Multi-person Location and Tracking Method Based on BP Neural Network*

Pan Wei /s per 1st Affiliation
Faculty of Cognitive Science Department
Xiamen University
Xiamen 361005, China
E-mail:wpan@xmu.edu.cn

Liu Zhizhan/s per 2nd Affiliation
School of Information Science and Technology
Xiamen University
Xiamen 361005, China
E-mail: xmulzz@yahoo.com.cn

Zou Yi /s per 3rd Affiliation
School of Information Science and Technology
Xiamen University
Xiamen 361005, China
E-mail: zouyimumu@gmail.com

***Abstract*—This paper focuses on the study of multi-person location and tracking in a complex scene created by 3ds max. To establish the complicated relationship between the 2D-image information that is obtained through the three-camera system and the 3D information of the target, an artificial neural network is proposed. In order to overcome the shortcomings of traditional BP algorithm as being slow to converge and easy to reach extreme minimum value, the model adopts LM (Levenberg-Marquardt) algorithm to achieve a higher speed and a lower error rate. Experiment results verify that the BP neural network improves the efficiency, accuracy and robustness of the method comparing with Traditional Binocular Location (TBL) methods.**

***Key words*—location; tracking; BP neural network; Matlab; simulation**

I INTRODUCTION

Computer Vision detection technology [1] which has lots of advantages such as non-contact, high-speed, high-precision, high degree of automation, is gaining more and more applications. Especially, Traditional Binocular Location (TBL) has been successfully applied in industrial inspection, object recognition, robot guidance, space objects 3D profile measurement, and other fields [2]. Multi-person location and tracking method is an important component of the study. In order to Compare TBL, in this paper a new method - Three-Camera System (TCS) is proposed. Through this new method, the coordinate mapping between the two-dimensional coordinates of the points in the image and the three-dimensional coordinates of the corresponding points is established through BP neural network, as neural networks have highly nonlinear mapping ability. The number of neurons and the type of transfer functions in the hidden layer of the neural network will be discussed in order to achieve the best results.

II THE ESTABLISHMENT OF 3D SCENE

3ds max is a powerful, object-oriented 3D modeling, animation and rendering software. Through the user-friendly interface of the 3ds max software, we create a three-dimensional scene (as shown in Figure 1), the origin is the red point in the ground (as shown in Figure 1). The direction of X-axis is from left to right; the direction of Y-axis is from back to front; for the Z-axis, the direction is from bottom to top. The area of the ground in the scene is 20 square meters, the height of the wall is 10 meter, and three cameras are respectively located at the point A, point B and point C which are 3-meter high.

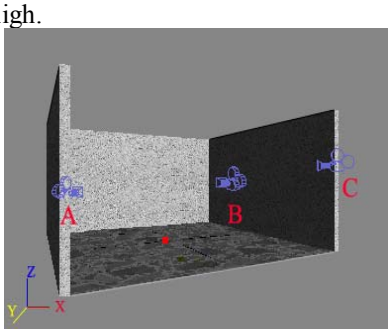

Figure 1. Three-Vision cameras system (3D)[1]

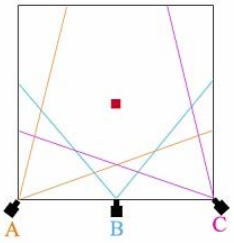

Figure 2. Three-Vision cameras system (2D)

Different from Traditional Binocular Location (TBL) methods based on neural networks, this paper use the 3ds max software to create three-vision cameras, which are located in the left, middle and right of the 3D scenes. The main purpose

*This project is supported by the Program of 985 Innovation Engineering on Information in Xiamen University (2004-2007)

of adding a video camera to the traditional binocular vision system is to solve the situation in which the figures overlap and improve the accuracy of location and tracking.

