Machine Learning in Computer Vision

Technology Description

In the era of big data, large quantities of data call for automatic methods of data analysis are known as machine learning (Murphy, 2012). Machine learning considers a set of methods that can detect patterns in data and help in transforming the knowledge discovery into decision making. These methods have generated huge technological and social impacts in a wide range of applications such as computer vision, speech recognition, natural language processing, neuroscience, and Internet of Things (Zhou et al., 2017). Machine learning typically goes through data preprocessing, learning, knowledge discovery, and evaluation phases (see Figure 1).

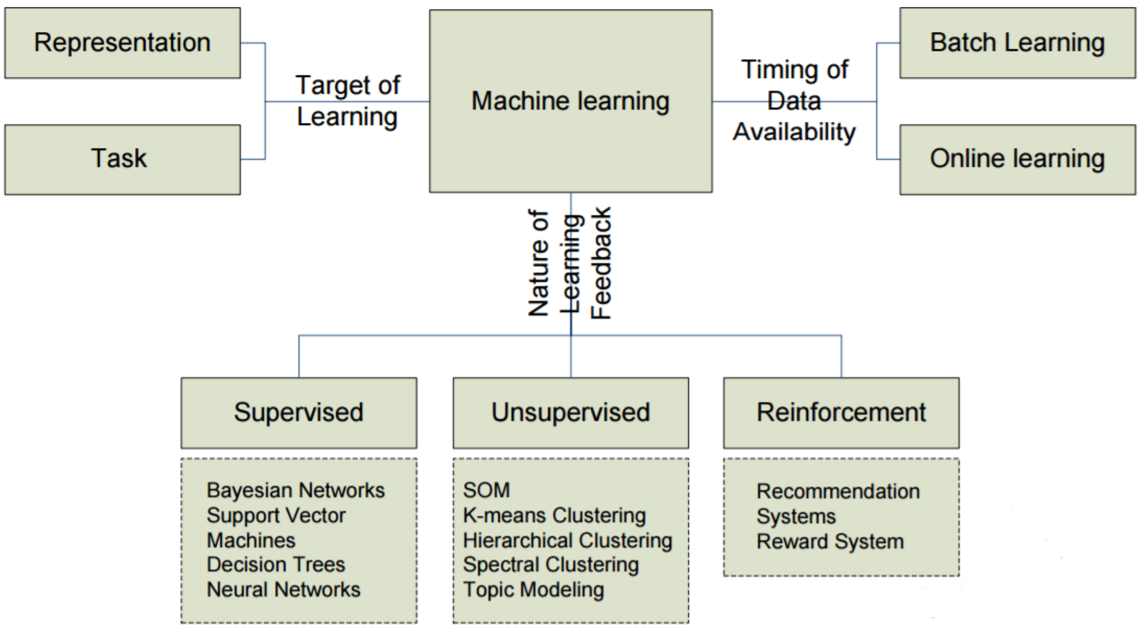

FIGURE 1. A MULTI-DIMENSIONAL TAXONOMY OF MACHINE LEARNING

Source: Zhou et al. (2017).

Computer vision is the process of using an image sensor to capture images, then using an algorithm to analyze these images to extract knowledge. Ballard and Brown (1982) described computer vision as a range of representations that connect input and output, in four parts:

- Iconic (what humans see),
- Segmented (finding the edges of an image),
- Geometric (three-dimensional data), and
- Relational (the position of objects relative to others).

Machine learning requires a lot of information to associate certain objects in a picture with other objects. For example, a picture of vehicles on a roadway should be analyzed in such a way that the vehicles can be differentiated from the roadway surroundings. To do this task properly, a computer requires sophisticated algorithms. Machine learning offers effective methods for computer vision to perform knowledge extraction. It helps to automate the model acquisition and process updating, adapt task parameters and representations, and use experience for generating, verifying, and modifying hypotheses. Some of the applications of machine learning in computer vision are: segmentation and feature extraction; learning rules; learning and refining visual models; indexing and recognition strategies; learning shape representation; and surface reconstruction strategies (Sebe et al., 2006). Since 2004, self-driving car technologies have started to use machine learning in computer vision for market ready automated technology (see Figure 2 for computer vision used in Google's Waymo[1] project).

