The Cooperative Vehicle Infrastructure System Based on Machine Vision

Daxin Tian
Beihang University
China
dtian@buaa.edu.cn

Chuang Zhang
Beihang University
China
285095869@qq.com

Xuting Duan
Beihang University
China
231665217@qq.com

Jianshan Zhou
Beihang University
China
jianshanzhou@foxmail.com

Zhengguo Sheng
University of Sussex
United Kingdom
steveshun@gmail.com

Victor Leung
The University of British Columbia
Canada
vleung@ece.ubc.ca

ABSTRACT

The information acquisition is a key procedure of cooperative vehicle-infrastructure system (CVIS). With the advancement of computer image processing technology, more and more researchers use image recognition as the source of information acquisition. On this background, the authors develop a CVIS based on machine vision, including vehicular subsystem, the roadside subsystem and the parking lot subsystem. The system uses improved Canny algorithm to detect road channelization, HOG+SVM method to detect pedestrian and Haar+Adaboost method to detect vehicle. The experiment result shows that the detection accuracy and real-time of system is relatively high. In addition, the test also prove that the system is significant in driving assistance.

KEYWORDS: CVIS, machine vision, road channelization detection, pedestrian detection, vehicle detection

ACM Reference format:
Daxin Tian, Chuang Zhang, Xuting Duan, Jianshan Zhou, Zhengguo Sheng, and Victor Leung. 2017. The Cooperative Vehicle Infrastructure System Based on Machine Vision. DIVANet'17, November 21-25, 2017, Miami, FL, USA, 5 pages. http://dx.doi.org/10.1145/3132340.3132347

1 INTRODUCTION

With the rapid growth of vehicle number in urban traffic, traffic problems are getting worse. In order to alleviate traffic pressure and improve traffic safety, more and more institutions began to study the cooperative vehicle-infrastructure system (CVIS) [3]. CVIS is a branch of the intelligent transportation system (ITS). CVIS obtain vehicle and road information based on wireless communication, sensor detection and other technology. It enables infrastructure and vehicle to cooperate intelligently, which can achieve the goals of optimizing the use of system resources, improving traffic safety and ease traffic congestion [12].

The information acquisition is a key procedure of CVIS [4]. With the advancement of computer image processing technology, more and more researchers use image recognition as the source of information acquisition [15]. Hur J [6] used conditional random fields to detect multi-lane. Sun [11] used Haar-SURF mixed features to detect vehicle. Sivaraman S [10] used support vector machines to detect vehicle. Becker D [2] used a combined motion detection and classification algorithm to detect pedestrian. This paper presents a cooperative vehicle infrastructure system based on machine vision. The system detects roads, pedestrians and vehicles at the same time, which can analysis traffic conditions more comprehensively.

This paper is organized as follows: In section 2, this paper analysis the system architecture and introduces system operation procedure. In section 3, this paper describes the image analysis method of road channelization detection, pedestrian detection and vehicle detection. In section 4, this paper shows the results of experiment. This paper concludes in section 5 with a discussion and future work.

2 SYSTEM DESIGN

2.1 System Architecture

The cooperative vehicle infrastructure system based on machine vision consists of three parts: the vehicular subsystem, the roadside subsystem and the parking lot subsystem. The system architecture is shown in Figure 1.

DIVANet'17, November 21-25, 2017, Miami, FL, USA

ACM ISBN 978-1-4503-5164-5/17/11...$15.00.
http://dx.doi.org/10.1145/3132340.3132347

