Mechanical movement data acquisition method based on the multilayer neural networks and machine vision in a digital twin environment [version 1; peer review: awaiting peer review]

Hao Li¹, Gen Liu¹, Haoqi Wang¹, Xiaoyu Wen¹, Guizhong Xie¹, Guofu Luo¹, Shuai Zhang², Miying Yang³

¹Henan Key Laboratory of Intelligent Manufacturing of Mechanical Equipment, Zhengzhou University of Light Industry, Zhengzhou, 450002, China
²University of Greenwich, London, SE10 9LS, UK
³School of Management, Cranfield University, Cranfield, MK43 0AL, UK

Abstract

Background: Digital twin requires virtual reality mapping and optimization iteration between physical devices and virtual models. The mechanical movement data collection of physical equipment is essential for the implementation of accurate virtual and physical synchronization in a digital twin environment. However, the traditional approach relying on PLC (programmable logic control) fails to collect various mechanical motion state data. Additionally, few investigations have used machine visions for the virtual and physical synchronization of equipment. Thus, this paper presents a mechanical movement data acquisition method based on multilayer neural networks and machine vision.

Methods: Firstly, various visual marks with different colors and shapes are designed for marking physical devices. Secondly, a recognition method based on the Hough transform and histogram feature is proposed to realize the recognition of shape and color features respectively. Then, the multilayer neural network model is introduced in the visual mark location. The neural network is trained by the dropout algorithm to realize the tracking and location of the visual mark. To test the proposed method, 1000 samples were selected.

Results: The experiment results shows that when the size of the visual mark is larger than 6mm, the recognition success rate of the recognition algorithm can reach more than 95%. In the actual operation environment with multiple cameras, the identification points can be located more accurately. Moreover, the camera
calibration process of binocular and multi-eye vision can be simplified by the multilayer neural networks.

Conclusions: This study proposes an effective method in the collection of mechanical motion data of physical equipment in a digital twin environment. Further studies are needed to perceive posture and shape data of physical entities under the multi-camera redundant shooting.

Keywords
digital twin; mechanical movement data; multilayer neural network; machine vision; data acquisition
Introduction

To meet and adapt to the manufacturing development requirements of informatization, personalization, intelligence, and service, various corresponding advanced manufacturing development methods have been proposed globally. One of the promising methods is digital twin (DT), which aims at realizing the interconnection and interoperability of the physical world and the information world, enabling intelligent manufacturing. However, one of the difficulties in achieving this goal is to maintain a high degree of virtual and physical synchronization of the shape, structure, motion, mechanics, electricity, temperature, liquid, airflow, and other states of physical devices in the physical world and the information world. As the physical and virtual synchronization is a simple simulation process, it involves the exchange of lots of data, especially in the process from a physical entity to a digital model, which requires intensive and all-around data collection.

The real-time and intensive data collection of physical equipment running state is an important step to realize intelligent manufacturing in a DT environment. To achieve the real-time collection and low-latency exchange of the mechanical motion data, as required by the above physical and virtual synchronization, programmable logic controller (PLC) is usually used to communicate data between software and hardware. For example, the Siemens PLC system cannot realize synchronous linkage between the digital model and physical entity, which mainly relies on PLC to accomplish the connection between software and hardware. However, the traditional way of mechanical movement data acquisition based on PLC cannot realize the total factor dynamic data collection perfectly. For example, various sensors are usually employed for the acquisition of motion state information. Angle sensors can be used to collect the angle data of relative moving objects. Photoelectric sensors can detect the positions of objects through the occlusion of light beams by objects. Changes in the rotation and displacement of objects can be detected by the inertial sensor. In the complex working condition of general manufacturing workshops like the automobile body-in-white welding production line, the shape, position, attitude, and motion state of the parts in the equipment are all visible. Based on the machine vision, the mechanical movement data of the parts, such as the coordinates of the movement space, movement tracking, and linear speed, can be obtained. This machine vision has become an indispensable and important means to acquire the mechanical movement data of the equipment. However, there are few types of research on the acquisition of mechanical movement data by machine vision. Combined with multilayer neural networks, machine vision technology has been greatly improved, which can be good support for data acquisition using the visual method in a DT environment.

