Preliminary Study of planar motion tracking control system based on deep learning object detection method

Ming-Chang Lin, Yu-Ming Hsu

***Abstract*—The purpose of this work is to use the neural-network based deep learning method from the planar image data to establish a two-dimensional displacement sensor and apply it to the tracking control of planar motion. Compared with the traditional one-dimensional displacement sensor, the two-dimensional sensor can increase the flexibility of system development. In addition to building the sensor, a brand-new two-dimensional mechanism instead of the traditional X-Y platform is proposed. The advantages of the mechanism are compact, small, and light, and it is suitable for applications with limited space and fast response. To verify the effectiveness of this method, this work uses human faces as the detection target and uses a motor to drive the face image to rotate at a constant speed with a radius of 0.1 meters. The image capture rate is 15 fps (frame per second). Preliminary experimental results show that the maximum detection errors of the image position are 0.075 m and 0.03 m under the rotation speed of 40 rpm and 20 rpm, respectively. In the future, it will be further used with a two-dimensional mechanism to achieve face tracking control.**

Keywords: deep learning, two-dimensional displacement sensor, two-dimensional mechanism, face tracking control

I. INTRODUCTION

In the development of a good planar tracking control system, in addition to the robust controller design, the design of the sensor and the driving device also play important roles. In the design of the sensor, the two-dimensional displacement detection capability is mainly considered. Meanwhile, the size of the mechanism and the driving ability are important considerations for the design of the driving device. This work proposes a brand-new approach and preliminary study for these two parts.

With the rapid development of microcomputers, artificial neural network technologies that used to be limited by computational constraints have made significant progress, and their applications have become larger and larger. In image processing, the operation of convolution processing can effectively extract image features, making image recognition and object detection greatly improved. Image recognition has been successfully used in medical diagnosis and factory fault diagnosis [1-3]. And object detection has also been successfully used in the intelligent driving assistance and collision avoidance design of the automotive industry[4]. This work hopes to use the deep learning method to realize the measurement of two-dimensional displacement and apply it to the tracking control of planar motion. Traditional displacement sensors are mostly one-dimensional, and only a few, such as cameras, can provide two-dimensional information. This work is based on the image of the camera to realize the two-dimensional displacement detection of the object. The displacement signal can be fed back to achieve motion tracking control. In addition, to improve tracking control efficiency, this work also proposes a brand-new two-dimensional mechanism to replace the traditional X-Y table. The advantages of the mechanism are compact structure, small and light weight, suitable for applications with limited space and quick response. To verify its effectiveness, this work intends to design a face tracking control platform to evaluate the accuracy and dynamic response of the control. This method can be applied to sport simulators in the leisure and entertainment industry, or fire control systems in the military industry, and automation of factory equipment.

The steps of deep learning in object detection of images can be divided into three stages: feature extraction, image recognition, and object positioning. In the past, the feature value was obtained by convolutional neural network (CNN) for classification, and then its position was obtained. For example, the regional convolutional neural network (R-CNN) model architecture [5-7] was proposed to draw 2000 possible regions of the image and put the possible regions into a CNN model to extract features; then, the position is obtained through linear regression. The disadvantage of this method is the large amount of calculation, so it cannot achieve real-time object detection. To reduce the computing time, an architecture called YOLO (You Only Look Once) [8] was proposed. The feature of this method is that it only needs to perform CNN processing on the picture once, so the recognition speed is greatly improved. Afterward, there have been revised versions of YOLO-v2, YOLO-v3[9], etc. The convergence speed, image adaptability, and detection accuracy have been greatly improved. This work is based on YOLO-v3 to establish an object training database for the development of two-dimensional object displacement detection.

