to each particle depending on likelihood measurements conditionally to predicted state. Resampling step selects particles proportionally to their weights: it eliminates particles with low weights and duplicates particles with high weights to keep the most likely hypotheses. Finally, weighted mean of particles gives an estimation of target’s state. This approach is chosen for our application. Indeed, presence of background clutter on fish tank sequences may disturb tracking algorithm from its target and thanks to multiple hypotheses handled, particle filtering methods are robust to this clutter. Besides, they are able to maintain tracking during short-lived occlusions.

We have to define state space, dynamical and observation models to apply this method. State is defined by the smallest rectangular containing all target pixels and is called "bounding box". State vector is

$$X_t = (x_t, y_t, \dot{x}_t, \dot{y}_t, w_t, h_t)$$

where \((x_t, y_t)\) and \((\dot{x}_t, \dot{y}_t)\) are respectively the rectangular’s center location and velocity and \(w_t\) and \(h_t\) represent rectangular’s width and height at time \(t\). Rectangular window is chosen for its simplicity but not really adapted to the shape of objects to track.

The dynamical model relies on positions and velocities estimated in the previous frames. Velocity is updated according to target estimated positions in the previous frames. Finally, motion model is given by

$$x_{t+1} = x_t + \dot{x}_t + \sigma \omega_t$$  \(1\)

$$\dot{x}_{t+1} = \dot{x}_t - x_{t-1}$$  \(2\)

and the same equations for \(y\) component. \(\sigma\) is the standard deviation of dynamical noise and \(\omega\) is unit zero-mean Gaussian noise. At each time step, we consider three size hypotheses namely, same scale or \(\pm10\%\), width and height varying independently to take different target sizes into account. This target scaling is first applied in (Comaniciu & al., 2000) who suggest mean shift algorithm which is a tracking procedure for non rigid objects relying on color distribution.

Observation model is based on color information as in (Nummiaro & al., 2003)(Pérez & al., 2002). Target model and candidate targets are characterized by their color distributions by means of histograms. We compare these distributions using Bhattacharyya distance. Histograms are evaluated in RGB space using \(8 \times 8 \times 8\) bins. Target model is a combination of histogram evaluated on frame 1 and histogram evaluated on previous image at each time step to take appearance variations into account.

Finally, weighted mean of particles is performed at each time step to evaluate target’s state.

The Condensation algorithm used is described in (Isard & Blake, 1998).

Tracking results obtained from Condensation method are presented on figure 2. Target estimated trajectory is represented on each frame. Target location is correctly evaluated despite occlusions of target by a rock occurring on frame 21 and by another fish on frame 60. Target size evolves along time according to fish orientation and coincide with fish size. Fish of interest undergoes a strong illumination from frame 163 to frame 173 which induced mistakes in estimation of location and size. However, thanks to multiple hypothesis handled by Condensation approach, track is maintained by some particles and target is recovered once the fish is under a normal illumination (frame 181 to the end). Our tracking scheme incorporating an update of target appearance from segmentation map is a solution to cope with variations due to illumination or orientation.

Figure 1: Proposed tracking scheme

\[\text{Tracking} \rightarrow \text{target’s estimated position} \]

\[\text{Local Motion Segmentation} \rightarrow \text{number of motion clusters} \]

\[\text{Spatial Support of motion clusters} \rightarrow t \rightarrow t + 1 \]

\[\text{Update?} \rightarrow \text{target’s model} \]
Figure 2: Frames from Aqu@thêque sequence. Tracking results obtained from Condensation method. Target estimated trajectory from starting location to current location is shown by successive red crosses.
3.3 Motion segmentation

The aim of segmentation module is to give informations about the environment of the target and occlusions. In this latter case, it also informs about the identity of the occluder object.

3.3.1 Background

Motion segmentation consists in decomposing each frame of an image sequence into independently moving objects. There are two kinds of approaches: motion based method and regions from a static segmentation merging method. Motion based approaches group together pixels with coherent motion. In (J. Wang & Adelson, 1994), Wang and Adelson generate hypotheses which consist in affine motion parameters estimated on arbitrary regions by means of least squares method. Then, they merge regions by employing k-means clustering algorithm in the affine parameter space. The main drawback is the inaccuracy in boundary estimation. An example of regions merging approach is the method proposed by Dufaux and al. in (Dufaux & al., 1995). The authors perform a merging procedure on regions obtained from a static segmentation by applying k-medoids method in affine motion parameter space. These methods make an assumption on the model of motion. In (Bergen & Meyer, 1997) motion segmentation is made by successive merging of regions with similar motions if the quality criterion induced by the merged region is comparable with this from each single region. A similarity criterion which is in accordance with quality measure is introduced. Recently, Wang and al. associate in (Y. Wang & al., 2003) motion based method and static regions merging method: they suggest to use a gray level segmentation followed by a merge or split stage according to motion information. They represent interactions between motion vector field, intensity segmentation and object segmentation field through bayesian network to estimate these three fields simultaneously. They estimate accurately boundaries thanks to gray level segmentation and as the regions can merge, this approach does not lead to an over segmentation.

