Event Recognitions from Traffic Images based on Spatio-Temporal Markov Random Field Model

Shunsuke KAMIJO Katsushi IKEUCHI Masao SAKAUCHI
Institute of Industrial Science, University of Tokyo
E-mail : kamijo@iis.u-tokyo.ac.jp

Abstract

One of the major interest on ITS is event recognitions from traffic informations gathered by image sensors or spot sensors. For that purpose, we employed image sensors due to its more rich information rather than spot sensors. And then, in order to gather precise information from traffic images, we have been developed occlusion robust tracking algorithm based on Spatio-Temporal Markov Random Field model. This success has led to the development of an extendable robust event recognition system algorithm based on the Hidden Markov Model. This system learns various event behavior patterns of each vehicle in the HMM chains and then, using the output from the tracking system, identifies current event chains. By this system, ordinary traffic activities at an intersection were able to be recognized at success rates of 90% *average, classifying observation sequences that are obtained from the tracking results. By this system, it becomes possible to classify ordinary traffic activities in detail and abnormal event would be found by distinguishing then from ordinary situations. Actually, we could reach a presumable idea of accident detection method. This method recognizes observation sequences similar to HMM model of accidents out of a lot of ordinary traffic activities. And then, combining with recognition results of activities among neighbor vehicles, this system successfully determined pairs of vehicles that are really involved in accidents.*

1 Introduction

For the past many years, one of the most important research efforts in ITS have been the development of systems that automatically monitor the flow of traffic at intersections. Such systems would be useful both in reducing the workload of human operators and in warning drivers of dangerous situations. Not only would the systems automatically monitor current situations at intersections but, if the systems could reliably assess those situations and predict whether they might lead to accidents, they might be able to warn drivers and thus reduce the number of accidents. Besides, if the systems could control periods of signals to according to information about masses of traffic flow, they might be able to reduce traffic jams and economic losses.

For event recognition, systems traditionally use spot sensors. Successful event recognition systems with spot sensors for traffic monitoring include Gangisetty's[7] Incident Detection System (IDS) with inductive loop sensors. Traffic monitoring system using loop detectors are quite popular and, in fact, in practical use in several cities including Toronto and Washington DC. Although such spot sensors are reliable, stable and in practical use, they have rather limited scope of usage in terms of event recognition; almost all of the spot sensors can obtain information only on whether a vehicle exists on a sensor spot; therefore, a large number of sensors are required to survey an area for event recognition.

On the other hand, one of the most important advantages of utilizing vision sensors for event recognition is their ability to collect rich information such as illegally parked vehicles, traffic jams, traffic violations and accidents. Some representative vision-based systems can be found in [10][11][12][4]. Rojas[10] and Zeng[11] developed methods or employed systems for tracking vehicles on highways from a fixed TV camera. Lai et al.[12]developed "Red light runners detection" at an intersection. This system is now in operation in Hong Kong. Grimson, et al.[4] are monitoring traffic by Forest of Sensors". The traffic activity is classified by clustering motion vectors and detecting abnormal events.

As a research on event recognition at an cross road, V.Kettnaker and M.Brad has performed interesting work[13]. However, since the cross road in this paper is small, events are very limited.

2 Primary ST-MRF model as Our Previous Works

In order to track vehicles against occlusions, we have develped the tracking algorithm based on Spatio-Temporal Markov Random Field model[18]. This algorithm sengments spatio-temporal images to define region boundaries among occluded vehicles by referring to labeling correlations, texture correlations, etc. between consecutive images. Since this successful work have led effective event recogintion, an idea of ST-MRF as our previous work should be briefly explained at first.

2.1 Initialization of Object-Map

In preparation for this algorithm, an image which consists of 640x480 pixels is divided into 80x60 blocks because a pixel is too small to be considered as one site in the ST-MRF and therefore, we need a larger group of pixels. Each block consists of 8x8 pixels. One block is considered to be a site in the ST-MRF. The algorithm classifies each block into vehicles or, equivalently, assigns one vehicle label to each block. Since the ST-MRF converges rapidly to a stable condition when it has a good initial estimation, we use a deductive algorithm to determine an initial label distribution. Then, these labels are refined through the ST-MRF. In this refining process, the algorithm considers correlation of blocks between consecutive images as well as neighbor blocks and then assign labels to them through the Markov Random Field model. A distribution of classified labels on blocks is referred to as an Object-Map. Here, since this algorithm can be defined to be applicable to gray scaled images, all the experiments has been performed by only using gray scaled images in this paper.

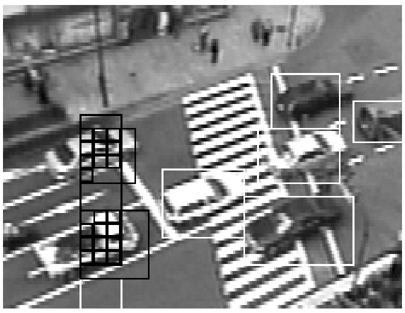


Figure 1: Object Generation

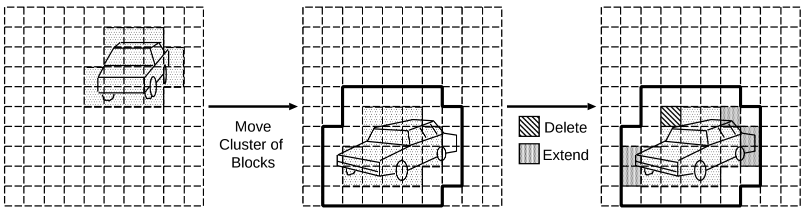


Figure 2: Object Moving and Re-mapping

2.2 Spatio-Temporal Markov Random Field Model

Some blocks may be classified as having multiple vehicle labels due to occlusion and fragmentation. We can resolve this ambiguity by employing stochastic relaxation with Spatio-Temporal Markov Random Field (MRF) model. 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. Here are the notifications:

- $G(t-1) = g, G(t) = h$: 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 $X_k(t-1) = x_k, X_k(t) = y_k$, where k is a block number.