In the mechanism design that realizes two-dimensional motion, the most general design is an X-Y table or five-bar linkage [10-11]. There are two designs on the X-Y table, one is a screw drive and the other is a belt drive. The advantage of the screw drive is that the accuracy is high but the transmission speed is relatively low. On the contrary, the speed of belt transmission can be higher, but the gap and elasticity of the belt will reduce the accuracy. In a five-bar linkage design, the degree of freedom of the link is two. The trajectory of the

Ming-Chang Lin is with the Department of Mechanical Engineering, Chung Yuan Christian University, Taoyuan, 32023 Taiwan ROC (phone: +886-3-265-4345; fax: +886-3-265-4399; e-mail: mclin@cycu.edu.tw).

Yu-Ming Hsu is with the Department of Mechanical Engineering, Chung Yuan Christian University, Taoyuan, 32023 Taiwan ROC (e-mail: :tommyhsu03@gmail.com).

coupling point can be obtained by kinematic analysis. However, the orientation is hard to be guaranteed and the transmission requires a larger space. To meet the considerations of fast transmission speed, high positioning accuracy, and small size, this work intends to propose a brand-new two-dimensional mechanism based on the design of eccentric cam [12] to meet the design requirements. The proposed mechanical structure is simple and compact, and the moving trajectory is a simple harmonic curve, which will play an important role in the tracking and control of planar motion. The contributions of the work can be summarized as follows:

- Use the deep learning method to construct a two-dimensional object displacement sensor.
- The design of a new two-dimensional mechanism is proposed. The advantage of the mechanism is its compact structure, suitable for applications with limited space and fast response.

II. Two-dimension Position Sensor

This work is to use the CNN model to distinguish objects and backgrounds through the method of region random cutting, and corrects the position of the objects through linear regression. This method is called R-CNN and its architecture is shown in Figure 1. However, due to a large amount of calculation and time-consuming, it was replaced by a method called YOLO. The concept of YOLO is to cut the input picture into 7*7 squares. Each square will predict 2 bounding boxes and several categories. Each bounding box contains five values representing the X coordinate, Y coordinate, width, and height, and the confidence value that the bounding box contains the target. The confidence value is the product of the P(Object) and IoU as shown in equation (1), where P(Object) means probability of whether the bounding box has a target and IoU represents the fraction of the intersection of the predicted box and the real box relative to the union set.

$$\text{Confidence} = P(\text{Object}) \times \text{IoU} \quad (1)$$

Since YOLO only predicts two bounding boxes for each square, this limits the number of overlapping or adjacent objects that can be predicted, and its square division is also small, so it cannot detect smaller objects. To improve this shortcoming, YOLO-v3 uses Feature Pyramid Network (FPN) as shown in Figure 2. It is a multi-scale architecture. It outputs 13*13, 26*26, 52*52 feature maps for prediction where Darknet-53 is used as the main backbone network and ResNet is used to solve the gradient problem. Since there are more divided blocks and more bounding boxes predicted, the detection ability of small objects and neighboring objects is improved. In this work, the two-dimensional displacement sensor is based on YOLO-v3 and is used with the face database required for the experiment to develop the object's two-dimensional displacement detection.

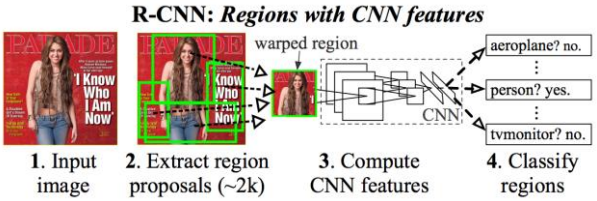

Figure 1. Schematic diagram of R-CNN architecture [5]

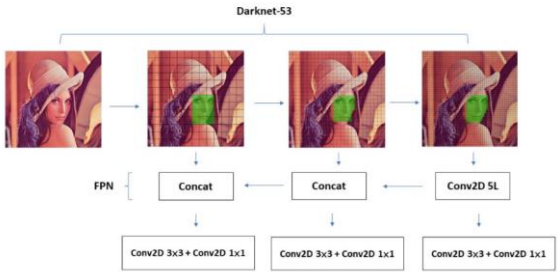

Figure 2. Schematic diagram of YOLO-v3 architecture

The trained convolutional neural network model is like a two-dimensional displacement sensor. When the face image is input, the model will output the predicted X and Y coordinate, as shown in Figure 3. The coordinate information can be fed back to the control system for tracking control.

