

Safety Quantification of Intersections Using Computer Vision Techniques

Mohammad Shokrolah Shirazi^(✉) and Brendan Morris

University of Nevada, Las Vegas, Las Vegas, USA
shirazi@unlv.nevada.edu, brendan.morris@unlv.edu

Abstract. Vision-based safety analysis is a difficult task since traditional motion-based techniques work poorly when pedestrians and vehicles stop due to traffic signals. This work presents a tracking method in order to provide a robust tracking of pedestrians and vehicles, and quantify safety through investigating the tracks. Surrogate safety measurements are estimated including TTC and DTI values for a highly cluttered video of Las Vegas intersection and the performance of the tracking system is evaluated at detection and tracking steps separately.

1 Introduction

Intersection safety is one of the most important transportation concerns due to complex behavior of vehicles and pedestrians and their interactions which might lead to accidents. Around 2 million accidents and 6,700 fatalities in the United States occur at intersections every year which constitutes 26% of all collisions [1, 2].

Although safety is emerging as an area of increased attention and awareness, it is difficult to assess due to the lack of good predictive models of accident potentials and lack of consensus on what constitutes a safe or unsafe facility. There are two major methods for intersection safety analysis. The basic method uses data mining techniques on real accident datasets to find the contributing reasons. Since there are availability and quality problems regarding collision data, some studies rely on traffic conflict analysis as an alternative or complementary approach to analyze traffic safety.

The correlation between accidents and conflict-based safety measurements [3] encourage researches to use surrogate safety measurements for safety quantification. Surrogate safety measurements are reliable and consistent in definition, and they are proven to be a practical metric for safety analysis [3]. The surrogate safety measurements such as time to collision (TTC), distance to intersection (DTI), and time to intersection (TTI) are usually estimated through the video frames. TTC is defined as a time for two vehicles (or a vehicle and pedestrian) to collide if they continue at their present speeds on their paths [3].

Vision-based techniques are used for automatic detection, tracking, and safety analyses of vehicles and pedestrians from the sequence of images (i.e. videos) [4]. For instance, Sayed et al. [5] assessed vehicle-vehicle conflicts by calculating TTC values. Low TTC value indicates the severity of the near-accident

events. Further, computer vision techniques are the useful tools to analyzing vehicle-pedestrian conflicts since the collisions involving pedestrians are less frequent than other collision types. As an example, Zaki et al. [6] showed the maturity of computer vision techniques to estimate pedestrians' conflict and violation by calculating TTC values.

The performance of a vision-based safety quantification method is directly affected by the underlying detection and tracking algorithm. Vision based methods usually use motion as cue such as optical flow and background subtraction; however, they are not robust to track the temporarily stopped vehicles and pedestrians at intersections due to traffic signals [7]. As a result, DTI and TTI values can not be efficiently estimated and appearance-based methods are required to improve safety analyses results.

This work presents a robust tracking system in order to provide the reliable vehicle and pedestrian trajectories at intersections. The system benefits contextual fusion of appearance and motion at detection step followed by improved version of optical flow tracking for pedestrians and vehicles. Since stopped participants are tracked, more accurate estimation is provided in comparison with traditional methods (i.e., optical flow). The proposed system was evaluated separately at detection and tracking steps, and DTI, TTI, and TTC values for one of the Las Vegas intersections were estimated through a video. The remainder of the paper presents more details regarding tracking system and safety quantification process.

The paper is organized as follows: Sect. 2 presents the proposed system and Sect. 3 shows the scene preparation process. Section 4 shows experimental evaluations, and finally Sect. 5 concludes the paper.

2 Detection and Tracking System

Vehicle and pedestrian detection systems are performed through the two different modules since pedestrians require enhanced method to provide the robust trajectories. The enhancement is conducted through the detection and tracking steps.

Gaussian mixture model (GMM) [8] is used to create an adaptive background model for background subtraction method. Moving vehicles are detected in the motion area to initialize the tracks (see Fig. 1a). Since detection by motion is prone to occlusion or blob merging [7], the motion area is placed in a location in which vehicles do not usually stop even when the queue line is created behind the red signal. In addition, motion area should be close enough to the camera as it ensures stable moving objects (i.e. blobs) obtained by background subtraction.

Pedestrian detection is performed separately through another module since it is more challenging than vehicles, and it worsens for intersection videos. Pure motion-based method for pedestrian detection fails for different scenarios:

1. Although motion is widely used for highway scenarios, it is not consistent at intersection since traffic signals force pedestrians to stop. It also affects waiting time estimation of pedestrians.