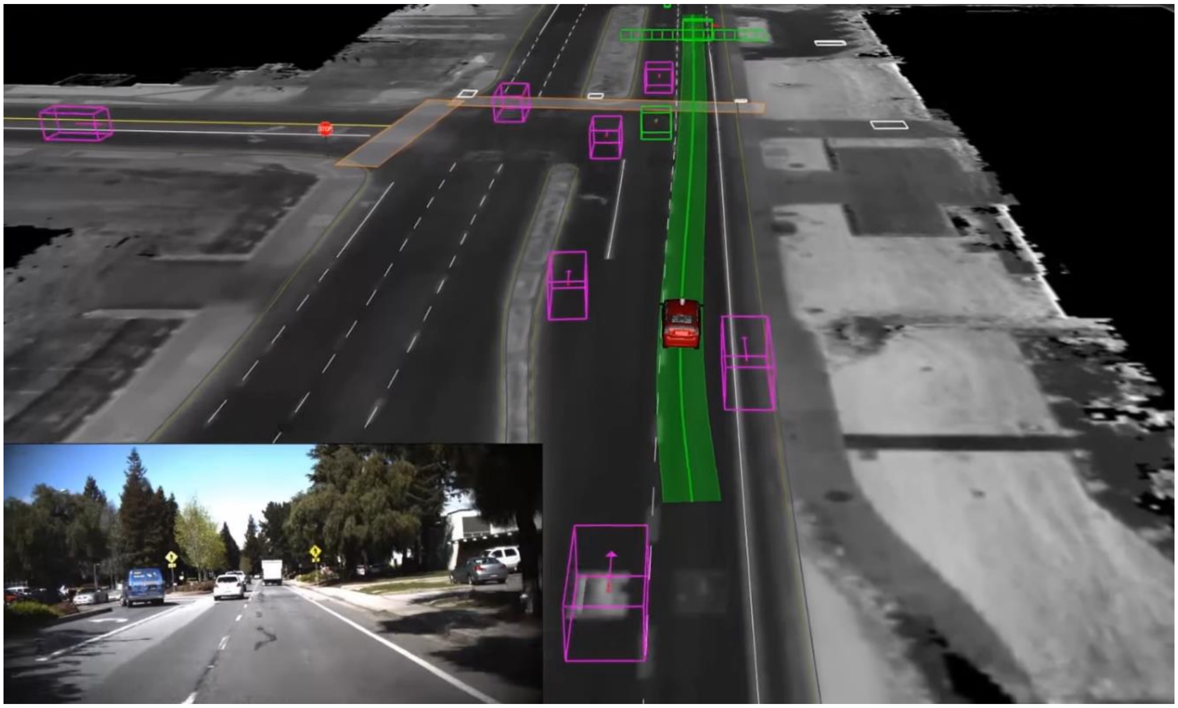

FIGURE 2. COMPUTER VISION USED IN GOOGLE'S AUTOMATED VEHICLE.

Source: Google's Waymo (2017).

Technology Capabilities (for enabling automation)

Computer vision is applicable for a much wider range of transportation engineering problems. It has capability of transforming an image into information, which is particularly useful for roadway transformation. Computer vision systems are currently used for the following applications:

- Measure distance between vehicles
- Detect lane markings, signs, and signals
- Classify and categorize vehicles
- Detect obstacles and animals
- Monitor driver behavior using facial recognition
- Detect surface cracks and other abnormalities on roadways
- Enable Advanced Driver Assistance Systems (ADAS)

Many studies were conducted to improve different functionalities of computer vision. A short review on some of the key functions is described below:

Lane Detection

Many computer vision algorithms emerged as components of fully automatic vehicle navigation systems (McCall and Trivedi, 2006). Earlier studies focused on well-paved roadway that is easily separated from its surroundings. After the DARPA Grand Challenge (DARPA, 2004), a competition between automated off-road vehicles, many stud-

