Applying Computer Vision Techniques to Traffic Monitoring Tasks

Jorge Badenas and Filiberto Pla

Dept. Informática, Universitat Jaume I, Castellón 12071 (SPAIN)
Telf: +34-964-345680 Ext. 4714
(badenas,pla)@inf.uji.es

Abstract. This paper presents a method[1] for tracking and segmenting vehicles in a traffic scene. The approach is based on a frame to frame segmentation followed by a tracking process. As opposed to usual segmentation methods, our method feedbacks segmentation with tracking information for improving results. Several facilities are provided for traffic monitoring such as vehicles trajectories surveillance, segmentation of vehicle shape, measuring the mean velocity of the traffic, counting the vehicles that are moving on the lanes of a road or a motorway, counting the vehicles that stop at a junction and detection of events such as a vehicle stops on a road or a possible accident.

Key Words : Region tracking, motion analysis, motion segmentation, traffic monitoring, vehicle surveillance

1 Introduction

In recent years, as result of advances in information techniques both in terms of computational power and cost, it has become possible to use computer vision to perform many everyday tasks. This paper shows how it can be applied to traffic monitoring substituting existing methods such as Inductive Loop Detectors and Microwave Vehicle Detectors which are more expensive, and have more limited usefulness.

A typical traffic scene consists of untextured objects with a regular shape moving over a surface which is also untextured. In this context, a traffic monitoring system has to survey moving objects detecting anomalous behaviors and situations, and producing measures.

Several works have been reported on the application of computer vision to traffic monitoring. One of the early works [1] describes a system with motion analysis based on simple frame differencing. This simple approach can not provide enough information and produces poor results due to it is very sensitive to noise. Dubuisson and Jain [2] proposed a technique for segmenting vehicles also based on image substraction, but combined with color segmentation of regions.

[1] The present work was partially supported by the projects ESPRIT EP-21007 IOTA, CICYT TIC95-0676-C02-01 and Generalitat Valenciana GV97-TI-05-26

This technique can not deal with occlusions since it does not make motion estimation. Hence, it distinguishes moving regions from stopped regions, but not regions with different motions. Badenas et al [3] presented another method based on regions. This method is a combination of several techniques which produces motion estimation and segmentation of vehicles. Thus, this approach can cope with occlusions and carry out several monitoring tasks such as counting vehicles and measuring traffic fluency.

Another group of methods fits models to moving vehicles. These methods are suitable for very structured scenes, such as roads, but require a high computational cost which reduces their usefulness in traffic monitoring. We can distinguish two main tends on these methods: methods which fit 3-D models of generic vehicles [4], [5], and methods which track 2-D contours onto the image plane [6], [7]. The method [4] fits wireframe models to vehicles which can move with three degrees of freedom. A similar approach is shown in [5], where the edges extracted from the image are matched to 3-D segments of a generic car model that is projected onto the image plane. This matching is only applied to edges moving with coherent motion, to avoid mixing up edges from different vehicles. 2-D models have a lower computational cost, but present problems when the size and the number of vehicles increase. In [7], B-splines with four control points are fitted to vehicles projection. Weber at al. [6] method describe the contours by a closed cubic spline and employs two Kalman filters, one for estimating the affine motion parameters and other one for estimating the shape of the contours.

A different group of methods for motion analysis tracks single tokens such as points or lines, that are extracted from the image,[8], [9]. These methods present problems due to they do not provide explicit grouping of tokens moving with coherent motion and are quite sensitive to noise.

We propose in this paper an approach based on tracking regions which integrates information recorded over a sequence on a frame by frame basis for improving segmentation. For every frame of the sequence, our method carries out a frame to frame motion segmentation that is followed by a tracking 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 recorded by using a mask strategy that enable us to correct frame to frame segmentation of the current frame. This correction avoids both sudden appearance and disappearance of regions or parts of regions from one frame to the next frame.

Our method provides traffic monitoring facilities such as counting vehicles that are moving on each road lane, counting vehicles that have stopped at junctions, surveillance of vehicles trajectories, estimation of mean velocity of the traffic and detection of traffic jams. In addition to these facilities that could be referred to normal facilities, anomalous or dangerous situations can also be detected such as vehicles that stop on the road, vehicles that cross to a forbidden lane, vehicles that go out the boundaries of the road or a crash of vehicles.

The rest of the paper is organized as follows: section 2 explains the parts of the algorithm that allows us to segment vehicles and survey their trajectories, section

3 describes the traffic monitoring facilities provided by the method, section 4 shows results over two traffic sequences, and section 5 presents conclusions.

