Journal of Korea Multimedia Society Vol. 21, No. 8, August 2018(pp. 888-896)
https://doi.org/10.9717/kmms.2018.21.8.888

Recognition of Car Manufacturers using Faster R-CNN and Perspective Transformation

Israfil Ansari†, Yeunghak Lee††, Yunju Jeong†††, Jaechang Shim††††

ABSTRACT

*In this paper, we report detection and recognition of vehicle logo from images captured from street CCTV. Image data includes both the front and rear view of the vehicles. The proposed method is a two-step process which combines image preprocessing and faster region-based convolutional neural network (R-CNN) for logo recognition. Without preprocessing, faster R-CNN accuracy is high only if the image quality is good. The proposed system is focusing on street CCTV camera where image quality is different from a front facing camera. Using perspective transformation the top view images are transformed into front view images. In this system, the detection and accuracy are much higher as compared to the existing algorithm. As a result of the experiment, on day data the detection and recognition rate is improved by 2% and night data, detection rate improved by 14%.

Key words: Car Logo Detection, Faster R-CNN, Perspective Transformation, Vehicle Manufacturer Detection

1. INTRODUCTION

In recent years, much deep learning based vehicle detection or recognition algorithms are being proposed. Vehicle logo is one of the essential vehicle highlight descriptors, which can furnish the astute activity framework with valuable data to distinctive vehicles. Therefore, vehicle logo recognition can be applied in the highway toll system and vehicle identification for public security and so on. To respond rising demand from the reasonable circumstance, it is increasingly important to find and perceive vehicle logo clearly. There are principal challenge to expand acknowledgment precision and productivity. Vehicle type classification is already addressed in previous work [1, 2]. The primary purpose of this paper is to locate and identify the vehicle brand to monitor the traffic better along with vehicle detection. In paper [3], reported algorithm for vehicle logo detection using day data but the same algorithm was unable to detect vehicle logos in some images. Whereas, here we report the extension of the proposed algorithm[3] by including night data during training the model. Along with night data we also propose selective transformation of the images according to their logo area. In previous work[3], the parameter of perspective transformation was same for all the images, in upward direction. Therefore, the logos on the top were moved more upward after trans-

※ Corresponding Author : Jaechang Shim, Address: (36729) 1375 Gyeongdong-ro, Andongsi, Gyenongsang-buk-do, Korea, TEL : +82-54-820-5645, FAX : +82-54-820-6164, E-mail : jcshim@anu.ac.kr
Receipt date : Jul. 13, 2017, Revision date : Aug. 9. 2018
Approval date : Aug. 21, 2018
† Department of Computer Eng., Andong National University (E-mail : israfila3@hotmail.com)
†† Department of Computer Eng., Andong National University (E-mail : yhsjh.yi@gmail.com)
††† School of Computer Science and Eng., Kyungpook National University (E-mail : vrjung@hanmail.net)
†††† Department of Computer Eng., Andong National University
※ This work was supported by "Cooperative Research Program for Agriculture Science and Technology Development" Rural Development Administration, Republic of Korea (Project No. PJ01384901).

formation, so the detection rate went down. In this paper we propose both upward and downward perspective transformation. The logos on the top area in the images are passed through downward perspective transformation and the logos in the lower area are passed through upward perspective transformation.

The main difficulty to detect vehicle models is due to a rapid change of vehicle brands designs based on their appearance for a competitive market. In today's competitive market some of the brands have almost the same front view with slight changes, so the only logo of the manufacturer is the option to differentiate vehicles. A vehicle manufacturer logo is a small object that should be unique to represent their brand and model. In most of the cases the logo is found especially in the front middle area and lays between the lights and above the number plate of the vehicle. The vehicle logo is mostly made with some unique features or geometries using different colors and textures to be distinguished as soon as it is seen.