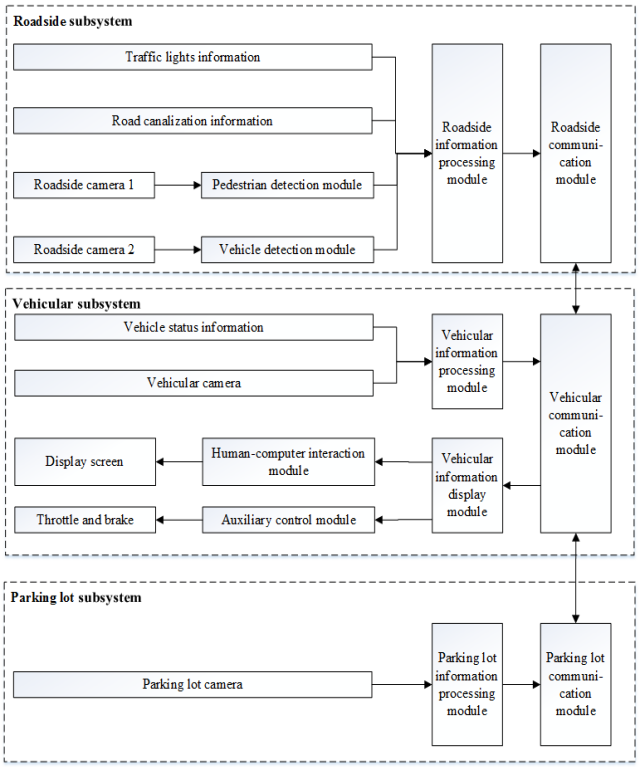

Figure 1: System architecture

2.2 System Operation Procedure

2.2.1 Information Acquisition. Firstly, there are two sets of cameras in the roadside subsystem. They are used to obtain pedestrian information and vehicle information in the intersection. Secondly, roadside subsystem can automatically read real-time signal lights information. Thirdly, on-board camera in the vehicular subsystem can obtain information about pedestrian, vehicles and roads. Finally, the camera in parking lot subsystem can obtain parking space information.

2.2.2 Information Processing. Information processing is the core step of the whole system, mainly including the following four parts: Recognizing the channelization information in the image based on the improved Canny edge detection algorithm. Detecting pedestrian information in image based on HOG (Histogram of Oriented Gradient) feature extraction algorithm and SVM (Support Vector Machine). Detecting vehicle information in image based on Adaboost machine learning algorithm. Identifying free parking space in image based on contrasting pixel changes.

2.2.3 Information Transmission. Information transmission mainly includes the following parts: firstly, roadside subsystem sends pedestrian and vehicle recognition results to the vehicular subsystem. Secondly, roadside subsystem sent intersection signal phase information to the vehicular subsystem. Thirdly, the parking lot subsystem send parking space information to the vehicular subsystem.

2.2.4 Information Fusion. When the communication module of the vehicular subsystem receives a plurality of information, it is necessary to integrate the information, such as signal lights information, pedestrian and vehicle information in the intersection. So as to consider various situations and make decisions.

2.2.5 Information Visualization. Information visualization mainly includes three parts. In roadside subsystem, the display connected to the camera can show the image results of pedestrian and vehicle detection. In parking lot subsystem, the display connected to the camera can show the image results of vehicle and empty parking detection. In vehicular subsystem, the display can show traffic lights information, the information of pedestrian and vehicle in the intersection, the free parking space information.

3 IMAGE ANALYSIS METHOD

3.1 Road Channelization Detection

In this paper, an improved Canny edge detection algorithm is used to detect channelization information. Canny edge detection is a multi-level edge detection, it has been widely used because of its high accuracy and sensitivity [14]. However, the Canny algorithm needs to manually input the image segmentation threshold, which reduces the automation degree of the whole system. Therefore, this paper combines the Maximum Between-Class Variance (Otsu's Method) with the Canny algorithm, which can input the adaptive threshold obtained by the Otsu algorithm as a parameter into the Canny algorithm, so as to improve the efficiency and accuracy of the road canalization information recognition.

Maximum Between-Class Variance (Otsu's Method) is an adaptive threshold segmentation method based on least squares algorithm [9]. The basic idea is that the image is divided into two parts according to the gray threshold, which can maximize the variance between classes and minimize variance within the class. The principle of the improved Canny edge detection algorithm is as follows: (1) Assuming that the number of pixels of the whole image is N, the gray value range is [0, L-1], the number of pixels corresponding to gray scale i is n_i For the number of pixels, the proportion of each gray level in totality is p_i. The mathematical expression is as follows.

$$p_i = \frac{n_i}{N}$$
$$\sum_{i=0}^{L-1} p_i = 1 \tag{1}$$

(2) Supposing a threshold T, which divides the image into class C_0 of the gray value in the interval [0, T] and class C_1 of the gray value in the interval [T+1, L-1]. Suppose that the average gray level of the whole image is u_t, then the relevant parameters of C_0 and C_1 are as shown in equation (2).

$$u_0 = \sum_{i=0}^{T} ip_i / \omega_0, u_1 = \sum_{i=T+1}^{L-1} ip_i / \omega_1$$
$$\omega_0 = \sum_{i=0}^{T} p_i, \omega_1 = \sum_{i=T+1}^{L-1} p_i = 1 - \omega_0 \tag{2}$$

