Proceedings of the 2009 IEEE International Conference on Systems, Man, and Cybernetics
San Antonio, TX, USA - October 2009

Non-rigid body Object Tracking using Fuzzy Neural System based on Multiple ROIs and Adaptive Motion Frame Method

Hyunsoo Lee
Department of Industrial and Systems Engineering
Texas A&M University
College Station, Texas, USA
hsl@neo.tamu.edu

Amarnath Banerjee
Department of Industrial and Systems Engineering
Texas A&M University
College Station, Texas, USA
banerjee@tamu.edu

***Abstract*— The proposed framework supports a new and efficient non-rigid body object tracking among objects with similar patterns. Human tracking is used as an example of a non-rigid body tracking. The main objective of this framework is to track the targeted person among people with similar patterns in a series of images frames. The targeted person is identified in the initial frame by the user. This framework consists of three stages: generation of panoramic images for a wider range, detection stage, and tracking stage. In detection stage, all multiple regions of interest (ROIs) are classified and the target is detected using multiple ROIs and fuzzy neural system. In tracking stage, the targeted person is tracked using an adaptive motion frame method that checks the target's movement. This suggested framework explains how multiple ROIs are used for detecting non-rigid body object and how the target can be tracked using previous trajectory and velocity. This algorithm contributes towards the tracking of a desired target that exists among many similar non-rigid body objects.**

Keywords*— *Non-rigid body detection and tracking, fuzzy neural system, multiple ROIs, adaptive motion frame method

I. INTRODUCTION

The main objective of this paper is to detect and track a non-rigid body among similar objects. In many detection and tracking problems, detection is usually considered as an identification of interest patterns, while tracking is considered as identification of an exact one among similar interests. In this context, the detection stage is a prerequisite element to the tracking process, and tracking algorithm is considered to be harder than detection algorithm. There are many existing detection and tracking algorithms. These algorithms and technologies have application areas, such as surveillance systems, traffic control systems and guiding applications. Table I describes some of the domains and specific application areas where object detection and tracking technology can be applied.

TABLE I. GENERAL DOMAINS AND APPLICATIONS OF OBJECT DETECTION TECHNIQUES

General Domains	Application Areas
Manufacturing System	1. Separation process - Separating target product from other products - Separating defective items in a lot of similar products. 2. Motion analysis - Time and Motion study in manufacturing process - Manufacturing Process Design
Traffic Control System	1. Traffic Signal control 2. Analysis of traffic volume 3. Analysis of traffic accident 4. Signaling in high-noise environment like airport
Surveillance System	1. Access Control 2. Control of various location (e.g. building, house, ATM, department store)
Virtual Applications	1. Interactive interface between real object and virtual object 2. Teleconferencing 3. game control

Since humans have non-rigid (flexible) body and movements, it is considered as the object for tracking in this paper. Non-rigid body transformation is caused by human actions such as walking, sitting, bending and turning. These characteristics make it difficult to detect and track humans as compared to rigid body objects. On the other hand, rigid body objects such as car and train can be detected using various template matching methods. Most of these template matching methods are very fast and accurate detection methods. However, most of these algorithms cannot be applied to detection and tracking of non-rigid body objects. To overcome this limitation, effective pattern features such as mean shift, Histogram of Gradient (HOG) and other filters have been tested [1]. These pattern features are combined with training methods such as different types of neural nets, Support Vector Machines (SVMs), regression models, and Hidden Markov Models (HMMs). While these approaches are considered as good detection algorithms of non-rigid body objects, these are some difficulties encountered in tracking the exact target among many similar objects. This paper provides an efficient tracking algorithm for non-rigid body object using fuzzy neural system based on multiple Regions of Interest (ROIs) and adaptive motion frame method.

In section II, some of the existing detection and tracking approaches are reviewed and compared. The detailed algorithm and framework are described in section III and IV. Section V shows the implemented results.

978-1-4244-2794-9/09/$25.00 ©2009 IEEE

II. BACKGROUND AND LITERATURE REVIEW

This section reviews some of the related human detection and tracking approaches. The detection and tracking approaches can be divided into three approaches: "trained model approach", "non-trained model approach" and "combined approach". Figure 1 shows how the detection and tracking algorithms can be classified.

