Proceedings of the World Congress on Engineering and Computer Science 2009 Vol II
WCECS 2009, October 20-22, 2009, San Francisco, USA

A Method for Calculating Future Behaviors of Vehicles toward Effective Collision Prediction

Noritaka Sekiyama, and Toyohide Watanabe *

Abstract— **There are many accidents in traffic intersections because of complicated situations. In order to prevent the accidents, collisions between vehicles should be predicted and drivers should be alerted by in-car systems as soon as possible. Our research viewpoint is to detect possible collisions by predicting vehicle behaviors. We propose a region-based approach that uses "attainable region" for predicting the behaviors. In our method, traffic situations are observed by analyzing video streams of the scenes. Vehicle behaviors are transformed into the scene plane as spatial regions. With attainable regions, possible collisions between vehicles can be estimated by checking overlaps between regions. Through evaluation experiments, we show the feasibility of prediction with attainable regions.**

Keywords: visual surveillance, traffic monitoring system, behavior prediction, collision prediction, attainable region

1 Introduction

In traffic scenes, there are many participants such as vehicles, motorbikes, bicycles and pedestrians. Their behaviors influence each other, and various chains of events can lead to traffic accidents. Especially in traffic intersections, many accidents are caused by drivers' carelessness. Drivers must pay attention to all the other participants in the scene, the drivers bear heavy loads to recognize situations. In order to reduce such loads, it is effective that in-car systems help drivers' recognition to providing useful information. Our research is the first step to construct the system which reduces the number of participants that drivers must pay attention to, and lessens drivers' loads. The system focuses on a certain vehicle in the scene and divides all the other participants into two groups: dangerous participants and the others. In order to estimate whether the participant is dangerous for the target vehicle, it is necessary to predict future behavior of the participants. To predict future behaviors of the participants is, therefore, our objective. Although all types of traffic participants must be considered, only vehicles are considered as participants for simplification.

*The authors are with the Graduate School of Information Science, Nagoya University, Furo-cho, Chikusa-ku, Nagoya, 464-8603, Japan. Tel/Fax: +81-52-789-2735/3808 Email: {sekiyama,watanabe}@watanabe.ss.is.nagoya-u.ac.jp

Dangerous vehicles for the target vehicle are defined as the vehicles which will collide the target vehicle in a few seconds. It requires that the system estimates possible collisions. Collisions can be defined as an observational situation in which occupied regions of two vehicles overlap each other. Future occupied regions are important factors to estimate possible collisions between the target vehicle and others.

In this paper, we introduce a concept of attainable regions for representing future behaviors. In our method, vehicle behaviors are transformed into the scene plane as spatial regions for representing future behaviors of vehicles. By representing physical elements as spatial elements, future collisions between vehicles are estimated effectively. An attainable region means a spatial range where a vehicle can attain in the near future. Our method can easily estimate possible collisions by checking whether the regions overlap each other.

We develop a vision-based system for predicting future behaviors of vehicles as attainable regions. Our system is a kind of traffic monitoring applications in the research area of visual surveillance. We assume the environment where a traffic monitoring camera is set in high points and overlooks the intersection. Traffic situations are observed from the video stream captured by the camera.

This paper is organized as follows: Section 2 briefly reviews the related work of behavior prediction. Section 3 describes our approach and system framework. Section 4 introduces a method for tracking vehicles in the scenes. Section 5 presents our method for modeling the scenes based on trajectory clustering. Section 6 covers our method for predicting vehicle behaviors as attainable regions. Section 7 describes experimental results. The last section summarizes our paper.

2 Related Works

Traffic monitoring systems have been addressed by many researchers in visual surveillance [1]. For example, anomaly detection [2] and collision prediction [3, 4] are the major applications. The researchers aim to develop driver warning systems for assisting drivers.

