Transportation Research Part C 73 (2016) 77–90

Contents lists available at ScienceDirect

Transportation Research Part C

journal homepage: www.elsevier.com/locate/trc

A robust approach for road users classification using the motion cues

Haider Talib *, Karim Ismail, Ali Kassim

Department of Civil and Environmental Engineering, Carleton University, 1125 Colonel By Drive, Ottawa, Ontario K1S 5B6, Canada

A R T I C L E I N F O

Article history:
Received 22 May 2015
Received in revised form 11 October 2016
Accepted 25 October 2016
Available online 3 November 2016

Keywords:
Road-users classification
Fuzzy sets
Traffic safety
Automated video analysis

A B S T R A C T

Video monitoring of traffic is a common practice in major cities. The data generated by video monitoring has practical uses such as traffic analysis for city planning. However, the usefulness of video monitoring of traffic is limited unless there is also a reliable way to automatically classify road users. This paper presents an automated method of road users' classification into vehicles, cyclists, and pedestrians by using their motion cues. In this method, the movement of road users was captured on sequences of video frames. The videos were analysed using a feature-based tracking system, which has returned the tracks of road users. The separate pieces of information gained from these tracks are hereafter called Classifiers. There are nineteen classifiers included in this method. The classifiers' values were assessed and integrated into a fuzzy membership framework, which in turn required prior configurations to be available. This led to the final classification of road users. The performance of this method demonstrated promising results. An important contribution of this paper is the creation of a robust approach that can integrate different classifiers using fuzzy membership framework. The developed method also uses parametric classifiers, which do not depend on the specific geometry or traffic operation of the intersection. This is a key advantage because it enables transferability and improves the practicality and usefulness of the method.

1. Introduction

Video monitoring of traffic is a common practice in major cities in order to meet practical needs such as exploring traffic conditions, identifying traffic violations, and detecting incidents (Ismail et al., 2011). Using tracking systems such as computer vision (CV) techniques, which enable the automated analysis of video data, is also growing in popularity. In these tracking systems the paths of road users are tracked in a sequence of video frames. A track is composed of a sequence of positions that represent the movement of a road user within the field of view recorded in one frame after another. From this sequence of positions, valuable information can be extracted such as speed, acceleration, the direction of movement, and variations thereof. While tracking road users is an important task, it does not completely meet all practical needs for traffic studies. This is because a mixture of road users such as pedestrians, cyclists, and vehicles typically use traffic facilities in urban environments. Therefore, it is important to associate the extracted track with corresponding types of road user. This becomes particularly important if information needed is for a specific road user or for interactions between specific pairs of road users. Without using automated classification, having a human observer review these tracks manually, along with the traffic videos, is necessary before further analysis becomes possible. While certainly less laborious than manual tracking, this

* Corresponding author.
E-mail addresses: haider.talib@carleton.ca (H. Talib), karim.ismail@carleton.ca (K. Ismail), ali.kassim@carleton.ca (A. Kassim).

http://dx.doi.org/10.1016/j.trc.2016.10.015

still can be time-consuming if a large quantity of video data are available. This paper attempts to use motion cues obtained from road user tracks to classify road users into one of the three categories (*vehicles*, *cyclists* and *pedestrians*). As such, this classification technique does not require additional video or image analysis alongside obtained road user tracks. An important contribution of this paper is the creation of a robust approach that can integrate different classifiers using fuzzy membership framework. The developed method also uses parametric classifiers, which do not depend on the specific geometry or traffic operation of the intersection. This is a key advantage because it enables transferability and improves the practicality and usefulness of the method.

2. Previous work

Due to the importance of road users' classification in various traffic applications, researchers have developed several techniques to achieve this goal. The most common classification approach relies on the features of the road user in the video images. These features are extracted from video images using various techniques, and then used to train a classifier. Histograms of Oriented Gradients (HOG) (Dalal and Triggs, 2005) and Haar-like features (Bilgic et al., 2010) are the most common techniques for feature extractions. Adaboost algorithm (Viola and Jones, 2001) was used to construct a strong classifier by combining the best weak classifiers. Finally, the strong classifiers are integrated into classification method for final classification of road users. Support Vector Machine (SVM) (Schölkopf and Smola, 2002) is the most common methods for objects classification. Such techniques were used in many studies such as (Xu et al., 2011; Chen and Ellis, 2013; Yang et al., 2013; Nigam et al., 2013; Somasundaram et al., 2013; Hong et al., 2015; Govardhan and Pati, 2014; Hariyono et al., 2014; Ng et al., 2014; Ballesteros and Salgado, 2014; Liang and Juang, 2014). Gupte et al. (2002) and Hsieh et al. (2006) have developed a classification approach that is capable of classifying road users by relying on the dynamic parameters of the road users such as the speed and dimensions. Messelodi et al. (2005) and Buch et al. (2010), developed classification system of road users by using 3D frames model. Ismail et al. (2010); Zaki and Sayed (2013), and Zaki et al. (2013) have developed a classification system that is capable of classifying road users by relying on their movement trajectories.

3. Data collection

3.1. Site description

For the purpose of studying road users classification a video camera was used to monitor road users "vehicles, cyclists and pedestrians" movements. The case study in this paper uses two four-leg signalized intersections in the Downtown of Ottawa, Ontario. The two intersections are located where the main street Laurier Avenue West crosses with Lyon Street North and Bay Street. Laurier Avenue is a two-lane two-way street with a segregated cyclist lane on each side of the street. Lyon Street is a three-lane one-way street with bike lane on the right side; the third left lane is used as a parking lane during off-peak. Bay Street is a two-lane one-way street with bike lane on the right side. In order to have a full-view of areas of interest in each intersection, the video camera was placed at elevations of the 15th and 17th floors of two buildings overlooking the intersections. The duration of observation was 6 days, the video sequences were collected from 8:00 A.M to 5:00 P.M on September & October of 2011, and on July of 2012. Video subsequences that display the three road users co-existing in the intersections were only used in this paper. Diagrams of the intersections are presented in Fig. 1. It is noteworthy that data considered in this paper was related to cyclist movement conflicting with right-turn vehicles. For more information, the reader is referred to (Kassim, 2014; Kassim et al., 2014).

3.2. Data description

The video subsequences were analysed using a feature-based tracking system. The outputs of the tracker system include:

1. The positions of the vertices of the bounding box which encapsulates grouped features in world coordinates. The bounding box is the rectangular shape that contains the features of the tracked object. A feature is defined as "*a salient point, or sub window, which can be tracked for an adequate number of frames while in motion*" (Kassim, 2014).
2. The sequence of the average displacement components dx and dy. The displacement is the change in position of the tracked feature inside the bounding box between successive frames. The average displacement is the sum of changes in positions for these features divided by the number of these features.

4. Methodology

The classification process begins by obtaining the classifiers values by utilizing the outputs of the tracking system. There are nineteen classifiers used in this paper, and divided into five groups as follows:

- Tracked object speed: include the average and maximum speed;
- Tracked object acceleration and deceleration: include the average and maximum acceleration; the average and maximum declaration;