2 The tracking algorithm

The purpose of the algorithm is to segment the shape of the vehicles and to track them along a sequence. Figure 1 shows the different parts of the developed algorithm. Initially, the first frame is segmented using the current frame and its consecutive one, and the segmented regions are inserted in a list called the *master list*. Then, several steps are repeated for each new frame of the sequence. First a frame to frame segmentation that generates a list of new regions is carried out, then tracking step matches each region from the master list (master region) to the most similar new region. Hence, matching establishes a correspondence between the apparitions of a same region in several frames of the sequence. When a new region is not matched, it is inserted in the master list, since it can be a new vehicle. The next step recovers lost regions that have regularly appeared in previous frames and that illogically have disappeared in the current frame to frame segmentation. Updating master list step adds for each master region information from current segmentation. For each region from the master region, a matrix is stored. This matrix is updated after each frame to frame segmentation and records for each pixel of the image how many times it has belonged to the region during recent segmentations. Hence, each of these matrixes contains the evolution of the shape of its corresponding region. At the next step, information stored for each region is used to produce the final segmentation of the current frame. The next step removes from master list, the regions that have not matched to new regions in recent iterations. The last step applies the motion estimated for each region to the matrix of each master region. Thus, the values stored in each matrix for the pixels of each region are moved to the place where they will be found in the next frame to frame segmentation.

2.1 Motion segmentation using two consecutive frames

This part carries out an initial segmentation for every frame of the sequence. The frame is segmented into regions of similar intensity by using a clustering algorithm.

The next step is region-based motion estimation. Motion estimation is performed with a robust kernel with subpixel accuracy and by a mutiresolution approach. This step involves finding for every region the translation parameters that minimises the sum of displaced frame differences, transformed by a robust kernel ρ. The summation is over all pixels from the reference region. In fact, we are minimising an error measure E defined as follows:

$$E_i(dx, dy) = \frac{1}{N_j} \sum_{(x,y)\in C_j} \rho(I_1(x,y) - I_2(x+dx, y+dy), \alpha, \lambda) \qquad (1)$$

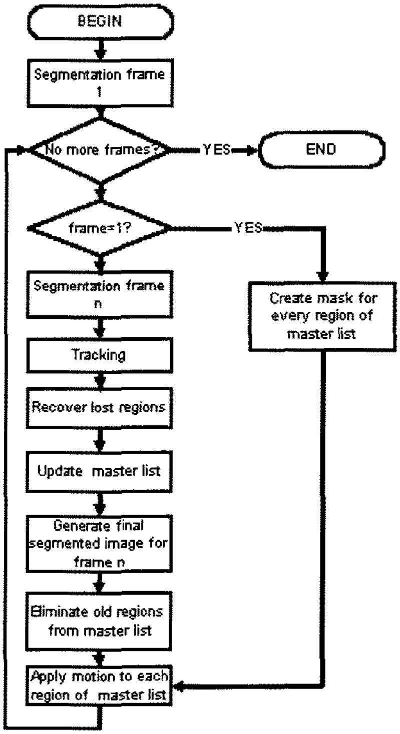


Fig. 1. Algorithm of the method

where C_j is the region j, $I_1(x, y)$, $I_2(x, y)$ are the pixel intensity values at location (x, y) in the reference and consecutive frames respectively. $\rho()$ is the robust redescending function and α, λ are the function parameters.

Motion segmentation is performed after estimating motion parameters. First, adjacent regions with coherent motion are merged. Next, for each pair of adjacent regions we calculate the error for each region when the motion of the other one is applied. If one of these errors for a region is lower or similar to the error that it had with its estimated motion, then both regions are merged. For a complete description of the frame to frame segmentation we refer the reader to [3].

2.2 Tracking regions

Tracking enables the system to follow the trajectories of the segmented regions. Tracking is based on matching regions from the master list to regions included in the list of regions segmented in current frame to frame segmentation.

Each master region is matched to the most similar new 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:

$$D^{mn} = (V^m - V^n)^t W (V^m - V^n) \tag{2}$$

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

2.3 Integrating information from several frame segmentations

The information that is recovered from each motion segmentation and related by tracking is used at two levels: pixel level and region level.

Pixel level. Every time a region R_m is inserted in the master list, a matrix M^m of image size [2] is created. Each matrix cell contains two values (v and o) that are initialized to zero. Each time that a master region R_m matches to a new region R_n, the values v of the matrix of R_m that coincide with the pixels that belong to the region R_n are incremented by one. Moreover, for these pixels, o is set to zero. The value v contains for each pixel the number of times that it has belonged to a region.