The trained model approach and non-trained model approach are classified based on whether they use previously trained model or not. There are a number of detection algorithms that belong to the trained model approach [2-5]. This approach can be classified based on algorithms used for extracting pattern features, training models and learning methods. HOG, Wavelet filter, PCA shift and particular shape/geometry are some of the popular pattern feature extraction methods. HOG has been widely used due to the ability to capture edge and gradient structure. It has characteristics of local shape and easily controllable degree of invariance to local geometric and photometric transformations [2]. It has been experimentally verified that HOG feature has better classification performance as compared to Wavelet feature. Table II shows the comparisons among some trained model approaches.

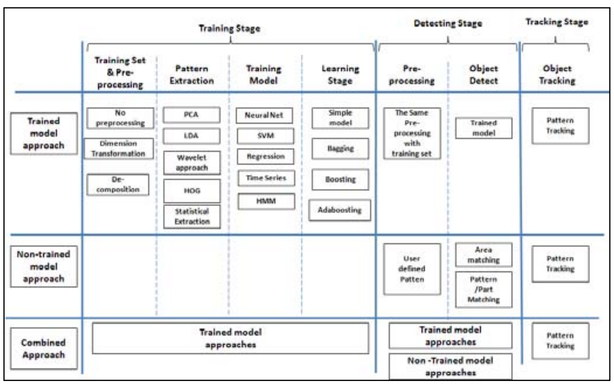


Figure 1. Classification of detection and tracking algorithms

TABLE II. TRAINED MODEL APPROACHES AND THEIR CHARACTERISTICS

	Number of target windows/ Number of ROI	Used Pattern features	Trained Model	Learning Method
Felzenszwalb, *et al*. [3]	Multiple ROIs	HOG	SVM	-
Dalal, *et al*. [2]	1	HOG	SVM	-
Papgeorgiou, *et al*.[4]	1	Wavelet	SVM	-
Zhu, *et al*. [5]	1	HOG	SVM	AdaBoost

In their approach, Felzenszwalb *et al* used a target's overall shape (human body) as well as some of the target's components (head, left and right arm, left and right thigh, and legs) for input to SVM. The output gives regions that represent not only overall shape but also shapes of components.

Many tracking algorithms including template matching belong to non-trained model approach [6, 7]. Their main objective is to find a user-defined pattern and track the target. Dasgupta and Banerjee suggested a pattern tracking method using morphological watershed algorithm [6]. In the initial frame, the user divides the frame into some segments and identifies the mask representing the boundary of the target. This mask represents the region of interest (ROI). In the following image frames, block matching is recursively executed to track the object and calculate the motion parameters. A non-trained model approach can be classified as template matching method or part based method. In template matching methods, the number of ROI is usually one. Multiple ROIs can be used in part based method. Many real object tracking applications use template matching methods due their smaller computation time compared to part based matching. However, tracking methods have some limitations. The first limitation is that it is hard to track a non-rigid body object. For example, a mask from a person's walking posture does not coincide with a mask from a person's bending posture. Another limitation is that it is difficult to track the exact target among similar objects. If many objects have similar pattern feature, a non-trained model approach's tracking ability may be decreased. A combined approach is suggested to overcome these limitations. The proposed algorithm combines the trained model approach and non-trained model approach. Using a trained model approach, multiple ROIs are acquired for the object of interest. A user identifies the object of interest among similar objects in the initial frame. The object is then tracked in the following frames. The following sections describe the algorithm and its implementation in detail.

III. NON-RIGID BODY OBJECT TRACKING ALGORITHM

This section describes the non-rigid body object tracking method using fuzzy neural system. Figure 2 shows the overall procedure of the suggested non-rigid body tracking algorithm. Image frames are taken by fixed multiple cameras and these multiple images are generated into one panoramic image using normalized DLT algorithm or RANSAC algorithm [8]. The usage of multiple cameras is not to acquire stereo image but to generate one panoramic image. The benefit of using panoramic image is that to track the target object in a wider range.

