Vision Based Traffic Conflict Analytics
of Mixed Traffic Flow

Yen-Lin Chiu¹, Albert Y. Chen² and Meng-Hsiu Hsieh³

¹) Graduate Research Assistant, Department of Civil Engineering, National Taiwan University, Taipei, Taiwan. Email: r03521502@ntu.edu.tw
²) Assistant Professor, Department of Civil Engineering, National Taiwan University, Taipei, Taiwan. Email: albertchen@ntu.edu.tw
³) Graduate Research Assistant, Department of Civil Engineering, National Taiwan University, Taipei, Taiwan. Email: r02521527@ntu.edu.tw

Abstract:
Traffic problems caused by the mixed traffic flow mainly derive from the complicated interactions among various vehicle types. The structural differences between automobiles and motorcycles caused distinct driving behavior. Moreover, most of the research in traffic flow analytics is done in the context of developed countries where vehicle follow lane markings. However, motorcycles have very different behaviors, and the traditional automobile-based traffic theory and transportation management cannot be applied to mixed traffic streams with automobiles and motorcycles. The purpose of this study is to observe the features of mixed traffic flow, driver behavior and traffic conflict between automobiles and motorcycles. The data was collected using video photography at the intersection in Taipei. The microscopic characteristics of mixed traffic flow such as vehicle types, velocity, acceleration, trajectories are observed through computer vision and image processing methods. The Histogram of Oriented Gradients (HOG) descriptor is adapted for the detection of vehicles utilizing a Support Vector Machine (SVM), and the Kalman Filter is employed for the tracking of the vehicles’ trajectory. The traffic conflict severity was conducted by calculating the time-to-collision (TTC) and invasion rate of safety space for vehicles. The results of this study can serve as a reference for roadway safety guidance and can potentially be utilized for traffic conflict detection.

Keywords: traffic conflict, roadway safety, image processing, histogram of oriented gradient, support vector machine, time-to-collision.

1. INTRODUCTION

The motorcycle is a common mode of transport in Southeast Asia especially in Taiwan. Unfortunately, motorcyclists suffer severe injury, sometimes fatal, when traffic accidents occur. Since without the protection of the vehicle, motorcyclists are particularly vulnerable to collisions. One feasible way of addressing this problem is to separate motorcycles from other vehicles through the arrangement of exclusive motorcycle lanes.

The purpose of this study is to observe the microscopic characteristics of mixed traffic streams, driving behaviors and traffic conflicts between motorcycles and automobiles. The traffic conflict severity was measured by computing the time-to-collision (TTC) values for vehicles (Hayward, 1971) and the rate of different vehicles that invade into the motorcycle’s safety space (Nguyen, 2012; Nguyen & Hanaoka, 2013). By doing so, we expect to assess whether the exclusive motorcycle lanes are capable of reducing potential traffic collisions.

The traffic data was collected using video photography in urban roadways. Computer vision and image processing techniques have been used to extract the traffic information such as vehicle types, trajectories, velocity and acceleration. For vehicle detection, we need image descriptors to represent these image data more efficient. Histograms of Oriented Gradient (HOG) descriptor has been widely used for object recognition and outperforms the state-of-the-art features for pedestrian and vehicle detection, since it is robust to cluttered backgrounds and illumination changes (Dalal & Triggs, 2005). In order to detect motorcycles in images, the Support Vector Machine (SVM) has been utilized (Burges, 1998; Joachims, 2008). Image processing techniques, including foreground segmentation (Pakorn & Bowden, 2001; Zivkovic, 2004; Zivkovic & Heijden, 2006), morphological operation (Gonzalez & Woods, 2007) and connected components (He et al., 2008; Suzuki, 1985), also improve the detection performance. Tracking algorithms (Welch & Bishop, 1995) have been applied for relating detected objects from frame to frame in the video.