At present, there are still three challenges in mechanical movement data acquisition based on the neural network and machine vision in a DT environment:

1. PLC data acquisition mode can’t complete the data acquisition of the shape, position, attitude, and motion state of parts quickly, which is required by the virtual and physical synchronization.

2. It is necessary to measure and calibrate the parameters of each camera when using the multi-vision to locate the target. At the same time, the multi-vision positioning algorithm becomes complex, with the number of cameras increasing. Moreover, the conventional multilayer neural networks have high requirements on the system hardware, while the training process needs manual supervision, with a heavy workload.

3. Without multilayer neural networks, the convex lens image of the camera will have some image distortion, which will affect the accuracy of the target location.

To solve the above problems, this paper presents an improved mechanical movement data acquisition method based on a multilayer neural network and machine vision, which can be used to collect the position, attitude, and motion information of components on the equipment required by the virtual and physical synchronization. The main features of our method are as follows:

1. A set of visual marks is designed to reduce the amount of image data that the neural network algorithm needs to process. The Hough transform method is used to recognize the shape of the shape of the target object and the histogram statistics is applied to recognize the objects’ color, and they are combined to recognize the object’s location in which the sign points are attached.

2. Using a multilayer neural network model and a dropout algorithm for problems of multi-eye location, we only need to train the algorithm of the multi-eye system and realize the multi-eye location automatically.

3. Through the training process of a multi neural networks algorithm, the algorithm can automatically acquire the ability to correct camera image distortion, which can improve the accuracy of the target location.

The outlines of this paper are as follows. In the background section, the research status of machine vision target recognition, location, and neural network algorithm is reviewed. The methods section describes the steps and methods of mechanical movement data acquisition based on multilayer

**Abbreviations**

DT: Digital twin; PLC: Product lifecycle management; RFID: Radio frequency identification; RGB: Red, green and blue; DCNN: Deep convolutional neural networks; 2D: 2-dimensional; 3D: 3-dimensional; ABB: Asea Brown Boveri Ltd; OpenCV: Open-source computer vision library
neural network and machine vision. Results and discussion section introduces the content of system development, and carries out case test. Lastly, the paper sums up finding in conclusions.

Background
Data acquisition method under the virtual and physical synchronization requirement
Data collection provides the underlying data guarantee for the interconnection and interoperation of the physical world and information world of manufacturing. The data collection layer is an important part of the technical system of the digital twin, which is one of the key technologies to achieve virtual and real synchronization of manufacturing equipment\textsuperscript{10}. Reference 11 discusses the data interaction and integration theory and implementation method of physical world and information world of the general manufacturing workshop. Reference 12 emphasizes the importance of data guarantee for the virtual and physical synchronization and introduces the new requirements for seamless data transmission between the physical world and the virtual world. Reference 13 introduces the key data of rocket attitude, deformation, load, take-off drift, etc. collected by level sensors, pull wire sensors, infrared laser sensors, etc. under the requirement of virtual and real synchronization. Reference 14 discusses the three characteristics of data in the virtual and physical synchronization, i.e. volume, variety, and velocity, and introduces the technology of real-time perception of manufacturing resources in the physical workshop using radio frequency identification (RFID) in the aircraft assembly workshop.