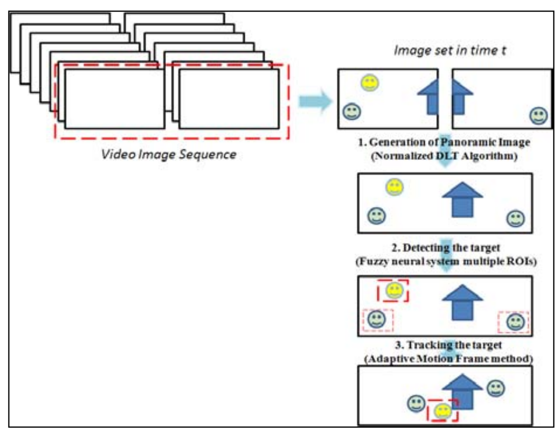


Figure 2. The overall tracking procedure.

This paper focuses on detection and tracking over three stages. After panoramic images are generated, the targeted person is identified using fuzzy neural system using multiple ROIs. A fuzzy neural system is a type of neural net model incorporating fuzzy logic [9, 10]. The suggested fuzzy neural system detects the target in five steps, which is summarized in figure 3. As a preprocessing step, multiple ROIs are extracted from all detected people using Felzenszwalb's method [3].

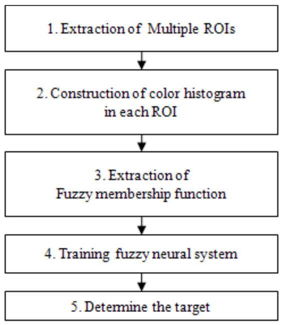


Figure 3. The procedure for non-rigid body object tracking.

The advantage of this method is that it can detect several parts which are defined using root filter and part filters. This multiple filters concept is used to incorporate multiple ROIs in this paper. Multiple ROIs can be classified as "*primary ROI*" and "*secondary ROI*". A *primary ROI* is similar to the root filter and *secondary ROI* is similar to the part filter in Felzenszwalb's method. The use of multiple ROIs can help to increase the detection accuracy. Here, one window box is used as a *primary ROI* (covering a targeted person's silhouette) and 6 small boxes are used as *secondary ROIs* (head, two arms, two legs, and foot). A Histogram of Oriented Gradient (HOG) is calculated using the multiple ROIs at the pattern extraction stage, and converted to a scored value. This value is used as the input to LSVM, which is the training model. After training LSVM, a neural net model for detecting all people in the image is acquired. Figure 4 shows the modified Felzenszwalb's approach for extracting multiple ROIs.

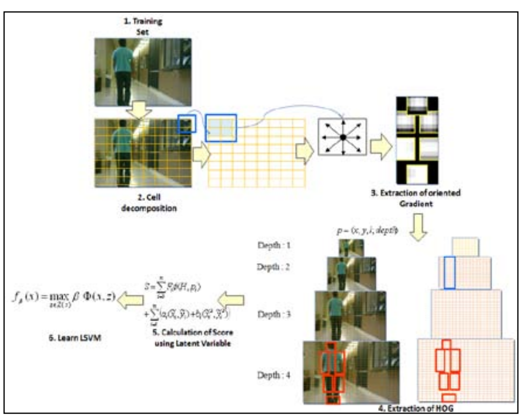


Figure 4. Step 1: Extracting primary and secondary ROIs using modified Felzenszwalb's method.

After acquiring multiple ROIs, the color (R, G, B) histograms are constructed for each *secondary ROI*. From these histograms, a median color value is acquired. This median value is used as a representative color for each secondary ROI. The representative color is fuzzified into fuzzy membership functions. The color and light in background environment influence the fuzzification. Finally, each *secondary ROI* is converted into three (R, G, B) fuzzy membership functions. Figure 5 shows the fuzzy membership functions of a secondary ROI (representing target's right thigh).

