International Journal of Grid and Distributed Computing
Vol. 11, No. 2 (2018), pp.111-122
http://dx.doi.org/10.14257/ijgdc.2018.11.2.10

A Study of Partial Image Classification of Vehicles Using Finger Gestures

Junho Jeong[1], Jun Young Lee[2] and Yunsik Son[2*]

[1]Electronic Commerce Institute, Dongguk University
[2]Department of Computer Science and Engineering, Dongguk University
{yanyenli, kuroneko0441, sonbug}@dongguk.edu

Abstract

There have been many studies done to recognize types of vehicles or license plates by image analysis. However, there have not been many studies conducted to determine the components of a vehicle using partial images of the vehicle. To efficiently analyze the components using partial images of vehicles, we added meta information about the state of the vehicle and partial images using specific gestures when recording the images. Finally, we propose a method to classify images with partial images of vehicles automatically.

Keywords: *partial image of vehicle, gesture, image analysis, bag-of-word*

1. Introduction

The existing recognition technology for vehicles is mainly used to recognize the type and license plates of vehicles. For this purpose, it is common practice to use the entire image of the vehicle or perform recognition utilizing the area corresponding to the license plates. However, there has not been much research conducted to determine the components of vehicles based only on a partial image of the vehicle, rather than the entire image of the vehicle [1-3].

In practice, this is similar to the process of classifying parts of a vehicle based on the human eye and determining its status. Recognition using only partial images can solve issues of mutual trust by recording the state of each vehicle before and after rental. It can be used very efficiently for managing rental history.

When analyzing the components of a vehicle to generate an image, a specific gesture can be used to add meta information about the state of the vehicle and partial images. That is, it is possible to classify an unclassified partial vehicle image automatically. Therefore, this paper proposes automatic image classification through gesture analysis. In Section 2, we study the existing vehicle and gesture recognition. Section 3 introduces the proposed method. In Section 4, we perform the experiment using Bag-of-Words.

2. Related Works

2.1. Vehicles Recognition

Many techniques for recognizing vehicles have been carried out in order to identify vehicles for reasons such as a customer trying to purchase a car or in the case of a car theft or an accident. Kang *et al.*, Proposed an algorithm to recognize the manufacturer and model name as well as the macro classification of the vehicle in real time [4]. After only extracting the front part of the vehicle using the vehicle characteristic, the PCA

Received (November 5, 2017), Review Result (January 24, 2018), Accepted (January 29, 2018)
* Corresponding Author

ISSN: 2005-4262 IJGDC

(Principal Component Analysis) technique calculates the unique vehicle model that shows the most general characteristics of the vehicle, Then, each learning vehicle and the input vehicle are projected in a specific vector space. This method has the advantage of lowering the image of the vehicle to the dimension of the desired feature. Then, we propose a method of recognizing the model of the vehicle by calculating the Euclidean distance between the input vehicle and the learning vehicle using the vector coordinates in the reduced dimension.

Qi *et. al.*, Proposed a method of recognizing a vehicle model using CNN learning algorithm [5]. Before learning, car model images were collected and pre-processed using Gaussian blur. The image is not fixed in resolution size, which increases the robustness of the neural network. If noise is present in the image, the cost function is difficult to converge and a problem may occur in the learning process.

M. Huzaifa *et al.*, Proposed a method of recognizing a car model using the Binary Robust Invariant Scalable Keypoint method [6]. This method solves the difficulty of feature extraction with high quality description with a small calculation time. First, we change the input image to grayscale, find the keypoint (feature point) using the BRISK algorithm, then store each keypoint in the array, and compare similarities with the training data set. The BRISK method acts as a detector to detect removal errors consistent with RANSAC (RANdom SAmple Consensus) and stores the matching function with Hamming distance in the database. That is, hamming distance is required by matching with keypoint, and RANSAC is required for error matching reduction. The proposed method has a reliable advantage in recognizing the vehicle models in frontal vehicle images of different sizes and brightness.