Target recognition and location based on machine vision
In recent years, the technology of object recognition and location based on machine vision has made rapid progress and will be widely used in the virtual and physical synchronization. In the aspect of target recognition, reference 15 uses different features acquired by the vision and laser-based sensors, and combines them to obtain the precise location of the robot relative to the cylindrical structure, to realize the location and detection of cylindrical objects. In reference 16, the generalized Hough transform algorithm is introduced to recognize the geometric features in the image, which does not need to give the analytical expression of the target figure to recognize any given shape. In reference 17, an image color segmentation method based on histogram statistics is proposed, which can recognize specific color regions from different backgrounds. In reference 18, a target tracking algorithm of the vision system is proposed, which introduces the background weighted linear RGB histogram target feature, reducing the weight of background features. In the aspect of the target location, the method of binocular or multi-eye vision is usually employed. In reference 19, a method of binocular target recognition and 3D reconstruction is proposed. The chessboard method is employed to calibrate and calibrate the dual cameras, and the Bayesian classifier is used to identify the target, and the scale-invariant feature transformation algorithm is applied for matching the feature points. According to the triangulation principle, the target distance can be calculated, and the 3D reconstruction of the target is achieved finally by parameters of the calculated target distance. In reference 20, a target recognition system based on trinocular vision is designed. The system obtains the edge features of the gray-scale image of the target through edge detection and recognizes the target object through feature matching.

Vision algorithm based on the neural network
The application of the neural network makes the machine vision algorithm able to learn, which can be found in recognition and classification. The neural network algorithm can automatically learn the ability to recognize images through a large number of image data after training.

In reference 21, the multilayer neural networks algorithm is introduced into the recognition of visual marks, which can accurately identify traffic lights, traffic warning signs, and other visual signs. Reference 22 extracts features such as color and texture to achieve the classification of the wheat-based on multilayer perceptron neural networks and image processing technology. In reference 23, the multilayer neural networks method is introduced, which combines the initial network with the recursive layer of deep convolutional neural networks (DCNN) architecture to improve the performance of target recognition. In the aspect of the target location, reference 24 employs the neural network algorithm to complete the calibration process of binocular vision camera. In the calibration process, the plane with uniformly distributed solid circles should be placed at different positions of the camera’s visual object. Then the data of the center of the circles should be used as the input of the neural network. The successful neural network algorithm can obtain the 3D coordinates of the objects without any complicated camera calibration operation. In reference 25, the traditional machine vision algorithm and the neural network algorithm are combined to realize the recognition and location of specific texture patterns at a lower cost. In reference 26, a new technology of signal fingerprint is used to locate the indoor target. The neural network algorithm can acquire the feature of the spatial location of the signal by training.

Methods
The architecture of mechanical movement data acquisition method
A method of collecting the state information of production equipment through machine vision is proposed in this paper, as shown in Figure 1. This method can provide a real-time data feed service for the digital twin model, helping the motion state of the digital twin model keep pace with the physical entity in real-time. The model is available from GitHub and is archived with Zenodo\textsuperscript{27}.

As shown in Figure 1, the method mainly includes three steps: (i) the design of visual marks, (ii) the recognition of visual marks, and (iii) the location of visual marks. A key step of the design of visual marks is to design the prominent shape and color characteristics. The proposed method uses circular visual markers coded with color blocks. On this basis, multiple cameras are arranged to take pictures of the objects from different positions and angles. When the object carries visual marks to move, the
cameras capture many images of the object. The location and size of the visual mark from the image can be found by the recognition algorithm, which can be as input data for the target location algorithm based on the neural network. In the process of the location of visual marks, the trained neural network algorithm can calculate the three-dimensional coordinate position of the visual marker on the target object based on the input data. Details of the design of visual marks, the recognition of visual marks and the location of visual marks are provided respectively in the following.

**Design of visual marks for equipment**

There are many kinds of industrial equipment, including a wide variety of visible moving parts. When using the machine vision directly to identify and locate these devices and components, we need to separate the object and background to extract the shape features, color features of the objects to identify such objects. This requires establishing a feature library of all pieces of equipment and components to be tested, in which complex visual recognition algorithms are necessary. Because of the different shooting angles and light conditions, the accuracy of recognition may be affected. If the neural network algorithm is used to identify the target object, the number of nodes in the input layer of the neural network is equal to the number of pixels in each frame of the camera image. Moreover, the number of nodes in other layers of the neural network is also close to this number. Substantial computing power is indispensable for such a huge neural network, which is uneconomical in practical application. In addition, the training process of the neural network is difficult to be carried out automatically. Thus, it is necessary to manually mark the location of the target object in each frame image. The amount of image data in the process of the neural network algorithm is too large, thus a powerful computer is required to process these data. In order to make the algorithm of visual recognition simple and accurate, a unified kind of visual mark is designed in this paper. Visual marks are attached to the moving objects. When the cameras capture the positions of the visual marks, the position and motion state information of the object can be obtained indirectly. After introducing the visual mark, the number of pixels in an image is reduced from millions to single digit. Because the visual marks are specially designed, they have obvious characteristics, and the recognition of them is simple and accurate. In the process of multi-vision location, the amount of data processed by the neural network algorithm is largely reduced because of the introduction of visual marks, which makes the target location method based on the neural network feasible.