The equation (3) can be deduced from the above four formulas.

$$u_r = \omega_0 u_0 + \omega_1 u_1 \tag{3}$$

(3) Calculating the variance between classes, as shown in equation (4). Let T take value in the range of [0, L-1] to maximize σ_B^2.

$$\begin{aligned}\sigma_B^2 &= \omega_0(u_0 - u_r)^2 + \omega_1(u_1 - u_r)^2 \\ &= \omega_0(u_0^2 + u_r^2) + u_r^2(\omega_0 + \omega_1) - 2(\omega_0 u_0 + \omega_1 u_1)u_r \\ &= \omega_0 u_0^2 + \omega_1 u_1^2 - u_r^2 = \omega_0 u_0^2 + \omega_1 u_1^2 - (\omega_0 u_0 + \omega_1 u_1)^2 \\ &= \omega_0 u_0^2(1-\omega_0) + \omega_1 u_1^2(1-\omega_1) - 2\omega_1\omega_0 u_0 u_1 \\ &= \omega_1\omega_0(u_0 - u_1)^2\end{aligned} \tag{4}$$

(4) For each point on the image, using the approximate formula of first finite difference method to calculate the partial derivative $P(\mathrm{i},\mathrm{j})$ of x and the partial derivative $Q(\mathrm{i},\mathrm{j})$ and y, as shown in equation (5).

$$\begin{aligned}P(\mathrm{i},\mathrm{j}) &\approx [g(\mathrm{i},\mathrm{j}+1) - g(\mathrm{i},\mathrm{j}) + g(\mathrm{i}+1,\mathrm{j}+1) - g(\mathrm{i}+1,\mathrm{j})]/2 \\ Q(\mathrm{i},\mathrm{j}) &\approx [g(\mathrm{i},\mathrm{j}) - g(\mathrm{i}+1,\mathrm{j}) + g(\mathrm{i},\mathrm{j}+1) - g(\mathrm{i}+1,\mathrm{j}+1)]/2\end{aligned} \tag{5}$$

(5) Calculating the gradient value $M(\mathrm{i},\mathrm{j})$ and azimuth $\theta(\mathrm{i},\mathrm{j})$, as shown in equation (6). If the $M(\mathrm{i},\mathrm{j})$ of the point (i,j) is larger than the two adjacent points in the direction of its gradient, and the angle difference between the two points is less than 45°, so then the point (i,j) is considered to be an edge pixel.

$$\begin{aligned}M(\mathrm{i},\mathrm{j}) &= \sqrt{P(\mathrm{i},\mathrm{j})^2 + Q(\mathrm{i},\mathrm{j})^2} \\ \theta(\mathrm{i},\mathrm{j}) &= \arctan\left(\frac{Q(\mathrm{i},\mathrm{j})}{P(\mathrm{i},\mathrm{j})}\right)\end{aligned} \tag{6}$$

(6) The final image segmentation result is determined by the double threshold method. Let the result T of third step be high threshold, and low threshold is half of the high threshold. In the image, the point with the gray value higher than the high threshold must be the edge point, and the point with the gray value below the low threshold must not be the edge point. For other points, removing the points around the point that is not the edge, leaving the points around the edge to form weak edges.

3.2 Pedestrian Detection

By inspection of the frequency slope of the modes, one can immediately understand the localization of modes into dots of different materials looking at their slope.

In this paper, the HOG (Histogram of Oriented Gradient) + SVM (Support Vector Machine) method is used to detect pedestrians. This method has high accuracy and robustness. When the false positive rate is 10^{-4}, the detection accuracy is as high as 89% [5]. The HOG+SVM method goes as follows:

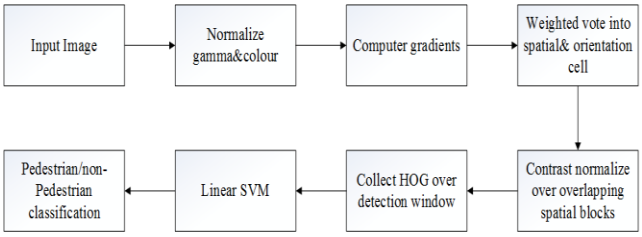

Figure 2: HOG+SVM method flow chart

(1) Normalize gamma & color

In order to reduce the influence of light factors, the entire image need to be normalized. In the texture intensity of the image, the local surface exposure contribution of the larger proportion, so this compression can effectively reduce the image of the local shadow and light changes. Because the role of color information is small, the image usually is converted to grayscale.