Many approaches for traffic conflict assessment have been proposed. Nguyen et al. developed a simulation model to measure traffic conflict by calculating deceleration rates at each time step when the density of motorcycle flow was changed (Nguyen et al., 2014). The concept of safety space was introduced as an approximate half ellipse in front of a motorcycle and a rectangular area in the both side to describe the driving behaviors of increasing or decreasing speed (Nguyen, 2012; Nguyen & Hanaoka, 2013). Lu et al. established a
method to quantify and classify the traffic conflict severity by analyzing time-to-collision (TTC) and non-complete braking time (TB) (Lu et al., 2012). In this study, we aim to extract the traffic information and evaluate the severity of traffic conflict with automation.

2. METHOD

The overview of the method is shown in Figure 1. First, we set up a camera on a high building to record the traffic flow video. Second, we divide images into two sets, vehicles and non-vehicles. Third, calculate the HOG features of samples and a detector is then trained by using the SVM. Before detecting vehicles, we can narrow the candidate detecting regions through foreground segmentation i.e. setting the region of interest (ROI) to reduce the amount of information to be processed by the detector. Next, Kalman Filter was adopted to form the trajectories of vehicles. Finally, kinematics information of vehicles is obtained and we can do the further analysis of traffic conflicts.

![Flow chart for research](image)

2.1 Video Clips

Since our objective is to evaluate the severity of traffic conflicts, the accuracy of traffic data of vehicles is considerably important. Hence, testing video clips were captured at nearly vertical angle with respect to the ground. Compared with lower camera angles, there are several advantages. We can obtain the wider observation region while significant reducing possible occlusions among vehicles. Moreover, in top-view the appearance of vehicles is comparatively simple and vehicle size is nearly fixed. In addition, errors caused by coordinate transformation can also be reduced.

2.2 Training of the Support Vector Machine Classifier

In the training phase of our work, all training samples were manually cropped from recorded videos and have a fixed size of 128 x 64 pixels (width x height) as suggested in Dalal and Triggs’ work (Dalal & Triggs, 2005). All the positive samples have 8 pixels of margin around image to include the surrounding information, which can improve the detection performance (Pham and Lee, 2015). Since automobiles and motorcycles have to be detected, about 700 positive and about 300 negative samples of the automobiles, and about 700 positive and about 2000 negative samples of the motorcycles were taken as input to train the corresponding classifier respectively. Re-training was performed by adding hard examples i.e. false positive detections from the SVM classifier into the negative samples set. This process greatly improves the performance of detectors.

2.3 Region of Interest Segmentation

For each frame in the traffic video clips, the pre-trained detectors scan the whole image via a detection window from the left-top to the bottom-right. This process is highly time-consuming. We restrict the detection regions through image processing. This action not only reduces the computing time on detection but also reduces the false positive (FP) rate.

(1) Background Subtraction

By using an appropriate threshold of pixel value, the background area i.e. static regions of sequential video frames can be eliminated. In this step, we employ the Adaptive Gaussian Mixture Model in the Open Computer Vision (OpenCV) Library (Intel, 2010) to extract the foreground regions. The foreground image was converted to a binary image in this process. Figure 2(a) shows the example of input image and Figure 2(b) shows the image after background subtraction.
(2) Dilation

As shown in Figure 2(b), there are a few gaps within the foreground regions, and the dilate operation can fill in the gaps and enlarge foreground regions as shown in Figure 2(c). The dilation was performed through adopting a 5*5 kernel element and reiterate 10 times.

(3) Connected Component

To set the Region of Interest (ROI) as shown in Figure 2(d), we surround the rectangle region containing the foreground regions. We use the algorithm of Suzuki (Suzuki, 1985) to find the connected component and use rectangles to wrap the component precisely. The size of Region of Interest (ROI) is not necessary to fit the vehicle size exactly because then it will be validated by SVM classifier to obtain the final detection result.

2.4 Vehicle Detection

For detection phase, the pre-trained SVM classifier scans the ROIs with a unique-scale detection window for each frame in the traffic video. Thanks to top-view observation, the size of automobiles and motorcycles are almost fixed, and therefore the detection window with a fixed size of 128x64 pixels is sufficient for the detection. This also help to reduce the computing time and eliminate false detections.