Image matching or feature matching is traditionally used by researchers for logo detection from vehicle images. However, these methods fail if the quality of images is not good enough or the images are taken from street CCTV cameras. In paper [2] authors propose SIFT-based vehicle logo detection and recognition method. Here, multiple images were merged to enhance the detection rate. Furthermore, hough transform was used for feature clustering, and geometric verification was applied for the affine transformation. In paper [4] authors propose a two-layer CNN method with a concept of pre-training model. During the pre-training step principal component analysis is used to improve the accuracy and speed of the system. In paper [5], authors propose the CNN and SVM combination method. In the initial stage, CNN is applied to select candidate areas that are likely to be the manufacturer logo. After the first stage, the pyramid of the histogram of orientation for smooth changes between two points to improve detection rate of vehicle logo, leads to improvement of overall performance of the system. It is then applied to check the truth of the candidate area and eliminate the false areas that do not contain logo. The methods of using SVM as the classifier to manually selection of the feature. To select a feature with clear effect on the classification result, lighting up a weak feature, noise and other harsh surrounding conditions can cause extreme ineffective. At the same time, CNN classification does not need to select feature manually. In paper [6] the authors propose vehicle detection and recognition models using modest Adaboost algorithm and radial Tchebichef moments. The system can recognize vehicle logos regardless of variation in viewpoints such as rotation, scaling, translation, and skewing. To detect the vehicle logo, Haar-like features are extracted to represent parts of the image and modest Adaboost algorithm is used to classify them. When the logo has been detected, the system normalizes and shifts the logo back to the normal front viewpoint. Then, the image pattern is represented by radial Tchebichef moments and is recognized with k-Nearest-Neighbor (k-NN) classifier.

This paper proposes R-CNN based pre-trained model to train our custom data. For training data, we have used 1000 images from the different brand of vehicles and level them before training data. After training, the trained file used to detect and recognize vehicle logo.

2. PROPOSED WORK

In this study the data for testing were taken from top view street CCTV cameras. It was then passed through image preprocessing part to increase the detection rate of the vehicles logo. In the front facing camera the logos of the vehicles are clearly visible but when the images are taken from top view, the quality of the image and the logo area are not clearly visible. To overcome this problem propose a perspective transformation method [7]. Perspective is called when human eyes see near things they look bigger as compared to far away and transformation is known as the transfer of an object from one state to another. For perspective transformation, a 3×3 transformation matrix is required for processing. To find this transformation matrix, 4 points are required from the input image and corresponding points on the output image. Among these 4 points, 3 of them should not be collinear. During perspective transformation, it does not change the image content but deform the pixel grid and map this deformed grid to the destination image. To avoid sampling artifacts, the mapping is done in the reverse order, from destination to the source. That is, for each pixel (x, y) of the destination image, the functions compute coordinates of the corresponding “donor” pixel in the source image and copy the pixel value.

Direct linear transformation algorithm used to transform $X_i = (x_i, y_i)$ to $X_i' = (x_i', y_i')$ with projection matrix M.

$$X_i^{'} = MX_i \tag{1}$$

$$t\begin{bmatrix} x_i' \\ y_i' \\ 1 \end{bmatrix} = \begin{bmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{31} & m_{32} & 1 \end{bmatrix} \begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix} \tag{2}$$

where t is the weight of the feature points. Equation (1) is the scalable variable. We can derive equation (3) from equation (2)

$$x_i' = \frac{m_{11}x_i + m_{12}y_i + m_{13}}{m_{31}x_i + m_{32}y_i + 1},\ y_i' = \frac{m_{21}x_i + m_{22}y_i + m_{23}}{m_{31}x_i + m_{32}y_i + 1} \tag{3}$$

In projection matrix M, we have to find 8 elements. If we know 4 pairs of source and destination point, then we can find projection matrix M.

Fig. 1(a) depicts the upward perspective transformation of images, where vehicle logos are in downward position. As seen in the figure, after transformation the logos are clearly visible. Fig. 1(b) depicts the downward perspective transformation of the vehicle images. This transformation helps to increase the detection rate of the vehicle logos. In case of recognition of logo types, with the normal dataset, there was some wrong and miss recognition which was solved by the perspective transformation.

3. CONVOLUTIONAL NEURAL NETWORK

CNN for training data is divided into 3 parts. Fig. 3 depicts the full flow of the proposed system. Initially, the dataset is level with their corresponding classes. After leveling all the dataset, it is then forwarded for training with the faster R-CNN model. After training an inference graph is saved, the graph used to localize and identify the logo. However, before detecting the logos from the captured data, all the data are pass-through image processing for the better result.