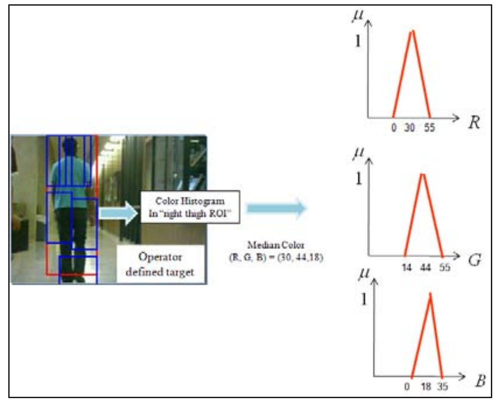


Figure 5. Conversion from secondary ROI to fuzzy membership functions.

The fuzzy membership functions are used as input to the fuzzy neural system. The architecture of the suggested fuzzy neural system consists of preprocessor, input layers, hidden layers and output layer. In the preprocessor, R, G, B fuzzy sets are defuzzified and entered into the input nodes of the fuzzy neural system. The input layer has six input nodes which represent the six *secondary ROI*s. A binary MLP (multi-layer perceptron) with three hidden layers is used. Figure 6 shows the architecture of the fuzzy neural system for detecting the targeted person.

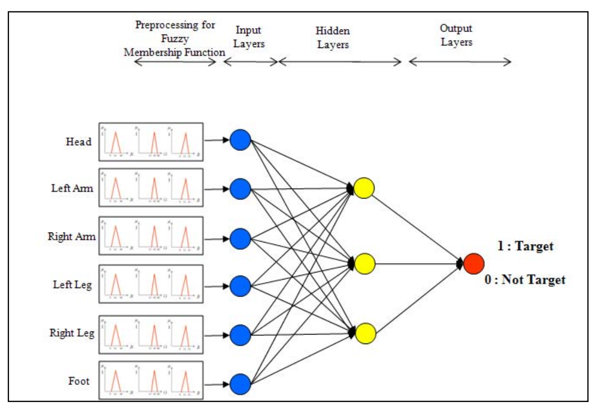
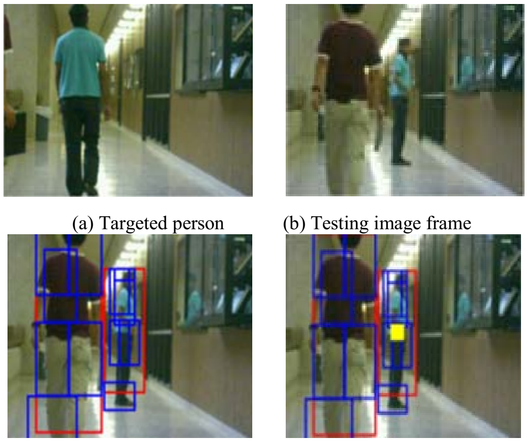


Figure 6. The architecture of fuzzy neural system detecting the target.

As the case with most neural nets, including fuzzy neural systems, a good and comprehensive training set is likely to increase the detection ability. The main characteristic of this system is in the combined model of LSVM using HOG and a fuzzy neural system. The LSVM using HOG, can detect all similar non-rigid body objects (people). Subsequently, only the targeted person can be selected using fuzzy neural system. Figure 7 shows the detection result.



(a) Targeted person (b) Testing image frame

(c) Extraction of Secondary ROIs (d) Detecting target using fuzzy neural system

Figure 7. Detecting the targeted person using fuzzy neural system.

IV. ADAPTIVE MOTION FRAME METHOD

The suggested fuzzy neural system can detect the targeted person using color fuzzy membership functions based on *secondary ROI*s. This proposed system has limitations similar to the ones that are encountered with many color based detection and tracking algorithms. The main limitation is that it is hard to detect and track the targeted person among people with the similar (color) patterns. Suppose that all people wear the same clothes and have similar body sizes. In this situation, most existing algorithms experience difficulties detecting and tracking the target.

To solve this limitation, an adaptive motion frame method is suggested using the target's moving trajectory. From the previous panoramic images, the moving trajectory of the tracked target can be acquired. This trajectory is formulated into a regression model. Using this regression model, the target's position in the next frame can be predicted. Figure 8 shows the concept of the adaptive motion frame method.

