Vehicle Segmentation by Edge Classification Method and the S-T MRF Model

Hiroshi Inoue, Mingzhe Liu, and Shunsuke Kamijo
{Student, Student, Associate Professor}, The University of Tokyo

Abstract—In this paper, we propose a tracking algorithm, which is based on the collaboration of the S-T MRF model and a dedicated segmentation algorithm. Although the S-T MRF model was designed to be robust against occlusion, it regards vehicles that move in parallel occluding each other from the beginning to the end of the traffic images as a single region. In order to compensate such a defect of S-T MRF, we have developed a dedicated segmentation algorithm which decides boundaries of vehicles contained in such a single region by referring to the difference of edge patterns among the vehicles. By the experiments using traffic video from three different angles at different locations, our method was proved to be very successful.

I. Introduction

Recently, vehicles spread so widely in the world that traffic accident and traffic congestion have been problems. Traffic surveillance is indispensable in order to solve these problems. In order to carry out traffic surveillance, various sensors such as image sensors, ultrasonic sensors, loop detectors are employed. Both ultrasonic sensors and micro loops can only get information that a vehicle is on the road or not. In addition, since you have to install these for every lane of a driveway, the cost may be high. On the other hand, real-time video surveillance becomes popular because image contains rich information than the other sensors, and it can be shared by human investigation. However, image sensors have defects such as instability against environmental changes and occlusion problem. In order to solve the occlusion problem, related works have been done in [1] and [2]. However, neither is almighty to occlusion. Although our laboratory also tackled this problem and achieved the results [3] [4], this work is not almighty to occlusion either. Therefore, we should combine with pattern matching to solve it.

II. System Overview

Our system is implemented as software. Video clip from the monitoring camera on the street is transferred to a PC on which this system runs. The operational flow consists of 7 processes, which is as shown in Figure 1.
III. VEHICLE SEGMENTATION BY THE S-T MRF MODEL

A. Outline of S-T MRF Model

Our system adopts the S-T MRF model based tracking algorithm. So now we will explain about the S-T MRF model. Usually, the spatial MRF segments an image by each pixel. However, since the usual video cameras do not have such high frame rates, objects typically move for ten or twenty pixels among consecutive image frames. Therefore, neighbor pixels within a cubic clique will never have correlations of either intensities or labeling. Consequently, we defined our Spatio-Temporal Markov Random Field model (S-T MRF)\(^5\) as to divide an image into blocks as a group of pixels, and to optimize labeling of such blocks by referring to texture and labeling correlations among them, in combination with their motion vectors. Combined with employing stochastic relaxation method, our S-TMRF optimizes object boundaries precisely, even when serious occlusions occur. Here, a block corresponds to a site in the S-TMRF, and only the blocks that have different textures from the background image are labeled as one of the object regions.

B. Adaptive Tracking Algorithm

Because vehicles appear in various kinds of forms, we cannot employ any models of vehicle shapes. And because vehicles move in random manners, occlusions and confusions situations occur in complicated manners. In such situations, confused vehicles cannot be segmented by a method which employs contour models. As we can see from above discussion, the most robust way for this purpose must not require any information other than images themselves. The way requires some kind of image segmentation. But, it is very difficult to segment images with a measure of pixel.

In order to resolve these problem, we developed a dedicated tracking algorithm utilizing what we call Spatio-Temporal Markov Random Filed model. In this algorithm, an image of 640x480 pixels is divided into 80x60 blocks, where each block consists of 8x8 pixels. And it is determined to which object each block should belong at each image. The determination process estimates correlation of blocks between consecutive images as well as neighbor blocks with in an image, and segment the blocks into most possible objects by the stochastic process.

Then, we will explain how to track vehicles, which can be adapted to changing of size and shape of vehicles, below.

1. Generate new vehicle IDs: The algorithm sets up a slit at each entrance and also constructs a background image. The algorithm decides that it detects a vehicle, generates a new vehicle ID, and assigns it to the block.

