Segmentation based on region-tracking in image sequences for traffic monitoring

Jorge Badenas and Filiberto Pla
Dept. Informática, Universitat Jaume I, Castellón 12071 (SPAIN)
(badenas,pla)@inf.uji.es

Abstract

This paper [1] presents an algorithm for segmenting and tracking moving objects in a scene. Temporal information provided by a region tracking strategy is integrated for improving frame to frame motion segmentation. The method has been applied to a traffic monitoring system and it provides facilities such as estimating trajectories of vehicles, detecting stopped vehicles and estimating the mean velocity of the traffic.

1 Introduction

The method presented here is part of a traffic monitoring system. It carries out surveillance of the vehicles trajectories and vehicles shape segmentation by using the information provided by tracking. We can find in the literature several trends and tracking strategies. Most of these tracking methods can be divided into four groups: optic flow methods [3], feature-based methods [6], [7], model-based methods [8],[4], [2] and region-based methods [5].

Our approach is based on a regions tracking algorithm since, in traffic scenes, this type of methods present advantages with respect to the rest. Region-based methods define groups of connected pixels that are detected as belonging to a single object that is moving with a different motion from its neighbouring regions. Regions tracking is less sensitive to occlusion due to the extensive information that regions supply. Characteristics such as size, shape or intensity can be directly obtained from them. Moreover, regions are very suitable for scenes with a stationary background(road scenes), since motion estimated for regions, allows to separate moving regions from stationary regions.

The rest of the paper has the following structure. The next section describes the method proposed. Section 3 explains some of the facilities that the method provides to traffic monitoring. Section 4 shows results, and section 5 presents conclusions.

[1]This work was partially supported by the projects ESPRIT PROJECT EP-21007 IOTA and CICYT TIC95-0676-C02-01

2 Region segmentation and tracking

For every frame of the sequence, our method carries out a frame to frame motion segmentation that is followed by a matching process. This process matches regions segmented in previous frames to regions segmented in the current frame. Matching allows the system to know the evolution of the regions. This evolution is stored in a *master list*, where the regions segmented in the current frame that are not matched to regions segmented in previous frames are inserted.

Frame to frame motion segmentation (F.F.M. segmentation)

This process carries out motion segmentation for every frame of the sequence. It consists of three steps: segmentation based on grey levels, motion estimation for each resulting region, and motion segmentation. The first step divides the image into regions of similar intensity. The second step estimates translational motion for each region, and third step merges adjacent regions with coherent motion parameters. A complete description of this step can be found in [1].

Matching regions Tracking is based on matching regions from the master list to regions segmented in the current F.F.M. segmentation. Matching is an iterative procedure which repeats two steps:

1. *Matching.* Each master list region is matched to the most similar region considering five features: x and y region means, intensity mean and motion parameters. These features define a vector $V = (m_x, m_y, m_z, v_x, v_y)$ for each region that is used for looking for that region that minimizes a weighted squared Euclidean distance measure:

$$D^{mn} = (V^m - V^n)^t W (V^m - V^n) \qquad (1)$$

where V^m is the features vector of the master list region R_m, V^n is the vector of the new region R_n, and W is a 5x5 diagonal matrix that contains a weight for each feature $(w_x, w_y, w_z, w_{v_x}, w_{v_y})$.

2. *Updating weights.* For every iteration, the weights are recalculated in order to minimize the total sum of the distances D_n^m, where R_m and R_n are two corresponding regions, and K^m is the number of regions that we are matching.

$$\min_{w_x, w_y, w_z, w_{v_x}, w_{v_y}} D_T = \sum_{m=1}^{K^m} D^{mn} \qquad (2)$$

To avoid the uncontrolled increasing of weight values the constraint $w_x w_y w_z w_{v_x} w_{v_y} = 1$ is introduced and used for calculating weights by applying Lagrange Multipliers.

$$D_T = \sum_{m=1}^{K^m} D^{mn} + \lambda(w_x w_y w_z w_{v_x} w_{v_y} - 1) \qquad (3)$$

These two steps of the matching process are repeated until a minimum is reached for D_T.