(2) Computer gradients

Calculating the gradient of the image in x and y directions, and calculating the gradient direction values for each pixel position, as shown in equation (7) and (8). This derivative operation not only captures the contours, silhouette and some texture information, but also further weakens the influence of illumination.

$$\begin{aligned}G_x(\mathrm{x},\mathrm{y}) &= H(\mathrm{x}+1,\mathrm{y}) - H(\mathrm{x}-1,y) \\ G_y(\mathrm{x},\mathrm{y}) &= H(\mathrm{x},\mathrm{y}+1) - H(\mathrm{x},y-1)\end{aligned} \tag{7}$$

$$\begin{aligned}G(\mathrm{x},\mathrm{y}) &= \sqrt{G_x(\mathrm{x},\mathrm{y})^2 + G_y(\mathrm{x},\mathrm{y})^2} \\ \theta(\mathrm{i},\mathrm{j}) &= \arctan\left(\frac{G_x(\mathrm{x},\mathrm{y})}{G_y(\mathrm{x},\mathrm{y})}\right)\end{aligned} \tag{8}$$

(3) Weighted vote into spatial & orientation cell

According to the coordinate (x,y) of the pixel and the gradient vector angle θ of the pixel, each pixel block is subjected to three-dimensional linear interpolation and accumulated to form a 36-dimensional vector histogram. As shown in equation (9). The purpose of this step is to provide a coding for the image area while maintaining sensitivity to the posture and appearance of the pedestrian object in the image.

$$\begin{aligned}&h(\mathrm{x}_i,\mathrm{y}_j,\mathrm{z}_k) \leftarrow h(\mathrm{x}_i,\mathrm{y}_j,\mathrm{z}_k) + \\ &w\left[1-(\mathrm{x}-\mathrm{x}_i)/\mathrm{b}_x\right]\times\left[1-(\mathrm{y}-\mathrm{y}_j)/\mathrm{b}_y\right]\left[1-(\mathrm{z}-\mathrm{z}_k)/\mathrm{b}_z\right]\end{aligned} \tag{9}$$

(4) Contrast normalize over overlapping spatial blocks

Due to the change of local light and foreground-background contrast, the range of gradient intensity is very large. This requires normalizing the gradient intensity. Normalization can further illuminate light, shadows, and edges. The normalized block descriptor (vector) is called the HOG descriptor.

(5) Collect HOG over detection window

All the overlapping blocks in the detection window are collected for the HOG feature [7] and combined into the final feature vector for classification.

(6) Linear SVM

The training module uses SVM [1] to train the pedestrian detection database to obtain a coefficient w and a threshold φ.

Then, the results of the HOG eigenvalue multiplication are compared with φ to determine whether there are pedestrians in the image.

3.3 Vehicle Detection

In this paper, the Haar+Adaboost method is used to detect vehicle. The Haar+Adaboost method goes as follows:

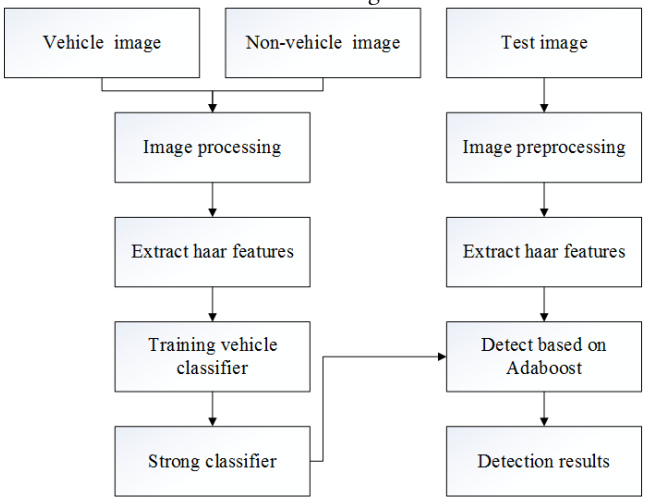

Figure 3: Haar+Adaboost method flow chart

Haar features provide information about the grey-level distribution of two adjacent regions in an image. The detection method based on Haar feature is better than other methods that directly deal with pixels [8]. That's because it is easy to encode a program with relatively limited training data. What's more, the calculation speed of feature detection is much quicker than pixel-based ones.