[1] https://waymo.com/

ies attempted to investigate automated driving on different types of roadways. However, little progress has been made in developing a generalized algorithm that can be applicable for different types of roads. To determine lane markings effectively, some of the key algorithms are: color cue (He et al., 2004; Chiu and Lin, 2005; Sun et al., 2006), Hough Transform (Yu and Jain, 1997; Southhall and Taylor, 2001), steerable filters (McCall and Trivedi, 2006), spline model (Jung and Kelber, 2004; Wang et al., 2004), and AdaBoost based segmentation (Alon et al., 2006). These algorithms are applicable for roadways with clear lane markings. Yi et al. (2015) improved Hough Transformation algorithm in a computationally efficient manner that is suitable for real-time lane detection even at night (see Figure 3).

FIGURE 3. LANE MARKING PREDICTION AT LOW LIGHT CONDITION

Source: Yi et al. (2015).

Object Detection

In many cases, connected and automated vehicle (CAV) technologies require deeper understanding of the roadway environment for accurate decision making. Object detection is an important computer vision task. It adopts many functions of computer vision technologies such as image search, image auto-annotation, and image perception. However, many studies have focused on refining the algorithms due to the complexity of object classes and images. The studies focusing on object detection can be divided into three categories: top-down, bottom-up, and combination of the two.[2] Top-down approaches usually consider a 'train modeling approach' to determine object classification (Borenstein and Ullman, 2002; Felzenszwalb and Huttenlocher, 2005; and Dalal and Triggs, 2005). Bottom-up approaches start from low-level or mid-level image features (Ferrari et al., 2006; Ren et al, 2005; Mori et al., 2005; Srinivasan and Shi, 2007). These approaches develop hypotheses from image features, extend them by association rules and then evaluate them by certain cost functions. The third category of approaches combines both methods by taking advantage of both aspects.

Use of an 'Image segmentation' algorithm helps convert the undifferentiated image plane into some measure of discrete objects and many studies have been developed to refine this task over the years. The scene flow segmentation or optical flow utilizes temporal correlation between different frames of a scene captured by stereo cameras to classify obstacles that are in motion (Wedel et al., 2009; Franke et al., 2005; Lenz et al., 2011). This approach thus naturally handles tracking moving obstacles. Figure 4 illustrates pedestrian detection by using computer vision.

[2] Some relevant studies focusing on object detection include: Viola and Jones, 2001; Borenstein and Ullman, 2002; Levin and Weiss, 2005; Leibe et al., 2005; Ferrari et al., 2006; Kokkinos et al., 2006; Zhao and Davis, 2005; Ren et al, 2005; Mori et al., 2005; Srinivasan and Shi, 2007; Felzenszwalb and Huttenlocher, 2005; Dalal and Triggs, 2005; and Hariyono and Jo, 2017.

FIGURE 4. COMPUTER VISION IN PEDESTRIAN DETECTION

Source: Hariyono and Jo (2017)

Vehicle Detection

Computer vision is an important tool for vehicle identification and classification. To minimize computational complexity, many studies considered segmenting video data into a background image and a moving object image. Background images tend to be motionless over a long period of time, and moving object images only contain foreground objects. Change detection (Kim et al., 2001; Foresti et al., 1999) is the simplest method for video segmentation. Jung et al. (2001) proposed an adaptive background update method to collect background images. He et al. (2004) applied the Gaussian distribution to model background images. A few studies have used spectral features (colors at each pixel) to model background images (Stauffer and Grimson, 2000; Haritaoglu et al., 2000; Wren et al., 1997). Some spatial features have also been exploited to improve performance in different illumination conditions (Li and Leung, 2002; Javed et al., 2002; Paragios and Ramesh, 2001). Chen et al. (2011) investigated the effectiveness of state-of-the-art classification algorithms to classify vehicles for urban roadways. Chen et al. (2004) proposed a statistical algorithm to efficiently extract color backgrounds and moving vehicles. Daigavane et al (2011) developed an application based on neural network for vehicle detection and classification. Pang et al. (2004) applied a cubical model of the foreground image to detect occlusion and separate merged vehicles from a monocular image. Song and Nevatia (2005) developed a model-based vehicle segmentation method to detect vehicles. Figure 5 shows detection and tracking of moving vehicles by using computer vision. A deep learning tool like the convolution neural network[3] could assist in improving the current performance. For example, one study used training data

[3] http://deeplearning.net/tutorial/lenet.html

from less than a hundred hours of driving to train the car to operate in diverse conditions on different roadways during sunny, cloudy, and rainy conditions by using the convolution neural network (Bojarski et al., 2016).