We propose a motion based technique which makes no assumption on the model of motion and works directly on motion vectors to group them according to their similarity. Our similarity measure is based on norm and direction of the geometric vectors.

3.3.2 Motion estimation

Our approach is performed in two steps: estimation of optical flow and grouping by means of fuzzy relational clustering. First, optical flow is evaluated by matching technique on color images. This method is adapted for non rigid entities. However, it induces erroneous motion vectors if there is no texture information available in regions to match. This method is also computationally expensive. That is why we only perform a local segmentation in the vicinity of the target. As in our application the objects of interest are non rigid objects, motion vectors are not necessarily uniform which make the partitioning task very complex.

3.3.3 Grouping motion vectors by applying fuzzy relational clustering

A relational clustering approach and precisely, the Fuzzy c-Medoids algorithm or FCMdd proposed in (Krishnapuram & al., 2001) by Krishnapuram and al. for Web Mining applications is applied on motion vectors to group them in coherent regions. It handles relational data for which we only have a degree to link objects of pairs of data. In contrast to object clustering, relational approach is applied when object can not be described by numerical vectors. In our case, this relational approach is justified because of lack of distance measure to compare two vectors. Euclidean distance between two vectors represents distance between heads of vectors when they are drawn to the same origin. This information does not take into account direction and norm of vectors which both represent the geometrical meaning of vectors. Consequently, Euclidean distance is not judicious for our application.

We propose a similarity measure depending on our application. A similarity degree represents the link between two objects of a group. The link is as strong as the similarity value is important. A similarity measure is a numerical application R between two objects x_i and x_j:

R: Ω × Ω → R^+ (3)

R: (x_i, x_j) → R(x_i, x_j) (4)

which verifies:

1. ∀(x_i, x_j) ∈ Ω × Ω, R(x_i, x_j) = R(x_j, x_i)
2. ∀(x_i, x_j) ∈ Ω × Ω, x_i ≠ x_j, R(x_i, x_i) ≥ R(x_i, x_j)

It is less restricting than a distance because it does not check triangular inequality. To define our similarity measure, we rely on the definition of equality between geometric vectors: two vectors are equal if they have the same direction and norm. So, vectors of significant norms are similar if the following criteria are satisfied:

1. the angle they form is roughly equal to zero,
2. their norms are roughly the same.

We must take into account the special case of two vectors of small (i.e. insignificant) norms. Because of noise and image resolution, orientations of small vectors are often poorly estimated by optical flow methods and then not reliable. Consequently, it is quite natural to consider only their norms, so that they are defined as similar. Practically, these vectors may concern in most cases background of image. According to those criteria, we propose the similarity measure given in algorithm 1.

This similarity measure evaluates degree of resemblance according to norm and direction of vectors. Besides, this general framework will allow us to integrate multiple cues like color or texture to enforce the degree of resemblance between pair of data.
FCMdd consists in selecting \( c \) representative objects called Medoids from the data set. The medoids minimize the total fuzzy dissimilarity within each cluster thanks to an objective function

\[
J_m(V;X) = \sum_{j=1}^{n} \sum_{i=1}^{c} \mu_{ij}^m r(x_i;v_j)
\]

where:

- \( X = \{x_i | i = 1,2,\ldots,n\} \) is the set of \( n \) objects we have to classify.
- \( V = \{v_1,v_2,\ldots,v_c\}, v_i \in X \) is a \( c \)-subset of \( X \). It is the set of medoids.
- \( r(x_i, x_j) \) represents the dissimilarity between object \( x_i \) and object \( x_j \):
  \[
r(x_i, x_j) = 1 - R(x_i, x_j)
\]
- \( \mu_{ij} \) the fuzzy membership of \( x_j \) in cluster \( i \)
- and \( m \) the fuzzy degree, which is set to 1.2 in our experiments.