The number of features in different size and scale is huge. Viola [13] proposed a method of using the integral graph to calculate eigenvalues, which can improve the calculation speed. The concept of the integral graph can be represented by Figure 4: The integral graph $A(x,y)$ of the coordinates (x,y) is the sum of all the pixels in the top-left corner(the shaded portion in the Figure 4). And its mathematical definition is shown in equation (10), $i(x,y)$ is the pixel intensity in the original image of the position (x,y).

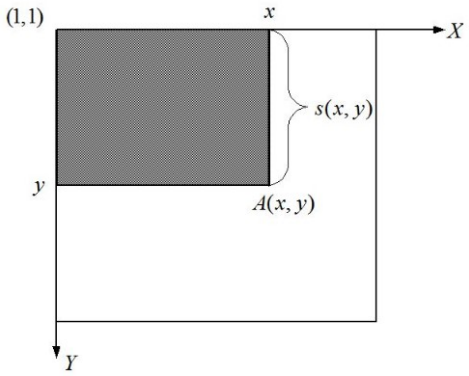

Figure 4: The concept of the integral graph

$$A(x,y) = \sum_{x'<x,y'<y} i(x',y') \tag{10}$$

4 EXPERIMENT RESULT

In this paper, a large number of tests are carried out on road channelization detection, pedestrian detection and vehicle detection algorithm based on the real time videos shot by cameras in the road, intersection and parking lot. The test hardware operating environment is the PC of Intel i7 cpu, HIKVISION 1221D Camera and TECLAST pad. The software operating environment includes Windows 10 system, Android 6.0 system, visual c++ integrated development environment, eclipse java development environment and OpenCv open source computer vision library. The test is carried out in a variety of circumstances such as strong light, obscured by obstacle, multi-lane and other complex scenes. In all cases, the detection accuracy and real-time of system is high.

In this paper, multiple real-time video clips for about 15 hours are used to calculate the detection accuracy and real-time. The average accuracy of road channelization detection based on improved Canny algorithm is 92.73%, and the average false alarm rate is 9.7%. The average accuracy of pedestrian detection based on HOG+SVM is 87.11%, and the average false alarm rate is 6.8%. The average accuracy of vehicle detection based on Haar+Adaboost is 90.24%, and the average false alarm rate is 7.3%. In order to reduce the video image delay caused by image processing, this paper adopts the method of multi-thread programming, and the average processing time on the PC platform is about 52ms for each frame in the test video.

Some test results are shown below.

Figure 5 shows the result of road channelization detection. In the left image, there is a vehicle on the road. In the right image, there are many lanes of intersection. The system can deal with these situations.

Figure 5: Road channelization detection result

Figure 6 shows the result of pedestrian detection. In the left image, the circumstance is normal. In the middle image, the light is high. In the right image, there are cars and pedestrians at the same time. The system can deal with these situations.

Figure 6: Pedestrian detection result

Figure 7 shows the result of vehicle detection. In the left image, the circumstance is normal. In the middle image, the light condition is not good. In the right image, there is a tree blocking the car. The system can deal with these situations.

Figure 7: Vehicle detection result

The detection result is sent from the PC to the vehicular tablet, Figure 8 shows the application developed for this system. There are five parts of application. Part A shows the information of signal lights. Part B and C shows the results of pedestrian and vehicle detection. Part D shows the information of parking lot. Part E shows the real-time map.

Figure 8: The application of vehicular tablet

5 CONCLUSIONS

This paper presents a cooperative vehicle infrastructure system based on machine vision, which includes vehicular subsystem, the roadside subsystem and the parking lot subsystem. The system uses improved Canny algorithm to detect road channelization, HOG+SVM method to detect pedestrian and Haar+Adaboost method to detect vehicle. The system is tested in lots of circumstances, and the result shows that the detection accuracy and real-time of system is relatively high. In addition, the test also prove that the system is significant in driving assistance.

However, there are also some drawbacks in this system. One is the adaptability of pedestrian detection is low. When the roadside camera shooting height is high, the portrait will be deformed, and the detection accuracy will be reduced. So our future work will be devoted to the improvement of pedestrian detection in some complex circumstances.