Because the size of the automobiles and motorcycles is obviously different, it is accessible to distinguish them from the ROIs size. However, the background subtraction sometimes temporarily failed due to illumination changes and camera instability. Besides, if two vehicles are too closer to each other, they may be regulated in the same ROI. For the reasons mentioned above, the SVM classifier is still indispensable for vehicle detection.

2.5 Tracking

The Kalman Filter is utilized to relate the detections between neighboring frames. When a vehicle is first detected, the Kalman Filter will be triggered to predict the estimated location of the vehicle in the subsequent frames. Then, the estimated location will be assigned to the closest detected object. The lateral and longitudinal displacement of the same vehicle between the previous frame and the current frame should not exceed given thresholds respectively so as to prevent unreasonable mismatches.

Sometimes the vehicles are not detected by reason of camera oscillation, changes in illumination and surrounding background. When the vehicle is missing, the predicted position is referred to as the probable detected position. If the prediction still cannot match up with a detection over 10 frames, it is assumed that the vehicle has passed and Kalman Filter will stop tracking.
2.6 Time-to-collision (TTC)

After getting the kinematic parameters of vehicles, the further analysis of traffic conflicts is conducted. The time-to-collision (TTC) can reflect the interaction intensity among vehicles and has often been used to assess the safety and risk in road traffic. Time-to-collision (TTC) was defined by Hayward (1971) as “the time required for two vehicles to collide if the speed and path of the both vehicles remained unchanged.” This is to say, with the current state to project the future trajectories so as to predict whether a collision will occur and when to happen. It is straightforward that lower TTC values correspond to higher traffic conflict severities. According to the definition mentioned above, the calculation of TTC can be conducted by using a simple mathematical formula.

\[
TTC_i = \frac{X_i(t) - X_j(t) - l_i}{\dot{X}_i(t) - \dot{X}_j(t)} \quad \forall \dot{X}_i(t) > \dot{X}_j(t)
\]

(1)

Where \(\dot{X}_i\) and \(\dot{X}_j\) indicate the speed of vehicle \(i\) and vehicle \(j\) respectively; \(X_i\) and \(X_j\) indicate the position of the speed of vehicle \(i\) and vehicle \(j\) respectively; \(l_i\) indicates the length of vehicle \(i\) (Minderhoud, M. M. and Bovy, P. H. L., 2001).

![Figure 3. Concept of time-to-collision (TTC)](image)

3. RESULTS

The testing video is recorded from a Panasonic Lumix DMC-GF6 camera and with the resolution of 1920*1080 pixels. The camera was mounted at 40 meters above to record the traffic streams at almost vertical angle with respect to the roadway. The video is 25 fps and the length is around 15 minutes. The presented methodology is implemented in C++ with the OpenCV Library (Intel, 2010) on a computer running OS X Yosemite 10.10.5 operating system with 2.4 GHz Intel Core i5 processing unit and 4GB of DDR3 memory. The SVM being utilized in this work is SVMLight (Joachims, 2008).

![Figure 4. Vehicle Trajectories](image)
Some of the vehicle trajectories obtained from the traffic video are shown in Figure 4. These trajectories are color-coded centroid point of the bounding boxes which was regarded as the same vehicle between frames. This also demonstrates why we need to relate the detections between neighboring frames through Kalman filter. Labels attached to each trajectory denote vehicle type and ID number. As shown in Figure 4, the vehicles are tracked consecutively over time and space. The visualization of vehicle trajectories can contribute to comprehend the tracking accuracy and subsequent improvements.

Our main purpose is to obtain the traffic information from the recorded video to conduct the following analysis instead of vehicle counting and classifying. Therefore, the definition of a valid detection indicates that the trajectory length of individual detection should be greater than the given threshold \( T_{\text{valid}} \). On the other hand, since bicycles and pedestrians will also be detected because of their similar appearance to the motorcyclists, we attempted to eliminate these false detections by checking their speed. The detection results and performance of detectors tested on a 15 minutes traffic video are shown in Table 1.