Recent works in predicting behaviors of vehicles are lim-

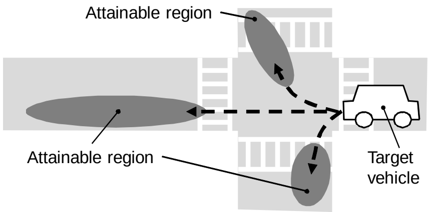

Figure 1: Attainable regions of a certain vehicle. Dashed arrows mean possible directions of the vehicle. Gray eclipses represent attainable regions.

ited in some ways. Atev, et al. [3] presented a vision-based system that issues warnings about imminent collisions on the assumption that velocities of vehicles stay constant. Although this assumption is reasonable in case of short-term prediction, it is unreasonable in case of long-term prediction. It is because the velocity and acceleration of vehicle are changeable at each frame. Saleemi, et al. [5] proposed a method for modeling and learning the scene activity based on Kernel Density Estimation (KDE) to obtain a priori knowledge in the scene. Future positions of vehicles are estimated based on their velocities and a priori knowledge. This approach, however, ignores the sizes of vehicles and is not appropriate for predicting possible collisions between moving vehicles. Hu, et al. [6] developed a method for anomaly detection and behavior prediction based on statistical learning. The system predicts the moving directions of vehicles as multiple candidate trajectories. This method enables probabilistic prediction of vehicle behavior. However, since the system cannot predict positions of vehicles, it is difficult to predict future scenes concretely.

3 Our Region-based Approach

Our approach is to use attainable regions for representing future behaviors of vehicles. An attainable region is defined as a region where a vehicle can attain in a few seconds. Figure 1 shows the attainable regions of a certain vehicle. In this figure, the vehicle is about to enter the intersection. There are three major possibilities of vehicle's behavior: turn right, turn left, and go straight. All of the possibilities must be represented, therefore, the attainable regions reflect the possibilities. The attainable regions also depend on information about the vehicle in a certain time stamp. The information includes positions, velocities, accelerations, and sizes. If the vehicle goes straight, it can be considered to attain farther because its velocity and acceleration probably become greater. Thus, the attainable region in front of the vehicle is larger than others.

In our approach, future behaviors of vehicles are projected to the scene plane. If attainable regions of two vehicles overlap each other, it means that the vehicles may collide each other in a few seconds. By estimating attainable regions of all vehicles in the scene and comparing all pairs of them, the system can predict possible collisions in the scene.

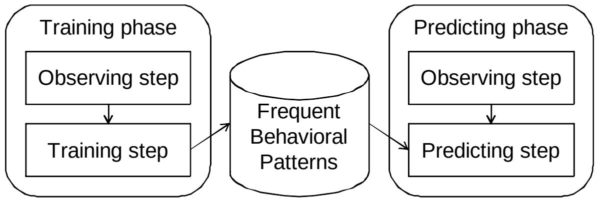

Figure 2: Our framework.

In order to predict the behaviors in a particular scene, frequent behavioral patterns provide useful knowledge. The patterns represent a priori knowledge about the scene. The future behavior of the vehicle can be considered to be similar to some of frequent behavioral patterns. The patterns are obtained by learning behavioral patterns of vehicles statistically. In addition, observed information is also important to predict future behaviors. The information includes positions, velocities, accelerations, sizes, and so on. Future behaviors of vehicles are represented as attainable regions by combining observed information and the priori knowledge.

The framework for our system is shown in Figure 2. It is divided into two phases: training phase and predicting phase. In the training phase, frequent behavioral patterns are calculated by learning vehicle behaviors in a target scene statistically. Outputs of this phase are stored in a database and are used in the predicting phase. In the predicting phase, future behaviors of vehicles are estimated by integrating the frequent behavioral patterns with observed information. Both phases have the observing step in which vehicles in the scene are tracked at each frame.

4 Vehicle Tracking

In both of the training phase and the predicting phase, the system must obtain information such as positions, velocities, accelerations, sizes, and so on. We introduce the traditional visual tracking method into our system. A flow of the observing step is shown in Figure 3.