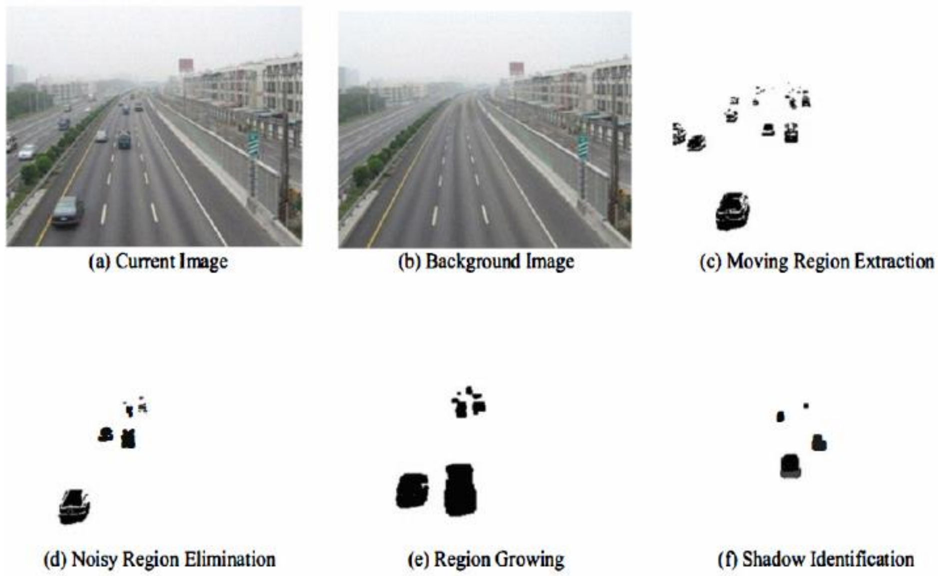

FIGURE 5. DETECTION AND TRACKING OF MOVING VEHICLES

Source: Hadi et al. (2014)

Sign and Signal Detection

Color Based

The prevalence of detecting traffic signs based on color has been used in many studies. One of major difficulties in this approach is that color is unreliable - depending on the time of day, weather conditions, and different illumination criteria. Since red-green-blue (RGB) color space is very sensitive to illumination, many studies have carried out the color-based segmentation in other color spaces. Different kinds of approaches were developed to refine color based sign and signal detection: detect and recognize a small subset of traffic signs that contain red components (Estevez and Kehtarnavaz, 1996), determine influences of daily illumination changes (Benallal and Meunier, 2003), detect red in hue intensity saturation (HIS) color space (Escalera et al., 2003), classify colors based on their similarity with pre-stored hues (Fang et al., 2003), overcome the color dependency on the light source (Broggi et al., 2007), perform color-based segmentation as a starting stage in traffic sign recognition (Ruta et al., 2008).

Shape Based

Current literature provides several approaches for shape-based detection of traffic signs. One of the most used techniques is to use some form of Hough transform. Generalized Hough transform finds arbitrary shapes in an image. Other common approaches are: (a) corner detection followed by reasoning, and (b) simple template matching. Gavrila (1999) used distance transform based template matching for shape detection. Different approaches were developed to perform shape based sign and signal detection: develop a general regular polygon detector to detect traffic signs (Loy and Barnes, 2004), apply the Harris corner detector to identify triangular and rectangular signs (Paulo and Correia, 2007), develop a variant of histogram of oriented gradients (HOG) that exploited the symmetry shape of traffic sign images for classification (Kassani and Teoh, 2017).

