Video Road Vehicle Detection and Tracking based on OpenCV

Wei Hou[1]* , **Zhenzhen Wu[2]** , and **Hoekyung Jung[3]*** , *Member*, *KIICE*

[1]Department of Electrical Engineering, Shaanxi Polytechnic Institute, Shaanxi 712000, China
[2]Weifang University of Science and Technology, Shandong 262700, China
[3]Department of Computer Engineering, PaiChai University, Daejeon 35345, South Korea

Abstract

Video surveillance is widely used in security surveillance, military navigation, intelligent transportation, etc. Its main research fields are pattern recognition, computer vision and artificial intelligence. This article uses OpenCV to detect and track vehicles, and monitors by establishing an adaptive model on a stationary background. Compared with traditional vehicle detection, it not only has the advantages of low price, convenient installation and maintenance, and wide monitoring range, but also can be used on the road. The intelligent analysis and processing of the scene image using CAMSHIFT tracking algorithm can collect all kinds of traffic flow parameters (including the number of vehicles in a period of time) and the specific position of vehicles at the same time, so as to solve the vehicle offset. It is reliable in operation and has high practical value.

Index Terms: Background model, CAMSHIFT algorithm, OpenCV, Static background, Vehicle tracking, Video surveillance

I. INTRODUCTION

Computer graphics, image processing, image processing, pattern recognition, artificial intelligence, artificial neural network, neurophysics and cognitive science, mathematics and physics are based on the intersection of information in a multi-disciplinary field [1]. In recent years, with the rapid development of sensor technology and multimedia technology, video monitoring system, vehicle automatic navigation, medical automatic diagnosis, product welding and inspection, map drawing, physical 3d reconstruction and recognition, intelligent man-machine interface and other fields are widely developed.

The intelligent transportation system is a comprehensive application of traffic information acquisition technology, information communication technology, computer control technology, sensor technology and computer information technology to effectively integrate the whole transportation system, which is established in a large range of efficient, accurate and real-time transportation integrated management system. The aim is to improve the traffic environment, make people, cars, roads effectively cooperate, harmonize, improve the efficiency of the transportation system and ensure the safety of traffic.

The research of intelligent video traffic system has very important theoretical significance and practical value for traffic safety and traffic control. Common vehicle detection methods are passive RADIO frequency identification, ultrasonic detection, microwave detection, video vehicle detection and so on.

The basic content of video vehicle analysis is to use real-time vehicle video or existing vehicle video, process video image frame by frame, extract moving vehicle, identify and track the extracted moving vehicle, and understand and describe its behavior.

Compared with traditional vehicle detection technology,

---

video vehicle detection has the following advantages: almost no impact on the road environment, signal transmission between visual systems will not interfere; Installation does not need to excavate the road surface, close the road, will not affect the normal traffic, low maintenance cost; A common CCD camera can simultaneously detect and collect multi-lane vehicle information within a few hundred meters, providing detailed traffic data. The vehicle detection technology based on video sequence is very important for the development of intelligent transportation system. Great promotion, for People's Daily life and national traffic road development has great practical significance [2].

In this paper, OpenCV and Visual C++6.0 are used to build an experimental platform, and a stationary camera is used to shoot a video of AVI for testing [3].

II. SYSTEM OVERALL PLAN

In this chapter, we look at the human body size data provided by the National Institute of Technology and Standards and related research using it.

Video image is composed of color image frame sequence, but the computer cannot process such data, so the transformation of color model, image grayscale, morphology processing and so on become the indispensable premise of computer image processing.

First convert the video recording to OpenCV support video format, for example, can handle have Uncompressed AVI video coding format RGB, 24 or 32 bit, Uncompressed YUV, 4:2:0 chroma subsampled, Identical to 1420, or install a decoder xVID on your computer. Secondly, frames are extracted from video sequences and images are preprocessed for background modeling. After that, the moving objects were detected and tracked, and the tracking objects were numbered. At the same time of tracking, constantly detect whether the target is beyond the warning line. If it is beyond the warning line, cancel the number of the object and update the number of the tracking object entering the warning line.