Figure 3. Two-dimensional displacement sensor

III. Two-dimension Mechanism Design

The idea of the transmission mechanism in this research comes from an eccentric cam, as shown in Figure 4. The contour of the cam is circular, the center *O* of the rotating shaft is not on the center of the circle, the radius is *R*, and the eccentricity is *e*. When the shaft rotates, the cam moves the follower up and down, and the span of the movement of the follower is twice the amount of eccentricity.

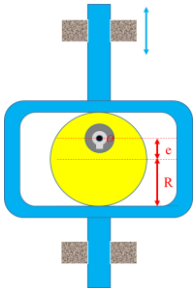

Figure 4. Eccentric cam

The transmission mechanism used in this work is to extend the above-mentioned eccentric cam into a two-dimensional motion, as shown in Figure 5 where the shafts of the cam are rotated by two motors. Using two rectangular frames, respectively driven by two eccentric cams, the output of the two-dimensional eccentric cam is simple harmonic motion, so the motion equations in the X and Y directions are as follows:

$$X = e_x \cos(\theta_x) \quad (2)$$

$$Y = e_y \cos(\theta_y) \quad (3)$$

Where e_x ,e_y , θ_x , θ_y respectively represent the eccentricity of the two eccentric cams in the X and Y directions and the angle of the rotation axis in the X and Y directions; the advantage of this mechanism is that it has a compact structure and is suitable for applications with limited space and fast response.

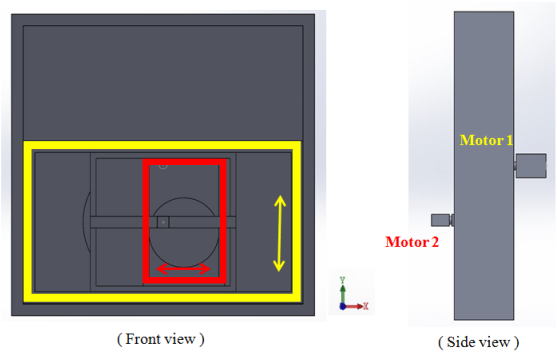

Figure 5. Two-dimensional eccentric cam

IV. Experimental Study

A. *Control loop design*

The DC motor drive system is shown in Figure 6. Assuming that the connecting shaft is rigid, according to the free body diagram of the motor, its dynamic equation can be expressed as follows:

$$\frac{d^2\theta_m(t)}{dt^2} = -\frac{B_m}{J_m + J_L}\frac{d\theta_m(t)}{dt} + \frac{1}{J_m + J_L}T_m(t) \quad (4)$$

Through Laplace transform, the relationship between the output angular position θ_m and the input torque T_m is as follows:

$$\frac{\theta_m}{T_m} = \frac{1/(J_m+J_L)}{s^2 + B_m/(J_m+J_L)s} \quad (5)$$

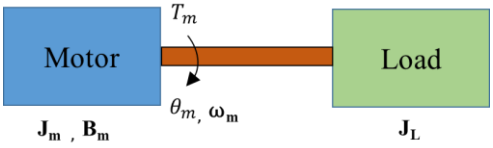

Figure 6. Schematic diagram of DC motor drive

Where

T_m = Motor torque (N-m)

θ_m = Angular displacement of motor shaft (rad)

ω_m = Angular velocity of motor shaft (rad/sec)

J_m = Equivalent inertia of the motor (kg-m^2)

B_m = Equivalent viscous friction coefficient or damping coefficient (N-m sec/rad)

J_L = Equivalent inertia of load (kg-m^2)

From equation (5), it can be seen that the transfer function of the motor drive system is second-order. Then, the camara for position sensing is mounted on the moving two-dimensional cam frame, and the overall tracking control architecture diagram is shown in Figure 7.