We will determine the most likely $X(t) = y$ so as to have the MAP(Maximum A posteriori Probability) for given $G(t-1) = g, G(t) = handX(t-1) = x$, previous and current images and a previous object map, and a previous object map $X(t) = y$. A posteriori probability can be described using the Bayesian equation:

$$\begin{gathered} P(X(t) = y | G(t-1) = g, X(t-1) = x, G(t) = h) = \\ \frac{P(G(t-1) = g, X(t-1) = x, G(t) = h | X(t) = y) P(X(t) = y)}{P(G(t-1) = g, X(t-1) = x, G(t) = h)} \end{gathered} \tag{1}$$

$P(G(t-1) = g, X(t-1) = x, G(t) = h)$, a probability to have previous and current images and a previous object map, can be considered as a constant. Consequently, maximizing a posteriori probability is equal to maximizing $P(G(t-1) = g, X(t-1) = x, G(t) = h | X(t) = y) P(X(t) = y)$.

$P(X(t) = y)$ is a probability for a block C_k to have $X_k(t-1) = y_k$ (for all ks). Here, y_k is a vehicle label. For each C_k, we can consider its probability as a Boltzmann distribution. Then, $P(X(t) = y)$ is a product of these Boltzmann distributions:

$$P(X(t) = y) = \prod_k exp[-U_N(N_{y_k})]/Z_{Nk} = \prod_k exp[-\frac{1}{2\sigma^2_{N_y}}(N_{y_k} - \mu_{N_y})^2]/Z_{Nk} \tag{2}$$

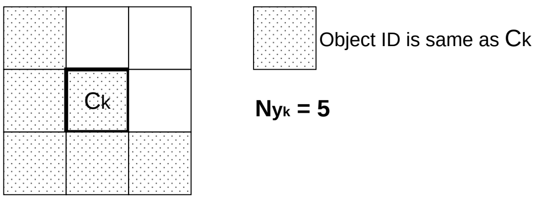


Figure 3: 8 neighbor blocks

Here, N_{y_k} is the number of neighbor blocks of a block C_k(Figure.3) that belong to the same vehicle as C_k. Namely, the more neighbor blocks that have the same vehicle label, the more likely the block is to have the vehicle label. Currently, we consider eight neighbors as shown in Figure. 3. Thus, $\mu_{N_y} = 8$, because the probability related to block C_k has maximum value when block C_k and all its neighbors have the same vehicle label. Therefore, the energy function $U_N(N_{y_k})$ takes a minimum value at $N_y = 8$ and a maximum value at $N_y = 0$.

We also consider the probability of $G(t-1) = g, G(t) = h, X(t-1) = x$ for a given object map $X(t) = y$ as a Boltzmann function of two independent variables:

$$\begin{aligned} P(G(t-1) = g, X(t-1) = x, G(t) = h | X(t) = y) &= \prod_k exp[-U_{pre}(M_{xy_k}, D_{xy_k})]/Z_{DMk} \\ = \prod_k exp[-U_M(M_{xy_k})]/Z_{Mk} &\cdot \prod_k exp[-U_D(D_{xy_k})]/Z_{Dk} \\ = \prod_k exp[-\frac{1}{2\sigma^2_{M_{xy}}}(M_{xy_k} - \mu_{M_{xy}})^2]/Z_{Mk} &\cdot \prod_k exp[-\frac{1}{2\sigma^2_{D_{xy}}}(D_{xy_k} - \mu_{D_{xy}})^2]/Z_{Dk} \end{aligned} \tag{3}$$

M_{xy_k} is a goodness measure of the previous object map $X(t-1) = x$ under a given current object map $X(t) = y$. Let us assume that a block C_k has a vehicle label O_m in the current object map $X(t)$, and C_k is shifted backward in the amount of estimated motion vector, $-\overrightarrow{V_{O_m}} = (-v_{mi}, -v_{mj})$ of the vehicle O_m, in the previous image (Figure.4). Then the degree of overlapping is estimated as M_{xy_k}, the number of overlapping pixels of the blocks with the same vehicle labels. The more pixels that have the same vehicle label, the more likely a block C_k belongs to the vehicle. The maximum number is $\mu_{M_{xy}} = 64$, and the energy function $U_M(M_{xy_k})$ takes a minimum value at $M_{xy_k} = 64$ and a maximum value at $M_{xy_k} = 0$. For example, when a block is determined to which of vehicle O_1, O_2 it belongs, $U_M(M_{xy_k})$ will be estimated as follows. First, assuming that a block belongs to O_1, the energy function is estimated as $U_M(M_{xy_k}) = U_{M1}$ by referring to $-\overrightarrow{V_{O_1}} = (-v_{1i}, -v_{1j})$. Then assuming that a block belongs to O_2, the energy function is estimated as $U_M(M_{xy_k}) = U_{M2}$ by referring to $-\overrightarrow{V_{O_2}} = (-v_{2i}, -v_{2j})$. As result of these estimations, when U_{M1} is less than U_{M2}, this block more likely belongs to vehicle O_{M1}.