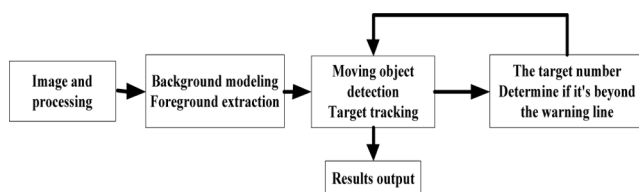The overall scheme block diagram of the system is shown in Fig. 1.

Fig. 1. Overall block diagram of the system.

In this paper, a moving vehicle tracking system is established by using the function library of OpenCV to track

moving objects, which is used for detecting and tracking moving vehicles on the road [4]. The composition of moving vehicle tracking system is shown in Fig. 2.
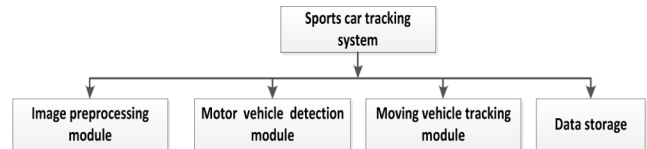

Fig. 2. Video vehicle analysis system.

A. Image Preprocessing Module

Most of the video files analyzed by the video vehicle tracking system are from color CCD cameras, which are often affected by various external light and shadow, and also caused by noise caused by the imaging error of the camera's own sensor and system circuit distortion. Therefore, the video sequence is preprocessed first, and an appropriate algorithm is selected to remove the noise dry [5,6]. Then image segmentation, edge detection, feature extraction, pattern recognition.

In this experiment, image grayscale change, image denoising, image binarization, mathematical morphological filtering were used for image preprocessing.

B. Image Greyscale

The process of image graying is to remove the color information of the image and only retain its brightness information. The three color components of a color image based on the RGB color space model are equivalent to each other, so the color image becomes a grayscale image. The phase equivalents of the three components are called grayscale values. In image processing, image graying is to change the values of R, G and B channels into single channel values in RGB model. The brightness value of each pixel in the grayscale image ranges from 0 to 255. 0 is the darkest (all black) and 255 is the brightest (all white). In THE RGB model, if R=G=B, even the data of three channels is also a grayscale color, and the grayscale value is the value of R, G and B.

There are many grayscale methods for color images [7]. In this paper, the weighted average method is used for grayscale of color images, which can be expressed by Formula (1):

$$f(i, j) = aR(i, j) + bG(i, j) + cB(i, j) \qquad (1)$$

Through searching for relevant materials and learning and a lot of experimental demonstration, it is more reasonable to obtain grayscale images when the weighted coefficients of A, B and C are 0.3, 0.59 and 0.11. When using this algo-

rithm, the coefficients of A, B and C are 0.3, 0.59 and 0.11, respectively, for gray conversion, and the conversion formula is shown in Eq. (2):

$$Gray = 0.3R(i, j) + 0.59G(i, j) + 0.11B(i, j)$$
$$Gray = R = G = B \tag{2}$$

In the upper type, gray indicates the gray value of the pixel in the image, and the red component of the pixel, which represents the green component of the pixel, and the blue component of the pixel. Image grayscale can be implemented by the function of the OpenCV function, which is implemented by the src-input of the original color floating point or 8 bit images, dstes-after the processing of the floating point or 8 bit floating point images, code-color space transformation, which is defined here as cv_rgb2gray-7. The original color image is grayscale through the above method, as shown in Fig. 3.

(a) The original chart   (b) Grayscale

Fig. 3. Grayscale of color chart.

C. Image Duality

Image segmentation is used in image segmentation, and the image segmentation method is usually based on the image segmentation method based on the area, the image segmentation method based on the edge detection, the image segmentation method based on the threshold value. Image domain value segmentation is a widely used image segmentation method.

In the gray image of the color image, the process of image bination is one of the two values of each pixel point in the original grey scale image, 0 or 255, which is set to 2555 in the image, and the background target is set to 0, so that after two values, the image will show that there are only black and white colors. In the process of image two value, the selection of the threshold value is determined by the effect of the second value, which also affects the accuracy of the further image processing and analysis, so the selection of the two value threshold is important. In terms of the difference between the target and the original background of the grey image, the two pixels are divided into two points, and then a proper threshold is divided into two sets of each of the previous two, which is shown in the original grayscale image, which completes the two-value process of the gray image. In

the processing of the graying map, the value of the vehicle is set to white, and the pixel value of the vehicle is set to black.