III IMAGE FEATURE EXTRACTION

Create a video document (301 frames total) in the environment of 3ds max software. In the video, several persons (with different height) walk into and out of the 3D scene along a route that is nonlinear one by one, in order to train the neural network. The images of Camera A, B and C in the 150th frame were shown respectively as follows:

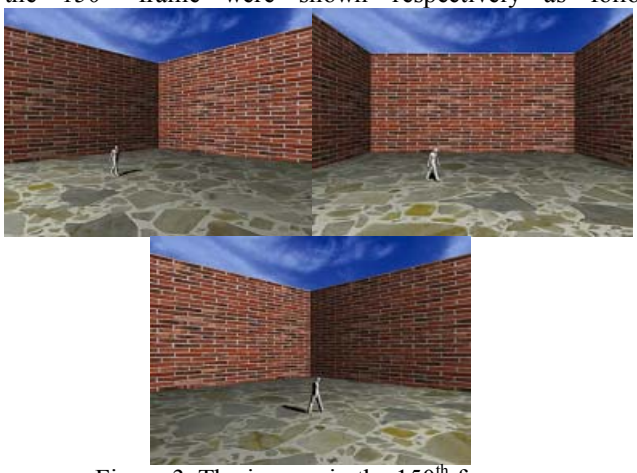

Figure 3. The images in the 150th frame

The motion detection method based on temporal difference is of strong self-adaptive. Lipton [3] used two-frame difference method to detect the moving objects from the actual video images, and to track the locomotor target for further. In the experiment of simulation we use three-frame difference method to substitute the two-frame difference method to judge the current pixel point $I_n(x, y)$ whether it is moving or not [4], by the giving threshold $T_n(x, y)$,which is shown as follow:

$$\left|I_n(x,y)-I_{n-1}(x,y)\right| > T_n(x,y) \quad (1)$$

$$\left|I_n(x,y)-I_{n-2}(x,y)\right| > T_n(x,y) \quad (2)$$

If the requirements (1), (2)are met, then determine the points are mobile, and record the two-dimensional coordinates of the head of the moving target.

The detected images will be divided into M×N sub-windows which are n×n (pixels) large, and we use the Color Moments (CM) brought forward by Stricker and Orengo [5] to describe and calculate the color space of each sub-windows respectively. Record the CM of the locomotor area (a rectangular area around the body), because skewness would be too sensitive to noise, here only use mean (formula 3) and variance (formula 4) to compute CM:

$$\mu_k = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} P_k(i,j) \quad (3)$$

$$\sigma_k = \left(\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}(P_k(i,j)-\mu_k)^2\right)^{\frac{1}{2}} \quad (4)$$

$$\hat{x} = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (5)$$

P_k (i, j) is the k^{th} color component of the pixel, which coordinates is (i, j) in the n×n sub-window.

Dealing with these three videos obtained by each camera through Matlab, we get the two-dimensional coordinate in each frame and create a two-dimensional array for the storage of these data. The two-dimensional coordinates in camera A will be recorded as follow: (XA_1, YA_1), (XA_2, YA_2),…,(XA_{301}, YA_{301}); the two-dimensional coordinates in camera B will be recorded as follow: (XB_1,YB_1), (XB_2,YB_2),…, (XB_{301},YB_{301}); the two-dimensional coordinates in camera C will be recorded as follow: (XC_1,YC_1),(XC_2,YC_2),…,(XC_{301},YC_{301}). Normalize the two-dimensional features before they are simulated by the BP neural network, here use formula 5 to translate the data into closed interval [0,1].