3.1 Labeling dataset

For labeling dataset, there are two things to be considered. First, RGB images are required which are encoded as jpeg or png. Second, a list of bounding boxes (Xmin, Ymin, Xmax, Ymax) is necessary for the image and the class of the object in the bounding box. Here 1000 images are taken to label them with their respective classes. Fig. 3 shows some of the labeling datasets. As shown in the figure 3 the red bounding box in the vehicle image labeled as brand Hyundai, the blue bounding box represents the brand Daewoo, and the yellow bounding box represents the brand Ssangyong.

3.2 Training data with faster R-CNN model

A faster R-CNN is a complex network, and the training with this model is not an easy task. Fig. 4 depicts the architecture of faster R-CNN. Faster R-CNN [8] has two networks: (a) region proposal network (RPN) for generating region proposals and (b) a network using these proposals to detect objects. An RPN takes an image of any size as input and outputs as set of rectangular object proposals, each with an object score. This method en-

(a)

(b)

Fig. 1. The perspective transformation (a) Upward perspective transformation, (b) Downward perspective transformation.

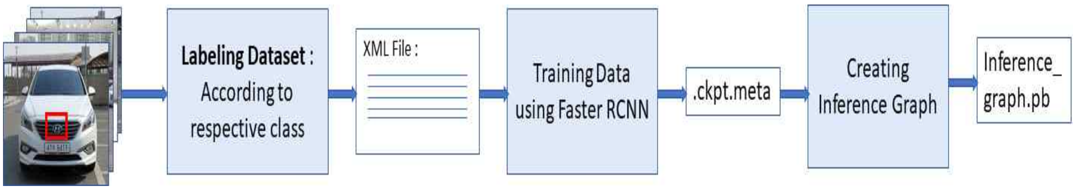


Fig. 2. Training Data Sequence.

ables a unified, deep-learning-based object detection system to run at near real-time frame rates. The learned RPN also improves region proposal quality and thus improves the overall object detection accuracy. The main difference between *faster* R-CNN with fast R-CNN is that the later uses selective search to generate region proposals. The time cost of generating region proposals is

Fig. 3. Labeling dataset.

much smaller in RPN than selective search, when RPN shares the most computation with the object detection network. Similar to Fast R-CNN, a faster R-CNN network may be built on a existing networks. Mainly, RPN ranks region boxes also known as anchors and proposes the ones most likely containing the object. To make a robust system in translation and scale, the RPN uses an algorithm based on anchors. For each position of the sliding window on the feature map, 9 anchors are placed. Then all the anchors are centered on a sliding window with the changes of their scale and ratio to generate 9 anchors. Each anchor is processed through the convolutional layers of the RPN, and the networks give an output probability that this anchor represents an object which is trained in the mode and potentially an offset to correct the anchor dimensions.

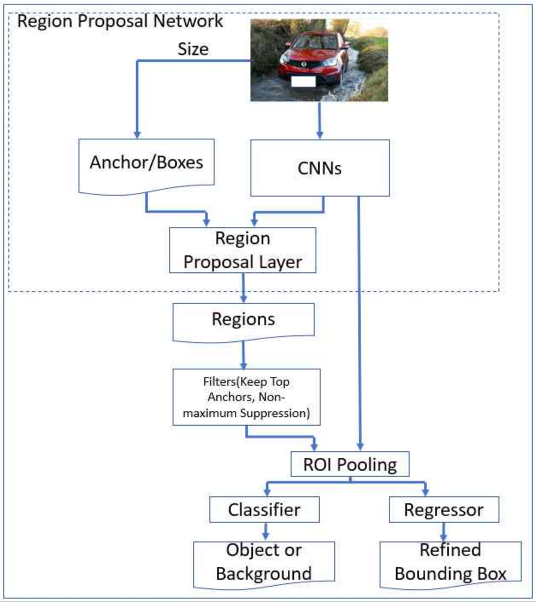


Fig. 4. Architecture of Faster R-CNN.