A method of distinguishing a vehicle's model based on the outline of the vehicle has been attempted, but it is difficult to recognize the specific vehicle and model. Unlike the contours, the pattern of the headlamps and the radiator grill at the front of the vehicle has a unique appearance for each vehicle. Therefore, the center coordinates of the license plate are obtained and then moved to the appropriate coordinates for finding the area of interest. Figure 1 shows the area of interest of the vehicle. The model can then be identified by extracting the region of interest at an appropriate size.

There are various research methods such as recognizing plate numbers of vehicles, and also using a geometric pattern vector of brightness change value or a heuristic division algorithm to recognize plates in real time [7-9].

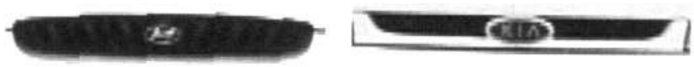


Figure 1. Extracted Region of Interest for Texture Features

Neural network learning such as CNN is effective for recognizing and classifying specific parts in one image. However, the Bag-of-Words method is effective because the image dataset that we will use is a picture of only a specific part of the vehicle without empty space. Shuyuan Yu *et al.*, Have proposed a method of recognizing a vehicle's logo by Bag-of-Words algorithm [10]. Andrés F. Gómez *et al.*, Used Bag-of-Words Respectively [11]. In contrast, our field of activity is to classify vehicles into images of all parts of the vehicle, not just the logos or specific images of the vehicle.

2.2. Gestures Recognition

Gestures are the most basic means of communication, and research has continued to explore them. Among them, various studies have been carried out in order to process the motion of fingers using gestures, and a method of obtaining a central point and a contour of a hand by using a Markov model or using a contour is typical.

The method of using contour is used to recognize the number of fingers by using the distance between the center of hand and the contour of the hand [12]. Since contours are composed of contours, the coordinates of each contour can be used. In order to grasp the number of fingers, if the second value among the three consecutive values is the largest, the point becomes the finger candidate. At this time, contours below the center of the hand are excluded from the calculation. When the search is completed, the number of finger candidates is recognized as the number of fingers.

Figure 2. Hand Center Point and Contour Detection

Figure 2 shows the outline and center point of the palm. Since the number of finger candidates is 5, the gesture displayed on the current screen can be confirmed that five fingers are spreading.

The recognition rate of identifying gestures and counting the number of fingers is superior to conventional algorithms. However, in this study, the gesture cannot be recognized through the features of the shape of the hand, and can only discriminate the number of the fingers.

Yoon *et al.*, Proposed a hand feature extraction algorithm based on the curvature analysis for recognition of various handlers [13]. The number of gestures that can be defined is more than that of other algorithms because it is possible to distinguish fingers based on curvature.

Starner *et al.*, Proposed a method of recognizing gestures by applying HMM (Hidden Markov Model) [14]. The study was aimed at understanding sign language, so we did not focus on the detailed shape of the finger, but rather perceived in which direction both hands were pointing. This paper trained the sign language of 295 sentences and showed good results.

Another research was carried out to classify sign language ASLs using Hough transform and neural networks [15]. ASL is able to represent simple words in simple handwritten letters such as A to Z in a single hand shape, analyzing hand shapes through images, and studying what it means in ASL.

Thus, various studies on model recognition and finger gesture through the image of the vehicle have been conducted. However, there have been no studies to determine the components of the vehicle with partial image information of the vehicle. In this paper, we propose a method to use the finger gesture information as metadata to effectively classify the partial images of the vehicle and set the region of interest.

3. A Classification by Partial Images of Vehicles and Select Region of Image(RoI) using Gestures

In this paper, we proceed by inputting the image of the vehicle, automatically classifying it based on gestures, and presenting classification information to the user as shown in Figure 3. Next, if the gesture specifies an area, the area is analyzed, and the information is presented to the user.

The types of gestures used in this study are not limited. Users can specify a gesture by using parts of the body, such as hands, feet, or objects. When it recognizes a gesture, it associates the type with the mapping table of the gesture, allowing the image to be classified.