Fig. 1. Defining areas (a) Motion area for vehicle detection (b) Mix area for pedestrian detection through contextual fusion

2. Pedestrians usually cross together in group and using motion is prone to detect a moving object. Detecting individuals is more appropriate for behavior analysis of individuals [7].

Appearance-based detection methods can significantly improve the pedestrian detection performance in video surveillance [7]. They recognize a pedestrian directly from an image by evaluating pixel values. Therefore, they do not need sequence of frames; instead, they use positive and negative samples to train the classifiers.

The local binary pattern (LBP) is used as an appearance-based detection method to improve the detection performance. LBP is the particular case of the texture spectrum and it has been found to be a powerful feature for texture classification [9]. LBP value of each pixel is computed by constructing 3×3 sub-window thresholded by the center value. Histograms of decimal numbers for each pixel are concatenated to provide a feature vector.

The key role of the proposed system is the contextual fusion which pools the best detections from several positives. The proposed contextual fusion [7] works at the decision level to combine the outputs from the GMM, and LBP detections in special mix areas where both detectors are active. In this method, appearance-based detections are limited to smaller processing regions for speed and reliability. The mix area is defined around signals and crosswalks and appearance-based detections out of this area are removed. Bounding box detections are grouped if they have more than 50% overlapping and the bigger one is removed for each two detections. The process continues until there are no multiple detections with more than 50% overlapping. More details are found in [7] regarding contextual fusion.

Optical flow is enhanced for vehicle and pedestrian tracking. Further, pedestrian tracking system benefits cooperation with bipartite graph to handle temporarily miss detected pedestrians. Figure 2 shows the tracking system. Final detected pedestrians (i.e., through contextual fusion in mix area) and vehicles

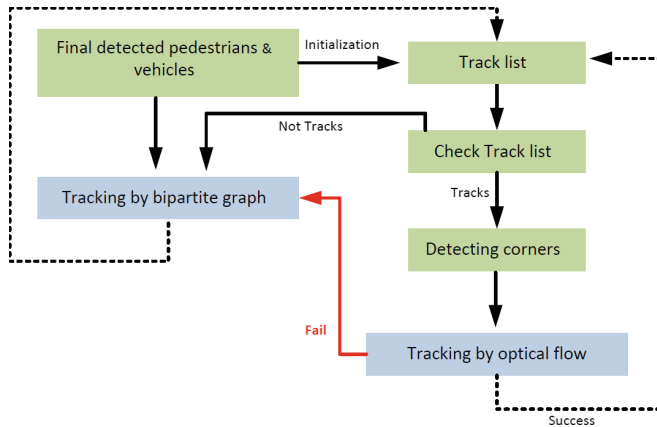


Fig. 2. Pedestrian and vehicle tracking system, red arrow only exists for pedestrian tracking (Color figure online)

(i.e. through GMM in motion area) are given to the tracking system which uses bipartite graph at first to initialize the tracks. The initialized tracks use enhanced optical flow which relies on detected corner and texture-based features inside tracks. The initialized tracks use optical flow as long as the tracking process is successful. The likelihood of successful tracking is determined based on the quality of the detected matches and the estimated bounding box around the features. If these values are less than predefined thresholds, which means the optical flow tracker failed, detected pedestrians are used by bipartite graph tracker to handle the problem. This is due to running appearance-based classifiers at each frame. If contextual fusion does not provide any detection for the bipartite tracker at the time of optical flow failure, the track of a pedestrian is finally lost.

2.1 Enhanced Optical Flow

Optical flow is a default tracking method used for initialized tracks since it is robust against partial occlusion and its effectiveness has been shown for behavior analysis of vehicles and pedestrians [5,6]. The optical flow is enhanced in this work through the three steps of enriching feature points, filtering features, and bounding box estimation.

The reduction problem of feature points is prevalent for small size objects with low quality. This worsens during the optical flow tracking for stopped or slow moving vehicles and pedestrians. The idea is to sample each vehicle and pedestrian with more feature points, called enriching feature points, to tackle this problem. Filtering features is a process of rectifying by removing falsely detected features inside a track. The filtering process is performed by determining the state of the track and filtering the opposing features. For example, when a vehicle or pedestrian state is waiting, the static features are predominant and moving

features are removed. The way of determining the state of the track is based on the average of the displacement vector. Bounding box estimation is a crucial part of the tracking process since it helps to keep the track of stopped vehicles in a queue. The bounding box estimation of stopped vehicles and pedestrians is a challenging task which leads to a tracking drift. As a result, the fixed bounding box around the fixed position is leveraged when the waiting state is determined.