The value o contains the number of consecutive times in which a pixel has not belonged to a region. Thus, for every region R_m, each time that a pixel with a value v larger than zero in the matrix M^m does not belong to the region R_n that has been matched to R_m the value o is incremented for that pixel. When a value o of a pixel reaches a value MAX_O, its corresponding value v is also set to zero.

The maximum value that each v can reach is limited by a constant MAX_V. In order to avoid that the background takes all the pixels of the image, for this region MAX_V is lower than for the rest of regions.

After updating the matrixes of the regions the final segmentation of the frame is carried out. Each image pixel is assigned to that region with a maximum value v in that pixel.

Region level. In some frames, a region that has been appearing for several frames unlogically 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 *vehicle region*. It means that a vehicle region can not disappear except when it goes out the image boundaries. The method uses the information stored in the matrixes for recovering its shape. This fact happens when a vehicle stops or when it is going away from the camera, its motion tends to be similar to the one of the neighbouring regions (cars or

[2] Actually only a matrix of the minimum rectangle size that encloses the region is required. For the sake of simplicity, we will consider that its size is the same that the image size.

background). In this case, a frame to frame segmentation can not separate the vehicle region from its neighbouring regions.

If a *vehicle region* R_m is not matched to another region, it is considered as *lost region.* The algorithm looks for what segmented region has taken its pixels and what master region has been matched this new region to. The region that has taken the pixels can be found out by analysing the matrix of the *lost region.* For each new segmented region R_n an accumulator A_n is created:

$$\forall I_{ij} \in Image/I_{ij} \in R_n, \quad A_n = A_n + M_{ij}^{R_m}.v \tag{3}$$

The largest A_n corresponds to the region R_n that has taken the pixels that would belong to the lost region if a good segmentation had been made. Supposing that R_n was matched to R_k, then a new region R_t that is matched to the lost region R_m is created in the segmented frame. R_t recibes the pixels (i,j) that were assigned in the frame to frame segmentation to R_n with $M_{ij}^{R_m}.v > M_{ij}^{R_k}.v$, where $M_{ij}^{R_x}.v$ is the value v in the matrix M of the region x.

When a region do not reach that value q in a few frames, it is removed, since the system considers that it appeared due to a bad segmentation.

3 Traffic monitoring facilities

For a traffic monitoring system the following basic facilities are provided: counting vehicles(traffic mean velocity), vehicle trajectories surveillance, detection of stopped vehicles and detection of anomalous events. To carry out some of these tasks it is needed to delimit the road and lane boundaries. This can be automatically achieved by applying the technique exposed in [10].

3.1 Vehicle trajectories surveillance and counting the number of vehicles

Trajectories surveillance is directly achieved from tracking. This allows us to know the position of every vehicle at each frame and to count the number of vehicles. In order to count the number of 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. Moreover, when the motion direction indicates that the vehicle is moving far, a counter C_u is increased, and when the vehicle is coming to the camera C_d is incremented.

With these counters C_T, C_u, and C_d the algorithm can measure the number of vehicles that are using the road per unit of time (traffic mean velocity). With more counters we can count in addition to the number of vehicles that are moving on each direction, the number of vehicles that are moving on each lane (for instance, in a motorway).

When traffic jams are produced the mean traffic velocity decreases, but a decreasing of this mean velocity does not always indicate a traffic jam. To detect these facts we need to count the number of vehicles that are present in the image. A large number of vehicles and a little mean velocity indicates a traffic jam.

3.2 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 from the background. Motion parameters of a stopped vehicle are set to zero. Therefore its shape always remains stopped thanks to the information stored in previous frames until the vehicle moves again. This facility enables the system to count the number of vehicles that stop at the traffic lights when the light is red, and thereby, to detect immediately traffic problems at certain junctions or hard-shoulders.