2. Estimate Motion Vectors of Vehicles: Once a vehicle region leaves the slit, the algorithm updates its shape along the time sequence. For this updating, the algorithm estimates motion vectors among blocks in a vehicle region. At each block, a block matching method is employed to estimate its motion vector between the time \(t\) and \(t+1\). Similarity between one block at time \(t\) at \((x(t),y(t))\) and one at time \(t+1\) at \((x(t+1),y(t+1))=x(t)+u(t),y(t)+v(t))\) is evaluated as equation (1); here, \(I(x; y; t)\) is a gray-scaled intensity of a pixel \((x; y)\) at time \(t\). Then the motion vector of a vehicle is approximated by the most frequent motion vector among those blocks of the same vehicle region.

   \[
   D = \sum_{0\leq di<8,0\leq dj<8} I(i+di+j+dj; t+1) - I(i+di+j+dj; t)
   \]

3. Update vehicle regions (Figure.3): By using the motion vector, all the blocks of the same vehicle region shift at the time \(t+1\) from the locations at the time \(t\). After a block has shifted, if the intensity difference between the current and the background at the new location is smaller than a threshold value, the algorithm does not consider the block as belonging to the vehicle region. On the other hand, if a neighbor block of the new vehicle region has a larger difference, the algorithm defines the block to be the vehicle region, and assigns the same vehicle ID.

4. Divide Vehicle Blocks: In some cases, multiple vehicles simultaneously pass through a slit and they may be considered to be a single object. In order to divide such vehicles thereafter, the algorithm examines connectivity and the distribution of motion.
vectors over object blocks. If multiple parts in one vehicle region have several different motion vectors, those parts are separated and assigned to different vehicle IDs.

The result obtained in this subsection will be utilized in the next step (C. Optimization utilizing Spatio-Temporal Markov Random Field Model) as the initial object map for the stochastic relaxation process based on the Spatio-Temporal Markov Random Field model. Because a motion vector is estimated by a measure of pixels, not blocks, in Step.2, a fragmentation problem will occur in renewing the Object-Map. When a block is considered to belong to two different blocks in Step.4, it must be determined to which object it is likely belong. These two problems can be optimized by the stochastic relaxation method, S-T MRF.

C. Optimization utilizing Spatio-Temporal Markov Random Field Model

When a block is estimated as belonging to two different objects, we must determine to which object such a confused block belongs. But, it is very difficult to determine this deductively in such a case. Therefore, we decided to apply Markov Random Field (MRF) model combined with stochastic relaxation method to the optimization of such confusions, fragmentations and shape adaptations inclusively.

MRF is an analogy from a statistical mechanics system. There are some studies to apply MRF to image restoration[6], image compression[7] and image segmentation[8]. In our case, some extension of usual MRF is required. This is because those time-series images have correlations between consecutive images along a time axis. Therefore we extended usual MRF to apply to the Spatio-Temporal image which we named Spatio-Temporal Markov Random Field Model. We applied this Spatio-Temporal MRF to vehicle tracking in traffic images. Our Spatio-Temporal MRF estimates a current object map (a distribution of vehicle labels) according to a previous object map, and previous and current images.

The optimization problem results in a problem of determining an object map \( X(t) = y \) which minimizes the following energy function.

\[
U(y) = \sum_k U(y_k) \tag{2}
\]

\[
U(y_k) \equiv U_N(N_{yk}) + U_D(D_{yk}, M_{yk})
= U_N(N_{yk}) + U_D(D_{yk}) + U_M(M_{yk}) \tag{3}
= a(N_{yk} - \mu N_y)^2 + b(M_{yk} - \mu M_{xy})^2 + cD_{yk}^2
\]

Here are the notations:

- \( G(t-1) = g; G(t) = h \): An image G at time \( t-1 \) has a value g, and G at time t has a value h. At each pixel, this condition is described as \( G(t-1; i; j) = g(i; j); G(t; i; j) = h(i; j) \).
- \( X(t-1) = x; X(t) = y \): An object Map \( X \) at time \( t-1 \) is estimated to have a label distribution as \( x \), and \( X \) at time \( t \) is estimated to have a label distribution as \( y \). At each block, this condition is described as