Gestures are divided into two main categories: gestures to identify the components of a vehicle and gestures to find specific areas such as scratches. If the gesture to find a particular area is recognized, the area is analyzed based on the proposed vector by identifying the element indicating the direction of a specific position.

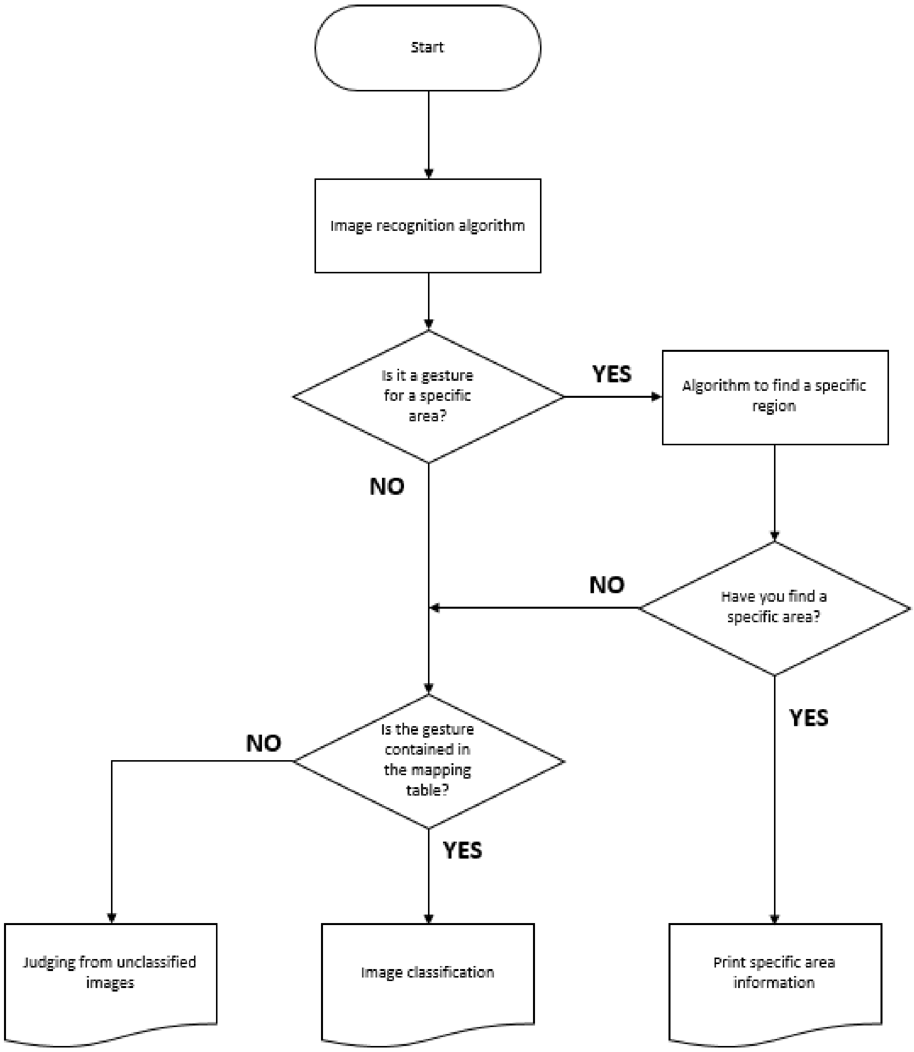


Figure 3. Overall Algorithmic Structure

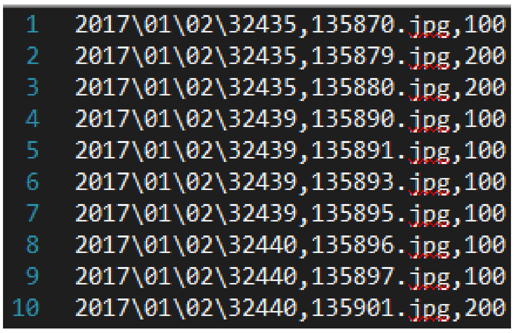
3.1. A Classification by Partial Images of Vehicles

In the image pattern classification, a step-by-step classification method was used in which a simple classification was performed before classification for all regions, and individual classification was performed for detailed regions. The classification method is shown in Table 1. Outdoor photography / indoor photography, and outdoor photography again divided into full / detail photographs. We left some leeway between label values for future classification.