In our situation, there are complex backgrounds in the scene. Therefore, we employ a FG/BG detector proposed by Li, et al. [7] and it is capable for our situation. The blob detection is developed using a connected component tracker [8]. For tracking objects in the scene, the object position and size are provided frame-by-frame. We use a hybrid object tracker proposed by Chen, T.P., et al. [9] in our system. It consists of two components: a connected-component tracker, and mean-shift and particle filtering

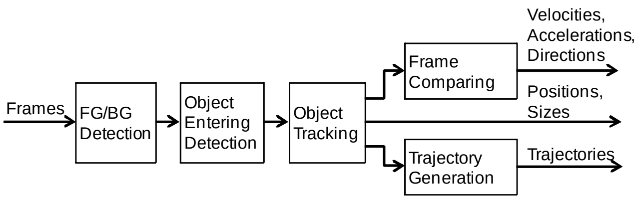

Figure 3: Observing step.

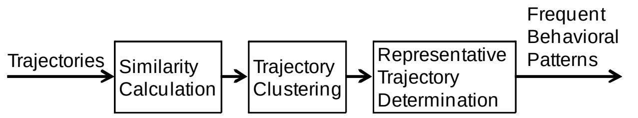

Figure 4: Training step.

based tracker. Choice of two components is dependent on existence of overlap between blobs.

Positions and sizes of objects are acquired from the tracking result directly. Velocities, accelerations, and directions are estimated by comparing each frame with its previous frame. Therefore, observed information consists of time stamps, positions, velocities, accelerations, directions, and sizes. The information of vehicle i at frame n is defined as 8-dimensional vector $\boldsymbol{O_i} = (t_n, x_n, y_n, v_n, a_n, \theta_n, width, height)$. t_n is the time stamp at frame n, x_n and y_n are the local coordinates of vehicle i, v_n is the velocity, a_n is the acceleration, θ_n is the direction, and *width* and *height* are the size of vehicle i. Time-series location trails, called trajectories, are generated by storing all positions of the same vehicle. Trajectory vectors $\boldsymbol{T_i}$ are, then, obtained from observed vectors $\boldsymbol{O_i}$ in the following definition:

$$\boldsymbol{T_i} = \{(x_b, y_b), (x_{b+1}, y_{b+1}), \cdots, (x_e, y_e)\} \qquad (1)$$

Here, b is the frame index at which the vehicle i entered the scene, and e is the frame index at which the vehicle i exited the scene.

5 Learning Scene Model

In this section, we discuss a priori knowledge of traffic scenes and propose a model for learning behavioral patterns of vehicles in the training phase. Figure 4 shows the flow of the training step.

5.1 Modeling a Target Scene

In order to predict future behavior, frequent behavioral patterns in a target scene must be considered. They are defined as patterns which are often observed in the scene. The patterns also reflect a constraint about the scene such as road alignment, traffic regulation, and so on. We focus on the patterns of vehicles as a priori knowledge of the scene. Our system obtains this information by training trajectories of vehicles.

Hu, et al. [6] proposed an algorithm for learning trajectories using fuzzy K-means clustering. Each cluster has a centroid which means a representative trajectory. The system can detect anomaly in the scenes and predict behaviors of vehicles by using these cluster centroids. This related research is based on the assumption that general behavior of vehicles can be represented by multiple representative trajectories. The system allows anomaly detection and behavior prediction by clustering trajectories of moving vehicles and using representative trajectories. The representative trajectory of each cluster reflects road alignment and traffic regulation (includes *"no U-turns allowed"*, *"contraflow"*, and so on). Automatic scene modeling reduces labor hour about environment settings for system administrators and contributes an adaptive traffic monitoring system. In our method, the system learns a traffic scene model by clustering trajectories, and obtains a priori knowledge.