To be used in industrial scenes successfully, certain conditions must be fulfilled by visual marks. A good feature should be able to be distinguished from the background environment. Firstly,
the visual marks should have significant characteristics. The classic features of the image contain color, texture, and shape. The relatively complex features contain sift, hog, surf, gist, and features based on the neural network. We choose the shape and color features to design the visual marks, which makes the visual marks have bright colors and regular shapes to simplify the algorithm’s complexity and increase the accuracy of recognition. Secondly, the visual marks should be able to adapt to the changes in light conditions in a certain range. Lighting is an important factor affecting the input of the machine vision system because it directly affects the quality of the input data and at least 30% of the application effect. Under the condition of an ordinary fluorescent lamp and natural light, the visual mark should be able to reflect a stable color. Finally, the specular reflection should not be possessed by the visual marks. This is because the specular reflection is white in the cameras, and the real color of the visual marks is covered. Thus, a diffuse material or materials with fluorescent properties should be chosen.

According to the recognition algorithm of the circle feature in the image in machine vision, the concentric circle visual mark is designed in this paper. The visual mark is a white circle with a black circle inside or a black circle with a white circle inside. The advantage of this design is that it can provide a solid color substrate for the center circle to obtain a large color gap. Thus, it can be distinguished from the environment easily, and more accurate recognition can be obtained by this algorithm. Based on the recognition characteristics of the Hough transform, the visual mark is designed as a circle, and the edge contour and the center position of the circle can be accurately found. The two-color mark points are represented “0” and “1”, respectively, which can be easily implemented into the graphic code. For example, the outer white circle with a black circle inside is denoted as “01”, while the outer black circle with a white circle inside denoted as “10”. The visual marks can be used alone or in combination. The visual marks used in groups can mark different parts of a moving object. As shown in Figure 2, the mechanical arm is marked with the visual marks.

Visual mark recognition based on Hough transform and histogram features

For the visual mark illustrated in Figure 1, the corresponding recognition algorithm is designed. This algorithm extracts the shape and color features of the visual mark, respectively, and determines the position of the visual mark in the image through matching the two features. The Hough transform method is used to recognize the shape feature of the visual mark to find the center position and radius of the visual mark. The histogram statistics method is used to recognize the color feature of the visual mark. The center and radius of the visual mark are the centers of the circle and radius obtained in the shape feature extraction, respectively. Local histogram statistics are carried out in this field. The visual marks have special colors. In the histogram statistics, the number of pixels in this color range is the largest. The region in accord with this property can be identified as a visual mark.

Firstly, the image captured by the camera is transformed into gray image through gray processing, and RGB color components are integrated into a value between 0 and 255. Gray scale conversion uses the mean value formula:

\[
\text{Gray} = \frac{\text{Red} + \text{Green} + \text{Blue}}{3}
\]  
(1)

In Equation (1), “Gray” is the gray value, “Red” is the pixel value of red component, “Green” is the pixel value of green component, and “Blue” is the pixel value of blue component. After the gray-scale image processing, the edge detection algorithm is used to find out all the edge feature areas of the gray-scale image, and remove the parts outside the edge. The function is \(f(x,y)\) used to represent the image taken by the camera, where \(x, y\) are the pixel coordinates of the image and subscript 1 is the corresponding camera number. There is a large gap between the image edge and the surrounding pixels.