Table 1. Image Classification and Labels and Meanings of each Classification

Name	Value	Information
Not Classified	100	Uncategorized pictures
Whole Picture	200	Front: Includes both headlights Side: Including front and rear wheels
Detail Picture	300	All outdoor pictures except Whole Picture
Indoor Picture	400	Pictures taken indoors

First, we need a label for each picture of text in the map learning for the image set of our car. We created a label of csv file format, as shown in Figure 4, with the labeling program of Figure 5. Each line represents one label of a picture, and it consists of "folder name, photo name, label".



```
2017\01\02\32435,135870.jpg,100
2017\01\02\32435,135879.jpg,200
2017\01\02\32435,135880.jpg,200
2017\01\02\32439,135890.jpg,100
2017\01\02\32439,135891.jpg,100
2017\01\02\32439,135893.jpg,100
2017\01\02\32439,135895.jpg,100
2017\01\02\32440,135896.jpg,100
2017\01\02\32440,135897.jpg,100
2017\01\02\32440,135901.jpg,200
```

Figure 4. The Format of the Generated Label File

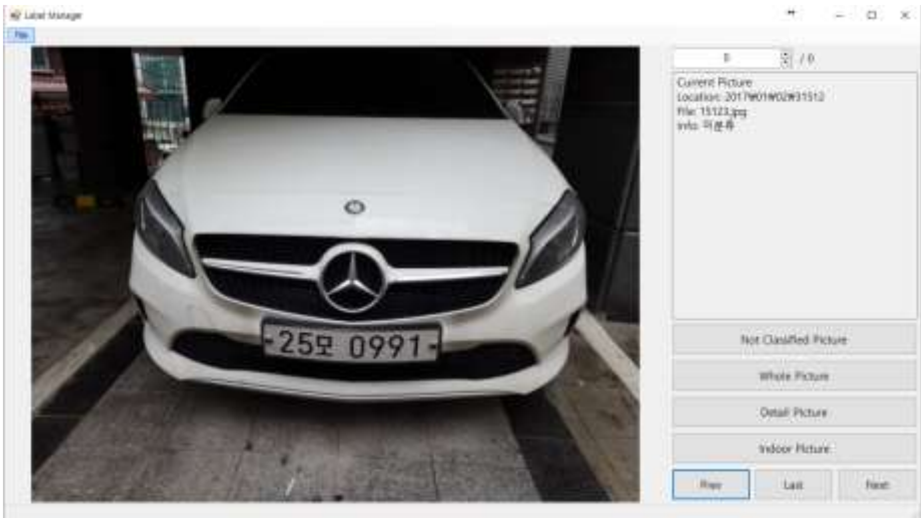


Figure 4. Label Generation Program

Bag-of-Words learning was performed with the generated label and image set. The Bag-of-Words framework is a C # based Accord.NET. In order to manage memory, 4K image quality was reduced to 512-pixel size. Figure 6 shows a sample Bag-of-Words learning program provided by the NET Framework. This program displays the thumbnails of the currently used images as labels and has a panel where you can assign various settings for learning.

However, this program is very memory consuming because it continuously displays image thumbnails by using large amounts of images, consequently it is unsuitable when a large amount of learning must be performed at the same time. Therefore, memory management is necessary.

We have developed a program that includes memory management and multi-learner functions. The program is shown in Figure 7 and consists of several tabs. Each tab carries out one learning, and there is a log indicating the progress of the learning and a table showing the result of classification after learning.