5.2 Calculating Similarities between Trajectories

For clustering trajectories, similarities between $\boldsymbol{T_i}$ and $\boldsymbol{T_j}$ must be calculated. It is important to choose distance function between two trajectories. The Euclidean distance is widely used for measuring similarity between two time-series data. However, the distance cannot be applied to our problem because it can be used only if two time-series data are of equal length. More generalized similarity measurements include Dynamic Time Warping (DTW), the Longest Common Subsequence (LCSS) [10], Edit Distance on Real sequences (EDR) [11], and the Sequence Weighted Alignment (Swale) [12]. Swale can achieve greater accuracy than DTW, LCSS, and EDR. Moreover, Morse, et al. [12] have presented the Fast Time Series Evaluation (FTSE) method which can be used for evaluating LCSS, EDR, and Swale quickly. Therefore, we employ Swale as distance function for clustering trajectories and FTSE as a speed-up algorithm.

5.3 Estimating Frequent Behavioral Patterns

The group-average clustering method [13] which is a kind of aggregative hierarchical clustering methods was used in our method. The method enables effective clustering when only similarities between vehicles can be observed and cluster centroids cannot be calculated. All trajectories are clustered with Swale [12] hierarchically based on this method. A medoid obtained in the training step indicates a representative trajectory in each cluster and is used in the predicting step. Each medoid represents a representative trajectory which reflects a priori knowledge in the scene. The medoid of cluster c is defined as

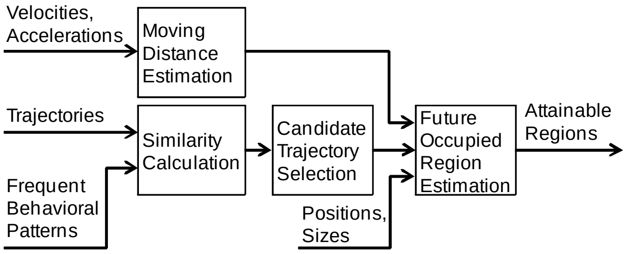

Figure 5: Predicting step.

follows:

$$m_c = \arg\max_{i \in V_c} \sum_{j \in (V_c - \{i\})} Swale(\boldsymbol{T_i}, \boldsymbol{T_j}) \qquad (2)$$

We express the vehicle index which represents medoid in cluster c as m_c. V_c is a set of vehicles included in cluster c. These medoids are acquired a training phase and used in the predicting phase as representative trajectories.

For accurate clustering, we divide a target scene into some zones. In group-average clustering, two clusters are combined only if both the entrance and the exit of each cluster's representative trajectory are the same. It reduces incorrect calculation of similarity and improves clustering accuracy.

6 Behavior Prediction

In this section, we explain a method for estimating attainable regions from representative trajectories and observed vectors (defined in Section 4) in the predicting step. Figure 4 shows the flow of the predicting step.

6.1 Calculating Similarities between Trajectories and Frequent Behavioral Patterns

When a target vehicle and a starting time of prediction are determined, the system acquires a partial trajectory from the time when the vehicle entered the scene to the starting time of prediction. Once a target vehicle is determined, a partial trajectory $\boldsymbol{P}$ of the vehicle from frame b to frame $curr$ is obtained in Equation (3).

$$\boldsymbol{P} = \{(x_b, y_b), (x_{b+1}, y_{b+1}), \cdots, (x_{curr}, y_{curr})\} \qquad (3)$$

We define a current frame index at a starting time point of prediction as $curr$ and a frame index at time point when the vehicle entered the scene as b. The partial trajectory $\boldsymbol{P}$ is, then, compared with all representative trajectories. Only representative trajectories whose similarity to $\boldsymbol{P}$ exceeds a threshold are selected as candidates of future trajectory. In the predicting step, the point is not dissimilarity but similar segments between two sequences. When two sequences are compared, one is an incomplete trajectory and the other is a complete trajectory. If we employ $Swale$, dissimilar segments between two sequences gets gap cost and a value of $Swale$ becomes larger improperly. The number of similar segments between two sequences should be counted up. LCSS is, therefore, a better choice than DTW, EDR, and Swale.