\[
X_k(t-1) = x_k, X_k(t) = y_k \text{, where } k \text{ is a block number.}
\]
- \( U \) is the energy function.
- \( N_{yk} \) is the number of neighbor blocks that belong to the same vehicle.
- \( D_{yk} \) represents texture correlation between \( G(t-1) \) and \( G(t) \). And the estimation motion vector is \( -v_{mi} \). \( M_{xy} \) is a goodness measure of the previous object map \( X(t-1) = x \) under a given current object map \( X(t) = y \).
- \( \mu N_y = 8 \)
- \( \mu M_{xy} = 64 \)

\[ U(y_k) \] is considered to be the energy function for Spatio-Temporal MRF, and \( U(y_k) \) will be minimized by the relaxation process. In the Step.3 of the deductive process, blocks are move temporally into Object-Map in the next frame by referring not to a motion vector characteristic of each block but to the representative motion vector of the cluster to which these blocks belong. And then, Object-Map of confusing blocks will be optimized by also referring to the representative motion vector of the cluster. Therefore, we call this optimization method as primary ST-MRF compared to advanced optimization method which refers to a motion vector characteristic of each block themselves.

D. Results from S-T MRF model Tracking

Figure 4 and Figure 5 shows an example of Object-map and Motion-Vector-map as tracking results. Object-map indicates the result of segmentation, and M-V-map indicates distribution of motion vector for each block.
IV. VEHICLE SEGMENTATION BY EDGE PATTERN CLASSIFICATION IN PARALLEL WITH THE WAY

A. Judging width of vehicle regions

However, since the S-T MRF model was defined as a segmentation algorithm for spatio-temporal image, this algorithm cannot divide the region of the two vehicles that move from the entrance to the exit of images occluding each others. In addition, for the same reason, the S-T MRF model cannot divide shadow from the vehicle. In order to solve these problems, some dedicated algorithms based on vehicle pattern recognition should be defined in order to segment such region correctly, and the segmentation result should be fed back to the S-T MRF model. Then, the S-T MRF model will perform segmentation in forward and backward direction along the temporal axis by referring to the correct information from the algorithm for vehicle pattern recognition.

First, we defined the lane region as shown in Figure 6.

![Fig.6. Lane Region and Region Number](image)

Generally, lane regions are provided double number of the lane in order to correspond flexibly to the size and course of a vehicle. And a number is given sequentially from the left. We call this number as lane region number. Then we focus on the rectangle of the tracking result by S-T MRF model which is as shown in Figure 7.

![Fig.7. The Result of Tracking by S-T MRF Model](image)

Analyzing the rectangle, we noticed when the camera is set up at the left side of the road, a lower right point of rectangle is affected by only the position of a vehicle, but a lower left point is affected by not only the position of a vehicle but also the size of a vehicle. Considering these conditions, we set up the lane regions.

Our system judges there are two vehicles in one vehicle region or not based on the lane regions. However, there is a problem that two vehicles occupy the width for only one vehicle caused by occlusion. We show an example that is as shown in Figure 8 and Figure 9.

![Fig.8. A region should not be divided](image)

![Fig.9. A region should be divided](image)

Although Figure 8 should not be divided, Figure 9 should be divided. Therefore, the system should distinguish one situation from another. By considering an offset, we solved this problem. If one vehicle region has two vehicles in Figure 8, there is an offset. Therefore, if we can find an offset by scanning the edge of vehicle, we can recognize there are two vehicles.
B. Boundary determination

When our system recognizes that there are two vehicles in a single region, our system proceeds to the following step. First, our system extracts edges that are orthogonal to vehicle’s traveling direction from an original image. Then our system scans the edges as shown in Figure 10, and obtains edge patterns. Finally, our system calculates humming distance of edge patterns to decide boundaries of vehicles contained in a single region. Thus, our system can divide the vehicle region at the boundary, and corrected Object-map will be fed back into the ST-MRF model.

Fig. 10. Scanning line

C. Noise Reduction

However, some images include shadows or headlights of vehicles. The system would regard such the noises as two vehicles included in a single vehicle region, and this leads errors in tracking results. Therefore, such the noises should be distinguished from correct vehicles. Fortunately, shadows and headlights tend to have edge patterns drastically different from edge patterns of vehicles as shown in Figure 11 and Figure 12. Therefore, those three will be distinguished by evaluating edge patterns of the images.