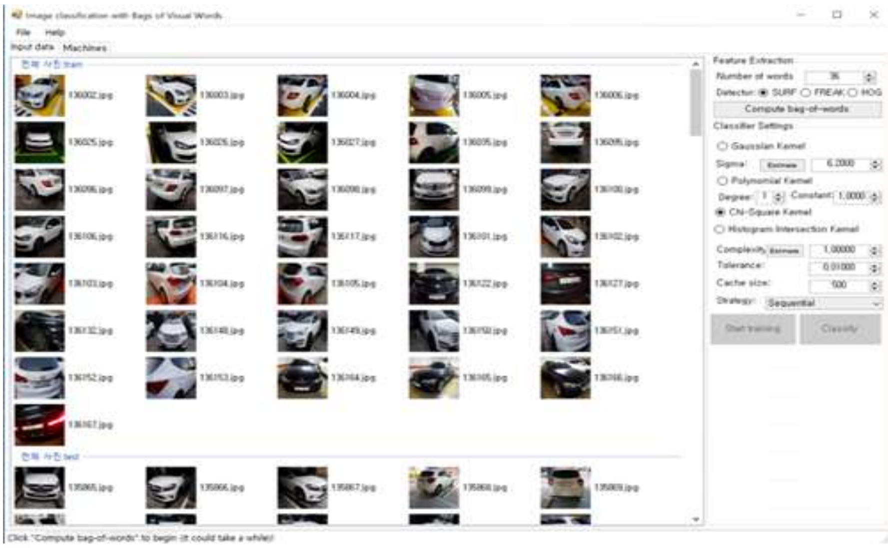


Figure 6. Accord. NET Learning Program Sample

Figure 7. Multi-tap Learning Program

3.2. Select Region of Image(RoI) using Gestures

The components for classifying the partial image of a vehicle for a specific area are roughly divided into four parts as shown in Table 2. Not classified, whole picture, detail picture, indoor picture. In addition, the finger gesture as shown in Table 2 was used as a method of simply dividing the components by the number and shape of the fingers. In this way, more effective vehicle parts can be distinguished by adding meta information to basic vehicle recognition methods.

Table 2. Vehicle Component Classification Criteria

Label	Component	Information	Gesture
0	Not Classified	Image before classification	None
1	Whole picture	The whole image of the vehicle	
2	Detail picture	All outdoor photographs except for the entire vehicle	
3	Indoor picture	Dashboard, indoor images other than mileage	

4. Analysis of Experimental Results

Image analysis experiment using Bag-of-Word. The Bag-of-Word configuration used in the experiment is the same as Table 3, and the number of images and the number of words are different.

Table 3. Bag-of-Words Settings

Property	Values
Number of layers	7 (16,384(input) – 512 – 256 – 128 – 10 – 10 – 4(output))
Number of learning	Unsupervised learning 10,000 times, Supervised learning 10,000 times
Learning rate	0.1
Momentum	0.5
Decay	0.001

The experiment was carried out with two kinds of image samples. The information of each sample is shown in Table 4. The first sample was a total of 3,007 images, consisting of 7 not classified photographs, 1,000 images of whole, detail and indoor photographs. The second sample consisted of 5,962 images, 7 not classified photographs, 1955 whole photographs, 2000 detailed photographs, and 2000 indoor photographs.

Table 4. Information from Two Samples

Sample	Not Classified	Whole picture	Detail picture	Indoor picture	Total
A	7	1000	1000	1000	3007
B	7	1955	2000	2000	5962

Figure 8 shows the results of Bag-of-Word test for sample A. The horizontal axis indicates the number of words, and the vertical axis indicates the accuracy according to the number of words. Accuracy is about 70% when the number of words is 500, and it increases gradually as the number of words increases. However, when the number of words is more than 30,000, the accuracy converges to about 80%, and it does not increase.