Technology Limitation (for enabling automation)

Computational Resources

Computer vision acquires tremendous amounts of image data. Image processing and pattern recognition in computer vision systems are extensive and require large amounts of computational resources and memory. Although there is availability of resources and memory, processing larger amount of data for real-time decision making requires faster and efficient algorithms. Moreover, current computer vision technology suffers from the problem of robustness for inclement weather and illumination issues. This limitation can be overcome by integrating computer vision information with data from other external sensors. This process is known as sensor fusion. Sensor fusion helps in combining data from different sensors to produce multiple inferences for effective decision choice (Guan et al., 2012). For example, a radar or lidar can provide supporting information while determining distance of a target vehicle independently from the computer vision system.

Environmental Conditions

The most significant and fundamental technical limitation to computer vision is its robustness during different weather conditions. Many transportation applications operate outdoors and are very susceptible to illumination variation such as shadows or other low lighting conditions. Although some vision-based transportation applications (for example, driver's infrared night vision) can overcome illumination issues, many other computer vision technologies still lack robustness in tackling environmental conditions.

Suppliers and Costs

There are a larger variety of computer vision technologies in the market starting from a $30 camera to a complex computer vision infrastructure. There is a vast amount future market potentials for many of the computer vision tools used in intelligent transportation systems, Figure 6 illustrates global market for computer vision systems in transportation and traffic management for 2015-2021.

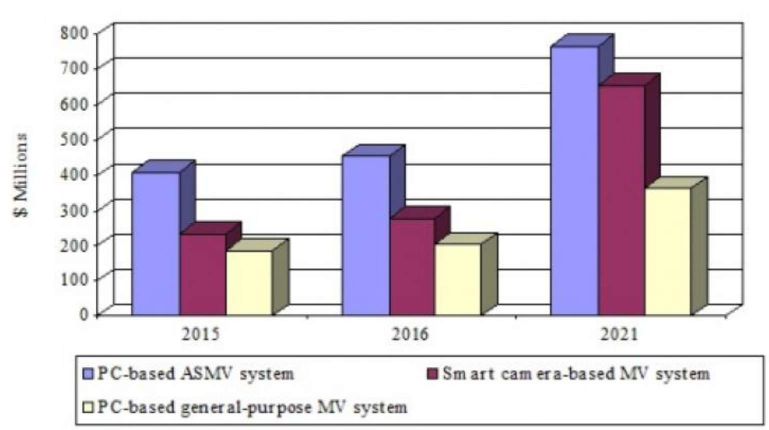

Figure 6. Global Market for Computer or Machine Vision (MV) Systems in Transportation

Source: Rajaram (2012)

Despite its technology limitations, the market potential of computer vision systems is still very significant. Dedicated computer vision-based devices are cheaper compared to other related technologies. For example, according to the National Highway Traffic Safety Administration (NHTSA) estimation, an automotive grade camera would cost in between $58-$88 for an embedded system.

According to CB Insights, thirty-three companies are currently working on self-driving cars. Some of the startup companies attempted to provide cheap computer vision tools for self-driving cars. For example, Comma.ai[4] aimed to sell their start-up kit product for $999 with a $24 per month software update. This venture did not work well due to NHTSA non-compliance issues (Techcrunch, 2017). Morgan Stanley (2013) reported that the various hardware components (computer vision, radar, lidar, and other accessories) needed to achieve full automated driving capability could cost less than $5,000 per car, which means that, together with other associated costs, the customer would pay a premium less than $10,000. For transit vehicles, installation of these devices would be more costly.

Overall Outlook for Application to Transit Automation

Computer vision techniques, like optical flow, texture recognition, and stereo vision are useful in current applications of intelligent transportation systems and CAV technologies. Since 2000, bus automated steering systems has been deployed in Toyota's the Intelligent Multimode Transit System (IMTS) (Aoki, 2000). European Commission's CityMobil2[5] has been working on Automated Road Transport Systems (ARTS). The core activity of CityMobil2 was the demonstration of ARTS in seven European cities between June 2014 and June 2016 (Alessandrini, 2017). There are two broad categories of transit riders- captive riders (riders with no cars), and choice riders (riders with cars). Self-driving cars (for example, Uber) would be beneficial for transit captive riders by providing car sharing. In April 2016, the National Center for Transit Research (NCTR) published a report "Evaluation of Automated Vehicle Technology for Transit – 2016 Update". This report provides status of several U.S. ventures on using computer vision in transit. This report includes Nova Bus/Volvo's pedestrian/bicyclist warning system, the Minnesota Valley Transit Authority (MTVA) maintained global position system (GPS) based lane-keep assist and collision avoidance system, Lane Transit District (Eugene, Oregon) maintained magnetic guidance system for precision docking of its Emerald Express (EmX) bus rapid transit (BRT) system, autonomous shuttle used by Contra Costa Transportation Authority (CCTA) in northern California (Pessaro, 2016).