IV MULTI-PERSON LOCATION AND TRACKING

In the video, we use the trained neural network to locate and track the figure in the video files. In order to better promote the network, first of all, create a two-dimensional array $a_{m\times n}$ for tracking, and create a three-dimensional locating array to store the three-dimensional coordinates of the characters in the video files.

$$a = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}, \quad (6)$$

$$b_{ij} = [b_1, b_2, b_3, b_4], \quad (7)$$

(i=1, 2,…, m; j=1, 2,…, n),

Among the variables, m is the number of the figures in the video, n is the number of the frames of the video, b_{ij} denotes the real space coordinate of the i^{th} figure in the j^{th} frame, which is obtained through BP neural network, b_1 denotes the x-coordinate of the i^{th} figure in the j^{th} frame, b_2 denotes the y-coordinate of the i^{th} figure in the j^{th} frame, b_3 denotes the z-coordinate of the i^{th} figure in the j^{th} frame, b_4 denotes the CM of the i^{th} figure in the j^{th} frame; the matrix $a_{m\times n}$ denotes the index of array b_{ij} which is used to store the space coordinates of the m^{th} figure in the n^{th} frame.

Thus, the realization of the process of multi-person moving object tracking described below:

1) Detect the locomotor regions through motion detection method based on temporal difference;

2) Through the detection, if there are several locomotor regions, go to Step 4); otherwise, proceed as follows:

3) Calculate the two-dimensional coordinates of head through Matlab in each camera, and go to step 6);

4) According to different situations of the Color Moments (CM) features in the moving regions, deal with the three

images of the moving object separately with same CM and extract six-dimensional characters, go to Step 6);

5) Refresh memory cells, if a region with one kind of CM no longer exits for more than a certain period of time (set as 3 frames), then the memory cells b_{ij} corresponding to the CM will be released; if a region with one kind of CM exits for more than a certain period of time (set as 3 frames), then a memory cells b_{ij} corresponding to the CM will be created, the creation of memory cell will be used to the space coordinates corresponding to the CM;

6) Turn the six features- (XA_n, YA_n), (XB_n, YB_n), (XC_n, YC_n) into space coordinates through BP neural network, and the first three rows of b_{ij} will be to store the space coordinates of moving object and the CM features of the figures will be stored in the fourth rows of b_{ij};

7) Reach the end of the video, and close the process; otherwise go to 1) to continue the circulation.

V TRAINING AND SELECTING NETWORK WITH MATLAB

In recent years, neural network technology has been widely adopted in several areas [6]. The model created in this paper is a model with a single hidden layer. As theoretically proven, BP neural network can approach any nonlinear function with limited interruptions at any accuracy as long as the neurons in the hidden layer of the model are sufficient [7]. Assumed that the actual output of the neuron j in the output layer is $y_j(t)$ at time t and the expected output $d_j(t)$, so the network error function E(t) at the time t will be defined as formula :

$$E(t)=\frac{1}{2}\sum_{j-1}^{q}(y_j(t)-d_j(t))^2, \qquad (8)$$

q is the number of neurons in the output layer; ε is a pre-set error margin. The model that stops testing when E(t) （as shown in formula 8）is less than ε is the desired model[8].

Create a BP neural network through Graphical User Interfaces (GUI) in MATLAB (as in Figure 4).While traditional BP algorithm is a gradient descent algorithm, which computes rather slowly due to linear convergence, LM (Levenberg-Marquardt) algorithm, improved from BP algorithm, is much faster since it adopts the method of approximate second derivative[9].Select TRAINLM as the Training Function in the simulation.

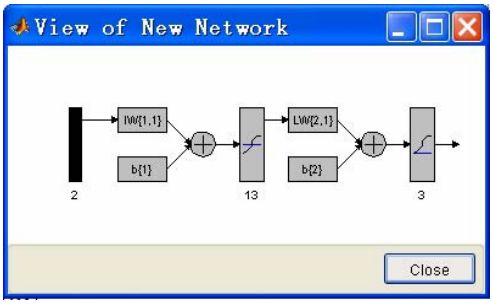

Figure 4. Network structure 1

Since the number of neurons in the hidden layer can be selected among 10 options between 11 and 20 and the transfer functions from either TANSIG or LOGSIG, 20 models can be totally created subject to various numbers of neurons and transfer functions in the hidden layer. Each model will be trained and tested with sample data separately for three times. Select the lowest error rate as the minimum value that the model can achieve. The training results show that the lowest error rate is *0.0003218*. Therefore, the corresponding network will be the desired model, of which the number of neurons in the hidden layer is 14 and transfer function TANSIG.