We use the FCM membership given by

\[
\mu_{ij} = \left( \frac{1}{r(x_i, v_j)} \right)^{1/(m-1)} / \sum_{k=1}^{c} \left( \frac{1}{r(x_i, v_k)} \right)^{1/(m-1)}
\]

The Fuzzy \( c \)-Medoid algorithm is described in algorithm 2 and we refer to (Krishnapuram & al., 2001) for details.

We choose FCMdd preferably to other relational clustering methods such as the Relational Fuzzy C-Means (RFCM) (Hathaway & al., 1989) because of its better efficiency. Complexity of FCMdd is \( O(n^2) \) in worst case where \( n \) is the number of data to classify but Krishnapuram and al. also propose in (Krishnapuram & al., 2001) a linear version. Moreover, by selecting medoids from the input data set it guarantees that cluster centers lie inside image objects even if they are not convex (as often are fishes tracked in our application). Means based algorithms clearly do not provide such a property. Test results show that it converges in a reduced number of iterations (in most cases less than 5 iterations). As for all relational clustering algorithm, the main drawback of FCMdd is the memory space required to store the dissimilarity matrix \( R \). However, in subsampling motion vectors, we overcome this difficulty without practical drawbacks.

Once FCMdd algorithm applied, we obtain a partition of motion vectors which enables distinction between occluder and occluded object.

\section{Experiments}

We have tested the proposed approach on fish tank sequences. Initialization of tracking is performed manually according to the application of the Aqu@thèque project in which the user selects the fish of interest by pointing out it on a touch screen (Semani & al., 2002). Standard deviation of dynamical noise is fixed to 5 pixels and of observation noise 0.1. Number of particles is 100.

We have focused on the fish located in the middle of the frame 103 on figure 3 first row and moving from right to left. At the beginning of the sequence, the fish of interest is partially occluded by another fish until frame 108. After that, it moves alone from right to left and from frame 118 to the end, it is occluded once again as can be seen on figure 3 last row. Target’s state which consists in target’s position and size is correctly estimated and tracking is maintained during the occlusions on frames 103 and 121. Motion segmentation is performed in the vicinity of the target. As results of clustering depends on initialization, we use medoids obtained at current image to initialize medoids.
Figure 3: Frames 103 and 121 from Aqu@thèque sequence. Left: results of tracking. Middle: local optic flows. Right: results of local motion segmentation. In left frame, white rectangular represents estimated target’s state and red square estimated target’s location. In right image, target motion blob is represented by red crosses and occluder object by blue squares. Background is shown by dots.

Figure 4: Frames 134,140,156 from Aqu@thèque sequence.
in the next image. This segmentation allows us to distinguish different motion blobs as can be seen on right images. The fish of interest is divided in two parts because of an occlusion by another fish on frame 121. The two parts of the occluded fish belong to the same motion cluster in accordance with optic flow and reality. Occluder fish and background correspond to the other clusters. An analogous event occurs on the second processed sequence. One difficulty in motion segmentation is to determine number of clusters. We introduce a prior knowledge: one cluster corresponds to the image background, a second cluster represents target and if the target is occluded, a third cluster is present.

Results are presented for another target on figure 4. The target is occluded by two other fishes from frame 120 to frame 150. The fishes belong to the same species and have similar color distribution. Hence, motion information is the only one which permits to distinguish the fishes. Tracking is maintained as shown on left images on frames 134, 140 and 156. Motion segmentation gives also satisfying results according to estimated motion vectors: we are able to distinguish the target and the occluder object as can be seen on right images. Clustering on motion vectors converges rapidly thanks to the temporal consistence introduced in the initialization step of medoids.

5 Conclusion

We propose a cooperation between segmentation and tracking. Condensation method is applied to track fishes thanks to target model provided by a segmentation step. This cooperation will allow us to avoid drift problem of the tracking procedure induced by the update of target model. Motion segmentation is made by estimation of motion vectors and grouping together similar vectors by means of fuzzy relational clustering. This segmentation performed on fish tank sequence helps to disambiguate complex situations such as occlusions and to distinguish occluded and occluder entity.

Segmentation is performed by estimation of motion vectors and the original application of fuzzy relational clustering to cluster coherent vectors in accordance with norm and direction. The dissimilarity measure reflects all the meaning given by motion vectors. The other interest of the dissimilarity measure is that it allows us to take into account multiple cues to compare relational data. Thus, we investigate the integration of color and texture information in dissimilarity measure to improve results and to build approach relying on both color, texture and motion information. We are also working on the determination of number of motion clusters to finish our implementation of tracking scheme.

References


Wang, Y., Tan, T., & Loe, K.-F. (2003). Video segmenta-