An appropriate threshold is selected to classify each pixel point in the image, that is, whether the pixel point belongs to the target or the background area, so as to generate the corresponding binary image and obtain the target to be detected [8].

The key of binarization processing is to select the threshold T, the principle is shown in formula (3) and (4) as follows. Let the input image be $f(x, y)$, and the output image be $f'(x, y)$.

$$f'(x, y) = \begin{cases} 1, f(x, y) \leq T \\ 0, f(x, y) > T \end{cases} \tag{3}$$

$$f'(x, y) = \begin{cases} 1, f(x, y) \geq T \\ 0, f(x, y) < T \end{cases} \tag{4}$$

The purpose of binarization is to divide the image into object and background. In practical processing, 0 is often used to represent the target object and 255 the background [9]. In OpenCV, the binary image is realized by the function Voide cv Threshold.

In this experiment, when the threshold value is 15, a lot of useless information is also extracted to become the target, while when the threshold value is 60, the target segmentation is distorted, so the selection of threshold value plays a crucial role in detecting the target. Experimental verification for vehicle detection, threshold 20 is the best value. On the basis of the original gray image, the threshold value is calculated by the maximum inter-class difference method and the binarization processing effect is shown in Fig. 4.

Fig. 4. Binarization rendering chart.

D. Image Denoising

This paper deals with urban traffic video sequence images from cameras. Due to the influence of the shooting machine itself and the signal transmission process, the image signal processed is often affected by the interference noise, which will be blurred by the interference noise points, which will cause great interference to the detection and tracking of the next image target. The key role of image denoising in this section is to reduce the interference of noise signals as much as possible on the premise of not changing the image target

1) Average Filtering

In the process of image denoising, the neighborhood average method is commonly used in mean filtering, which has a certain effect on gaussian noise removal in the processed image. The idea of the filtering algorithm for image noise processing is to select a point in the graph, set it as (x, y), and take this point as the center within the range of $M \times N$ to obtain the average value of pixel values of all points within the whole range, using formula 2.5.

$$g(x, y) = \frac{1}{m} \sum f(x, y) \qquad (2.5)$$

where m is the sum of all pixels in the range, and then take this average value as the pixel value of point (x, y).

The steps of the filtering algorithm are relatively simple, which has a certain smoothing effect on the noise in the processed image, but cannot completely eliminate the noise.

2) Median Filtering

Median filtering is an algorithm based on sorting statistics theory. Researchers have found that it can achieve good results in image denoising. Later, it is often used in image noise removal. The thinking of the algorithm for image noise removal as follows: the first randomly in the processed image for a bit, this point as the center, calculate the range of all pixels of the pixel value, and then put all the order of pixel values, among them pixel values as a point of pixel values, to eliminate the noise in the image signal.

In order to compare the removal effects of the above two filtering algorithms on Gaussian noise and salt-and-pepper noise, this paper added gaussian noise and salt-and-pepper noise to the images in the video sequence, and then carried out the filtering experiment. The experimental results are shown in Fig. 5.
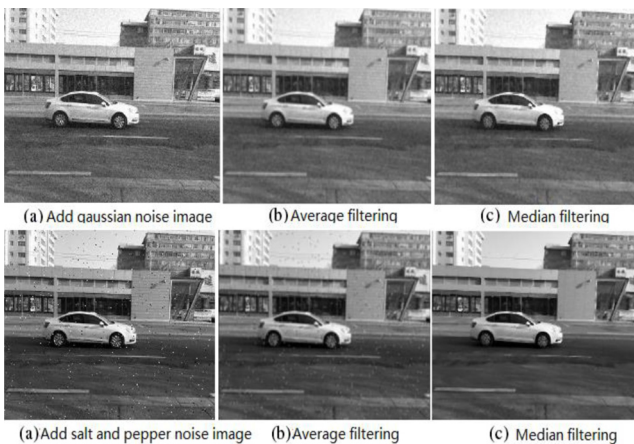

(a) Add gaussian noise image   (b) Average filtering   (c) Median filtering

(a) Add salt and pepper noise image   (b) Average filtering   (c) Median filtering

Fig. 5. Filtering effect chart.