**Figure 2.** The mechanical arm with the visual marks.
Calculating the derivatives of the image data in \( x \) and \( y \) directions, the image edge is located where the local maximum value of the first derivative, and the zero value of the second derivative.

According to this principle, Laplace operator is used to obtain the image derivatives. The expression is as follows:

\[
f'(x, y) = -4f(x, y) + f(x - 1, y) + f(x + 1, y) + f(x, y - 1) + f(x, y + 1)
\]

(2)

Extracting coefficient matrix from the derivation formula (2) is
\[
\begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix}
\]
The matrix is used as an operator to cover the gray-scale image. The image pixels are multiplied by the corresponding position values of the operator, and then sum all the values to obtain the values of the pixels corresponding to the central position of the operator. After edge detection, the original image \( f(x, y) \) is transformed into edge image \( F(x, y) \).

Then based on the edge image \( F(x,y) \), Hough transform methods for circle finding is employed for confirming the region with circular features in the edge image. The general equation of a circle is \((x-a)^2 + (y-b)^2 = r^2\), where \(a\) and \(b\) are the coordinates of the center, and \(r\) is the radius. Each point of circle boundary in \( x-y \) coordinate system is mapped to \(a-b\) coordinate system. A circle in the \( x-y \) coordinate system corresponds to multiple circles in the \(a-b\) coordinate system, and these circles will intersect at a point in the \(a-b\) coordinate system, which is the center of the circle in the \( x-y \) coordinate system. In \(a-b\) coordinate system, the equation of a circle is \((a-x)^2 + (b-y)^2 = r^2\), \(x\) and \(y\) are the coordinates of the center, and the radius \(r\) is set as a given value. Adjusting the radius \(r\) and repeating this algorithm, the position of the center of the circle can be found under different radius conditions. The center and radius recognized by the edge detection and Hough transform algorithm can’t be determined as the position of the visual mark in the image, because these centers and radii may be generated by all kinds of objects which are approximately circular. Thus, we must further use the color features of the visual mark to find the one which really belongs to the visual mark from all the centers.

Histogram statistics are used to extract color features. The statistical region is the circular region located by the Hough transform algorithm. The histogram statistics abscissa is the pixel value of the image, whose range varies from 0 to 255. The ordinate is the normalized value of the number of pixels with different pixel values in the image. As shown in Figure 3, it can be found that from the histogram the value of the visual mark has a significant increase in the range from 21 to 65 and from 205 to 244. Note that the range from 21 to 65 is the center dot of the visual mark and the range from 205 to 244 the background color of the visual mark. Through these two obvious features, we can distinguish the area of the visual mark in the image from other areas.

**Location of visual marks based on the multilayer neural networks**

The image acquired by each camera is planar. In order to obtain the position and motion state of the target object through the visual mark, it is necessary to obtain the position of the visual mark in three-dimensional space. The 3D coordinates of the same visual mark can be obtained from the 2D coordinates of the same visual mark in multiple 2D images, which can be realized in theory. The target location is realized by the binocular vision and binocular vision based on the geometric principle of the triangle localization algorithm. In this paper, the visual mark location method proposed is different from the conventional binocular and multi-eye visual location methods, which is based on multilayer neural networks. The core algorithm for location is to calculate three-dimensional coordinates from multiple two-dimensional coordinates, which is a mapping relationship. According to the training data, the neural network fits the mapping relationship by adjusting the weights of the nodes in each layer, which can be solved by the neural network algorithm. Thus, the image distortion caused by convex

![Figure 3. Histogram features of visual marks.](image)
lens imaging can be dealt with. And the calibration process of multiple cameras and multi-camera positioning algorithm can also be simplified.