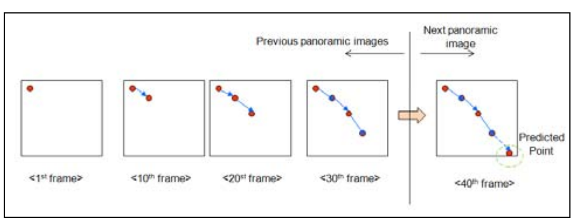


Figure 8. Concept of regression model from previous target's trajectory.

If many people with similar pattern are detected regardless of the fuzzy neural system, the distance between each identified person's position and predicted position using regression model is measured. Then, one identified person with minimal distance is selected as the targeted person. Figure 9 shows the selection process of the targeted person.

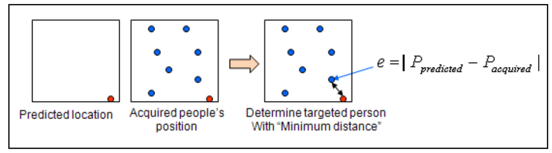


Figure 9. The selection of the targeted person.

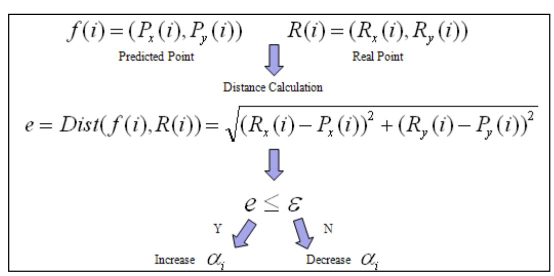
This regression model can be a linear model or a nonlinear model based on the type of target and frame intervals. If the interval between two consecutive image frames is decreased, the linear regression model can be used as a regression model. In this paper, we use a nonlinear regression model considering the fact that a person can exhibit drastic motion and velocity change. Equation (1) describes the regression model.

$$\begin{aligned} f(t_{i+1}) &= (P_x(t_{i+1}), P_y(t_{i+1})) \\ &= (\mathrm{Reg}_x(t_{i+1}), \mathrm{Reg}_y(t_{i+1})) \end{aligned} \tag{1}$$

$$\text{where } t_{i+1} = t_i + \alpha$$

In equation (1), $f(t_{i+1})$ represents the estimated position of target in frame time t_{i+1}. $f(t_{i+1})$ has each position $P_x(t_{i+1})$ and $P_y(t_{i+1})$ with respect to x and y axis. $P_x(t_{i+1})$ and $P_y(t_{i+1})$ are calculated from the regression model $\mathrm{Reg}_x(t_{i+1})$ and $\mathrm{Reg}_y(t_{i+1})$ respectively. α is the frame interval. As an adaptive method, α_i can be used instead of α. Using a threshold value (ε), if the distance between "predicted position" and "acquired position with minimal distance" is less than ε, we can conclude that the regression model is accurate with respect to current frame time α_i. It means that the prediction model gives good performance. So, α_{i+1} can be increased from previous frame interval α_i to save on computation time. Table III summarizes the adaptive motion frame method.

TABLE III. THE ADAPTIVE MOTION FRAME METHOD



As an example, suppose that two people wear similar clothes (figure 10). At the end of detection stage, two people are detected. To determine and track the targeted person, previous trajectories of the target are investigated. Table IV shows X, Y coordinate point acquired from previous image frames.

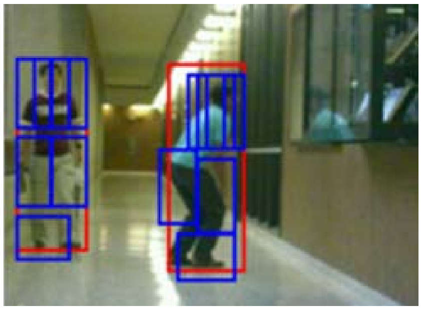


Figure 10. Two identified people using detection algorithm.

TABLE IV. THE TARGET PERSON'S PREVIOUS POSITION INFORMATION

No.	Panoramic Image frame	X coordinate	Y coordinate
1	415	305	246
2	435	304	263
3	455	297	271
4	475	294	280
Target Frame	495	?	?