In the image of the filtering effect of the image above, the filter effect of the two filtering algorithms is not very good for the filtering of gaussian noise, and the phenomenon of the edge blur is present in the filter processing. In the filter processing of pretzels noise, the effect of the median filter is better than the mean filter, and the median filter can filter the pretzels noise in the image, while the salt noise in the back image of the mean filter points out the obvious expansion of the hollow point phenomenon. Since the video image taken by the camera in this paper is more than the pretzels, this paper USES the median filtering algorithm in the case of image noise processing.

E. Mathematical Morphological Processing

In order to obtain a better vehicle test effect, the pixel points of all the targets in this image are expressed in white, and the points other than the target pixels are represented in black, except for the target pixels. After the test color image is processed, it not only causes the loss of target information in the process, but also is affected by some noise in the process, and the image that is processed has a sharp point or a small fine seam within the target of the vehicle, making the outline of the target edge uneven. Such images are applied directly to the later processing, which will affect the robustness of the subsequent algorithm. In this way, the problem of mathematical morphological filtering can be used to solve the above problems, and the operation of morphology can effectively fill the hollow inside the target, remove the isolated interference noise point outside the target, and achieve the function of smoothing the target contour, making the target of the detection clearer [10].

Mathematical morphology is a new method and theory in the process of extracting image components, which is very helpful in representing and describing the target region. It is only a general term, for the image processing can be carried out in many ways, for example, corrosion can effectively remove the isolated interference noise points in the image to be detected; Expansion can fill the small cracks in the target contour of the vehicle to be detected; As well as the combination of the two operations produced by the open and close operation, can effectively process the image to achieve the effect of smoothing the vehicle target contour. Based on these operators, a variety of mathematical morphology algorithms can be combined to analyze and process image structures.

In this experiment, cvMorphologyEx advanced morphological transformation, the function related to morphological operation in OpenCv, was used to combine expansion and corrosion operation, and then combined with set operation to complete some advanced morphological transformation: open operation and closed operation. Although the background is static in a broad sense, it is inevitable that the

camera itself will shake when collecting video, so the background of the sequence of frames will inevitably have background deviation, so that there will be many white spots in the segmentation image. If the closed operation is carried out, that is, expansion first and then corrosion, many white points will be merged, and it is difficult to eliminate corrosion. In the open operation of the image, the error information can be removed very well.

In this experiment, many white dots were merged in the closed operation, and the error information was removed in the open operation.

III. ALGORITHM FLOW DESIGN

A. Moving Vehicle Tracking Module

The CAMSHIFT algorithm is adopted in this target tracking experiment, which can effectively solve the problem of target deformation and partial occlution, and the operation efficiency is high.

The CAMSHIFT algorithm, which is full of self-aided adaptive-mean-shift, is a improvement in the mean shift algorithm, which automatically adjusts the size of the search window size to adjust to the target size, and can track the target of the size of the video. It is also a semi-automatic tracking algorithm that requires manual calibration tracking targets. The basic idea is to make the color information of the motion object in the video image, and each frame is a mean-shift operation, and the target center of the previous frame and the search window size (the core function bandwidth) are the initial values of the next frame mean shift algorithm, and the initial value of the search window size, so that the target is tracked by this iteration. Because the location and size of the search window are set to the location and size of the target current center before each search, the target is usually near the area and shorten the search time. In addition, the color changes in the target movement, so the algorithm has good robustness. It has been widely used in sports human tracking, face tracking and other fields

B. CAMSHIFT Algorithm Process

Figure 6 shows the flow chart of CAMSHIFT algorithm. The algorithm process is as follows:
(1) Initialize the search window and set the window size to S.
(2) Set the processing area. The center of the area is 1.1 times the size of the search window.
(3) Calculate the color histogram of the search window.
(4) Run the Mean-Shift algorithm to obtain the new size and position of the search window.
(5) Reinitialize the size and position of the search window