Fig. 11. Scanning and distinguishing vehicle from shadow

Fig. 12. Scanning and distinguishing vehicle from light

V. VEHICLE SEGMENTATION BY EDGE PATTERN CLASSIFICATION INVERTICAL WITH THE WAY

A. Judging length and width of vehicle regions

The S-T MRF model based tracking algorithm sometimes cannot divide the region precisely during heavy traffic congestion as shown in Figure 13. Investigating length and width of the region, the algorithm will determine whether the region should be divided or not as shown in Figure 14.

Fig. 13. One Region has Many Vehicles during heavy traffic congestion

Fig. 14. The matrix of dividing decision

B. Boundary determination

In order to divide the region precisely, the following three different kinds of gaps should be taken into account as shown in Figure 15. The first one is the gap of width, by which boundaries between large trucks and vehicles can be found. The second one is the gap of main axis, by which even vehicles of the same type can be divided. The third one is density gap of an edge interval, by which boundaries between any types of vehicles can be found.
First, our system detects the vehicle's edges from an image. Next, our system scans every line and finds the boundary.

\[ \text{SizeDiff}() \vee \text{CenterDiff}() \vee \text{EdgePattern}( ) \] (5)

Here, we explain how to decide the boundary using the density pattern of an edge interval. When we looked at Figure 15, we noticed that vehicle's back and roof can be seen, and vehicle's back has many edges and vehicle's roof has few edges. Therefore, we can say that the place which the edge density is changed from non-dense to dense is the boundary of two vehicles. The concrete method of determining a boundary line is as follows.

1. Level edge is extracted.
2. Count the number of edge.
3. A threshold is calculated by the following formula.
   \[ \text{threshold} = \frac{\text{height}}{\text{number_of_edge}} \] (6)
4. Edge is clustered using threshold which is calculated by the upper formula as shown in Figure 16.
5. The distance between clusters is measured and an evaluation value is calculated.
6. A cluster is selected which has the max evaluation value.
7. The region is divided at the center of the two edges as shown in figure 18.

### VI. RESULT

#### A. Dividing a region in parallel with the way

We will show the result of dividing a region in parallel with the way as follow.

Next, we will show the recall rate and precision rate of the result. We experimented four locations, thirty hours. The number of vehicles which passed by is about 18,000 and the number of regions which has two or more vehicles is 1,487.

- recall rate : 83.2%
- precision rate : 92.3%

#### B. Noise Reduction

We will show the result of noise reduction as follow. Figure 20 shows that our system recognized light, and Figure 21 shows that our system recognized shadow.
Next, the accuracy of distinction with shadows, headlights, and other noises, and vehicles is shown. The denominator of the upper row of a table expresses the number containing two or more vehicles to one region as a result of S-T MRF tracking. The denominator of the lower berth of a table contains only one vehicle to one region as a result of S-T MRF tracking, but the width of result is very wide because of shadows, headlights, and other noises.

<table>
<thead>
<tr>
<th></th>
<th>Divided</th>
<th>Not Divided</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ONLY S-T MRF</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Two Vehicles</td>
<td>73.7% (14/19)</td>
<td>26.3% (5/19)</td>
</tr>
<tr>
<td>Shadow, Light, Noise</td>
<td>19.7% (15/76)</td>
<td>80.3% (61/76)</td>
</tr>
</tbody>
</table>

**C. Dividing a region in vertical with the way**

The result of dividing a region in vertical with the way is as follow.

Next, we will show the recall rate and precision rate of the result. We experimented six hours in heavy traffic. The number of vehicles which passed by is about 10,000 and the number of regions which has two or more vehicles is 390.
- recall rate : 80.5%
- precision rate : 75.1%

**D. Total Accuracy**

Finally, we will show the total accuracy result by the combination algorithm of S-T MRF model and edge classification method. Here, “the total accuracy” means the rate of tracking each vehicle precisely. The number of vehicles which passed by is about 28,000 and the total accuracy is 96.4%

**ACKNOWLEDGEMENT**

This study was supported by “Industrial Technology Research Grant Program” in ‘04 from New Energy and Industrial Technology Development Organization (NEDO) of Japan. In addition to that, we are very grateful to Metropolitan Expressway Public Corporation and Japan Highway Public Corporation for their providing data, evaluating our experiment and their helpful comments.

**REFERENCES**