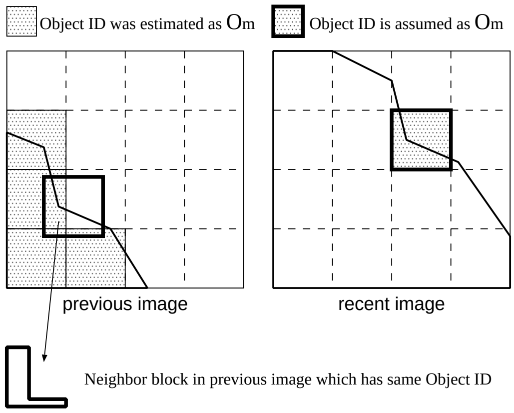


Figure 4: Neighbor condition between Consecutive Images

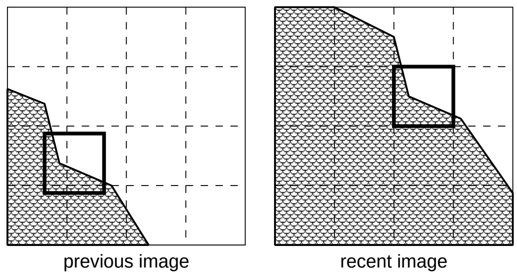


Figure 5: Texture Matching

D_{xy_k} represents texture correlation between $G(t-1)$ and $G(t)$. Let us suppose that C_k is shifted backward in the image $G(t-1)$ according to the estimate motion vector $-\overrightarrow{V_{O_m}} = (-v_{mi}, -v_{mj})$. The texture correlation at the block C_k is evaluated as(See Figure.5):

$$D_{xy_k} = \sum_{0 \le di < 8, 0 \le dj < 8} |G(t; i + di, j + dj) - G(t-1; i + di - v_{mi}, j + dj - v_{mj})| \tag{4}$$

The energy function $U_D(D_{xy_k})$ takes maximum value at $D_{xy_k} = 0$. The smaller D_{xy_k} is, the more likely C_k belong to the vehicle. That is, the smaller $U_D(D_{xy_k})$ is, the more likely C_k belong to the vehicle. For example, when a block is determined to which of vehicle O_1, O_2 it belongs $U_D(D_{xy_k})$ will be estimated as follows. First, assuming that a block belongs to O_1, the energy function is estimated as $U_D(D_{xy_k})) = U_{D1}$ by referring to $\overrightarrow{V_{O_1}} = (v_{1i}, v_{1j})$. Then assuming that a block belongs to O_2, the energy function is estimated as $U_D(D_{xy_k})) = U_{D2}$ by referring to $\overrightarrow{V_{O_2}} = (v_{2i}, v_{2j})$. As result of these estimations, when U_{D1} is less than U_{D2}, this block most likely belongs to vehicle O_{D1}.

Consequently, this optimization problem results in a problem of determining a map $X(t) = y$ which minimizes the following energy function.

$$\begin{aligned} U(y_k) &\equiv U_N(N_{y_k}) + U_{pre}(D_{xy_k}, M_{xy_k}) = U_N(N_{y_k}) + U_D(D_{xy_k}) + U_M(M_{xy_k}) \\ &= a(N_{y_k} - \mu_{N_y})^2 + b(M_{xy_k} - \mu_{M_{xy}})^2 + cD_{xy_k}^2 \end{aligned} \tag{5}$$

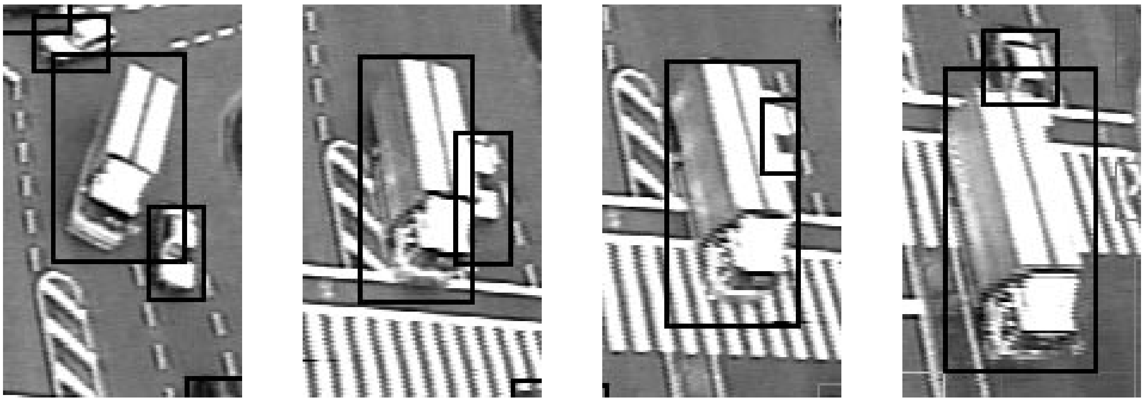
$U(y_k)$ is considered to be the energy function for Spatio-Temporal MRF, and $U(y_k)$ will be minimized by the relaxation process.

2.3 Experimental Results

Figure.6 shows a sequence of tracking two vehicles that caused an occlusion situation. These images are obtained at the rate of 10 frames/second, and a frame number is attached to each image. Although a car is partly occluded behind a truck, the two vehicles have been successfully segmented. We applied the tracking algorithm utilizing the Spatio-Temporal MRF model to 25 minute traffic images at the intersection. Three thousand, two hundred and fourteen vehicles traversed the intersection; of these, 541 were occluded. As a result, the method was able to track separated vehicles that did not cause occlusions at over 99% success rate, and the method was able to segment and track 541 occluded vehicles at about 95% success rate.