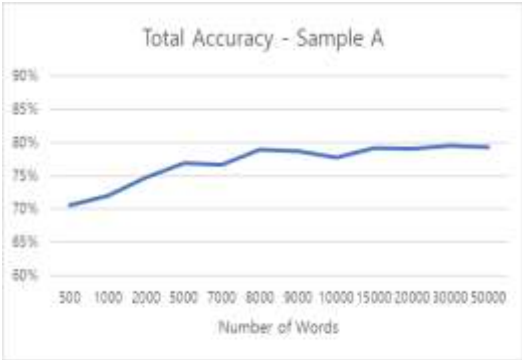


Figure 8. Overall Photo Accuracy of Sample A (Unit: %)

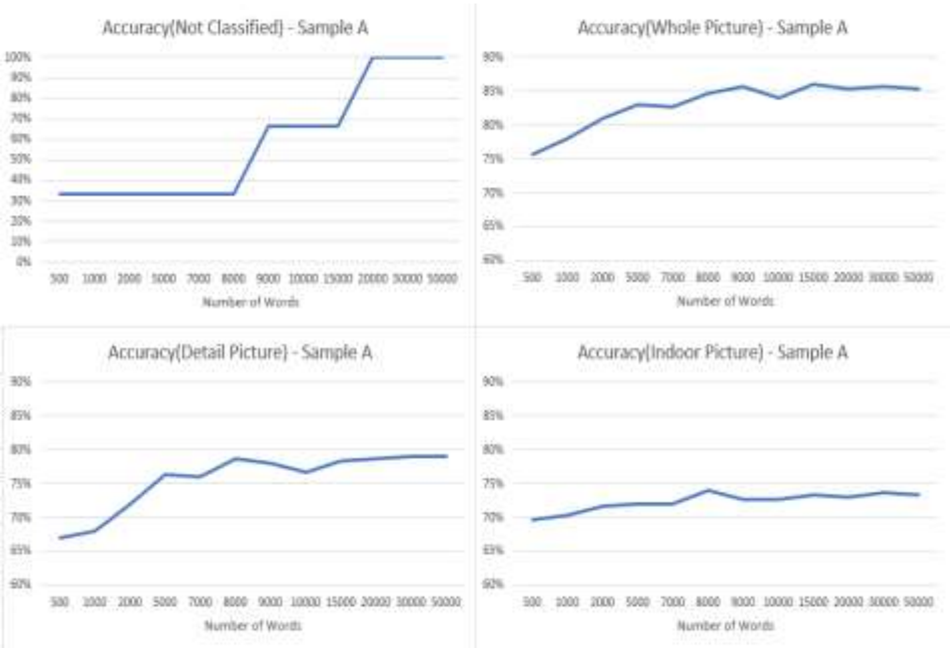


Figure 9. Photo Accuracy of Sample A by Category (Unit: %)

Figure 9 is a graph of accuracy for each classification of sample A. Since there are only 3 pictures of unclassified images, the accuracy is 33%, 66%, and 100%, and the remaining categories are 300 samples each. The whole images are the most accurate, with at least 75% to 85% accuracy.

The most inaccurate category are the indoor images, due to the fact that internal features are more varied than the exterior features of a car.

Figure 10 is a graph showing the time required for the learning of the sample A according to the number of words. After the learning, the experiment quickly finishes in about one minute, but the time required exponentially increases with the number of words, so if the number of words exceeds a certain level, another optimization method will be needed.

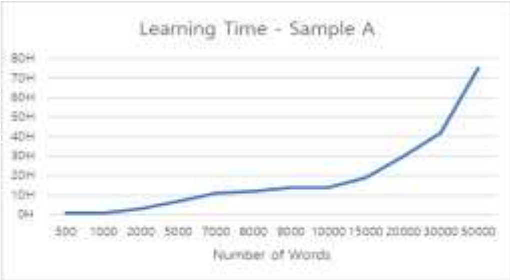


Figure 10. Time required for Bag-of-Word experiment using sample A (Unit: hour)

Figure 11 shows the overall accuracy of the experiment using sample B. The accuracy of sample B is growing faster than that of sample A and increasing the number of sample images appears to have a greater accuracy synergy than increasing the number of words.

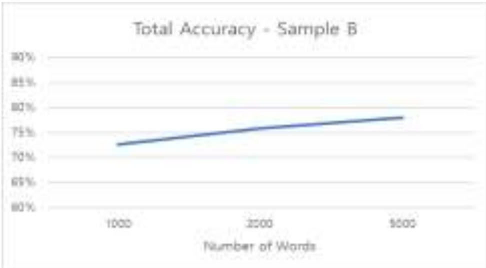


Figure 11. Overall Photo Accuracy of Sample B (Unit: %)

Figure 12 shows the accuracy of each type when using sample B. As in the case of using Sample A, the accuracy of the whole images was the highest and the accuracy of the indoor images were the lowest.