The structure of the neural network. The problem of the location described in this paper is a regression prediction problem, in fact, which can be solved by the multilayer neural networks method. In this method, a neural network with multiple hidden layers is pre-set, including the input layer, multiple hidden layers, and an output layer. Each layer consists of many nodes. Each node constitutes a perceptron, which has several inputs and one output. The output of the perceptron is processed by the activation function to keep the output data within a certain range, and the discrete values become continuous. The number of the input layer and output layer nodes of the multilayer neural networks is fixed. The input layer node represents the coordinates of the center and radius of the circle obtained from the feature extraction of the visual mark taken by each camera. The center coordinate is \((x, y)\), the radius is \(r\), and the number of parameters is 3. If the number of deployed cameras is \(n\), the number of input layer nodes is \(3 \times n\). The output node represents the 3D coordinates \((x, y, z)\) of the visual mark. The number of output nodes is 3. It is a complex mapping function from the two-dimensional image of the camera to the three-dimensional coordinate of the visual mark. In order to make the neural network fit this function better, we need to increase the number of hidden layers and nodes. According to the performance experience of the multilayer neural networks, the number of hidden layers is set to 2. In order to train the neural network, this paper uses the dropout algorithm, which discards some hidden layer nodes during the training process; thus, the number of hidden layer nodes is set to be more than that of the input layers. The initial connection between different layers of the neural network is a full connection, and there is no connection between the same layer.

Generation of the training samples. The multilayer neural network is a supervised neural network. Training supervised neural network needs to prepare enough labeled sample data. The sample data consists of two parts: the input variable and the output variable. The mark of the training sample is the output variable. The input variable is the 2D coordinate, and the radius of the visual mark and the output variable is the 3D coordinate of the visual mark.

The format of a training sample is \([(x_1, y_1, r_1, x_2, y_2, r_2, \ldots, x_n, y_n, r_n), (X, Y, Z)]\). From the data format, it can be found that the data format that the training data of neural networks has been reduced through the simplification of visual marks, which only several values of the data are needed, compared with millions of pixels of the data without simplification. The number of training samples depends on the sampling frequency of the camera and the time required for a moving cycle of the target object. Generally, the collection of training samples needs a manual operation. The mapping relationship between the input data and output data of training samples must be determined by human intelligence. Lots of neural network training samples are manually marked. In fact, a great number of samples are required by a successful neural network training. If the training samples are all manually marked, it is a lot of work, which is a limit of the practical application of the neural network. However, the problem of the target location in this paper can generate training samples automatically. Most of the motion of the machinery and equipment or its components in the industry is circular motion, moving back and forth in a fixed orbit, and the speed can be measured. When the device is started, the camera continuously shoots the object in a motion cycle of the target object. Each frame of the image will correspond to a spatial position of the target object or a position in its orbit. The spatial coordinates of this position can be calculated by the movement speed and the period of the camera shooting each image frame. This avoids the need for space measurement at every location where an object appears. The training samples for the neural network can be obtained only by running the equipment for one cycle.

Training the neural network. In this example, we assume 80% of all samples is a training set and the remaining 20% is a test set. The training set is used to train the neural network, while the test set is employed to test and evaluate the neural network after training.

The training set is expressed as:

\[
\left( \begin{array}{c}
(x_{1 \text{train}}, y_{1 \text{train}}, z_{1 \text{train}}), \\
(x_{2 \text{train}}, y_{2 \text{train}}, z_{2 \text{train}}), \\
\vdots \\
(x_{n \text{train}}, y_{n \text{train}}, z_{n \text{train}})
\end{array} \right)
\]

The test set is expressed as:

\[
\left( \begin{array}{c}
(x_{1 \text{test}}, y_{1 \text{test}}, z_{1 \text{test}}), \\
(x_{2 \text{test}}, y_{2 \text{test}}, z_{2 \text{test}}), \\
\vdots \\
(x_{n \text{test}}, y_{n \text{test}}, z_{n \text{test}})
\end{array} \right)
\]

The activation function of the neural network is set as:

\[
h(x) = \max(0, x) = \begin{cases} 
    x(x \geq 0) \\
    0(x < 0)
\end{cases}
\]

Each node adds the offset value to the weighted sum of the input data:

\[
a_j = \sum w_j x_j + b_j
\]