3 Classifying Ordinary Traffics by using Hidden Markov Model

By using such precise tracking results, detailed traffic monitoring will become possible. Although this traffic monitoring system intends to detect abnormal events such as accidents or dangerous



(a) frame 678 (b) frame 696 (c) frame 702 (d) frame 710

Figure 6: Tracking results by ST-MRF

manners, it is also significant to classify situations that occur in ordinary traffics.

Since motion sequences of vehicles that are involved in abnormal traffic events would be very different from motion sequences in ordinary situations, the system would detect abnormal events efficiently by classifying motion sequences among learning model of such abnormal events and ordinary situations. Therefore, at first, the classification method of such ordinary situations is discussed in this section.

3.1 Relative Motions in Ordinary Traffics

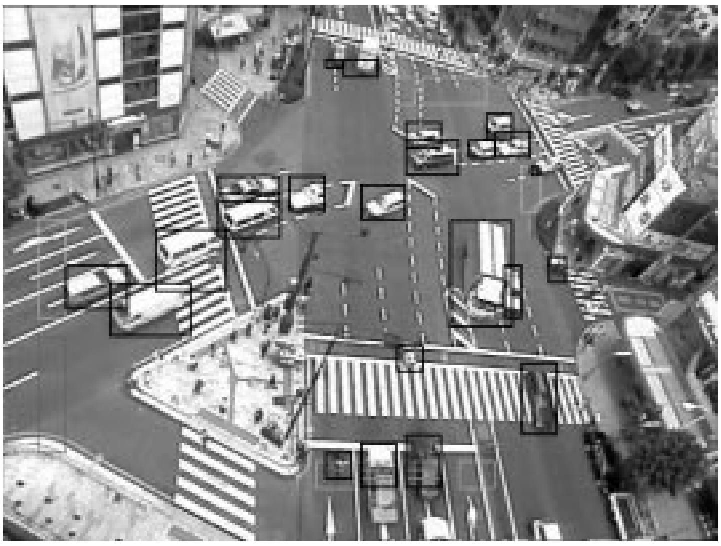
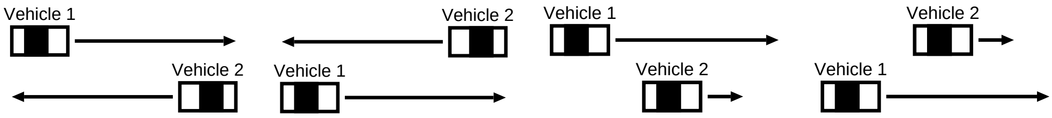


Figure 7: Entire Image of the Intersection

Figure.7 shows an entire image of the intersection, and horizontal traffics are active in this image. As shown in this figure, motion sequences between a pair of vehicles in ordinary situations are typically classified Figure.8.

Since vehicles drive on the left in Japan, these figures represents situations as follows. Figure.8(a) represents two vehicles on the opposite side passing each other, and such states are observed in cases of through traffics. Figure.8(b) represents two vehicles on the opposite side passing each other, but the locations of the two vehicles are different from Figure.8(a). Such states are observed in cases when both vehicles make right turns simultaneously. Figure.8(c) represents a vehicle passing by the other vehicle from the left lane of the same side. Such states are observed while the other vehicle is waiting in order to make right turns. Figure.8(d) represents a vehicle passing by the other vehicle from the right lane of the same side. Such states are observed while the other vehicle is waiting for pedestrians in order to make left turns.



(a) PAS-00 (b) PAS-11 (c) PAS-02 (d) PAS-13

Figure 8: Relations between the two Vehicles of Normal Traffics

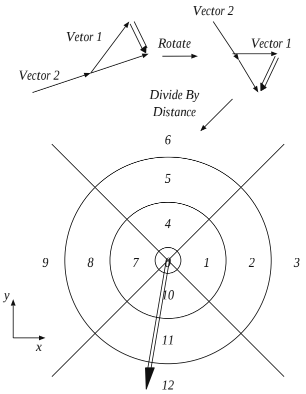


Figure 9: Feature Extraction about Relative Motion Vector

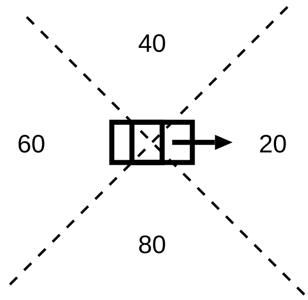


Figure 10: Feature Extraction about Relative Location

3.2 Feature Extraction about Relative Motion Vectors

By using the tracking results, we developed a method to recognize situations between a pair of vehicles which is generally applicable to intersections of any geometry.

We design our algorithm so as to be independent of geometric factors, such as geometry of the intersection, angle of video camera and position where the accident occurred. Yamato[14] shows successful results in gesture recognition of a tennis player. In this paper[14], images themselves were divided into blocks and observations were extracted directly from images within blocks. Though successful, positions of a video camera and tennis player are supposed to be fixed. Considering our purpose, vehicles run in so many ways and accidents may occur everywhere in a image frame. Therefore such geometric dependencies increase the amount of training data for such supervised learning methods as HMM. Thus, we intentionally avoid using image intensity itself for the features, because it depends on the color of vehicles. We also avoid employing motion vectors themselves with the features. We then defined features for HMM as follows (Figure.9).

$$\overrightarrow{V_n} = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} (\overrightarrow{V_2} - \overrightarrow{V_1})/d_{12} \tag{6}$$

3.3 Feature Extraction about Relative Locations

In order to classify situations with respect to a pair of vehicles, relations between the locations of the vehicles are necessary as well as relative motion vectors of the vehicles. Here, we call such relative locations between a pair of vehicles as 'relative locations'.