VI EXPERIMENTAL RESEARSH

Once the best working model is obtained through network training, the three-dimensional coordinates of the points will be worked out and compared with actual measured values to evaluate the veracity of the model. The comparison between the values of X-coordinate、Y-coordinate and Z-coordinate obtained by BP neural network and the values obtained by 3ds max is shown in the figures respectively.

It can be seen that the two curves almost overlap each other in those figures. Pearson Correlative Analysis between the two groups of data indicates that X-coordinate correlative coefficient is 0.999995 (see Figure 5), Y-coordinate correlative coefficient is 0.999916 (see Figure 6) and Z-coordinate correlative coefficient is 0.999932 (see Figure 7).

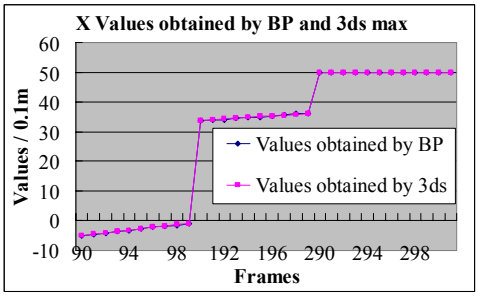

Figure 5. X-coordinate Value

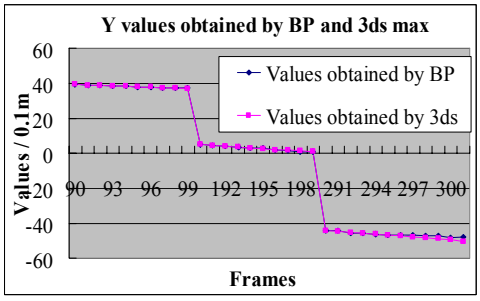

Figure 6. Y-coordinate Value

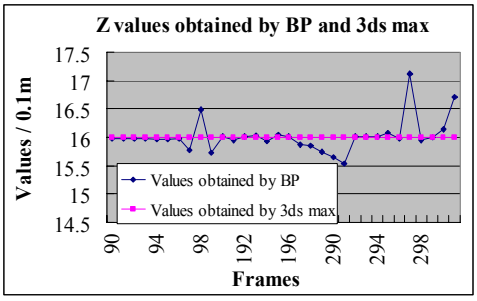

Figure 7. Z-coordinate Value

It can also see from the analysis that the average error of X-coordinate is 0.907%, with maximum at 29.0798% and minimum at 0.0011%.The average error of Y-coordinate is 0.6317%, with maximum at 12.4582% and minimum at 0.0201%. The average error of Z-coordinate is 0.2968%, with maximum 7.1032% and minimum at 0.0063%, as in Figure 8.

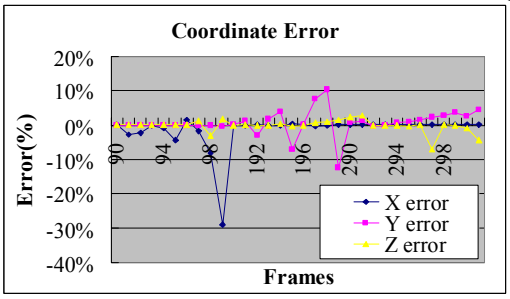

Figure 8. Coordinate Value error rate

It can be seen that error rates of some frames are relatively high. Nevertheless, the average error rate indicates that the overall results are fairly good with error rate controlled within an acceptable range, proving the viability of the network model.

Compared with the Traditional Binocular Location (TBL), the Three-Camera System (TCS) has a lower error rate. And the Convergence Time (CT) of TCS is shorter than that of TBL.