In the equation (4), subscript \(j\) represents the \(j\)th node of the neural network in this layer, subscript \(i\) represents the \(i\)th node of the neural network in the previous layer, \(w_j\) is the weight of the connection between the \(j\) node and the \(i\) node of the previous layer, \(x_i\) is the \(i\)th input of the node, and \(b_j\) is the offset value of the node. The final output of the node needs to be processed by the activation function. Take the above formula into the activation function:

\[
Z_j = b(\sum w_j x_j + b_j) m_j, m_j \sim \text{Bernoulli}(1 - P)
\]

In equation (5), where \(m_i\) is the mask parameter. This parameter is in accord with Bernoulli probability distribution, \(m_i\) varies according to the value of probability. When \(m_i=0\), the output of the node is 0, and the node will be deleted. The purpose of adding mask parameters is to discard some nodes in the
multiple hidden layers to avoid the overfitting in the neural network. The final output of the neural network is the three-dimensional coordinate $(X, Y, Z)$ of the visual mark, where $X = G_i(W, B, M)$, $Y = G_j(W, B, M)$, and $Z = G_k(W, B, M)$, i.e., coordinates $X$, $Y$, $Z$ are functions of parameters $W$, $B$, and $M$. $W$ is the vector of weight $w$, $B$ is the vector of offset value $b$, and $M$ is the vector of mask value $m$. The training process of the neural network is a gradient decreasing process. The training errors are as follows:

$$
E_x = \frac{1}{n} \sum (X - X_{\text{train}})^2
$$
$$
E_y = \frac{1}{n} \sum (Y - Y_{\text{train}})^2
$$
$$
E_z = \frac{1}{n} \sum (Z - Z_{\text{train}})^2
$$

From Equation (6), the gradients of errors are the partial derivatives of parameters $W$ and $B$. The mask parameter $M$ is a preset parameter. The formulas of the gradients are as follows:

$${\nabla}_w E = \frac{1}{n} \| G_i(W, B, M) - X_{\text{train}} \|^2_2 = 0$$
$${\nabla}_b E = \frac{1}{n} \| G_i(W, B, M) - X_{\text{train}} \|^2_2 = 0$$
$${\nabla}_M E = \frac{1}{n} \| G_i(W, B, M) - X_{\text{train}} \|^2_2 = 0$$

After bringing the training samples into the iterative operation, the weight vector $W$ and offset value vector $B$ are determined finally, and the neural network training is completed. The validity of the neural network can be verified by test samples.

An experimental study was carried out to evaluate the proposed data acquisition method based on multilayer neural networks and machine vision. In this experiment, an Asea Brown Boveri Ltd. (ABB) manipulator (IRB 1600-10/1.45) was taken as the object of action acquisition. The visual mark is pasted on the joint of the mechanical arm. Two cameras (Kingcent KS4A418-D) were distributed around ABB’s robotic arm. The positions of the two cameras were adjusted to ensure that the visual mark on the manipulator could be captured by the two cameras when the arm is moving. This experiment can be divided into the recognition experiment and the location experiment.

Experiment of the visual mark recognition

According to the recognition algorithm of the circle feature in the image in machine vision, we developed programs for the visual mark recognition based on OpenCV 3.1.0 in the first place. When the manipulator carries the visual mark to move in the space, two cameras shoot at a rate of 30 frames per second, and the the algorithm program is run to check whether it can accurately find out the visual mark in the image.

As shown in Figure 4, the images are taken by two cameras from different positions and angles. The edge of the image is highlighted by edge detection, and finds the circular feature in the edge by Hough transform. Note that some circular features in the environment may be detected by mistake. It is necessary to further screen out the circular area where the visual mark is located by histogram statistics in the circular area. The coordinates of the center and radius of the visual mark in the image are final outputs, as circled in the image.