<table>
<thead>
<tr>
<th></th>
<th>Automobiles</th>
<th>Motorcycles</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ground Truth (GT)</td>
<td>213</td>
<td>203</td>
</tr>
<tr>
<td>True Positive (TP)</td>
<td>195</td>
<td>191</td>
</tr>
<tr>
<td>False Positive (FP)</td>
<td>13</td>
<td>8</td>
</tr>
<tr>
<td>False Negative (FN)</td>
<td>18</td>
<td>12</td>
</tr>
<tr>
<td>Precision (P) = TP / (TP+FP)</td>
<td>93.8%</td>
<td>96.0%</td>
</tr>
<tr>
<td>Recall (R) = TP / (TP+FN)</td>
<td>91.1%</td>
<td>94.1%</td>
</tr>
</tbody>
</table>

The major detection errors result from the vehicles’ colors. After background subtraction, the regions of vehicles whose color is close to the roadway will be severely eroded or even broken into pieces, causing incomplete foreground areas and unexpected Region of Interest (ROI). It is supposed that lowering the threshold value for background subtraction should help to decrease the false positives and false negatives.

![Figure 5](image_url)

Figure 5. Composition of Safety Space and Relative Spatial Position

By accumulating the relative position of vehicles around any motorcycles in the test video, heat map was generated as shown in Figure 5(a). The white area represents the position of motorcycle and its proceeding direction is rightward in the figure. The length and width of motorcycle in the middle is set as 1.8 meter and 0.6 meter respectively that reference the specification of GT 125 Super II of SANYANG. The red blobs can be regarded as the areas that vehicles appear most frequently. In contrast, the regions with navy blue are the areas that motorcycles appear most sparsely.

Especially, there is a sparse area in the middle of Figure 5(a) where surround the motocycle, and those area can be described as the safety space of local motorcyclists keeping with other vehicles. One of the objectives is to compare the result with the safety space proposed by Nguyen (Nguyen, 2012) with our patterns. We substitute
the average speed, 30 kilometer per hour, measured by our system to draw out the safety space and superimpose with the relative spatical figure together in Figure 5(b). The black curve in front of motorcycle is the boundary of safety space proposed by Nguyen. It can be seen that the right front region of the motorcycle is a little invaded but the other part of the safety space is almost clear.

4. DISCUSSION

There are few roads with exclusive motorcycle lanes in Taiwan, and hence it is difficult to find an appropriate location to capture the traffic video in top-view. To avoid the occlusions from street lamps, trees and others facilities, we will take advantage of the unmanned aerial vehicle (UAV) to conduct the future data collection. In our preliminary tests, UAV was able to record the stable and proper perspective of traffic video.

From equation (1) in section 2.6, it can be seen that the TTC evaluation is applied only in one-dimension, and solely consider the case of rear-end collision in the specific lane. Unfortunately, since the vehicles driving in different lanes might also have a crash such as side collision and turn collision, it is necessary to expand TTC calculation into a 2-dimension problem (Hou et al., 2014). As a result, the kinematic parameters are vectors rather than scalars. Besides, the trajectories collected in the tracking phase assume vehicles as points, but we need to consider the body size of vehicles in TTC calculation.

5. CONCLUSIONS

In this work, we established a reliable method for traffic conflict analysis by utilizing computer vision and image processing techniques. In contrast to manual observation, the proposed approach provides an alternative and more automated way for traffic information extraction under mixed traffic flow. The detection results on the test video shows superior performance with precision of 93.8% and 96.0% for automobiles and motorcycles, respectively.

More video clips are to be recorded for the following analysis. Especially, roadways with different traffic conditions, including different number of lanes, time and traffic volume are to be conducted so as to compare the changes of traffic conflict severity under those conditions. In addition, we hope to provide a meaningful index of potential traffic conflict and help transportation engineers to make better decisions.

REFERENCES

Pakorn, K. and R. Bowden (2001). An improved adaptive background mixture model for real-time tracking with
shadow detection, *Video-Based Surveillance Systems*, 135-144.