Accumulating evidence Every time a region R_m is inserted in the master list, a matrix M^m of image size is created. Each matrix cell i, j contains two values ($M_{ij}^m.v$ and $M_{ij}^m.o$) that are initialized to zero for all the cells except for those ones that correspond to pixels that belong to the region. In this case the value $Mij^m.v$ is initialized to one. $M_{ij}^m.v$ accumulates the number of frames in which a pixel has belonged to the region R_m. $M_{ij}^m.o$ stores the number of continuous frames in which the pixel (i,j) has not belonged to the region R_m.

After every matching between the master regions and the regions generated by the F.F.M. segmentation, each matrix M^m of each master region R_m that has a corresponding region R_k is updated in the next way:

```
∀ I_ij ∈ Image
  if I_ij ∈ R_k then
     if M^m_ij.v ≤ MAX_V then
        M^m_ij.v = M^m_ij.v + 1
     M^m_ij.o=0
  else
     M^m_ij.o=M^m_ij.o + 1
     if M^m_ij.o ≥ MAX_O then
        M^m_ij.v=0
        M^m_ij.o=0
```

MAX_V and MAX_O (0<MAX_O≤MAX_V≤7) are two constants that define the maxima that can be reached by $M_{ij}^m.v$ and $M_{ij}^m.o$. When $M_{ij}^m.o$ reaches the value MAX_O, it is considered that the pixel $I_{i,j}$ has given up belonging to the region R_m. In order to avoid that the background takes all the pixels of the image, for this region MAX_O = MAX_V.

After updating the matrixes of the regions the final segmentation of the frame is carried out. Each image pixel (i,j) is assigned to the region with a maximum value $M_{ij}^m.v$.

The last step of the general algorithm consists of moving the values $M_{ij}^m.v$ and $M_{ij}^m.o$ with the motion estimated for each region R_m.

Recovering lost regions Sometimes, a region that has regularly been appearing for several frames suddenly disappears. Our method counts the number of frames in which each master region is matched to another region, and when this number is larger than certain threshold q, the region is declared as a *permanent region*. It means that a *permanent region* can not disappear except when it goes out the image boundaries. Therefore, we use the information stored in the matrixes for recovering its shape.

A *permanent region* should be always matched to another region. When this does not happen, the method considers that F.F.M. segmentation has *lost* the region to which a *permanent region* R_p should have been matched. Then the method generates artificially the *lost region* R_l and matches it to R_p.

In order to generate R_l, first, the algorithm finds out to what region R_n was merged in the F.F.M. segmentation. This region can be found by analysing the matrix of the *permanent region* M^{R_p}. For each region R_s that was segmented in the current frame an accumulator A_s is created:

$$\forall I_{ij} \in Image/I_{ij} \in R_s, \quad A_s = A_s + M_{ij}^{R_p}.v \qquad (4)$$

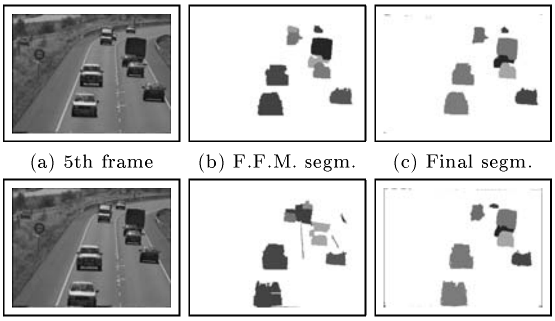
The largest accumulator corresponds to the region R_n that contains the pixels that should belong to the lost region. These pixels have to be returned to R_l. Thus, supposing that R_n was matched to the master region R_k, then R_l is matched to R_p. R_l recibes the pixels (i,j) that were assigned in the F.F.M. segmentation to R_n with $M_{ij}^{R_p}.v > M_{ij}^{R_k}.v$.

This step of the algorithm is made before generating the final segmented frame and before updating the matrix of each region from the master list. The same counter that allows the system to declare a region as permanent when it reaches a value q, enables it to remove regions that do not reach that value in a few frames. In these cases, the system considers that they appeared due to a bad F.F.M. segmentation.

3 Traffic monitoring facilities

For a traffic monitoring system the following basic facilities had to be provided: counting vehicles(traffic mean velocity) , trajectories surveillance and detection of stopped vehicles.