Algorithm 1 Build Rectangle List

```
Ensure: Rectangle List L
  i ← curr;
  j ← arg min dist(p_i, c_j);
  proceed ← 0;
  while proceed < proceedMax do
    move ← c_{j+1} − c_j;
    p_{i+1} ← p_i + move;
    proceed ← proceed + move;
    if proceedMin < proceed < proceedMax then
      Obtain rect_{i+1} from p_{i+1}, width, height, and θ_n;
      Insert rect_{i+1} into L;
    end if
    i ← i + 1;
    j ← j + 1;
  end while
```

6.2 Selecting Candidate Trajectories

An index set C_P of the candidate trajectories is given in Equation (4).

$$C_P = \{m_c \mid LCSS(\boldsymbol{P}, \boldsymbol{T_{m_c}}) > threshold\} \qquad (4)$$

In order to compare a partial trajectory with each representative trajectory accurately, we use zone information in this calculation. Only if the entrance of two trajectories is the same, the candidate trajectory is employed. Note that multiple candidate trajectories can be obtained. They represent possibilities of vehicle behaviors which depend on prediction. The candidate trajectories are used to determine future moving direction of vehicles. Occupied regions of vehicles at future frames can be calculated based on the trajectories.

6.3 Estimating Future Occupied Regions

The technique used to obtain future occupied regions of vehicles is shown in Algorithm 1. $\boldsymbol{P}$ is the partial trajectory of the target vehicle, $\boldsymbol{C}$ is the candidate trajectory, p_i is the i-th element included in $\boldsymbol{P}$, c_j is the j-th element included in $\boldsymbol{C}$, $curr$ is the frame index at which the current element of the target vehicle is, θ_n is the direction, and $width$ and $height$ are the size of vehicle. In this algorithm, $dist(a, b) = \sqrt{(b_x - a_x)^2 + (b_y - a_y)^2}$. $proceedMax$ and $proceedMin$ are calculated based on predicting time t_p and predetermined variable Δa which means a maximum changing rate of acceleration. $proceedMax$ and $proceedMin$ represent the range of the moving distance which depends on prediction. Each attainable region is calculated as a set of rectangle lists along the candidate trajectories.

7 Experiments

In the following, some samples of our method are first demonstrated. Performance of the algorithm for estimating attainable regions is then evaluated. In this experiment, we use the Next Generation Simulation (NGSIM) [14] data sets. We define t_p as the time range from a starting point of prediction to the target future time stamp.

7.1 Successful Examples

In the training phase, we used the video stream data between 8:30am and 8:45am. 1504 trajectories were learned and 96 clusters were calculated. In the predicting phase, we used the video stream data between 8:45am and 9:00am. In this case, predicting time t_p was 1 sec. and acceleration changing rate was 0.3 pixel/frame.

Successful examples of experimental results are shown in Figure 6. In these examples, future behaviors of vehicles are predicted accurately using attainable regions. Vehicle position, velocity, acceleration, shape, and trajectory are reflected in the prediction result. However, some failures still remain, and accuracy of whole predictions is not good because the tracking module often mistakes. In this case, observed information of vehicles is not acquired. Moreover, discontinuity of a vehicle trajectory causes that the system cannot learn vehicle behavioral patterns accurately.

7.2 Performance Evaluation

In order to break down the influence of tracking errors and evaluate the predicting algorithm purely, we use the vehicle trajectory data sets [14] which have been already tracked and contains vehicle position, velocity, acceleration, and so on. In the training phase, we used the vehicle trajectory data between 8:30am and 8:45am. Trajectories of 1210 vehicles were learned and then 270 clusters were calculated. In the predicting phase, we used the vehicle trajectory data between 8:45am and 9:00am.

We use an area ratio of the actual occupied region in predicted attainable regions as an evaluation function. If whole of the actual occupied region of the vehicle is included in the attainable regions, the area ratio is 1.