3.3 Detecting anomalous events

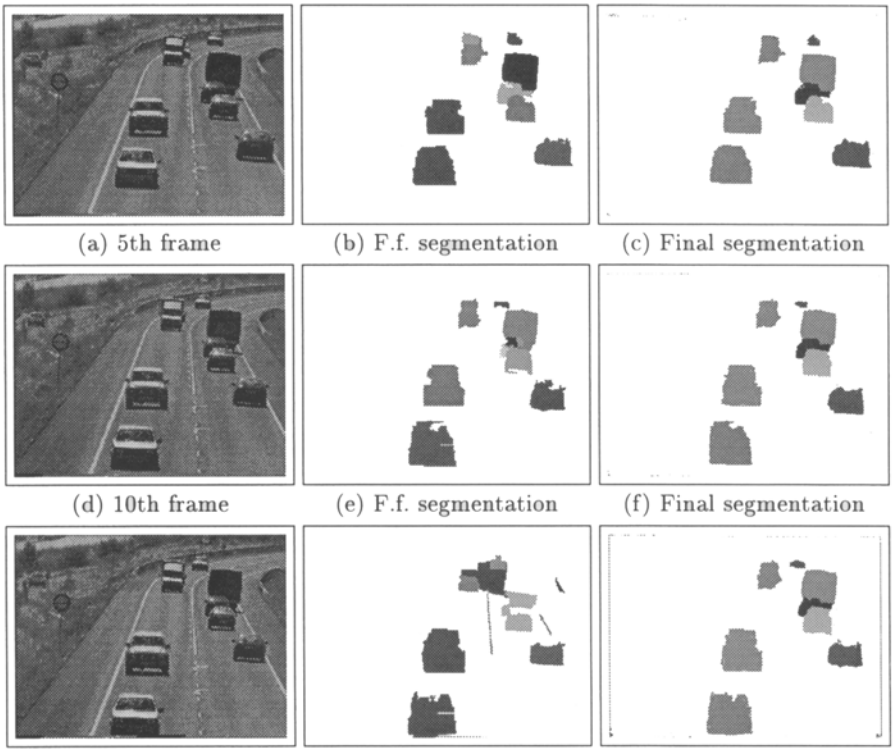
Dangerous situations such as accidents have to be detected by a traffic monitoring system and provoke an instantaneous warning. Our system can detect two types of accidents. On one hand, vehicles which cross danger lines. This virtual lines are defined in the system and delimit the part of the image on which the vehicles are moving. Danger lines are defined on the road boundaries or separating the two groups of lanes with different traffic direction in a motorway. When a vehicle cross a danger line, a warning is triggered which alerts of a dangerous situation. On the other hand, when two vehicles are moving with different directions and are only separated by a few pixels, it means that a vehicle crash is likely going to happen. Then another warning is produced which alerts to the traffic operator.

4 Results

This method has been tested with several traffic sequences. In figures 2 and 3 we present two of these sequences. First column of these figures presents several frames of the sequences. Second column contains the frame to frame segmentation of each frame, and third column is the final segmentation of the frame applying the whole algorithm, that is the result of integrating the segmentation information through several frames.

Figure 2 shows the segmentation of several frames of a sequence recorded at a road. These frames have been chosen due to their frame to frame segmentation presents some problems. Subfigure 2.a is the 5th frame of the sequence, and subfigure 2.b is its frame to frame segmentation. Here, the farthest truck has been splitted into two regions, but the final segmentation corrects this problem since in the previous segmentations the truck was formed by a single region.

In the segmentation of the 10th frame(subfigure 2.e), the intermediate car that is in the group of three vehicles is badly segmented. The region corresponding to this car was matched to the clearest region, which only contains part of the pixels of the car. The rest of pixels were assigned to the region of the car that



(a) 5th frame (b) F.f. segmentation (c) Final segmentation
(d) 10th frame (e) F.f. segmentation (f) Final segmentation
(g) 14th frame (h) F.f. segmentation (i) Final segmentation

Fig. 2. Road sequence. First column: original image. Second column: frame to frame segmentation. Third column: final segmentation of the frame

its covering it partially, and to a new region. However, the final segmentation (subfigure 2.f) is again right. The pixels are returned from both regions to the intermediate car region due to the fact that the values $M_{ij}^{R_i}.v$ of these pixels in its matrix are the largest.

The segmentation of the 14th frame (subfigure 2.h) 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 correct 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.

Figure 3 shows a sequence in which two cars stop in a traffic lights. Frame to frame segmentation only can detect the first vehicle until the 37th frame(subfigure 3.b). In the rest of frames, the car is lost by frame to frame segmentation since

