The Design of Deep-Learning-Based Commodity Identification System for Smart Shopping Cart

Shin-Yeu Yeh, Ying-Jen Chen, *IEEE Member*

***Abstract*— Technology is developing very fast now. A technology called artificial intelligence is widely used in our lives. It saves us lots of time and manpower, and also makes many things more convenient. Moreover, the combination of Artificial Intelligence, AI, and the Internet of Things (IoT) has become a new technology AIoT that can be used in combination with robots. The main purpose of this study is to design a commodity identification system based on machine vision and deep learning technology for smart shopping carts. At first, we use Gaussian mixture model with image processing technology to the commodity image for segmenting the new commodity put into the cart, and then classify the new commodity based on the convolutional neural network. Finally, the identification result is presented on the user interface.**

I. Introduction

In recent years, with the rapid development of computer science, artificial intelligence (AI) technology has been widely used in various systems and environments. It not only saves lots of manpower and time, but also brings more convenient functions. AI has gradually changed our lifestyles, such as smartphone applications, tablets for ordering in restaurants, or navigation robots in exhibition halls.

AIoT, which integrates AI and the Internet of Things (IoT), has become one of the mainstream trends in future technologies. In industry, The smart factory integrates AIoT technology into the system to connect all devices in the factory into a network for obtaining the data. Through these data, the purpose of improving product quality, reducing costs, and predicting device dconditions can be achieved. In addition to AIoT, smart robots are also one of the most important applications of AI. With the assistance of different sensors and AI, robots can achieve variety purposes such as perceiving surrounding conditions, identifying objects [1], [2], making autonomous judgments and actions, interacting with humans and assisting work. Moreover, robots can be connected with other devices through IoT technology.

The combination of AIoT and intelligent robots will promote the industry development, and the future industry model will be different. In order to cope with this shift, we hope to design a smart shopping cart based on machine vision and deep learning technology. It will have functions such as commodity identification, automatic tracking the customer, shopping guide and identity recognition. This research mainly focuses on the commodity identification system based on image processing, Gaussian mixture model and deep learning technology. In our overall structure, we use video cameras to read color images of commodities as input, and perform image processing through high-performance laptops placed in the shopping carts. Moreover, we use neural network for commodity identification, and finally display the result in the user interface.

There are many studies on commodity identification [3], [4]. Our research method is to use a single neural network combined with a Gaussian mixture model, and also use a convolutional neural network for product feature point extraction and product classification. Before we input the neural network, we use the Gaussian mixture model to segment foreground images. By this way, only the image of commodity newly put on the platform rather than a complex image containing all commodities is obtained. After that, we let the segmented commodity images be the training data for the neural network for commodity identification.

II. Related Work

Our research method mainly uses two core technologies: Gaussian mixture model and GoogLeNet neural network. The Gaussian mixture model is used to extract commodity images, and the GoogLeNet neural network is used for object classification, and the purpose is to combine the two for real-time commodity identification.

A. Gaussian Mixture Model

The Gaussian Mixture Model [5] is a method commonly used to build background models, and it has been widely used to detect moving objects or changes in the screen in continuous images [4]-[8]. The background pixel values in the image will not be fixed, and they will change due to the movement of the object and the change of brightness, such as the moving vehicle on the road, the change of the direction of light irradiation or the shadow of the shadow. These reasons will cause the background pixel value to float around the original value, so it is very suitable to use Gaussian distribution to model the background of these images.

The Gaussian distribution is a combination of multiple normal distribution curves to approximate the gray-level probability histogram of a pixel, as shown in Fig. 1. If t represent any point in time, the latest historical value of the pixel can be defined as $\{\boldsymbol{X}_1, \boldsymbol{X}_2, \cdots, \boldsymbol{X}_t\}$, then the probability density function of each Gaussian distribution curve is given as

Research supported by the Ministry of Science and Technology of Taiwan (MOST 109-2221-E-305-008-).

S.-Y. Yeh is with the Department of Electrical Engineering, National Taipei University, New Taipei City 237303, Taiwan (R.O.C.) (e-mail: xxszcvfh@gmail.com).

Y.-J. Chen is with the Department of Electrical Engineering, National Taipei University, New Taipei City 237303, Taiwan (R.O.C.) (corresponding author to provide phone: +886-2-8674-1111#68819; fax: +886-2-26736500; e-mail: yjcheng@mail.ntpu.edu.tw).