For example, although the relative motion vector between Vehicle-1 and Vehicle-2 in Figure8(a) is equal to the one in Figure8(b), situations are different between the two figures. And also, although the motion vectors differentiation in Figure8(c) is equal to the one in Figure8(d), situations are different between the two figures. When Vehicle-2 is located behind Vehicle-1, the observation with respect to a location of Vehicle-2 relative to Vehicle-1 is '60', and when Vehicle-2 is located on the left side of Vehicle-1, the observation of Vehicle-2 against Vehicle-1 is '40'.

3.4 Combined Observations of Relative Motion Vectors and Relative Locations

Even if the relative motion vectors are equal, situations differ depending on difference in relative locations. Therefore, in order to classify situations with respect to a pair of vehicles, it is necessary to consider features obtained from both relative motion vectors and relative locations. It seems to be appropriate to define a combined observation as summation of an observation number obtained from relative motion vectors and an observation number obtained from relative locations. Since observation numbers obtained from relative motion vectors are from '0' to '12' that are less than 20, and since observations obtained relative locations are multiples of 20, such a combined observation are guaranteed to stand for unique situation.

For example as in Figure.8(a), an observation number of Vehicle-2 relative to Vehicle-1 according to relative motion vector between them is '7', '8' or '9', and an observation number according to relative locations will be '20'. Thus an observation number of Vehicle-2 relative to Vehicle-1 will be '27', '28' or '29'. In this case an observation of Vehicle-1 relative to Vehicle-2 is equal to an observation of Vehicle-2 relative to Vehicle-1. However in the case of 8(c), an observation number of Vehicle-2 relative to Vehicle-1 as '27' etc differs from that of Vehicle-1 relative to Vehicle-2 as '61' etc.

Hereby, combined sequences will be categorized into following three types:

[pas-0] 27 27 88 88 89 87 87 87 87 67 67 67
27 28 88 88 88 88 87 87 87 67 67
27 27 20 27 28 87 88 88 88 88 87 87 67 67

[pas-1] 27 27 47 47 47 47 47 47 47 67 67 67
27 27 27 28 47 48 48 48 48 47 47 67 67
27 27 27 48 48 48 49 68 68 68 68

[pas-2] 61 61 42 42 43 41 41 41 41 21 21 21
61 60 60 61 62 41 42 42 42 42 41 41 21 20 21
62 62 62 43 43 43 42 22 22 22 22

[pas-3] 61 61 82 82 83 81 81 81 81 21 21
61 60 60 61 62 81 82 82 82 82 81 81 21 21
62 62 62 83 83 83 82 22 22 22 22 21

Here illustrated three examples for each type of pas-0, pas-1, pas-2, and pas-3. Each sequence out of three examples appear slightly different with each others because of disturbances in observation quantizations or periods of situation occupancies. Therefore, the method which is robust against such disturbances is considered to be appropriate for classification of such sequences.

By using above three types of observation sequences, pas-0, pas-1, pas-2, and pas-3, situations between a pair of vehicles can be classified as follows:

[PAS-00] Figure.8(a) : The sequence of Vehicle-2 observed from Vehicle-1 is as pas-0, and the sequence of Vehicle-2 observed from Vehicle-1 is as pas-0.

[PAS-11] Figure.8(b) : The sequence of Vehicle-2 observed from Vehicle-1 is as pas-1, and the sequence of Vehicle-2 observed from Vehicle-1 is as pas-1.

[PAS-02] Figure.8(c) : The sequence of Vehicle-2 observed from Vehicle-1 is as pas-0, and the sequence of Vehicle-2 observed from Vehicle-1 is as pas-2.

[PAS-13] Figure.8(d) : The sequence of Vehicle-2 observed from Vehicle-1 is as pas-1, and the sequence of Vehicle-2 observed from Vehicle-1 is as pas-3.

In order to recognize ordinary situations, observation sequences of every pairs of vehicles will be examined by a classification method to determine into which class out of PAS-00, PAS-11, PAS-02 and PAS-13 they most likely to classified, as described in subsection.3.5.

3.5 Classification utilizing HMM

Classifying sequences is a class of recognition problems to classify time series observations. There are various techniques for time-series event-recognition, such as DP-Matching, Neural Network and Hidden Markov Model. At first, we prefer methods based on stochastic models to those based on concrete models, because an accident sequence consists of a large number of random processes and those can be described by using stochastic models. DP-Matching is a kind of strict recognition method which is not robust against disturbance in observations. On that point, stochastic methods utilizing Hidden Markov Model(HMM) or Neural Network(NN) are superior to DP-Matching. In addition, we prefer methods that are robust against disturbance in length of observation sequences. HMM is considered to be superior to NN on that point. Thus, we decided to apply HMM to accident detection. We then applied simple left-to-right HMM for accident detection, as shown in Figure.11. A HMM model has parameters as follows.

- a_{ij} : Transition probability from state i to state j. Where, $a_{ij} = 0(j \neq i, i+1)$.
- $b_{ij}(k)$: Probability to output observation k when transition from state i to state j occurred. Where, $b_{ij}(k) = 0(j \neq i, i+1)$.