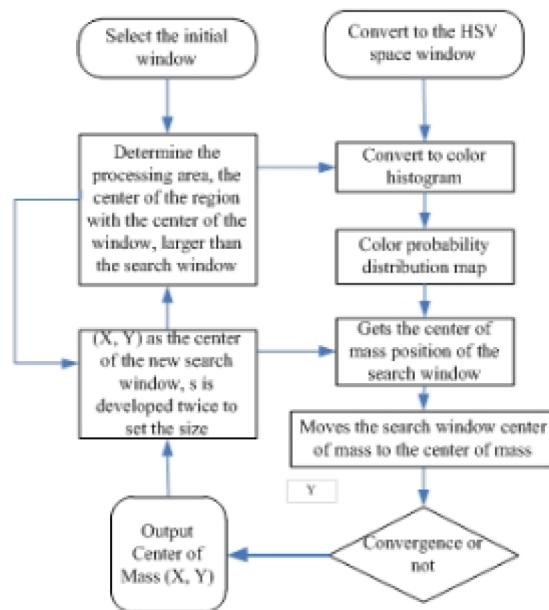with the value in (4) in the next video image frame, and continue after jumping to (2).

Fig. 6. CAMSHIFT algorithm flow chart.

C. Location of Vehicle

Mixed Gaussian background modeling is used in moving target detection. Its modeling is based on the background representation method of sample statistics, which uses statistical information such as probability density of a large number of sample values of pixels in a long time (such as the number of modes, mean and standard deviation of each mode) to represent the background, and then uses statistical difference to judge the target pixels. It can model complex dynamic background and has a large amount of computation [11].

In the mixed Gaussian background model, it is considered that the color information between pixels is not correlated, and the processing of each pixel is independent of each other. For each pixel in the video image, the change of its value in the sequence image can be regarded as a random process of constantly generating pixel value, which is to describe the color presentation rule of each pixel with Gaussian distribution.

Gaussian mixture model can separate the background with regular changes and update the background in real time. The color value of each pixel point in the image is regarded as a random process X, and it is assumed that the probability of the occurrence of the pixel value of this point follows the Gaussian distribution. Let (γrt, γgt, γbt) represent the pixel value at time t, μ, I, t and σ, I, t are the expected value and standard deviation of the Gaussian distribution of the pixel at time T, respectively.

Open CV is used to realize real-time regional location of vehicles, and only the frames saved after vehicle tracking and recognition are recognized, which improves the recognition efficiency. First you need to initialize the parameters. Then, on the basis of parameters, the gaussian mixture model processing process for the first frame image transmitted is as follows:

(1) Initialize the background model with the first frame image data, where *std_init* Set to 20 after test and debugging.

$$\sigma^2{}_0 = std\_init^2$$
$$\mu_0 = (\mu^r{}_0, \mu^g{}_0, \mu^b{}_0) \qquad (3.1)$$
$$\sigma_0 = std\_init$$

(2) Each pixel point in the current image is matched with the mixed Gaussian model. If successful, the point is judged to be the background point; otherwise, it is the foreground point. At time t, each pixel value X_t of the frame is matched with its corresponding mixed Gaussian model: foreground and background pixels are detected. If the distance between pixel value X_t and the mean value of the i-th Gaussian distribution in the mixed Gaussian model is less than 2.5 times of its standard deviation (λ=2.5), the Gaussian distribution Gi is defined to match pixel value X_t. Background pixel detection formula:

$$\left| X_t - \mu_{t-1} \right| < \lambda \sigma_{t-1} \qquad (3.2)$$

Foreground pixel detection formula:

$$\left| \overline{X}_t - \mu_{t-1} \right| \geq \lambda \sigma_{t-1} \qquad (3.3)$$

(3) If at least one Gaussian distribution in the pixel mixed Gaussian model matches the pixel value, the parameter updating rules of the mixed Gaussian model are as follows:

① For unmatched Gaussian distributions, their mean and variance remain unchanged;
② The matched Gaussian distribution, the parameters of the model μ_i, t, $\sigma^2_{i,t}$ are updated according to Equations (3.4) to (3.6).

$$\mu_{i,t} = (1-\rho) \times \mu_{i,t-1} + \rho \times X_t \qquad (3.4)$$

$$\sigma^2{}_{i,t} = (1-\rho) \times \sigma_{i,t-1} + \rho \times \left(X_t - \mu_{i,t} \right)^2 \qquad (3.5)$$

$$\rho = \alpha \times \eta \left(X_t \mid \mu_{i,t-1}, \sigma_{i,t-1} \right) \qquad (3.4)$$

Here, α is the learning efficiency of parameter estimation.