3.3 Creating inference graph

Inference graph is also known as a freezing model which is saved for further process. This will help to save time and also helps in real-time object detection. While training dataset with the model, each pair at different time steps, one is holding the weights ".data", and another is holding the graph ".meta". All together, we call it as meta-data. After training all the frozen files are saved as the trained model which is supportable to TensorFlow. This file is used further to detect logo area and recognize them.

4. EXPERIMENTAL RESULTS

This proposed system was successfully tested with the street live CCTV. After the training data-set, the frozen file was used to detect the logos of vehicle of newly received data. All the above experiments were performed using NVidia GeForce GTX 1070 with 8GB memory, Tensorflow framework and Windows OS with Intel Core i7 with 32GB RAM. The experiment mainly focused on vehicles available in S. Korea. In table 1 and table 2 we have shown the difference between the normal images and the preprocessed images. In both cases, we have used the same data. Initially, the original data was tested and then all the data were preprocessed using perspective transformation and tested with our trained model. Looking at table 1 and, 2 there is a significant difference in all the brands. However, when the images passes through perspective transformation, the wrong detection went to zero. Fig. 5 shows the final visualize output after preprocessing. It clearly shows the prediction percentage. The prediction percentage also differs from the processed data. In normal image when the prediction is 97 percentage after detection while, it is 99 percent in preprocessed data.

In Table 1 and 2 only day data were used. In Table 2 we can see the difference between the result of paper [3] and this algorithm. Furthermore, the average result of day data is increased from 97.5 [3] to 99.3. In this paper we have extended out algorithm by training model with night dataset. In Table 3 and 4 shows the result of night data taken from street camera. As shown on the table 2 and 4 perspective transformation helps to increase the detection rate as compared to unprocessed image.

5. CONCLUSION

Several methods are used to detect and recog-

Fig. 5. Result after perspective transformation.

Table 1. Result of day image

Vehicle Model	RESULT			
	Total images	Correct detection	Wrong detection	Result (%)
HYUNDAI	200	196	4	98
KIA	50	46	4	92
SAMSUNG	50	48	2	96
SSANGYONG	10	10	0	100
DAEWOO	10	6	4	60
Result	320	306	14	95.6

Table 2. Result of day preprocessed image

Vehicle Model	RESULT			
	Total images	Correct detection	Wrong detection	Result (%)
HYUNDAI	200	200	0	100
KIA	50	50	0	100
SAMSUNG	50	49	1	98
SSANGYONG	10	10	0	100
DAEWOO	10	9	1	90
Result	320	318	2	99.3

Table 3. Result of night image

Vehicle Model	RESULT			
	Total images	Correct detection	Wrong detection	Result (%)
HYUNDAI	186	156	30	83.8
KIA	19	14	5	73.6
SAMSUNG	14	9	5	64.2
SSANGYONG	2	2	0	100
DAEWOO	3	2	1	66.6
Result	224	183	41	81.6

Table 4. Result of night preprocessed image

Vehicle Model	RESULT			
	Total images	Correct detection	Wrong detection	Result (%)
HYUNDAI	186	184	2	98.9
KIA	19	18	1	94.7
SAMSUNG	14	12	2	85.7
SSANGYONG	2	2	0	100
DAEWOO	3	3	0	100
Result	224	219	5	97.7

nize vehicle logo by extracting a specific pattern. However, most of them are very complicated and has low detection. The rate goes even more lower while detecting it from street top view cameras.

This proposed system provides a practical approach for vehicle logo detection and recognition system using faster R-CNN model and perspective transformation. In case of perspective transformation both upward and downward direction transformation is proposed. By using this system, we can label all the required vehicle type logo with their respective classes and recognize them accordingly. This approach can help the traffic monitoring system to become smarter in classifying the vehicle data. As shown in the result, the perspective transformation in image preprocessing gives a better result as compared to the normal image which received from CCTV. We have successfully tested the proposed system on both day and night data.

There are some issues with night data as the image quality is very low. In the future, we plan to increase the image quality of the night images to detect and recognize vehicle logos.