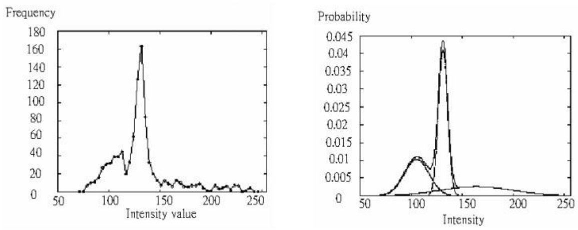


Figure 1. Gaussian mixture model.

$$\eta\left(\boldsymbol{X}_t,\boldsymbol{\mu},\boldsymbol{\Sigma}\right)=\frac{1}{2\pi\left|\boldsymbol{\Sigma}\right|^{\frac{1}{2}}}e^{-\frac{1}{2}\left(\boldsymbol{X}_t-\boldsymbol{\mu}\right)^T\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{X}_t-\boldsymbol{\mu}\right)} \tag{1}$$

where $\boldsymbol{X}_t$ is the (R, G, B) value of the pixel at time t, $\boldsymbol{\mu}$ is the average of the Gaussian distribution, and $\boldsymbol{\Sigma}$ is the standard deviation of the Gaussian distribution. If one of the pixels is modeled by K Gaussian distributions, the probability density function of the Gaussian mixture model is

$$P\left(\boldsymbol{X}_t\right)=\sum_{i=1}^{k}\omega_{i,t}\eta\left(\boldsymbol{X}_t,\boldsymbol{\mu}_{i,t},\boldsymbol{\Sigma}_{i,t}\right) \tag{2}$$

where $\omega_{i,t}$ is the weight of the i-th Gaussian distribution; $\boldsymbol{\mu}_{i,t}$ is the average of the i-th Gaussian distribution, and $\boldsymbol{\Sigma}_{i,t}$ is the standard deviation of the i-th Gaussian distribution. The larger the K is, the closer it is to the original image. However, the amount of calculation will also increase. Therefore, some adjustments need to be made according to the situation.

B. GoogLeNet

GoogLeNet [9] was proposed by Google in the 2014 ILSVRC competition. This network has a very large number of layers and the calculation effect is quite good. The new methods used by GoogLeNet is called Inception. The architecture of Inception network is shown in Fig. 2. When the model is trained, convolution processing of different sizes and maximum pooling will be used to link the data. The network architecture allows the data to take different levels of features in the same layer. In order to avoid the disappearance of the gradient, two softmax functions are added for auxiliary classification. The basic Inception only has 3x3 and 5x5 convolutional layers. An additional 1x1 convolutional layer can be introduced to reduce the dimensionality to solve the problem of excessive parameter volume. On the one hand, it increases the amplitude of the network, and on the other hand, it also increases the network. The adaptability to the scale improves the utilization of computing resources inside the network.

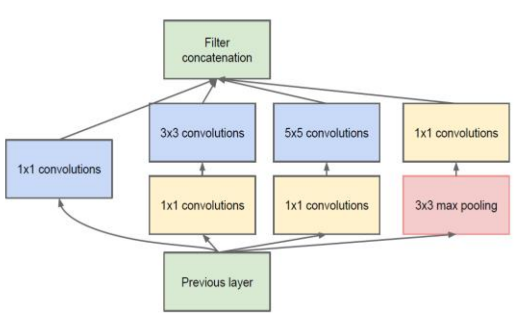


Figure 2. Inception network architecture [9].

III. RESEARCH METHOD

This section introduces how to extract commodity images and enable the images to be intercepted and identified correctly to achieve the purpose of commodity identification and display the results in the user interface. The flowchart of the commodity identification system is given as shown in Fig. 3. First of all, when we select a commodity, the camera is turned on and the Gaussian mixture model performed. At the same time the RGB image is converted to an HSV image to obtain the initial background image of the platform. Because of the influence of light, we must pre-process the camera image. When the commodity is placed on the platform, the Gaussian mixture model will perform background subtraction to extract the image of the newly-placed commodity. Since there is a lot of noise in the obtained image, it needs to be eroded and dilated. Moreover, the too large and too small parts should be excluded according to the edge of the image. After that the system count the time of the commodity staying. Then this commodity image is captured and stored. After the captured images are received by the system, the pre-trained neural network model is used for commodity identification. Finally commodity statistics are performed and displayed in the user interface.