Computer vision is considered as one of the significant tools for future transit automation. At the same time, it is difficult to operate fully autonomous transit vehicles in an urban roadway network due to the current technological limitations. Except being operated in a separate right-of-way, these vehicles have to react with the circumstances of an urban road network. One can hope that newer technologies would be able to make full transit automation possible in future.

References

Alessandrini, A. (2017). CityMobil2: Experience and Recommendations. Retrieved Jan 25, 2017 from http://www.citymobil2.eu/en/upload/Deliverables/PU/CityMobil2%20booklet%20web%20final_17%2011%202016.pdf

Alon, Y., Ferencz, A., and Shashua, A. (2006). Off-road path following using region classification and geometric projection constraints. Computer Vision and Pattern Recognition.

Aoki, K., and Suyama, T. (2000). A concept of intelligent multi-mode transit system based on automated bus. IEEE Intelligent Vehicles Symposium, pp. 590 - 59.

Ballard, D., and Brown, C. (1982). Computer Vision. Prentice Hall, First edition.

[4] http://comma.ai
[5] http://www.citymobil2.eu/en/

Benallal, M. and J. Meunier, J. (2003). Real-time color segmentation of road signs. CCECE 2003, pp. 1823–1826.

Bojarski, M., Testa, D., Dworakowski, D., Firner, B., Flepp, B., Goyal, P., Jackel, L., Monfort, M., Muller, U., Zhang, J., Zhang, X., Zhao, J., and Zieba, K. (2016). End to End Learning for Self-Driving Cars. arXiv.org.

Borenstein, E., and Ullman, S. (2002). Class-specific, top-down segmentation. European Conference on Computer Vision (2).

Broggi, A., Cerri, P., Medici, P., Porta, P., and Ghisio, G. (2007). Real time road signs recognition. Intelligent Vehicles Symposium, 2007 IEEE, pp. 981– 986.

Chiu, K., and Lin, S. (2005). Lane detection using color-based segmentation. IEEE Intelligent Vehicles Symposium.

Chen, C., Chiu, C., Wu, B., Lin, S., and Huang, C. (2004). The moving object segmentation approach to vehicle extraction. In Proceedings of IEEE International Conference on Networking, Sensing and Control, Vol. 1, pp. 19-23.

Chen, Z., and Ellis, T. (2011). Multi-shape Descriptor Vehicle Classification for Urban Traffic. International Conference on Digital Image Computing: Techniques and Applications.

Daigavane, P., and Daigavane, M. (2011). Vehicle Detection and Neural Network Application for Vehicle Classification. International Conference on Computational Intelligence and Communication Systems.

Dalal, N., and Triggs, B. (2005). Histograms of oriented gradients for human detection. Computer Vision and Pattern Recognition.

D. A. R. P. A. (DARPA). (2004). Darpa grand challenge. Retrieved Jan 10, 2017 from http://www.darpa.mil/grandchallenge.

Escalera, E., Armingol, J. and Mata, M. (2003). Traffic sign recognition and analysis for intelligent vehicles. Image and Vision Computing, vol. 21, pp. 247–258.

Estevez, L., and Kehtarnavaz, N. (1996). A real-time histographic approach to road sign recognition. Image Analysis and Interpretation. Proceedings of the IEEE Southwest Symposium on, pp. 95–100.

Fang, C., Chen, S., and Fuh, C. (2003). Road-sign detection and tracking. IEEE Transactions on Vehicular Technology. Vol. 52, no. 5, pp. 1329–1341.

Ferrari, V., Tuytelaars, T., and Gool, L. (2006). Object detection by contour segment networks. European Conference on Computer Vision.