To acquire the regression model, a 5-degree polynomial regression model and 4 previous frames are considered. Initial value α_i is set at 20. After applying adaptive motion frame method, the acquired prediction model is generated (table V).

TABLE V. ACQUIRED PREDICTION MODEL USING ADAPTIVE MOTION FRAME METHOD

$$R_x(t_{i+1}) = (9.9873\text{e-}010)t_{i+1}^5 + (\text{-}1.3128\text{e-}006)t_{i+1}^4 + (5.6636\text{e-}004)t_{i+1}^3 + (\text{-}0.0786)t_{i+1}^2$$

$$R_y(t_{i+1}) = (1.1502\text{e-}009)t_{i+1}^5 + (\text{-}1.5540\text{e-}006)t_{i+1}^4 + (6.9559\text{e-}004)t_{i+1}^3 + (\text{-}0.1018)t_{i+1}^2$$

Using this prediction model, the target's position is predicted and selected. Figure 11 shows the final result.

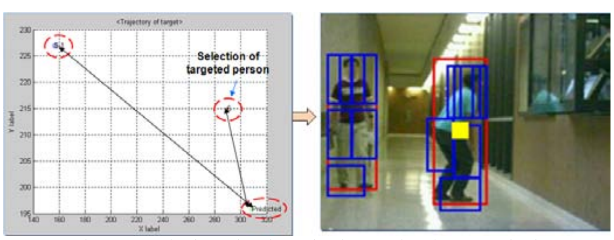


Figure 11. Final result using adaptive motion frame method

V. IMPLEMENTATION AND RESULTS

To show the effectiveness of the proposed approach, a person of interest is tracked in a series of image frames. Figure 12 shows some sampled image frames and results. In this example, our targeted person is the man in the 74th frame. Figure 12 (a) is the original image frame and 12 (b) shows the results after extracting multiple ROIs. Figure 12 (c) shows the successful target tracking using the fuzzy neural system and adaptive motion frame method. This approach is implemented and tested in a Unix/Linux environment using *Matlab* and *C++*. *C-Mex* method is used to link *Matlab* and *C++* files.

VI. DISCUSSION AND CONCULSTION

In this paper, a new effective non-rigid body object tracking approach is proposed. The suggested approach consists of three stages. In the first stage, panoramic images are generated for wide range detection. There is another advantage in using panoramic images. The target's moving trajectory in a wide range can improve calculation and increase the effectiveness of the adaptive motion frame method. After generating panoramic image, the non-rigid body objects are detected using LSVM. The result of LSVM is the extracted *primary ROI* and *secondary ROI*s of all identified non-rigid body objects. The representative color information are extracted from *secondary ROI*s and converted into fuzzy membership function. These fuzzy membership functions are used as input to a fuzzy neural system. In case of the existence of many similar non-rigid body objects, an adaptive motion frame method is applied to improve the tracking ability. From previous tracking results, the target's moving trajectory is calculated and used in a regression model. Using the regression model, the target's position is predicted. By measuring the distance between identified position and predicted position, the object with minimum distance (error) is selected as the target.

As further study, a better algorithm with low computation complexity is needed for real time application. Multi-target tracking can be considered using ensemble learning methods. Also, the suggested method can be applied for fast 3D reconstruction of the interest/target. Most 3D reconstruction algorithms are based on stereo/multiple images. While those methods need a lot of computation, this method can reconstruct a target's 3D model using extracted multiple ROIs. Since *primary ROI* and *secondary ROI*s have the structure information of the target, these can be used for good prior knowledge of 3D model.