IV. EXPRRIMENTAL RESULTS AND ANALYSIS

A. The Development Environment

In this paper, the motion target detection and tracking algorithm is studied, and the relevant experimental analysis is done. The paper hardware platform is based on the cool-processor, 1.87GHz, 2.00GB running memory, Windows 32 system notebook, using the Visual Studio 2010 development environment, using the C++ language and combining the OpenCV library function to write the program to realize the tracking of the motor vehicle.

This is a hybrid gaussian background modeling method, which is shown in Fig. 7.

Fig. 7. background image.

When the threshold 15 is selected, a lot of useless information is also extracted as the target. When the threshold 60 is selected, the segmented target is distorted, so the selection of the threshold plays a vital role in detecting the target. Experiments have verified that for vehicle detection, the threshold of 20 is the best value, as shown in Fig. 8:

Fig. 8. Target extraction diagram.

There are four stages of actual measurement on the expressway, first, target detection, target tracking, target overlap and target numbering.

Fig. 9. Inspection of cars No. 0 and No. 1.

After that, track 0 and 1. As shown below:

Fig. 10. Target tracking diagram.

When there are vehicles overlapping, as shown in Fig. 11(a)(b) below:
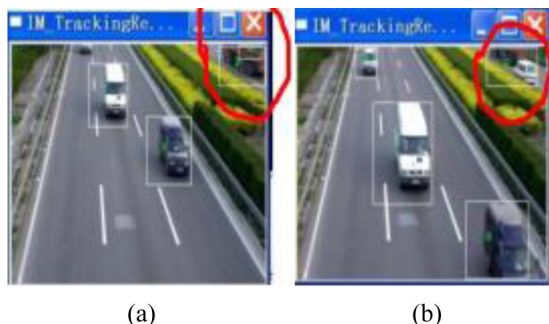

(a)   (b)

Fig. 11. (a) (b) target overlap map.

From the Figure (a) in the red circle, it can be seen that targets 2 and 3 overlap, but the numbering has not changed, and the correct numbering is still carried out. Picture (b) shows that when the red vehicle No. 2 walks out of the cordon, the white vehicle No. 3 immediately following is changed to No. 2.

This is a major advantage of CAMSHITFT algorithm tracking. Target occlusion, deformation, it can track well.

Number the targets that enter the cordon. For the vehicle numbered 2, the targets in front are 0, 1, ...

If the target numbered 0,1 goes out of the warning line, then the target numbered 2 is changed to number 0. The following is an experimental verification. Fig. 12(a)(b).
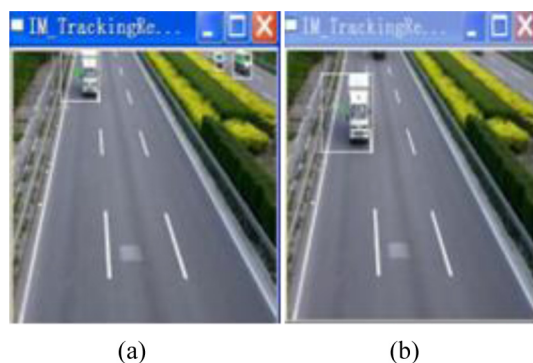

(a)   (b)

Fig. 12. (a) (b) target number

As can be seen from Figure A, the number of the third car entering the warning line is 2. Figure B shows that after the vehicle numbered 0, 1 walked out of the police line, the vehicle numbered 2 changed to 0. This paper has a high accuracy in vehicle detection and tracking, we can see the practicality and accuracy of the algorithm in this paper.