2.2 Bipartite Graph

Bipartite Graph uses a greedy approach to find a nearest detection for a track [10]. The nodes of the graph are tracks and frame's detections and a cost between each two nodes is the difference between appearance measurements as calculated in (1),

$$A(v_1, v_2) = A_{pos}(p_1, p_2)A_{size}(s_1, s_2)A_{appr}(h_1, h_2) \quad (1)$$

where A_{pos} , A_{size} and A_{appr} are affinities based on position, size and appearance defined as follows.

$$A_{pos}(p_1, p_2) = \frac{1}{2\pi\sigma_x\sigma_y} \exp\left(-\frac{(x_1 - x_2)^2}{2\sigma_x^2}\right) \exp\left(-\frac{(y_1 - y_2)^2}{2\sigma_y^2}\right) \quad (2)$$

$$A_{size}(s_1, s_2) = \frac{1}{\sqrt{2\pi}\sigma_s} \exp\left(-\frac{(s_1 - s_2)^2}{2\sigma_s^2}\right) \quad (3)$$

$$A_{appr}(h_1, h_2) = \sum_{i=1}^m \sqrt{h_1^u h_2^u} \quad (4)$$

Difference of distance and size plugs into Gaussian kernel for A_{pos} , A_{size} and A_{appr} is Bhattacharyya distance calculated separately for two histograms of colors and edges. When a detection does not find any match any tracks in the track-list, a new track is created. If an existing track does not find a detection, it is marked for deletion.

3 Scene Configuration

Scene configuration includes the essential steps required to feed the tracking system by defining mix area, typical paths and training appearance based classifier. Mix area, where LBP appearance detections are also performed, are defined in regions where stopped pedestrians are expected such as around the signal and crosswalk. The GMM is used across the entire scene to account for any visible motion (see Fig. 1b). Typical paths are defined to recognize vehicles and pedestrians regarding different lanes and crosswalk. This can be used for lane-based

Table 1. Pedestrian detection performance during traffic phases

| Type | Method | Green | | Red | | Total | |
|-------------------|--------------|-------|---------|------|---------|-------|---------|
| | | TPR | Jaccard | TPR | Jaccard | TPR | Jaccard |
| Motion | Optical flow | 0.02 | 0.01 | 0.04 | 0.02 | 0.03 | 0.01 |
| | GMM | 0.08 | 0.06 | 0.18 | 0.15 | 0.13 | 0.11 |
| Appearance | LBP | 0.64 | 0.18 | 0.61 | 0.20 | 0.64 | 0.19 |
| Contextual fusion | GMM+LBP | 0.52 | 0.41 | 0.58 | 0.41 | 0.55 | 0.41 |

behavior analysis of vehicles (e.g., flow, speed) and observing pedestrian crossing behavior (e.g., crossing speed, crossing count).

Datasets are required to effectively train appearance-based classifier. A large image dataset was created by collecting public available datasets listed in [7]. Finally, 60726 and 89798 positive and negative samples were collected and prepared to train LBP classifiers. The OpenCV implementation of the LBP cascaded classifier is used by Adaboost procedure to learns weak classifiers which are combined to form a strong classifier.

4 Experimental Results

Experimental results include two steps. Detection and tracking performance of the proposed system is evaluated which is followed by the safety quantification results. The vision-based tracking system was implemented by C++ using OpenCV 2.3 and it was run on Intel i7 quad core.

4.1 Detection and Tracking System Evaluation

Detection results are evaluated using true positive rate (TPR) and *Jaccard* coefficient shown in (5). *Jaccard* coefficient is way of accounting both false positive (FP) and false negative (FN) values with one indicator. *Jaccard* value is always less or equal to TPR and its value is increased when the wrongly detected pedestrians (FP) and missed detected pedestrians (FN) are reduced.

$$TPR = \frac{TP}{TP + FN}, \quad Jaccard = \frac{TP}{TP + FP + FN} \tag{5}$$

Table 1 shows the performance of the detected methods for 1000 frames. The higher values are shown with blue and red colors for TPR and *Jaccard* values respectively. Motion-based techniques indicate the lower detection rates for green and red traffic signals. Contextual fusion always have higher *Jaccard* values since appearance-based detections are constrained to the mix area. The GMM+LBP is finally used in this work since it has the higher *Jaccard* values for all traffic phases.