We show the performance of predicting behaviors of 100 arbitrary vehicles. The average area ratio is shown in Table 1. As the value of t_p is set shorter, the area ratio becomes greater. Our method had the best performance in case of $t_p = 0.5$. When t_p increased from 0.5, recall value did not decrease greatly. It shows the possibility for developing a long-term prediction system by using attainable regions. However, there are some prediction failures. These failures are mainly caused by three factors: incorrect clustering, error in estimating $proceedMin/proceedMax$, and abnormal behaviors of vehicles. Clusters should be allocated to each lane in the training phase. In the experimental result, however, there were negative examples including one cluster which crosses lanes. More detailed information than zone information may enable correct clustering. Moreover, accuracy about estimating moving distances of vehicles is insufficient. Our method is sensitive to tracking noises, we have to reduce the noises by using some filters. Additionally, abnormal behaviors are not able to be estimated by using our method in principle. Therefore, the abnormal behaviors should be detected and the system provides information about abnormal vehicles to the drivers.

Table 1: Average area ratio.

$t_p(sec)$	Average area ratio
0.5	0.72
1.0	0.67
2.0	0.63
3.0	0.56

8 Conclusion

In this paper, we proposed the method for predicting vehicle behavior with attainable regions. This method reflects frequent behavioral patterns of vehicles in the target scene. Experimental results showed the feasibility of prediction with attainable regions. We should solve the existing problems and improve our method.

References

[1] Hu, W., et al., "A Survey on Visual Surveillance of Object Motion and Behaviors," *IEEE Trans. on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, V34, N3, pp. 334-352, 8/04.

[2] Basharat, A., et al., "Learning Object Motion Patterns for Anomaly Detection and Improved Object Detection," *IEEE Conf. on Computer Vision and Pattern Recognition*, Anchorage, Alaska, pp. 1-8, 6/08.

[3] Atev, S., et al., "A Vision-Based Approach to Collision Prediction at Traffic Intersections," *IEEE Trans. on Intelligent Transportation Systems*, V6, N4, pp. 416-423, 12/05.

[4] Saunier, N., et al., "Probabilistic Collision Prediction for Vision-Based Automated Road Safety Analysis," *IEEE Conf. on Intelligent Transportation Systems*, Seattle, Washington, USA, pp. 872-878, 9/07.

[5] Saleemi, I., et al., "Probabilistic Modeling of Scene Dynamics for Applications in Visual Surveillance," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, (Accepted for future publication).