A. Real-time Image Processing

Before processing the commodity images, we need to deal with the light problem. Since the change of light and shade will affect the separation result of background and foreground, we convert the input image of the camera from RGB to HSV and adjust the parameters of V. The change of value affected by the light is reduced so that we can filter some problems caused by the light.

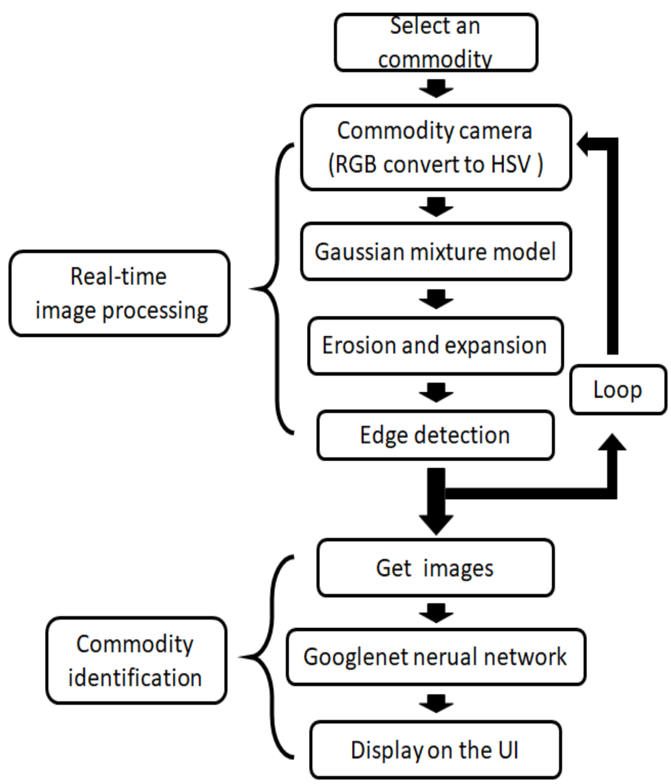


Figure 3. The flowchart of the commodity identification system

After the conversion is completed, the Gaussian mixture model is performed to set the storage platform as the initial background. When a commodity is placed into the storage platform and captured by the camera, the background subtraction will be executed to extract the newly-placed commodity. After the background subtraction, we will erode and dilate the subtracted image. In the process of erosion, small noises are filtered, and the blocks separated by complex colors are enlarged and connected during expansion to achieve the purpose of highlighting the complete commodity image. In order to avoid the situation where the captured image is too small to be distinguishable or too large to cause too much content, we set the threshold of edge to capture the processed images. By this way, filtering is completed and the commodity image can be segmented from the whole image as shown in Fig. 4.

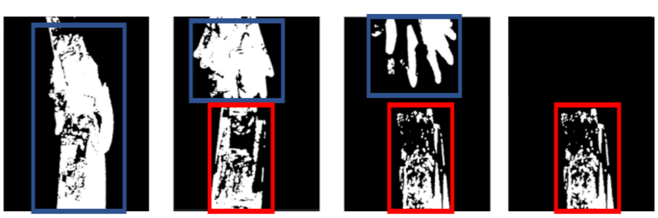
The commodity detection is performed in streaming images. Therefore, if you directly capture the segmented commodity images, there will be many same commodities being detected for only one commodity. For solving this problem, we use the number of frames that the image stays on the screen as the basis for judgment. If the image has almost no change in the screen for 10 frames, then we recognize it to be a newly-placed commodity image for sending to the identification neural network. On the contrary, if the detected item in the screen continues to move or disappears in a flash, no action is taken, and the item is judged to be in a state where the item has not been placed. The process for recognizing a newly-placed commodity is shown in Fig. 5 where the stationary object in the red rectangle is recognized as the newly-placed commodity.

B. Neural Network Classification and Training Data

We apply TensorFlow deep learning library for training and verification. The commodities are divided into 6 categories, and each category has 500 images for training, and the learning rate and number of drops corresponding to this neural network are set to 0.001 and 10000 times respectively.