Five criteria are defined to evaluate the performance of the tracking system quantitatively [11]:

Table 2. Comparison of the optical flow with the proposed tracking methods.

| Tracker | Tracking method | GT | MT | ML | FG | FT | IS | Duration |
|------------|----------------------|----|----|----|----|----|----|-------------|
| Vehicle | Optical flow [5, 12] | 42 | 18 | 5 | 12 | 0 | 6 | 2000 frames |
| | Proposed | 42 | 31 | 3 | 4 | 0 | 1 | |
| Pedestrian | Optical flow [6, 13] | 14 | 4 | 7 | 19 | 30 | 3 | 5000 frames |
| | Proposed | 14 | 8 | 4 | 5 | 17 | 2 | |

1. Number of mostly tracked (MT) trajectories: more than 80 % of the trajectory is tracked. Value should be high.
2. Number of mostly lost (ML) trajectories: more than 80 % of the trajectory is lost. Value should be low.
3. Number of fragments (FG) of trajectories: the generated trajectory is between 80 % and 20 % of the ground truth.
4. Number of false trajectories (FT): trajectories corresponding to no real object. Value should be low.
5. Frequency of identity switches (IS): identify exchanges between a pair of result trajectories. Value should be low.

The tracking system is compared against our implementation of the pure optical flow used in [5, 6, 12, 13]. The pure optical flow detects moving objects by clustering the features using motion magnitude and direction. Each moving object is initialized by features, and features find their match using KLT algorithm. The main difference in comparison with the proposed method is the lack of three introduced steps (i.e., enriching features, filtering features, bounding box estimation) by enhanced optical flow. Moreover, when pure optical flow fails due to dramatic reduction in the number of matched features, it loses its track since there is no assistance by another tracker through the detections by contextual fusion for pedestrians.

Table 2 shows the comparison of the pure optical flow with proposed tracking method. The proposed tracking method outperforms all criteria values for vehicles and pedestrians; it provides higher MT value and lower ML, FG, FT and IS values than optical flow.

4.2 Safety Evaluation

The intersection safety was evaluated by the proposed system for 18318 frames (i.e., 12:04 p.m–12:56 p.m). DTI and TTC values were calculated in this work for safety analysis. TTI and DTI are naturally conducted by drivers before taking turns to assess the level of threat posed by opposing traffic. These important safety measurements are used in safety systems including advanced driver assistant systems (ADAS) and decision support systems.

DTI is calculated based on the distance of the vehicle from the stop bar. Figure 3 shows the DTI plots of 4 vehicles. DTI value has a decreasing trend until a vehicle stops which causes DTI remains unchanged (i.e., vehicles 1 and 2).

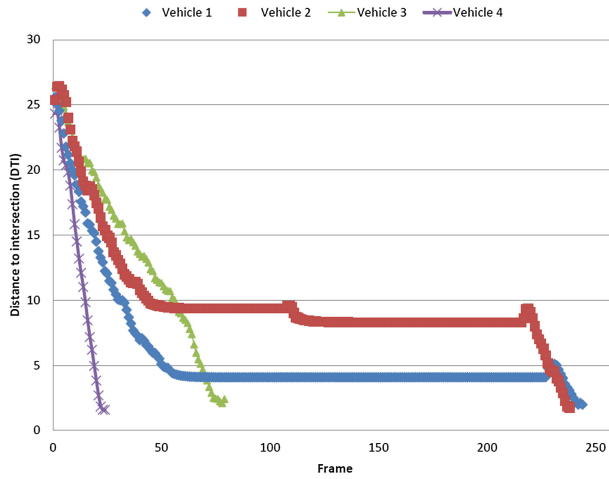
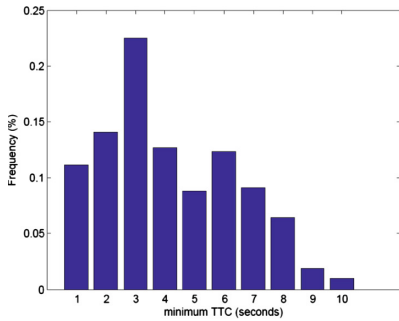


Fig. 3. DTI of four typical vehicles



(a)



(b)

Fig. 4. TTC evaluation (a) Probability density of TTC values (b) Vehicle-pedestrian conflict heatmap frequency

DTI values of vehicles 1 and 2 indicate different position of waiting vehicles in a queue behind the red signal. Vehicles 3 and 4 never face with the red signal and they continue their moving during their track life time. Vehicle 3 takes brake at some time since its slope is reduced but vehicle 4 moves with the same speed during its travel time. DTI plots can not be easily utilized with traditional vision-based frameworks since the trajectories of stopped vehicles are lost after a while.