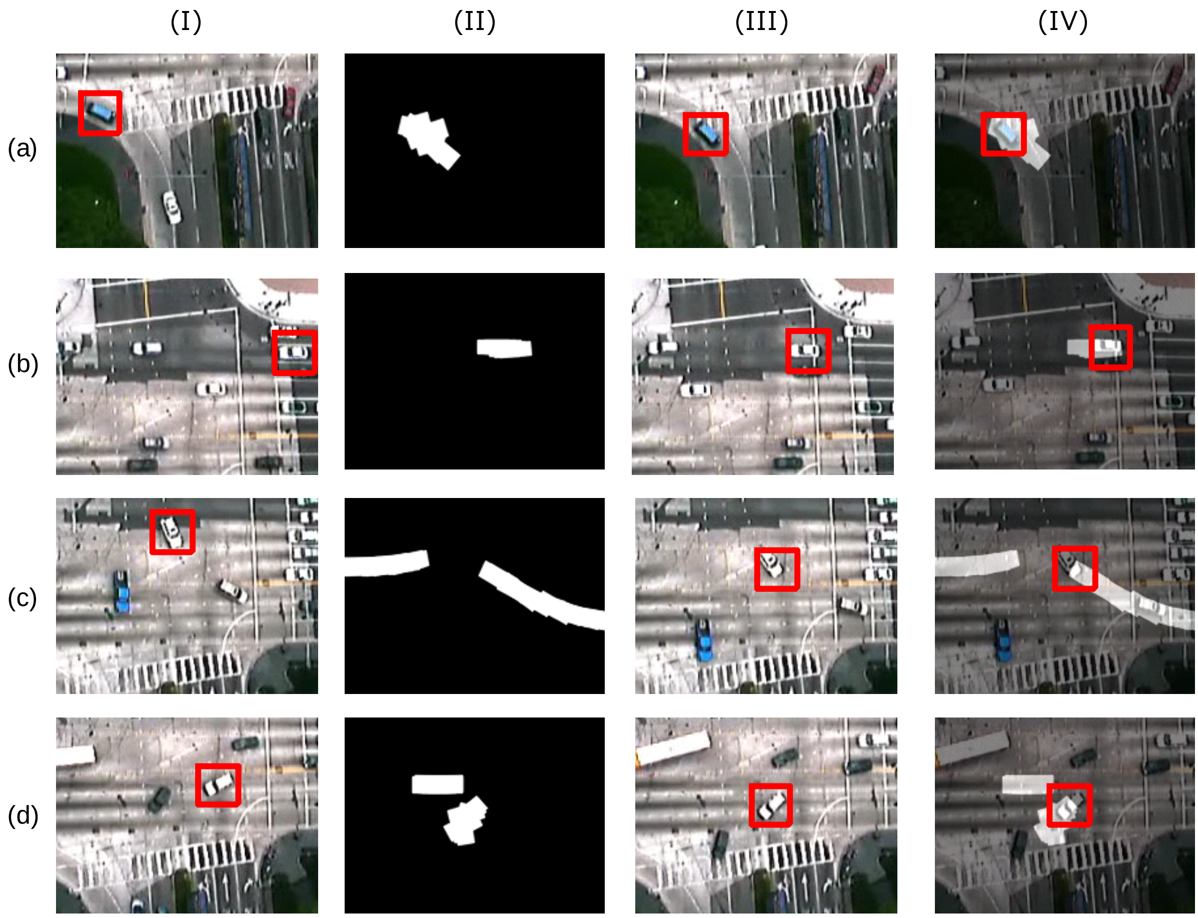

Figure 6: Four results of predicting attainable regions were shown in (a), (b), (c), and (d). In each image, the target vehicle is surrounded by a rectangle. (I) is the scene at a starting time of prediction. In (II), white regions represent attainable regions estimated in (I). In order to evaluate prediction results, we show the actual scene after t_p from the starting time of prediction in (III). The result is shown in (IV). We can compare attainable regions (estimated in a starting time of prediction) with actual occupied regions (after t_p from the starting time).

[6] Hu, W., et al., "A System for Learning Statistical Motion Patterns," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, V28, N9, pp. 1450-1464, 9/06.

[7] Li, L., et al., "Foreground Object Detection from Videos Containing Complex Background," *ACM Int'l Conf. on Multimedia*, Berkeley, California, USA, pp. 2-10, 11/03.

[8] Senior, A., et al., "Appearance Models for Occlusion Handling," *IEEE Workshop on Performance Evaluation of Tracking and Surveillance*, Kauai, Hawaii, 12/01.

[9] Chen, T.P., et al., "Computer Vision Workload Analysis: Case Study of Video Surveillance Systems," *Intel Technology Journal*, V9, N2, pp. 108-119, 5/05.

[10] Vlachos, M., et al., "Indexing Multi-Dimensional Time-Series with Support for Multiple Distance Measures," *ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining*, Washington, DC, USA, pp. 216-225, 8/03.

[11] Chen, L., et al., "Robust and Fast Similarity Search for Moving Object Trajectories," *ACM SIGMOD Int'l Conf. on Management of Data*, Baltimore, Maryland, pp. 491-502, 6/05.

[12] Morse, M.D., et al., "An Efficient and Accurate Method for Evaluating Time Series Similarity," *ACM SIGMOD Int'l Conf. on Management of Data*, Beijing, China, pp. 569-580, 6/07.

[13] Kamvar, S.D., et al., "Interpreting and Extending Classical Agglomerative Clustering Algorithms Using a Model-Based Approach," *Int'l Conf. on Machine Learning*, Sydney, Australia, pp. 283-290, 7/02.

[14] Federal Highway Administration, "Next Generation Simulation (NGSIM)," *http://ngsim.fhwa.dot.gov/*.

ISBN:978-988-18210-2-7