Counting vehicles. A virtual line is defined perpendicular to the road. Every time a centroid of a vehicle region cross the line, a counter C_T is increased.



(a) 5th frame (b) F.F.M. segm. (c) Final segm.

(d) 14th frame (e) F.F.M. segm. (f) Final segm.

Figure 1. Road sequence. First column: original image. Second column: frame to frame segmentation. Third column: final segmentation of the frame

Vehicle trajectories surveillance. Trajectories surveillance is directly achieved from tracking. It is more easily carried out because vehicles are not lost in a frame as a result of a bad F.F.M. segmentation and temporal occlusion is solved by means of the tracking strategy.

Detecting stopped vehicles. When a vehicle region has a velocity close to zero in a part of the image where usual speeds are quite different to zero and it is merged to the background for t continuous frames, this vehicle is considered as stopped. A stopped vehicle is never removed from the master list. Hence, the region of the object is always detected as a lost region and then recovered.

4 Results

This method has been tested with several traffic sequences. Figure 1 presents one of these sequences. It shows the segmentation of several frames of a sequence recorded at a road. These frames have been chosen due to their F.F.M. segmentations present some problems. Subfigure 1.a is the 5th frame of the sequence, and subfigure 1.b is its frame to frame segmentation. In this segmentation the farthest truck has been splitted into two regions, but the final segmentation corrects this problem since in the previous frames the truck was formed by a single region.

The segmentation of the 14th frame (subfigure 1.e) is the worst of the sequence. In this segmentation some static regions have been detected as moving regions, part of the farthest truck has been lost, the first car and the second car of the group of three vehicles have been merged, and the truck of this group has disappeared. Moreover, the shape of some vehicles appears corrupted. In spite of these errors, the method corrects the segmentation. The static regions are removed and the vehicles recover their shape. The lost truck of the group is detected by the algorithm as a lost vehicle region, and their pixels are extracted from the background region.

5 Conclusions

In this work we have presented a new approach for segmenting and tracking moving objects. Our method integrates segmentations provided by a frame to frame segmentation process, and accumulates the segmentation information to obtain an improved segmentation. As opposed to usual methods, our method not only uses segmentation to achieve tracking of objects, but once tracking has been performed, it is used for improving segmentation.

The experiments carried out have shown that our approach stabilizes the shape of the vehicles along the sequence and corrects errors occurred in the frame to frame segmentation. This correction avoids both sudden appearance and disappearance of regions or parts of regions from one frame to the next frame. Moreover, it enables us to store the vehicle shape and maintain it even when the vehicles stop.

References

[1] J. Badenas, M. Bober, and F. Pla. Motion and intensity-based segmentation and its application to traffic monitoring. In *Proceedings, International Conference on Image Analysis and Processing ICIAP'97, Florence, Italy*, pages 502–509. IAPR, 1997.

[2] A. Blake, M. Isard, and D. Reynard. Learning to track the visual motion of contours. *Artificial Intelligence*, 78:101–133, 1995.

[3] P. Burt, J. Bergen, R. Hingorani, R. Kolczynski, W. lee, A. Leung, J. Lubin, and H. Shvaytser. Object tracking with a moving camera. In *Proceedings, Workshop on Visual Motion, (Irvine, CA, March 20–22)*, pages 2–12. IEEE, 1989.

[4] D. Koller, K. Daniilidis, and H. Nagel. Model-based object tracking in monocular image sequences of road traffic scenes. *International Journal of Computer Vision*, 10:257 –281, 1993.

[5] D. Koller, J. Weber, and J. Malik. Robust multiple car tracking with occlusion reasoning. In J.-O. Eklundh, editor, *Proceedings, 5th European Conference on Computer Vision, (Berlin, 1994)*, pages 189–196. Springer-Verlag, 1994.

[6] P. McLauchlan and J. Malik. Vision for longitudinal vehicle control. In *Proceedings of the British Machine Vision Conference*, pages 260–269. BMVA Press, 1997.

[7] S. Smith and J. Brady. Asset-2– real-time motion segmentation and shape tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17:814–820, 1995.

[8] A. Worral, R. Marslin, G. Sullivan, and K. Baker. Model-based tracking. In *British Machine Vision Conference*, pages 310–318. BMVA Press, 1991.