(a) Original Image (b) Add New Commodity (c) Segmentation Image

Figure 4. Commodity segmentation.



(a) Moving Object (b) Moving and Stationary (c) Moving and Stationary (d) Stationary Object

Figure 5. The process for recognizing a newly-placed commodity.

Regarding the pre-training data, after we actually put the products on the platform, the processed captured images are used to be the training data. In addition, other commodities will be placed around to simulate the actual situation when a variety of items are actually placed. If the photos of general commodity are used as training data, there will be too large a gap with the actual situation such that the identification results will be inaccurate. These images do not need to be the same size. Both mobile phone photos and computer screenshots can be put into the training data, as long as it contains the characteristics of the commodity.

Because the pre-training requires a lot of data, we not only use images of complete commodities, but also images containing only some characteristics of commodities such as the trademark of the cookie brand. After that each image selected in previous step is rotated 5 times in a 60-degree interval. In this way, we can obtain a 6 times images to be the training data as shown in Fig. 6.

D. User Interface

As shown in Fig. 7., we have designed a user interface, for this system that can be used for manually switching the screen of the commodity camera lens and is connected to the back-end commodity database. The identification results are displayed on the screen and the prices of commodities are displayed at the same time. Finally, you can press the checkout button to proceed settlement.

IV. EXPERIMENT RESULTS

This section is divided into two parts to show the results of commodity image extraction and commodity identification. The goal we designed the system is to be used on smart shopping carts. Its core idea is to combine machine vision with deep learning. Therefore, through the experimental results, we can see the possibility and development of this idea.

Figure 6. Rotating the image to obtain 6 times training data.

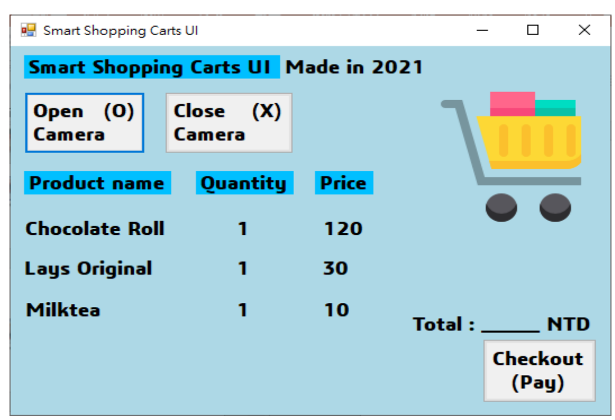


Figure 7. User interface.

A. Commodity Image Segmentation

For the commodity identification system, the most important step after placing the commodity is to correctly capture its image. We use the Gaussian mixture model for background subtraction, and then further erode and expand the image. Finally, a box that fits the size of the product is selected, and the selected image is sent to the pre-trained neural network for identification, as shown in the figure Fig. 9.

B. Commodity Identification

After the commodity image is obtained in the previous step, and the deep learning neural network is immediately run for commodity identification. Since other tasks are completed in pre-training, the results can be obtained after running the deep learning neural network and displayed in the user interface. If a new commodity is superimposed, the camera can continuously detect the newly-placed commodity on the storage platform. However, in order to enable the system to correctly identify the commodity items and quantity, the image capturing will be temporarily stopped when the deep learning model is run. The function is stalled and waiting for the Gaussian mixture model to set the newly-placed commodity as the background to facilitate the detection of the next commodity.

C. Commodity Recognition Accuracy Rate

We selected 5 products and performed 3 random sets of 100 commodity identification. The average accuracy was 89%. The biggest error was caused by the shadow such that the image changes too much. The system included images containing other commodities. The range of extraction makes the features of other commodities appear and cause the identification failure.