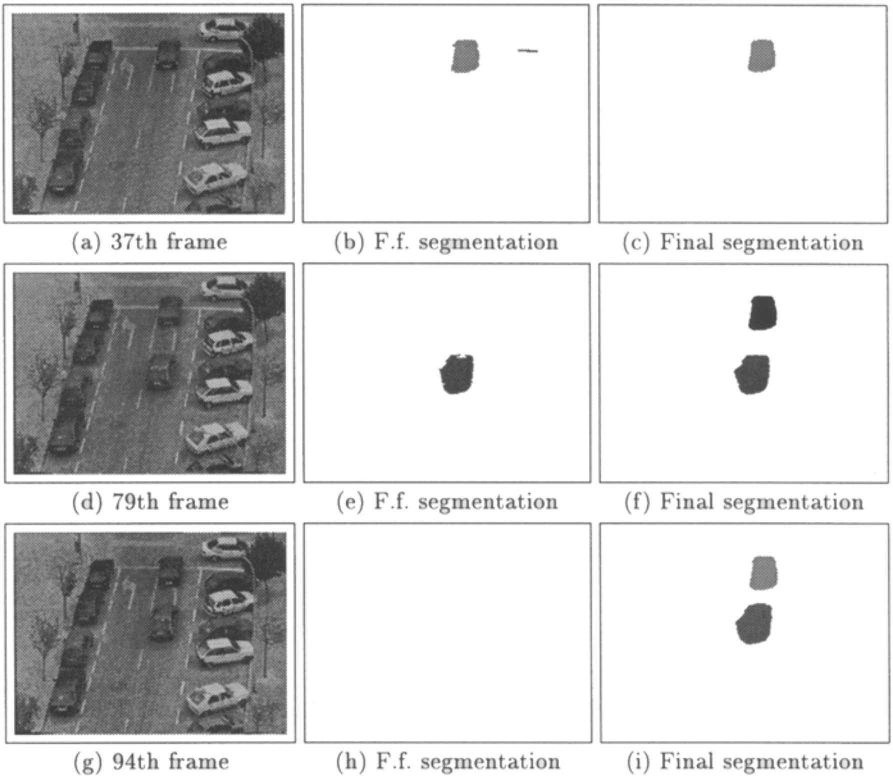


Fig. 3. Traffic lights sequence. First column: original image. Second column: frame to frame segmentation. Third column: final segmentation of the frame

it stops. However, the final segmentation rightly segments it since it is declared by the algorithm as a stopped vehicle.

In 79th frame of traffic lights sequence a new car is moving into the image. This new vehicle is segmented by frame segmentation (subfigure 3.e), but, of course, not the other one. However, the final segmentation shows both vehicles.

In 94th frame (subfigure 3.g), both cars have stopped, and can not be detected by frame to frame segmentation. The algorithm detects their shape recognizing the vehicles as stopped. Both vehicles stay in the master region list, waiting for a frame in which their motions are detected by frame to frame segmentation.

5 Conclusions

In this work we have presented a new approach for segmenting and tracking moving objects on sequences. Our method is based on tracking regions and integrates segmentations provided by a frame to frame segmentation process, and accumulates the segmentation information for obtaining 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 system provides useful facilities for traffic monitoring such as surveillance of vehicle trajectories, counting the number of vehicles that are using a road or a lane, counting the number of vehicles that have stopped at junctions or traffic lights, estimating the mean traffic velocity, or detecting traffic jams. Moreover, the system permits the detection of dangerous situations (accidents, stopped vehicles, vehicles that go off the road, etc.) that trigger an automatic warning.

References

1. R.C. Waterfall and K.W. Dickinson. Image procesing applied to traffic. *Traffic Engeneering Control*, 25:60 - 67, 1984.
2. M.P. Dubuisson and A.K. Jain. Contour extraction of moving objects in complex outdoor scenes. *International Journal of Computer Vision*, 14:83–105, 1995.
3. 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.
4. J.M. Ferryman, A. D. Worrall, G.D. Sullivan, and K.D. Baker. A generic deformable model for vehicle recognition. In *Proceedings of the British Machine Vision Conference*, pages 128–136. BMVA Press, 1995.
5. D. Koller, K. Daniilidis, and H.H. Nagel. Model-based object tracking in monocular image sequences of road traffic scenes. *International Journal of Computer Vision*, 10:257 –281, 1993.
6. D. Koller, J. Weber, and J. Malik. Robust multiple car tracking with occlusion reasoning. In Jan-Olof Eklundh, editor, *Proceedings, 5th European Conference on Computer Vision, (Berlin, 1994)*, pages 189–196. Springer-Verlag, 1994.
7. N.J. Ferrier, S.M. Rowe, and A. Blake. Real-time traffic monitoring. In *Proceedings, 2nd IEEE Workshop on Applications of Computer Vision*, pages 81–87, 1994.
8. P. McLauchlan and J. Malik. Vision for longitudinal vehicle control. In *Proceedings of the British Machine Vision Conference*, pages 260–269. BMVA Press, 1997.
9. S.M. Smith and J.M Brady. Asset-2– real-time motion segmentation and shape tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17:814–820, 1995.
10. B.D. Steward, I. Reading, M.S. Thomson, T.D. Binnie, K.W. Dickinson, and C.L. Wan. Adaptative lane finding in road traffic image analysis. In *Proceedings 7Th International Conference on Real Traffic*, pages 133–136, 1994.