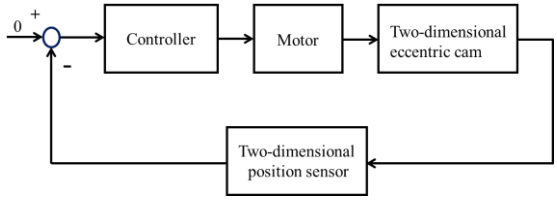

Figure 7. Block diagram of tracking control

After the training data is converged by the regression of YOLO-v3, a trained “model” will be obtained, as shown in Figure 8(a). The model is a two-dimensional displacement sensor, the model will be built on an embedded platform, when new image data is received, the model will output the predicted coordinate value of the displacement. The displacement signal is fed back to the controller for tracking control. The control effort of the microprocessor directly outputs to the motor to drive the two-dimensional mechanism, and the relationship is shown in Figure 8(b).

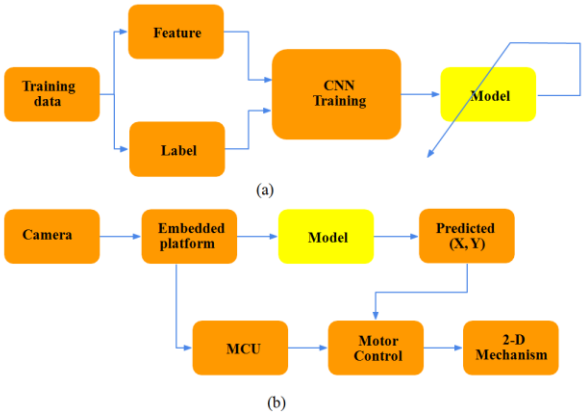

Figure 8 (a). Schematic diagram of CNN training model

(b). Schematic diagram of tracking control system

B. *Preliminary experimental results*

To verify the measurement error of the two-dimensional displacement sensor, the experimental method is to use a motor to drive the face image to rotate at a constant speed of 40rpm and 18rpm with a radius of 0.1 meters. The external fixed camera is used as the sensor, and the image of the sensor is directly output to the host computer for calculation with the Nvidia 2060s graphics card. The image capture rate is 15 frames per second. The experimental results at 40 rpm and 20 rpm are shown in Figure 9 and Figure 10, respectively. The orange line is the actual position, the blue line is the detection position, and the green dash line is the displacement error. Due to the constant speed rotation, the output of the displacement in the X direction is a sine wave. It is obvious from the figure that there is a phase delay between the detected position and the actual position. This is caused by insufficient detection speed. According to the current preliminary experiments, the maximum position error is 0.075 meters at 40 rpm and 0.03 meters at 20 rpm. This experimental data shows that at the same image capture speed, the error of the detected position is inversely proportional to the object's moving speed. This shows that the most effective way to reduce the displacement error is to increase the image sampling frequency.

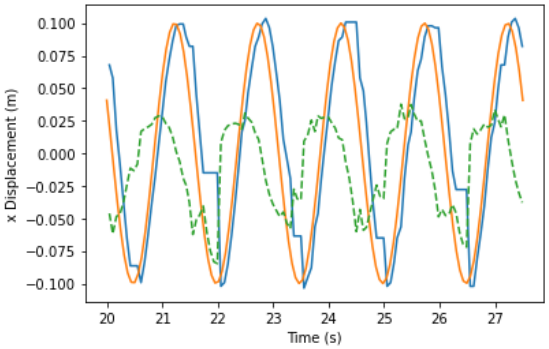

Figure 9. The relationship between X-direction displacement and time when the speed is 40 rpm