Felzenszwalb, P., and Huttenlocher, D. (2005). Pictorial structures for object recognition. International Journal of Computer Vision, 61(1).

Foresti, G., Murino, V., and Regazzoni, C. (1999). Vehicle recognition and tracking from road image sequences. IEEE Transactions on Vehicular Technology, Vol. 48, pp. 301-318.

Franke, U., Rabe, C., Badino, H., and Gehrig, S. (2005). 6dvision: Fusion of stereo and motion for robust environment perception. In Pattern Recognition, Springer, pp. 216–223.

Gavrila, D. (1999). Traffic sign recognition revisited. In DAGM-Symposium, pp. 86–93.

Guan, A., Bayless, S., and Neelakantan, R. (2012). Connected Vehicle Insights: Trends in Computer Vision. ITS America.

Guizzo, E. (2011). How Google's Self-Driving Car Works. Retrieved Jan 10, 2017 from http://spectrum.ieee.org/automaton/robotics/artificial-intelligence/how-google-self-driving-car-works

Hadi, R., Sulong, G., and George, L. (2014). Vehicle Detection and Tracking Techniques: A Concise Review. Signal & Image Processing: An International Journal (SIPIJ) Vol.5, No.1.

Haritaoglu, I., Harwood, D., and Davis, L. (2000). W4: Real-time surveillance of people and their activities," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 22, pp. 809-830.

Hariyono, J., and Jo, K. (2017). Detection of pedestrian crossing road: A study on pedestrian pose recognition. Neurocomputing, in press.

He, Z., Liu, J., and Li, P. (2004). New method of background update for video-based vehicle detection. In Proceedings of IEEE Conference on Intelligent Transportation Systems, pp. 580-584.

Javed, O., Shafique, K., and Shah, M. (2002). A hierarchical approach to robust background subtraction using color and gradient information. In Proceedings of IEEE Workshop on Motion Video Computing, pp. 22-27.

Jung, C., and Kelber, C. (2004). A robust linear-parabolic model for lane following. 17th Brazilian Symposium on Computer Graphics and Image Processing.

Jung, Y., Lee, K., and Ho, Y. (2001). Content-based event retrieval using semantic scene interpretation for automated traffic surveillance. IEEE Transactions on Intelligent Transportation Systems, Vol. 2, pp. 151-163.

He, Y., Wang, H., and Zhang, B. (2004). Color-based road detection in urban traffic scenes. IEEE Trans. on Intelligent Transportation Systems.

Kassani, P., and Tech, A. (2017). A new sparse model for traffic sign classification using soft histogram of oriented gradients. Applied Soft Computing, 52, pp. 231–246.

Kokkinos, I., Maragos, P., and Yuille, A. (2006). Bottom-up & top-down object detection using primal sketch features and graphical models. Computer Vision and Pattern Recognition.

Kim, J., Lee, C., Lee, K., Yun, T. and Kim, H. (2001). Wavelet-based vehicle tracking for automatic traffic surveillance. In Proceedings of IEEE International Conference on Electrical and Electronic Technology, Vol. 1, pp. 313-316.

Levin, A., and Weiss, Y. (2006). Learning to combine bottom-up and top-down segmentation. European Conference on Computer Vision.

Leibe, B., Seemann, E., and Schiele, B. (2005). Pedestrian detection in crowded scenes. Computer Vision and Pattern Recognition.

Lenz, P., Ziegler, J., Geiger, A., and Roser, M. (2011). Sparse scene flow segmentation for moving object detection in urban environments. In Intelligent Vehicles Symposium (IV), 2011 IEEE, pp. 926– 932.

Li, L., and Leung, M. (2002). Integrating intensity and texture differences for robust change detection. IEEE Transactions on Image Processing, Vol. 11, pp. 105-112.

Loy, G., and Barnes, N. (2004). Fast shape-based road sign detection for a driver assistance system. In IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 70–75.

McCall, J., and Trivedi, M. (2006) Video based lane estimation and tracking for driver assistance: Survey, system, and evaluation. IEEE Trans. on Intelligent Transportation Systems, pp. 20–37.