REFERENCE

[1] T. Kato, Y. Ninomiya, and I. Masaki, "Preceding Vehicle Recognition Based on Learning from Sample Images," *IEEE Transactions on Intelligent Transportation Systems*, Vol. 3, No. 4, pp. 252-260, 2002.

[2] A. Psyllos, C.N. Anagnostopoulos, and E. Kayafas, "M-SIFT: A New Method for Vehicle Logo Recognition," *Proceeding of IEEE International Conference on Vehicular Electronics and Safety*, pp. 261-266, 2012.

[3] I. Ansari, W. Shin, Y. Lee, Y. Jeong, and J. Shim "Vehicle Logo Detection and Recognition using Neural Network and Perspective transformation," *Proceeding of International Conference on Multimedia Information Technology and Application*, Vol. 12, pp. 12-15, 2018.

[4] Y. Huang, R. Wu, Y. Sun, W. Wang, and X. Ding, "Vehicle Logo Recognition System Based on Convolutional Neural Networks With a Pretraining Strategy, "*IEEE Transactions on Intelligent Transportation Systems,* Vol. 16, pp. 1951-1960, 2015.

[5] W. Thubsaeng, A. Kawewong, and K. Patanukhom, "Vehicle Logo Detection Using Convolutional Neural Network and Pyramid of Histogram of Oriented Gradients," *Proceeding of International Joint Conference on Computer Science and Software Engineering*, pp. 34-39, 2014.

[6] S. Kam-Tong and X. Lin Tian, "Vehicle Logo Recognition Using Modest AdaBoost and Radial Tchebichef Moments," *Proceeding o f International Conference on Machine Learning and Computing*, Vol. 25, pp. 91-95, 2012.

[7] S.H. Heo and S.G. Kwon, "OMR Sheet Recognition Algorithm Using QR Code Recognition and Perspective Transform," *Journal of Korea Multimedia Society,* Vol. 21, No. 4, pp. 464-470, 2018.

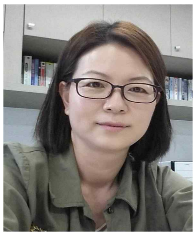
[8] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards Real-time Object Detection with Region Proposal Networks," *IEEE Trainsactions on Pattern Analysis and Machine Intelligence*, Vol. 39, No. 6, pp. 1137-1149, 2017.

Md Israfil Ansari

received his BCA degree in Computer Application from Punjab Technical University, Jalandhar, India in 2011 and the MCA degree in Computer Application from Visvesvaraya Technological University, Bangalore, India in 2015. His previous position was a Software Developer as an Intern at Xcelvations Consultancy Pvt Ltd, Hyderabad, India. Currently studies for Ph.D. at Department of Computer Engineering, Andong National University, Korea. He has interests in the areas of computer vision, pattern recognition, Deep Learning.

Yeunghak Lee

received his Ph.D degree from Yeungnam University, Korea, in 2003. He had one year experience at University of Cardiff as postdoc research fellow. He was a professor of department of avionic electronic engineering at Kyungwoon University. He is currently working as a senior researcher in the Department of Computer Science at Andong National University. He contributes as a management editor for the Journal of Multimedia and Information Systems. His research interests include pattern recognition, embedded system and computer vision.



Yunju Jeong

received her BS in Computational Statistics from Andong National University in 1991 and Master of Computer Engineering from Department of Computer Engineering, Andong National University in 2000. Currently, she studies for Ph.D. at Department of Computer Engineering of Kyungpook National University and also an Education Instructor at Department of Computer Engineering, Department of Electronic Engineering, Andong National University, Korea. Her research interests include computer vision, Deep Learning, and VR.

Jaechang Shim

received his B. S. degree in 1987 and Ph.D. degree from Kyungpook National University, Korea, in 1997. From 1997 to 1999, he worked as a researcher in IBM T. J. Watson Research Center. From 2005 to 2006, he was a visiting Fellow Professor at Princeton University. Since 1994, he has been a Professor of the Andong National University, Korea. He received the best Industrial-Related Paper Award form International Association for pattern Recognition in 1998. His research interests include computer vision, image processing, pattern recognition, and AI.