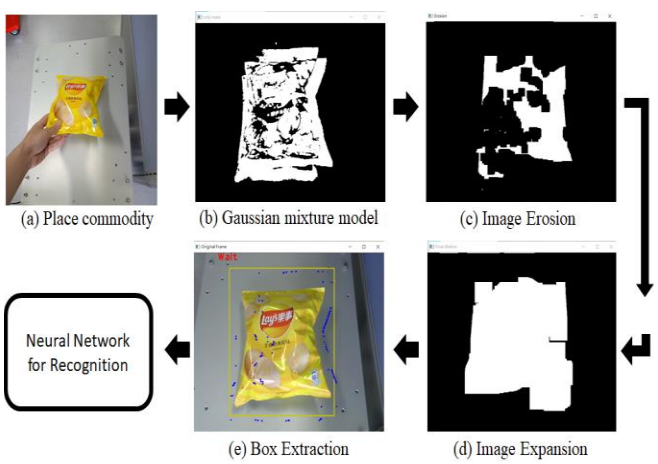


Fig. 8. Image processing for single commodity.

V. Conclusion

The purpose of this research is to design a real-time commodity identification system, which combines machine vision with deep learning. If the method is further combined with AIot technology, it can be used for smart shopping carts. In this research, how to correctly capture the commodity image is most important. A successful process is to successfully detect the commodity through machine vision. After the commodity image is selected, it can be identified by the deep learning neural network.

In this research, there are many things that can be refined and optimized, such as the accuracy of product image capture. The environment that causes machine vision to produce bad commodity images is also a problem worthy of discussion, such as the reflection and refraction of light.

References

[1] J. Li, X. Wang, and H. Su, "Supermarket commodity identification using convolutional neural networks," in Proceedings of 2016 2nd International Conference on Cloud Computing and Internet of Things (CCIOT), Dalian, 2016, pp. 115-119.

[2] L. Zhao, J. Yao, H. Du, J. Zhao and R. Zhang, "A Unified Object Detection Framework for Intelligent Retail Container Commodities," in Proceedings of 2019 IEEE International Conference on Image Processing (ICIP), Taipei, Taiwan, 2019, pp. 3891-3895.

[3] L. Zhao, J. Yao, H. Du, J. Zhao and R. Zhang, "A Unified Object Detection Framework for Intelligent Retail Container Commodities," in Proceedings of 2019 IEEE International Conference on Image Processing (ICIP), Taipei, Taiwan, 2019, pp. 3891-3895.

[4] E. Kiana, and S. Homayouni, and M. A. Sharifi, and M. Farid-Rohani, "Unsupervised change detection in SAR images using Gaussian mixture models," International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. XL-1/W5, pp. 407-410, 2015.

[5] C. Stauffer and W. E. L. Grimson, "Adaptive background mixture models for real-time tracking," in Proceedings of the 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Fort Collins, CO, USA, 1999, pp. 246-252 vol. 2.

[6] S. Li, H. Tsai, Y. Wang, T. Sun and Y. Chen, "Particle-filter-based intelligent video surveillance system," in Proceedings of the 2016 International Conference on System Science and Engineering (ICSSE), Puli, 2016, pp. 1-4.

[7] Gang Yang, Heng-Chao Li and Chi Liu, "Unsupervised change detection of remote sensing images using superpixel segmentation and variational Gaussian mixture model," in Proceedings of the 9th International Workshop on the Analysis of Multitemporal Remote Sensing Images (MultiTemp), Brugge, 2017, pp. 1-4.

[8] S. T. Ali, K. Goyal and J. Singhai, "Moving object detection using self adaptive Gaussian Mixture Model for real time applications," in Proceedings of the 2017 International Conference on Recent Innovations in Signal processing and Embedded Systems (RISE), Bhopal, 2017, pp. 153-156.

[9] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," arXiv:1409.4842 [cs.CV], 2014.

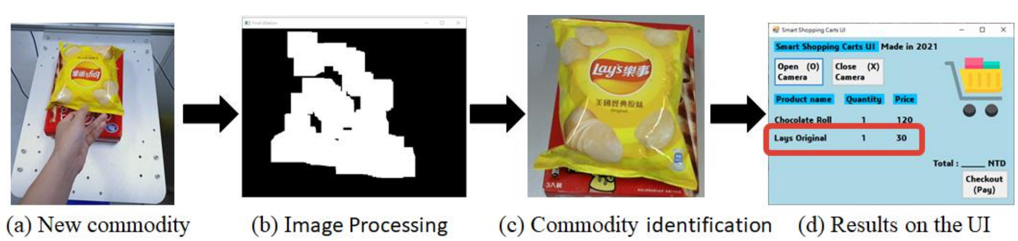


Figure 9. The process of commodity identification and the results on the UI.