Figure 13 shows the time taken to learn when using sample B. The sample A consisting of 3007 images took more than 50 hours, while the sample B consisting of 5962 images took more than 1000 hours. Increasing the number of images seems to require modest optimization because the time required exponentially increases compared to increasing the number of words.

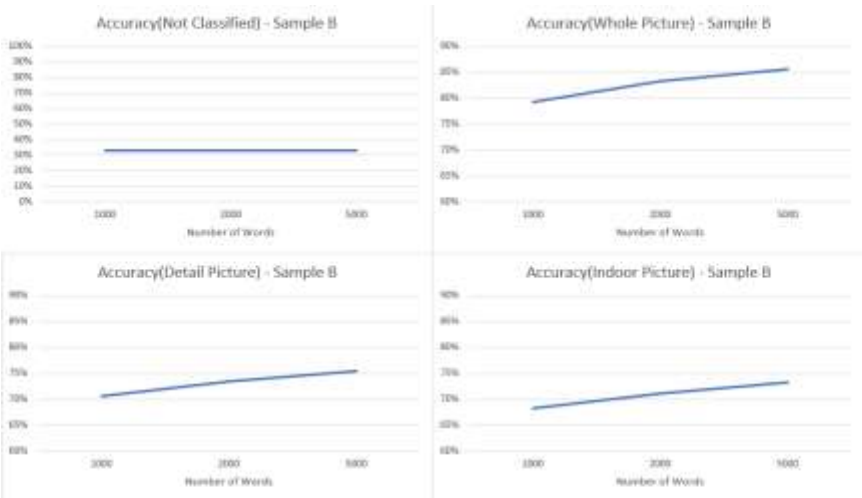


Figure 12. Accuracy of the Sample B's Classification (Unit: %)

Figure 13. Time Required for Bag-of-Word Experiment using Sample B (Unit: hour)

5. Conclusion

In this study, we proposed an automatic image classification method through gesture analysis. It is possible to automate the classification of the partial image of vehicles by using hand gestures. Also, the method of marking each gesture as metadata will not be limited to this study but used in various fields.

Classification of the vehicle images were also carried out using the Bag-of-Words method. Classification results have a performance of up to 80%, the disadvantage is that the learning time rapidly increases, but a more accurate classifier can be expected by increasing the number of images and words.

Acknowledgement

Following are results of a study on the "LINC+(Leaders in INdustry-university Cooperation +)" Project, supported by the Ministry of Education. And This paper is a revised and expanded version of a paper entitled A Method of Classification by Partial Image of a Vehicle using Gestures presented at 2017 1st International Workshop on Cultural and Technological Exchange and Mutual Development of The Pacific Rim Countries, Cheonan, Korea, December 17 2017.

References

[1] A. Chavez-Aragon, R. Laganiere and P. Payeur, “Vision-Based Detection and Labelling of Multiple Vehicle Parts”. Proceedings of the 14th International IEEE Conference on Intelligent Transportation Systems, Washington, DC, USA, **(2011)** October 5-7.

[2] M. T. Qadri and M. Asif, “Automatic number plate recognition system for vehicle identification using optical character recognition”, Proceedings of the International Conference on Education Technology and Computer, Singapore, Singapore, **(2009)** April 17-20.

[3] M.-Y. Jin, J. B. Park, D. S. Lee and D. S. Park, “Real-Time Vehicle License Plate Recognition System Using Adaptive Heuristic Segmentation Algorithm”, KIPS Transaction Software and Data Engineering, vol. 3, no. 9, **(2014)**, pp. 361--368

[4] W.-Y. Kang, T.-G. Lee, M.-H. Kwak and H.-J. Lee, “Development of Model Recognition Algorithm for Moving Vehicles”, Korean Information Science Society, vol. 39, no. 2C, **(2012)**, pp. 214-216.