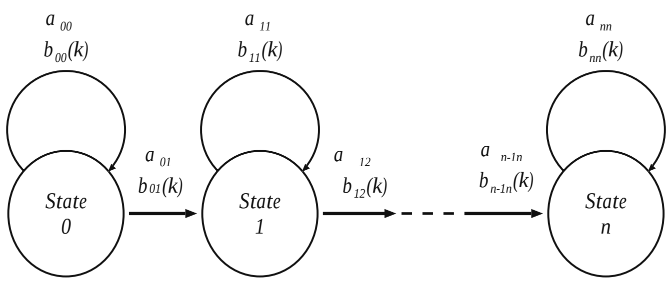


Figure 11: Left-to-Right HMM

Parameters $a_{ij}, b_{ij}(k)$ are trained by Baum-Welch algorithm[15][16]. For recognition, Trellis variables are calculated inductively as follows. Here we define forward variables $\alpha_t(i)$ and backward variables $\beta_t(i)$. $\alpha_t(i)$ is defined as the probability of the partial observation $o_1 o_2 \cdots o_t$ and state-i at time t, given the model λ.

[**Trellis Calculation**]

1. Initialization

$$\alpha_1(j) = a_{0j} b_{0j}(o_1) \qquad j = 0, 1 \tag{7}$$

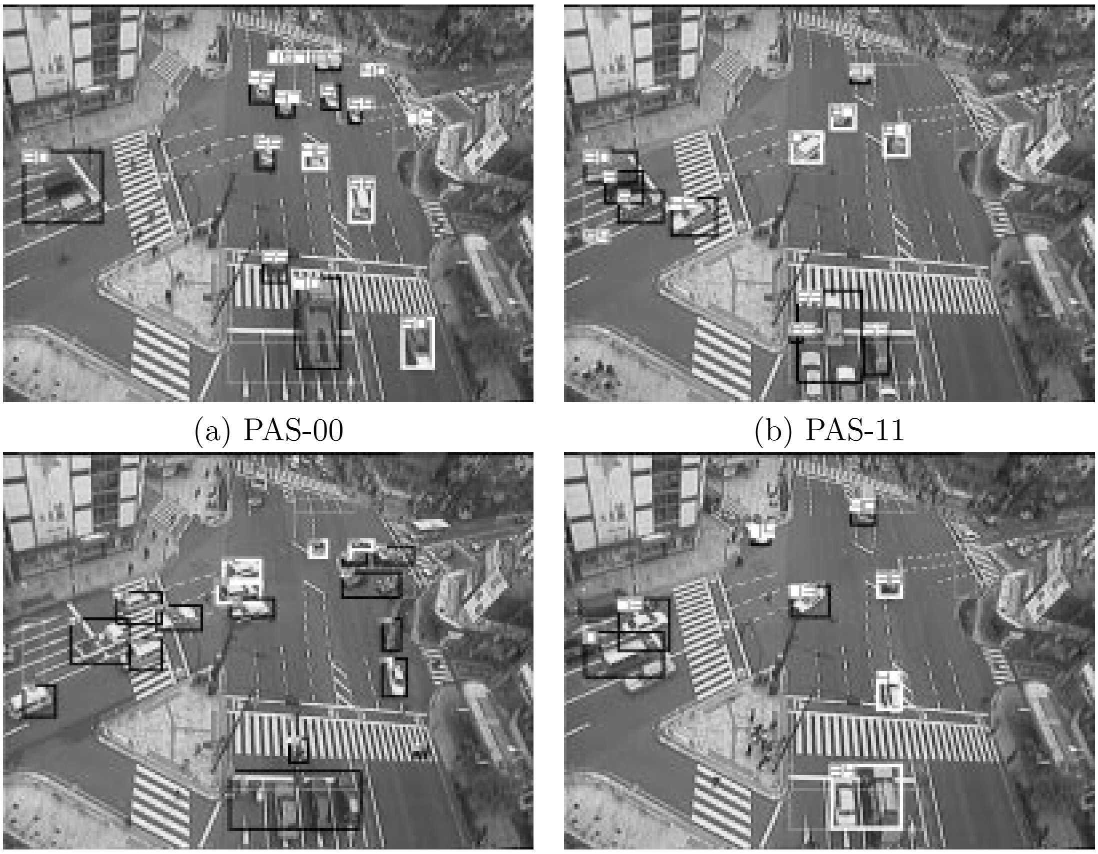
2. Induction

$$\alpha_{t+1}(j) = (\sum_{i=1}^{N} \alpha_t(i) \alpha_{ij}) b_{ij}(o_{t+1}) \qquad j = 0, 1 \tag{8}$$

3. Termination

$$P(o_1 o_2 \cdots o_T | \lambda) = \sum_{j=1}^{N} \alpha_T(j) \tag{9}$$

The model λ is provided with each category to be recognized. And the model which has the most likely probability with a test sequence is determined.



(a) PAS-00 (b) PAS-11

(c) PAS-02 (d) PAS-13

Figure 12: Recognition Results for PAS-00, PAS-11, PAS02, PAS13

3.6 Experimental Results

Four figures in Figure.12 show results of classifying combined observation sequences of a lot of pairs of vehicles into classes of PAS-00, PAS-11, PAS-02, PAS-13 by using HMM. In Figure.12(a), vehicles of ID-50,56 were detected as PAS-00 with respect to the vehicle of ID-46 and they were indicated in white rectangle. Here, vehicles other than ID-46 were out of examination for simplicity of the figure. In Figure.12(b), vehicles of ID-0,87 were detected as PAS-11 with respect to the vehicle of ID-91. In Figure.12(c), vehicles of ID-55,65,77 were detected as PAS-02 with respect to the vehicle of ID-23. In Figure.12(d), vehicles of ID-38,41 were detected as PAS-13 with respect to the vehicle of ID-34.