V. CONCLUSIONS

Moving vehicle detection and tracking is the main research content of intelligent transportation system. It is the most important research work to locate the target accurately in real time, analyze the behavior and finally track the moving vehicle. In recent years, through unremitting efforts, a large number of scholars have carried out extensive and in-depth research on target detection and tracking algorithms, and obtained effective algorithms for moving target detection and tracking in different environments. However, due to the variety of life scenes and the influence of various environmental disturbances, there are still many areas to be optimized and improved in target tracking. In this paper, the moving vehicle detection and tracking in the urban road environment is studied, mainly image pretreatment, detection and tracking of moving vehicles, system software design of these aspects. By processing the road traffic video shot by the fixed camera, the single vehicle and multiple vehicles can be tracked in the road environment with the influence of light and shadow. Through the system software test, the detection and tracking of moving vehicles can be better completed. The algorithm of Gaussian background model is improved continuously so that the noise suppression of the model will not be affected. In this way, the "shadow" caused by the original stationary object in the scene when it starts to move can disappear quickly, and finally absorb the tracking of multiple targets. It is the focus of future research to further combine motion estimation and structure information to improve multi-target tracking algorithm.

REFERENCES

[1] L. M. Sin, E. H. Lee, and S. Y. Oh, "The effect of creativity on job satisfaction and job performance in beauty service employees," *Journal of the Korean Society of Cosmetics and Cosmetology*, vol. 9, no. 3, pp. 339-350, Dec. 2019.

[2] S. Y. Go, "A comparative study of characteristics of the beauty major students," *Journal of the Korea Contents Society*, vol. 20, no. 3, pp. 336-344, Mar. 2020. DOI: 10.5392/JKCA.2020.20.03.336.

[3] S. H. Kim, Y. G. Seo, and B. C. Tak, "A recommendation scheme for an optimal pre-processing permutation towards high-quality big data analytics," *The Korean Institute of Information Scientists and*

Engineers, vol. 47, no. 3, pp. 319-327, Mar. 2020. DOI: 10.5626/JOK.2020.47.3.319.

[4] Li Zhaohui and Yu Yinglin, "A method for automatic location, tracking and recognition of video text," *Chinese Journal of Image and Graphics*, 2005, vol. 10, no. 4, pp. 457-462, Apr. 2015.

[5] J. O. Jung, I. Y. Yeo, and H. K. Jung, "Classification model of facial acne using deep learning," *Journal of The Korea Institute of Information and Communication Engineering*, vol. 23, no. 4, pp. 381-387, Apr. 2019. DOI: 10.6109/jkiice.2019.23.4.381.

[6] L. Lessig, *Remix: Making Art and Commerce Thrive in the Hybrid Economy*, Penguin Press, New York, 2008.

[7] J. Rifkin, *The Zero Marginal Cost Society: The Internet of Things, the Collaborative Commons, and the Eclipse of Capitalism*, St. Martin's Press, 2014.

[8] Seoul Metropolitan Government, "Report on the 2018 Sharing City Recognition Survey," 2018.

[9] C. Lidong, *Dix. Human-computer interaction [M]*, 3rd ed. Beijing: Electronic Industry Press, 2006.

[10] R. Botsman, "The Sharing Economy Lacks A Shared Definition," *Fast Company*, Nov. 2013.

[11] http://www.collaborativeconsumption.com/2013/11/22/th e-sharing-economy-lacks-a-shared-definition/

[12] B. Y. Han, "Deep learning: Its challenges and future directions," *Communications of the Korean Institute of Information Scientists and Engineers*, vol. 37, no. 2, pp. 37-45, Feb. 2019.

Wei Hou

He received his bachelor's degree in Computer Science from PaiChai University in 2020. Since 2020, he has been studying for a master's degree at PaiChai University. His current research interests are deep learning, machine learning, big data, and computer vision.

Zhenzhen Wu

She received an MS degree in 2010 from the College of Information Science and Engineering, Ocean University of China. She is currently enrolled in a PhD course in the Department of Computer Engineering of PaiChai University. Her current research interests include big data and artificial intelligence.

Hoekyung Jung

He received an M.S. degree in 1987 and Ph. D. degree in 1993 from the Department of Computer Engineering of Kwangwoon University, Korea. From 1994 to 1995, he worked for ETRI as a researcher. Since 1994, he has worked in the Department of Computer Engineering at Paichai University, where he now works as a professor. His current research interests include multimedia document architecture modeling, information processing, embedded system, machine learning, big data, and IoT.