[5] B. Qi, L. S. Lan, and P. Xu, “A CNN based car model recognition improvement”, Proceedings of the International Conference on Deep Learning Technologies, Chengdu, China, **(2017)** June 2-4.

[6] M. Huzaifa, and I. S. Suwardi, “Car model recognition from frontal image using BRISK”, Proceedings of Electrical Engineering and Computer Science (ICECOS), Palembang, Indonesia, **(2017)** August 22-23.

[7] H.-J. Lee, “Algorithm Based on Texture for the Recognition of Vehicles` Model”, The KIPS Transactions, vol. 12B, no. 3, **(2005),** pp. 257-264.

[8] E.-J. Lee and Y.-S. Suk, “A Vehicle License Plate Recognition Using Intensity Variation and Geometric Pattern Vector”, The KIPS Transactions, vol. 9B, no. 3, **(2002),** pp. 369-374.

[9] M.-Y. Jin, J.-B. Park, D.-S. Lee, and D.-S. Park, “Real-Time Vehicle License Plate Recognition System Using Adaptive Heuristic Segmentation Algorithm”, The KIPS Transactions on Software and Data Engineering, vol. 3, no. 9, **(2014)**, pp. 361-368.

[10] S. Yu, S. Zheng, H. Yang, and L. Liang, "Vehicle logo recognition based on bag-of-words". In Advanced Video and Signal Based Surveillance (AVSS), 2013 10th IEEE International Conference on, Krakow, Poland, **(2013)** August 27-30.
[11] A. F. Gómez, P. J. Hernández and B. Bacca-Cortes, "Vehicle classification based on a bag of visual words and range images usage". Ingeniería y competitividad, vol. 17, no. 1, **(2015)**, pp. 95-107.
[12] D.-Y. Lee, D.-K. Shin, and D.-G. Shin, "A Finger Counting Method for Gesture Recognition", Journal of Internet Computing and Services(JICS), vol. 17, no. 2, **(2016)**, pp. 29-37.
[13] H.-C. Yoon and J.-S. Cho, "Hand Feature Extraction Algorithm Using Curvature Analysis for Recognition of Various Hand Gestures", Journal of the Korea Society of Computer and Information, vol. 20, no. 5, **(2015)**, pp. 13-20.
[14] T. Starner, "Visual recognition of American Sign Language using hidden Markov models", Massachusetts Institute of Technology (MIT), **(1995)**, pp. 48-52.
[15] Q. Munib, M. Habeeb, B. Takruri, and H. A. Al-Malik, "ASL (American Sign Language) recognition based on Hough transform and neural networks", Proceedings of Expert Systems with Applications, vol. 32, no. 1, (2007), pp. 24-37.

Authors

Junho Jeong, received the B.S. degree from the Dept. of Computer Science, Dongguk University, Seoul, Korea, in 2007, and M.S. and Ph.D. degrees from the Dept. of Computer Engineering, Dongguk University, Seoul, Korea in 2009 and 2015, respectively. Currently, he is a Researcher of the Electronic Commerce Institute, Dongguk University, Gyeongju, Korea. His research areas include Privacy Preserving, Distributed System, Network Security and Secure Software.

Jun Young Lee, he is undergraduate student of Computer Engineering, Dongguk University, Seoul, Korea. His research areas include Artificial Intelligence, Embedded System.

Yunsik Son, Author's profile. He received the B.S. degree from the Dept. of Computer Science and Engineering, Dongguk University, Seoul, Korea, in 2004, and M.S. and Ph.D. degrees from the Dept. of Computer Science and Engineering, Dongguk University, Seoul, Korea in 2006 and 2009, respectively. He was a research professor of Det. of Brain and Cognitive Engineering, Korea University, Seoul, Korea from 2015-2017. Currently, he is an Assistant Professor of Dept. of Computer Science and Engineering, Dongguk University, Seoul, Korea. His research areas include secure software, programming languages, compiler construction, mobile/embedded systems, and u-Healthcare.