TTC is calculated based on the predicted arrival time and it is defined as the time for two objects to collide if they continue with their present speed on their paths. Each partially observed trajectory of vehicles is compared against typical paths to find its most probable path and its associated conflict point. The time to conflict point for a vehicle is compared to those pedestrians that are moving

toward it and the minimum TTC value is counted if both timings are in a same window.

Figure 4a shows the probability density of TTC values for vehicle-pedestrian conflicts. The intersection has a peak at 3 s which classifies intersection as a medium hazardous level since TTC values of less than 2 s indicate the high severity of conflicts. Figure 4b shows the high frequency of conflict on the cross-walk between pedestrians and vehicles on the right lane. The major reason is the higher flow of vehicles on this lane and higher number of pedestrians crossing from the right to left.

5 Conclusion

This paper presents a vision-based tracking system to provide the reliable safety measurements including DTI and TTC. The proposed system benefits contextual fusion (i.e. for pedestrians) at detection level and enhanced optical flow to provide long term track of moving and stopping vehicles and pedestrians at intersections. The system was evaluated at detection and tracking levels, and it analyzed safety of a Las Vegas intersection by estimating the surrogate safety measurements. The work can be further improved by running the system for longer time periods during the peak hours and finding the contributing factors (i.e. vehicle flow, pedestrian waiting time) to vehicle-pedestrian conflicts.

Acknowledgement. The authors acknowledge the Nevada Department of Transportation for their support of this research.

References

1. Liu, Y., Ozguner, U., Ekici, E.: Performance evaluation of intersection warning system using a vehicle traffic and wireless simulator. In: Proceedings of IEEE Intelligent Vehicles Symposium, pp. 171–176 (2005)
2. Shirazi, M., Morris, B.: Observing behaviors at intersections: a review of recent studies and developments. In: 2015 IEEE Intelligent Vehicles Symposium (IV), pp. 1258–1263 (2015)
3. Chin, H.C., Quek, S.T.: Measurement of traffic conflicts. *J. Comput. Sci.* **26**(3), 169–185 (1997)
4. Shirazi, M.S., Morris, B.: A typical video-based framework for counting, behavior and safety analysis at intersections. In: 2015 IEEE Intelligent Vehicles Symposium (IV), pp. 1264–1269 (2015)
5. Sayed, T., Zaki, M.H., Autey, J.: A novel approach for diagnosing cycling safety issues using automated computer vision techniques. Transportation Research Board Annual Meeting Compendium of Papers (2013)
6. Zaki, M.H., Tarek, S., Tageldin, A., Hussein, M.: Application of computer vision to diagnosis of pedestrian safety issues. *Transp. Res. Rec. J. Transp. Res. Board* **2393**, 75–84 (2013)
7. Shirazi, M.S., Morris, B.: Contextual combination of appearance and motion for intersection videos with vehicles and pedestrians. In: Bebis, G., et al. (eds.) ISVC 2014, Part I. LNCS, vol. 8887, pp. 708–717. Springer, Heidelberg (2014)

8. Stauffer, C., Grimson, W.E.L.: Adaptive background mixture models for real-time tracking, pp. 246–252 (1999)
9. Ojala, T., Pietikainen, M., Harwood, D.: A comparative study of texture measures with classification based on feature distributions. *Trans. Pattern Recogn.* **29**, 51–59 (1996)
10. Shirazi, M.S., Morris, B.: Vision-based turning movement counting at intersections by cooperating zone and trajectory comparison modules. In: *Proceedings of 17th International IEEE Conference on Intelligent Transportation Systems*, Qingdao, China, pp. 3100–3105 (2014)
11. Wu, B., Nevatia, R.: Detection and tracking of multiple, partially occluded humans by bayesian combination of edgelet based part detectors. *Intern. J. Comput. Vis.* **75**, 247–266 (2007)
12. Saunier, N., Sayed, T.: A feature-based tracking algorithm for vehicles in intersections. *Proceedings of 3rd Canadian Conference on Computer and Robot Vision*, Quebec, Canada, p. 59 (2006)
13. Ismail, K., Sayed, T., Saunier, N.: Automated analysis of pedestrian-vehicle: conflicts context for before-and-after studies. *Transp. Res. Rec. J. Trans. Res. Board* **2198**, 52–64 (2010)