TABLE I. COMPARISION BETWEEN TBL AND TCS

	TBL			TCS		
	1st	2nd	3rd	1st	2nd	3rd
X (%)	3.59	3.24	3.30	2.46	3.08	1.84
Y (%)	2.94	3.33	2.98	1.95	2.01	2.13
Z (%)	2.67	2.84	2.71	1.82	1.77	1.93
CT(Epochs)	97	45	176	40	33	56

VII CONCLUSIONS

This paper adopts the BP neural network technology and GUI function of MATLAB to achieve easier and faster multi-person location and tracking. The location model is created with application of neural network, in which LM algorithm is used for its faster convergence speed and lower error rate to overcome the shortcomings of traditional BP algorithm as slow to converge and easy to reach extreme minimum value.

Moreover, by changing the number of neurons and the type of transfer function of the hidden layer, the best working model was obtained. Experiments show that TCS method based on BP neural network can produce good location results in general and can be used for multi-person location and tracking.

Further more, this model can be extended to further applications, especially in the real scene. Nevertheless, due to the fact that real scene is more complicated and changeable, factors such as sunshine, need to be considered in order to promote the applicability of the BP model.

ACKNOWLEDGMENTS

Thanks for the support provided by the Program of 985 Innovation Engineering on Information in Xiamen University (2004-2007)

REFERENCES

[1] Songde Ma, Zhengyou Zhang, Computing Theory and Algorithm based on Computer Vision. Science Press, Beijing, 1998.

[2] Liu, Y. M., Ding, Y., Xu, X. H., Stereovision Measurement Method Based on Neural Network, Computer Engineering and Application, pp.33-35,Aug.2005.

[3] Lipton. A., Fujiyoshi. H., Patil, R. Moving Target Classification and Tracking from Real-time Video. IEEE Workshop on Applications of Computer Vision, Princeton, 1998.

[4] Collins, R., Lipton, A., Fujiyoshi, H., etal. A System for Video Surveillance and Monitoring. American Nuclear Society (ANS) Eighth International Topical Meeting on Robotics and Remote Systems, Pittsburgh, PA, Nisan, 1999.

[5] Stricker, M., Orengo, M., Similarity of Color Images. SPIE Storage and Retrieval for Image and Video Databases III, San Jose, California, 1995.

[6] Lee,J.H.W., Huang,Y., Dickman,M., Jayawardena,A.W., 2003. Neural network modeling of coastal algal blooms. Ecological Modelling, 159(2-3):179-201.

[7] Xu, D., Wu, Z., Neural Network-system Design and Analysis Based on MATLAB6.X.University of Xi'an Electronics Technology Press, Xi'an, 2002.

[8] Kuo, Y.M., Liu, C. W., Lin, K. H., 2004.Evaluation of the ability of an artificial neural network model to assess the variation of groundwater quality in an area of blackfoot disease in Taiwan. Water Research, 38(1):148-158.

[9] Wang, Q. H., Improvement on BP algorithm in artificial neural network. Journal of Qinghai University, Vol 22, No. 3:pp. 82-84. 2004.

[10] Zhao, Y., Nan, J., Cui, F. Y., etal. Water quality forecast through application of BP neural network at Yuqiao reservoir Zhejiang Univ Sci A, Vol 9, pp. 1482-1487,Aug 2007.

[11] Zhexue Ge, Zhiqiang Sun, Theory of Neural Network and Realization of MATLAB R2007, Electronic Industry Press, Beijing, 2007.

[12] Zhang, Y. H., Mastering MATLAB5.Tsinghua University Press, Beijing, 1999.

[13] Haritaohlu, I., Harwood, D., and Davis, L. W., real-time surveillance of people and their activities. IEEETras Pattern Analsis and Intelligence, 2000,22(8): 809-830.

[14] Meyer, D., Denzler, J., Niemann, H., Model based extraction of articulated objects in image sequences for gait analysis. In: Pro IEEE International Conference on Image Processing, Santa, California 1997:78-81.