Table.1 shows success rate of classifications among the four classes of PAS-00, PAS-11, PAS-02, PAS-13. Those success rate were obtained by investigating about tree hundred pairs of vehicles with respect to each class. The classification result of PAS-11 is 85.2% and it is lower than other classes of PAS-00, PAS-02, PAS-03. Here, since vehicles run in the left side in Japanese traffic rules, it is important to remind that PAS-11 situations are observed when pairs vehicles make right turns simultaneously. Since vehicles stop for long periods at the right turns lane in such cases, observation sequences tend to be short than other situations. Therefore, at beginning periods of passing situations, vehicles tend to move quite slowly, and this then caused lacks of beginning parts in observation sequences about PAS-11.

Class	Success Rate
PAS-00	93.1%
PAS-11	85.2%
PAS-02	92.0%
PAS-13	88.9%

Table 1: Classification Results about Four Situations

4 Idea of Accident Detection utilizing HMM

In the previous section, recognition results of ordinary traffic situations were successful by using Hidden Markov Models. Since observation sequences of vehicles that are involved in bumping accidents will be very different from those of ordinary situations the system is expected to be able to detect bumping accidents by distinguishing obtained observation sequences of bumping accidents from those of ordinary situations by using HMM.

Since data of bumping accidents are very few, evaluation of to be proposed algorithm would not be sufficient. However, an appropriate idea of accident detections is described in this section.

4.1 Observation Sequences of Bumping Accidents

Typical observation sequences of bumping accidents are expected to be as follows. Here, observation numbers as 61, 62, 63 in tdm-0 and 27, 28, 29 in tdm-1 means that a vehicle is getting closer to the other vehicle, observation numbers as 60 in tdm-0 and 20 in tdm-1 means just the moment of an impact, and observation numbers as 69, 68, 67 in tdm-0 and 23, 22, 21 means reaction after the impact.

[tdm-0] 61 62 62 60 60 60 67 68 67
61 62 62 63 60 60 60 60 60 69 68 67 67
61 60 61 62 60 60 60 68 68 67
[tdm-1] 27 27 28 28 20 20 20 21 22 21
27 28 28 29 20 20 20 20 22 23 22 21
27 28 27 28 20 20 20 20 23 23 22 20 21

Considering the case that Vehicle-2 bumped Vehicle-1 from then behind, an observation sequence of Vehicle-2 relative to Vehicle-1 appears to be as examples of tdm-0, and an observation sequence of Vehicle-1 relative to Vehicle-2 appears to be as examples of tdm-1. Here this class of sequence is defined as TDM-01.

However, consider the case of tandem in which a vehicle come closer to the other vehicle but with no impact such as bumping accidents, its observation sequence is very similar to the one of a bumping accident. As described in subsection.3.2, observation numbers as '9' and '3' of relative motions represent dangerous situation where relative motion vector is large and distance is very close between a pair of vehicles. However which observation out of 7, 8, 9 appears depends on absolute value of relative motion vector or relative distance of a pair of vehicles. Therefore, it seems to be very difficult to distinguish observation sequences between bumping accidents and tandem situations by only investigating observation sequences themselves.

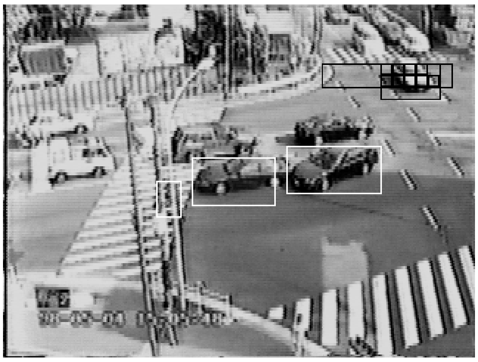
4.2 Referring to Knowledge about Traffic Rules

In order to resolve such a difficult problem, it appears to be effective to use knowledge about traffic rules.

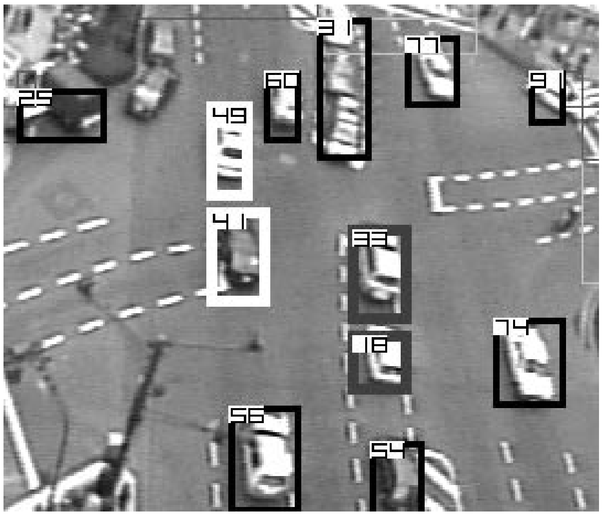
At first, if an observation sequence of a pair of vehicles is classified into TDM-01, this pair of vehicles likely caused a bumping accident. In the second, when a vehicle stopped inside the intersection, it likely to be involved in an abnormal event such as an accident or an engine failure. In the case of this section, if a pair of vehicles go through the intersection, it is not an accident but just a tandem situation. On the other hand, if a pair of vehicles stopped inside the intersection where they are not allowed to stop, they likely to have caused an accident. In the third, if other vehicles pass by the stopped vehicle, it more likely to be involved in an abnormal event. Such passing will be observed as classes of PAS-02 or PAS-13. In the case of this section, the stopped vehicles more likely to caused an accident.

4.3 Experimental Results

Since there are very few data of accidents, the HMM model for accidents and tandem situations was trained by fifty sequences of usual tandem situations including disturbance in observation numbers and length of sequences.