Experiment of the visual mark location

ABB Robot (IRB 1600-10/1.45) is a programmable device that can control the robot arm to carry the visual mark to traverse every position in the view space of the camera, from top to bottom, from left to right, from front to back a constant speed. In our experiment, the two cameras collect images at a fixed frame rate, keeping pace with the robot arm. Using the proposed vision recognition algorithm in the visual mark recognition section to extract the two-dimensional coordinate data of the visual mark from the image. The corresponding 3D coordinates of the visual marks can be calculated based on the shooting time and the manipulator’s movement speed. The two-dimensional and three-dimensional coordinate data are stored in a file as a training sample according to the format in the training the neural network section. According to this procedure, the training sample set in the whole process of the visual mark movement is generated.

Deeplearning4j is the neural network framework based on Java. Build the proposed neural network structure based on Deeplearning4j. The number of the input layer nodes is 6, the number of output layer nodes is 3, the number of second hidden layer nodes is set to 1100, and the learning rate is set to 0.01. 80% of all samples are used to train the neural network, while the remaining 20% are used to test the training effect of the neural network.

Results

In total, 1000 samples were selected from the training samples to test the recognition and location of the proposed method. The experiment results in Figure 5 shows that when the size of the visual mark in the image taken by the camera is larger than 6mm, the recognition success rate of the recognition algorithm can reach more than 95%. The size of the visual mark in the image taken by the cameras is less than 4mm, which leads to a more significant error.

In Figure 5, it can be seen that the recognition algorithm can run stably. For 1000 samples, if the location error is more than 1mm, the location fails. The results of the location experiment can be found in Figure 6. It is noted that the more nodes in the double hidden layers of the neural network within the recognition range of the visual marks, the higher accuracy of location through the neural network. As the density of nodes increases to a certain extent, the accuracy keeps stable. From our experiment, it can be found the calculation accuracy of the algorithm is the best when the number of nodes in the double hidden layers of the neural network is 1100.

The overall operation effect of the system is shown in Figure 7. With double cameras and single vision mark, cameras
Figure 4. Recognition of the visual mark.

Figure 5. Influence of number of hidden layer nodes on location accuracy.
adjust according to the movement of the manipulator to make the visual mark appear in the picture taken by cameras completely. In a certain distance range, the scale of the visual mark in the image is more than 6 mm. The neural network algorithm is used to process the location data. Using the data collected by this data acquisition method, the ABB manipulator model can be driven. Thus, the model and the ABB manipulator can maintain the synchronization of motion state. The experimental results verify the effectiveness of the proposed method.

**Conclusion**

This paper presents a mechanical movement data acquisition method based on the multilayer neural network and machine vision in a DT environment, which is applied to realize the accuracy of virtual and physical synchronization. In this method, visual marks are introduced to simplify the amount of data which needs to be processed. The position and attitude data of the attached equipment can be indirectly acquired through the identification and positioning of visual marks. To simplify the calibration process of double camera and multi-camera location of visual marks, and to solve the positioning error caused by camera lens imaging image deformation, a multilayer neural network algorithm is also introduced. Through the neural network algorithm’s training process, the distortion of the image is corrected. The location accuracy can also be improved, which avoids the camera calibration process. To reduce the calculation
amount of the neural network algorithm, the visual mark recognition problem can be solved by conventional image processing. The experimental results show that the method is effective in the collection of action information of physical equipment.

At present, only the data acquisition experiments under the condition of double cameras and single visual marks are carried out. We will further study and experiment with the application of multi-camera redundant shooting in the future. Furthermore, the visual mark will be improved to adapt to a more complex lighting environment and different shooting angles. Under the virtual and physical synchronization requirement working condition, with the help of the visual marks and the neural network-based visual algorithm, the proposed method will be the most potential data collection method for movement, posture, and shape data of physical entities.

Data availability

This project contains the following underlying data:
- Influence of number of hidden layer nodes on location accuracy.xlsx
- Sampling point.xlsx
- The image recognition results for spatial sampling points are 1000samples in total.xlsx
- The influence of the radius of the visual mark on the recognition accuracy1.xlsx

Data are available under the terms of the Creative Commons Attribution 4.0 International license (CC-BY 4.0).

Software availability
- Archived source code at time of publication: https://doi.org/10.5281/zenodo.5303105

License: MIT

References