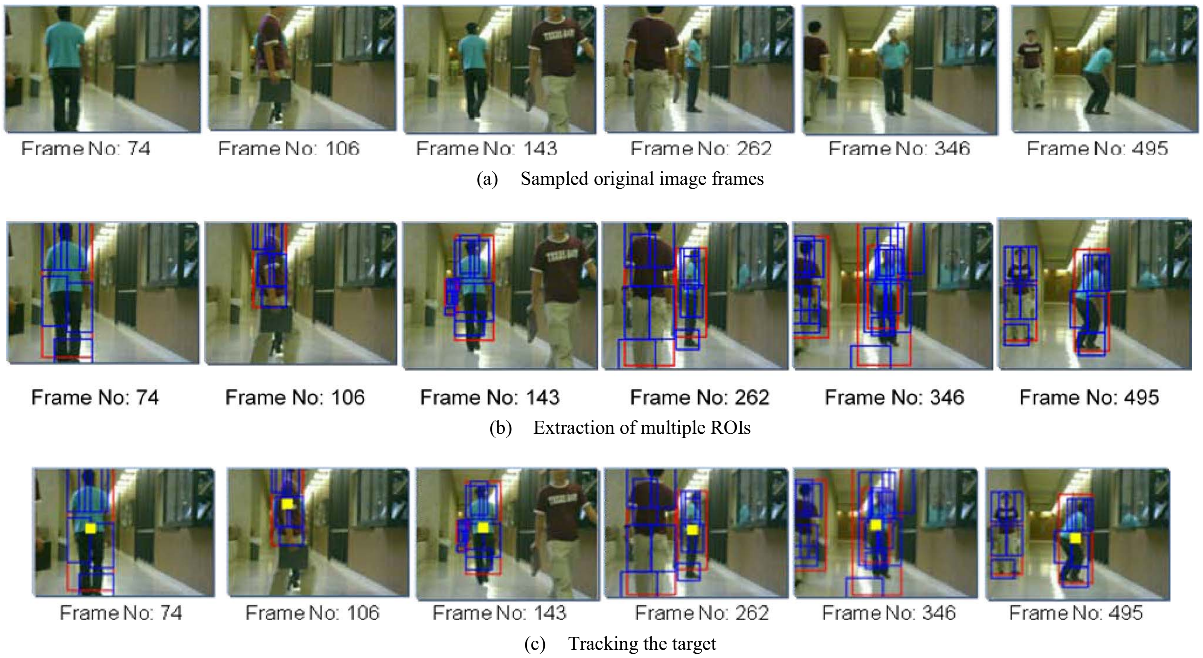


Figure 12. The tracking results using fuzzy neural system and adaptive motion frame method.

REFERENCES

[1] R. Patil, P. E. Rybski, T. Kanade, and M. M. Veloso, "People Detection and Tracking in High Resolution Panoramic Video Mosaic," *Proceedings of 2004 IEEE/RSJ International Conference on Intelligent Robots and Systems,* pp. 1323-1328, 2004.

[2] N. Dalal and B. Triggs, "Histograns of Oriented Gradients for Human Detection," *Proceedings of the Conference on Computer Vision and Pattern Recognition,* vol. 2, pp. 886-893, 2005.

[3] P. Felzenszwalb, D. McAllester, and D. Rammanan, "A Discriminatively Trained, Multiscale, Deformable Part Model," *Proceedings of IEEE CVPR,* 2008.

[4] C. Papageorgious and T. Poggio, "A Trainable System for Object Detection," *International Journal of Computer Vision,* vol. 38, pp. 15-33, 2000.

[5] Q. Zhu, S. Avidan, M. Yeh, and K. Cheng, "Fast Human Detection using a Cascade of Histogram of Oriented Gradients," *IEEE International Conference on Computer Vision and Pattern Recognition,* 2006.

[6] S. Dasgupta and A. Banerjee, "Pattern Tracking and 3-D Motion Recognition of a Rigid Body From a 2-D Image Sequence," *IEEE Transactions of sytems, man, and cybernetics,* vol. 35, pp. 116-125, 2005.

[7] S. Dasgupta and A. Banerjee, "An Augmented-Reality Based Real-Time Panoramic Vision System for Autonomous Navigation," *IEEE Transaction on Systems, Man and Cybernetics,* vol. 36, pp. 154-161, 2006.

[8] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2 ed.: Cambridge, 2003.

[9] A. D. Kulkarni, *Computer Vision and Fuzzy-Neural Systems*: Prentice Hall, 2001.

[10] J. Yen and R. Langari, *1998*: Prentice-Hall, Fuzzy Logic: Intelligence, Control, and Information.