(a) Yokohama:Harajuku(provide by AHSRA Japan)



(b) Tokyo:Suguradai

Figure 13: Bumping Accidents

Figure.13 shows results of accident detections. In Figure.13(a), the algorithm has detected a situation which seems to be an accident by classifying the observation sequence of a pair of vehicles. Although the sequence was classified into TDM-01, it is difficult to completely determine this situation as an bumping accident by only the sequence itself. Since we have only 7.2*sec* of images, situations as PAS-02 or PAS-13 related to other vehicles was not observed.

On the other hand in Figure.13(b), the algorithm certainly has detected an accident which is caused by the vehicle of No.49 and the vehicle of No.41. At first, the algorithm detected an observation sequence related to these pair of vehicles as a class of TDM-01. In the second these vehicles stopped for a while. In the third, since several vehicles including a vehicle of No.60 have passed by these two vehicle, the algorithm decided that vehicle of No.49 and the vehicle of No.41 have caused a bumping accident.

Here, according to same algorithm and detection rules, a pair of vehicle of No.33 and No.18 was also detected as a candidate which caused bumping accident. However, this pair of vehicles stopped at right turns lane while waiting for oncoming vehicles in on the opposite lane passing through. Therefore this pair of vehicles should be discarded from candidates for bumping accidents.

5 Conclusions

In this paper, we developed a method for event recognitions out of traffic images based on Hidden Markov Model. At first, in order to gather precise information with respect to each vehicle from traffic images, we had developed occlusion robust tracking algorithm based on Spatio-Temporal Markov Random Field model. Secondly, by using this successful tracking result an extendable robust event recognition system algorithm based on the Hidden Markov Model has been developed. Classifying observation sequences that are obtained from the tracking results by this system, ordinary traffic activities at intersections were recognized at 90% success rates in average. Finally, we could reach a presumable idea of accident detection method. This method recognizes observation sequences similar to HMM model of accidents out of a lot of ordinary traffic activities. And then, by considering activities of neighbor vehicles, this system determined whether such the detected vehicle is really involved in an accident. For the future work, we plan to make the system more reliable combined with knowledge about traffic rules.

References

[1] Natan Peterfreund, "Robust Tracking of Position and Velocity With Kalman Snakes" IEEE Trans. Pattern Analysis and Machine Intelligence(PAMI), Vol.21 No.6, 1999, pp.564-569.

[2] M.Kass,A.Witkin,and D.Terzopoulos, "Snakes: Active contour models" Int'l J.Computer Vision, Vol.1, 1988, pp.321-331.

[3] S.M.Smith and J.M.Brady, "ASSET-2:Real-Time Motion Segmentation and Shape Tracking", IEEE Trans. Pattern Analysis and Machine Intelligence(PAMI), Vol.17 No.8, 1995, pp.814-820.

[4] C. Stauffer and W.E.L. Grimson, "Adaptive background mixture models for real-time tracking", Proc. of CVPR 1999, Jun 1999, pp246-252.

[5] Holger Leuck and Hans-Hellmut Nagel, "Automatic Differentiation Facilitates OF-Integration into Steering-Angle-Based Road Vehicle Tracking", Proc. of Conputer Vision and Pattern Recognition(CVPR) '99, pp.360-365.

[6] Warren F.Gardner and Daryl T.Lawton "Interactive Model-Based Vehicle Tracking", IEEE Trans. PAMI, Vol.18 No.11, 1996, pp.1115-1121.

[7] Ramesh Gangisetty, "Advanced Traffic Management System on I-476 in Pennsylvania," Proc. of IEEE ITSC'97.

[8] M. Wener and W. von Seelen, "Using Order Statistics for Object Tracking," Proc. of IEEE ITSC'97.

[9] Parag H. Batavia, Dean A. Pomerleau and Cakles E. Thrope, "Overtaking Vehicle Detection Using Implicit Optical Flow," Proc. of IEEE ITSC'97.

[10] Juan Carlos Rojas and Jill D.Crisman, "Vehicle Detection In Color Images," Proc. of IEEE ITSC'97.

[11] Nan Zeng and Jill D. Crisman,"Vehicle Matiching Using Color," Proc. of IEEE ITSC'97.

[12] Andrew H. S. Lai and Nelson H. C. Yung, "A Video-Based System Methodology for Detecting Red Light Runners," Proc. of IAPR Workshop on MVA '98, pp23-26.

[13] V. Kettnaker and M. Brand, "Minimum-entropy models of scene activity," Proc. of CVPR 1999, pp281-286

[14] Junji Yamato, Jun Ohya and Kenichiro Ishii, "Recognizing Human Action in Time-Sequential Images using Hidden Markov Model," Proc. of CVPR 1992, pp379-385.

[15] Lawrence R. Rabiner, "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition, " Proc. of the IEEE, Vol.77, No.2, Feb 1989, pp257-286.

[16] S.E. Levinson, L.R. Rabiner and M.M. Sondhi, "An Introduction to the Application of the Theory of Probabilistic Functions of a Morkov Process to Automatic Speech Recognition", Bell Syst. Tech. J. Vol.62 No.4, 1983, pp1035-1074.

[17] S.Kamijo, Y.Matsushita, K.Ikeuchi, M.Sakauchi, "Traffic Monitoring and Accident Detection at Intersections", IEEE trans. ITS, Vol.1 No.2, June. 2000, pp108-118.

[18] S.Kamijo, Y.Matsushita, K.Ikeuchi, M.Sakauchi, "Occlusion Robust Vehicle Tracking utilizing Spatio-Temporal Markov Random Field Model", 7th World Congress on ITS, Torino, Nov.2000.