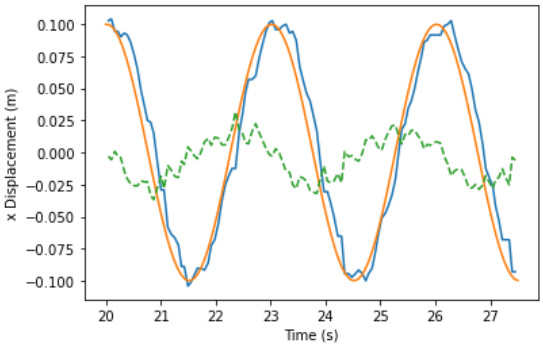

Figure10. The relationship between X-direction displacement and time when the speed is 20 rpm

Regarding the tracking control of the face contour, the experiment is currently in progress.

V. CONCLUSION

This work proposes a solution for the system design and development of planar motion tracking control. Traditional methods need to rely on high-performance arithmetic units to process planar image signals. Instead, this work attempts to build a two-dimensional sensor for fast displacement detection through deep learning, which can be used for tracking control. In addition, due to the large space requirements of traditional planar mechanism design, this work also proposes a brand-new two-dimensional mechanism, focusing on lightness and speed, which is conducive to building a fast tracking control system. Preliminary experimental results show that the speed of image capture has a great influence on the detection of object motion. Increasing the image sampling frequency and optimizing the image detection algorithm to achieve better measurement errors are under investigation. And the next step is to combine the displacement signal with the two-dimensional mechanism for real-time tracking control.

REFERENCES

[1] J. Song, L. Gao, F. Nie, H. T. Shen, Y. Yan, N. Sebe, “Optimized Graph Learning Using Partial Tags and Multiple Features for Image and Video Annotation,” *IEEE Transactions on Image Processing*, vol. 25, no. 11 , 2016.

[2] D. Tao, Y. Guo, Y. Li, X. Gao, “Tensor Rank Preserving Discriminant Analysis for Facial Recognition,” *IEEE Transactions on Image Processing*, vol. 27, no. 1, 2018.

[3] M.C. Lin, P.Y. Han, Y.H. Fan and C.H. G. Li, “Development of Compound Fault Diagnosis System for Gearbox Based on Convolutional Neural Network,” *Sensors*, vol. 20, 6169, 2020.

[4] H.S. Zhao, “Pyramid Scene Parsing Network (CVPR 2017),” [online]. Available: https://www.youtube.com/watch?v=rB1BmBOkKTw.

[5] R. Girshick, J. Donahue, T. Darrell, J. Malik, “Rich feature hierarchies for accurate object detection and semantic segmentation,” *The IEEE Conference on Computer Vision and Pattern Recognition* (CVPR), 2014.

[6] R. Girshick, “Fast R-CNN,” *The IEEE Conference on Computer Vision*, 2015.

[7] S. Ren, K. He, R. Girshick, and J.Sun, “Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks,” *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, 2017.

[8] J. Redmon, Santosh Divvala, Ross Girshick and Ali Farhadi, “You Only Look Once: Unified, Real-Time Object Detection,” *The IEEE Conference on Computer Vision and Pattern Recognition* (CVPR), pp.779-788, 2016.

[9] J. Redmon and A. Farhadi, ”Yolov3: An incremental improvement, “ *arXiv preprint arXiv:1804.02767*, 2018.

[10] Benedict CE, Tesar D, “Model Formulation of Complex Mechanisms with Multiple Inputs: Part I- Geometry,” *Journal of Mechanical Design*, vol. 100, pp.747-754, 1978.

[11] K Russell, RS Sodhi, “Kinematic Synthesis of Planar Five-Bar Mechanisms for Multi-Phase Motion Generation,” *JSME International Journal Series C*, vol. 47, pp.345-349, 2004.

[12] G.H. Martin, ”Kinematics and Dynamics of Machines”, McGraw-Hill, Inc.