REFERENCES

[1] Sebastian Bauer, Sebastian Kohler, Konrad Doll, and Ulrich Brunsmann. 2014. FPGA-GPU architecture for kernel SVM pedestrian detection. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops*. IEEE, San Francisco, CA, USA,61-68. https://doi.org/10.1109/CVPRW.2010.5543772

[2] Daniel Becker, Bernd Schaufele, Jens Einsiedler, Oliver Sawade, and Ilja Radusch. 2014. Vehicle and pedestrian collision prevention system based on smart video surveillance and C2I communication. In *Intelligent Transportation Systems (ITSC), 2014 IEEE 17th International Conference on*. IEEE, Qingdao, China, 3088-3093. https://doi.org/10.1109/ITSC.2014.6958186

[3] R. Bishop. 2000. A survey of intelligent vehicle applications worldwide. In *Intelligent Vehicles Symposium, 2000. IV 2000. Proceedings of the IEEE*. IEEE, Dearborn, MI, USA,, 25-30. https://doi.org/10.1109/IVS.2000.898313

[4] Chao Chen, L. U. Zhiyong, F. U. Shanshan, Qi Peng, and Wuhan. 2011. Overview of the Development in Cooperative Vehicle-Infrastructure System Home and Abroad. *Journal of Transport Information and Safety (2011)*. http://en.cnki.com.cn/Article-en/CJFDTOTAL-JTJS201101027.html

[5] Markus Enzweiler and Dariu M. Gavrila. 2009. Monocular Pedestrian Detection: Survey and Experiments. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 31, 12 (2009), 2179. https://doi.org/10.1109/TPAMI.2008.260

[6] Junhwa Hur, Seung-Nam Kang, and Seung-Woo Seo. 2013. Multi-lane detection in urban driving environments using conditional random fields. *In Intelligent Vehicles Symposium (IV), 2013 IEEE*. IEEE, Gold Coast, QLD, Australia, 1297-1302. https://doi.org/10.1109/IVS.2013.6629645

[7] K Negi, K Dohi, Y Shibata, and K Oguri. 2012. Deep pipelined one-chip FPGA implementation of a real-time image-based human detection algorithm. In *International Conference on Field-Programmable Technology*. IEEE, New Delhi, India, 1-8. https://doi.org/10.1109/FPT.2011.6132679

[8] Pablo Negri, Xavier Clady, Shehzad Muhammad Hanif, and Lionel Prevost. 2008. A cascade of boosted generative and discriminative classifiers for vehicle detection. *Eurasip Journal on Advances in Signal Processing* 2008, 1 (2008), 1-12. https://doi.org/10.1155/2008/782432

[9] Nobuyuki Otsu. 1979. A Threshold Selection Method from Gray-Level Histograms. *IEEE Transactions on Systems Man and Cybernetics* 9, 1 (1979), 62-66. https://doi.org/10.1109/TSMC.1979.4310076

[10] Sayanan Sivaraman and Mohan Manubhai Trivedi. 2009. Active learning based robust monocular vehicle detection for on-road safety systems. In *Intelligent Vehicles Symposium, 2009 IEEE*. IEEE, Xi'an, China, 399-404. https://doi.org/10.1109/IVS.2009.5164311

[11] Shujuan Sun, Zhize Xu, Xingang Wang, Guan Huang, Wenqi Wu, and De Xu. 2015. Real-time vehicle detection using Haar-SURF mixed features and gentle AdaBoost classifier. In *Control and Decision Conference (CCDC), 2015 27th Chinese*. IEEE, Qingdao, China, 1888{1894. https://doi.org/10.1109/CCDC.2015.7162227

[12] B Van Arem. 2007. Cooperative vehicle-infrastructure systems: an intelligent way forward? *Advanced Trac Management Systems* (2007). https://trid.trb.org/view/1154339

[13] Paul Viola and Michael J. Jones. 2004. *Robust Real-Time Face Detection*. Kluwer Academic Publishers.

[14] Bing Wang and Shao Sheng Fan. 2009. An Improved CANNY Edge Detection Algorithm. In *Second International Workshop on Computer Science and Engineering*. IEEE, Qingdao, China, 497-500. https://doi.org/10.1109/WCSE.2009.718

[15] Junping Zhang, Fei-Yue Wang, Kunfeng Wang, Wei-Hua Lin, Xin Xu, and Cheng Chen. 2011. Data-driven intelligent transportation systems: A survey. *IEEE Transactions on Intelligent Transportation Systems* 12, 4 (July 2011), 1624-1639. https://doi.org/10.1109/TITS.2011.2158001