Morgan Stanley. (2013). Autonomous Cars: Self-Driving the New Auto Industry Paradigm. 2013.

Mori, G., Ren, X., Efros, A., and Malik, J. (2004). Recovering human body configurations: Combining segmentation and recognition. Computer Vision and Pattern Recognition.

Murphy, K.P. Machine Learning: A Probabilistic Perspective. The MIT Press, 1 edition, August 24, 2012.

Pang, C., Lam, W., and Yung, N. A novel method for resolving vehicle occlusion in a monocular traffic-image sequence. IEEE Transactions on Intelligent Transportation Systems, Vol. 5, 2004, pp. 129-141. 13.

Paragios, N., and Ramesh, V. (2001). A MRF-based approach for real-time subway monitoring. In Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition, Vol. 1, pp. 1034-1040.

Paulo, C., and Correia, P. (2007). Automatic detection and classification of traffic signs. In Image Analysis for Multimedia Interactive Services, WIAMIS '07.

Pessaro, B. (2016). Evaluation of Automated Vehicle Technology in Transit - 2016 Update. National Center for Transit Research, Tampa, FL.

Rajaram, S. (2016). Global Market for Machine Vision Technologies. BCC Research, Wellesley, MA.

Ren, X., Berg, A., and Malik, J. (2005). Recovering human body configurations using pairwise constraints between parts. IEEE International Conference on Computer Vision.

Ruta, A., Li, Y., and Liu, X. (2008). Detection, tracking and recognition of traffic signs from video input. Proceedings of the 11th International IEEE Conference on Intelligent Transportation Systems, Beijing, China.

Sebe, N., Cohen, I., Garg, A., and Huang, T. (2006). Machine Learning in Computer Vision: 29 (Computational Imaging and Vision). Springer.

Song, X. and Nevatia, R. (2005). A model-based vehicle segmentation method for tracking. In Proceedings of IEEE International Conference on Computer Vision, Vol. 2, pp. 1124-1131.

Southhall, J., and Taylor, C. (2001). Stochastic road shape estimation. IEEE International Conference on Computer Vision, pp. 205–212.

Srinivasan, P., and Shi, J. (2007). Bottom-up recognition and parsing of the human body. Computer Vision and Pattern Recognition.

Sun, T., Tsai, S. and V. Chan. (2006). Hsi color model based lane marking detection. IEEE Intelligent Transportation Systems Conference, pp. 1168–1172.

Stauffer, C., and Grimson, W. (2000). Learning patterns of activity using real-time tracking. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 22, pp. 747-757.

Techcrunch. Comma.ai cancels the Comma One following NHTSA letter. Retrieved Jan 10, 2017 from https://techcrunch.com/2016/10/28/comma-ai-cancels-the-comma-one-following-nhtsa-letter/

Viola, P., and Jones, M. (2001). Rapid object detection using a boosted cascade of simple features. Computer Vision and Pattern Recognition.

Wang, Y., Teoh, E., and Shen, D. (2004). Lane detection and tracking using b-snake. Image and Vision Computing, pp. 269–280.

Wedel, A., Meiner, A., Rabe, C., Franke, U., and Cremers, D. (2009). Detection and segmentation of independently moving objects from dense scene flow. In Energy minimization methods in computer vision and pattern recognition, pp. 14–27. Springer.

Wren, C., Azarbaygaui, A., Darrell, T., and Pentland, A. (1997). Pfinder: Real-time tracking of the human body. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 19, pp. 780-785.

Yi, S., Chen, Y., and Chang, C. (2015). A lane detection approach based on intelligent vision. Computers and Electrical Engineering, 42, pp. 23–29.

Yu, B., and Jain, A. (1997). Lane boundary detection using a multiresolution hough transform. ICIP, vol. 2, pp. 748–751.

Zhao, L., and Davis, L. (2005). Closely coupled object detection and segmentation. IEEE International Conference on Computer Vision.

Zhou, L., Pan, S., Wang, J., Athanasios, and Vasilakos, V. (2017). Machine Learning on Big Data: Opportunities and Challenges. Neurocomputing. In press.