3-10

MVA2007 IAPR Conference on Machine Vision Applications, May 16-18, 2007, Tokyo, JAPAN

An Efficient Method for Human Behavior Identification

Fang-Hsuan Cheng
Department of Computer Science & Information Engineering, Chunag Hua University
Hsinchu, Taiwan 300
fhcheng@chu.edu.tw

Fu-Tai Chen
Department of Computer Science & Information Engineering, Chunag Hua University
Hsinchu, Taiwan 300

Abstract

This paper presents a recognition method for human behavior identification based on motion history image theory. The motion history image has the advantage that it can record the motions of object over time. It can save complete motions history of object and have less computation for reckoning. The action features used in our system are motion gradient magnitude histogram and global/local motion orientation obtained from the motion history image. An error back-propagation neural network is used to identify the human behavior. The experimental results prove the feasibility and usefulness of the proposed method.

1. Introduction

The video capturing device and the vast amount of storage have become cheaper and more popular. At present the surveillance system is utilized on many different fields concretely. The tendence on many traditional video data of surveillance use computer storage equipment, but it can't get the message of the event in real time. It still need artificially review the video data of surveillance effect. Therefore, we present the recognition method of real time human behavior identification, it can get the human behavior speedily, consequently it can be applied on intellectual products of video surveillance system to conquer the past queries on major systems for only analyzing on the aim of single object. The *wavelet transform* has been used in signal processing, image processing and video compressing field etc. In contrast with *Fourier transform*, the *discrete wavelet transform* can keep the information of both time and space domain after transforming. This is the reason why it has been adopted to implement extensively. We apply it by down-sampling the data to speed up the computation and filter out the high frequency noise, and use the object's color feature information to solve the problem of multiple objects identification [1]. Nowadays the most researches of human behavior identification utilize features acquired from single video image. But a meaningful human behavior is composed of a few sequent images. Therefore we employ the characteristic of *motion history image* to accumulate object motion history over a short span of time, and it can produce the *motion gradient magnitude histogram* and global / local motion orientation features. Finally, the *error back-propagation neural network* is used to identify the human behavior.

Recently the motion detection algorithms have been proposed. Generally, they can be classified into three methodologies. The first one is *background subtraction* method. This method can get complete foreground feature, but it suffers a great influence upon movement background such as light changing, the wind blowing and the defoliation etc. The W^4 [2] propose a statistical cycle to observe the variant range of every pixel in the frame that has no moving objects. After the task they can build a background model by the observed information, and compute the inter-frame difference of the current frame and background model to detect the moving pixels. But the background can't be moved or the background model need to be re-established. The second on is *optical flow* method [3, 4, 5]. It utilizes the motion parts of object in light change to detect the moving object. It can do a good job to detect the moving object even the camera is moving, but need a considerable complicated computing. It can't be applied in real time system without any special hardware support. The last one is *temporal difference* method. *Lipton* [6] using a temporal difference algorithm, the motion features are extracting from the neighbor on two frames' difference. They even use three frames' difference to capture the motion features as *VSAM* [3]. They use a threshold value to separate motion pixel and motionless pixel. A good threshold value will concern with the results. This methodology has a very fast and high throughput and fits for real time system. The only drawback of this method is that it can't gain complete motion features, there exist many cavities or imperfect foreground features.

Black and *Yaccob* [7] present the cardboard people model which indicates the human limbs and trunk by the several connecting planes and the variations of parameters of planes are concerning with the human's motion. *Rohr* [8] applies 14 elliptic cylinders for exhibiting the human structure. The origin of the coordinates is set on the center of body, and all of the motion of limbs and trunk are changing the coordinates which will characterize the human behavior. *Bobick* and *Davis* [9] analyze the human behavior by *motion energy images (MEI)* and *motion history images (MHI)*. The *MEI* takes down the bi-tonal image of moving area which will be accumulated by time, and reckons the proportion of the value of each moving pixel to the motion continued time that can acquire the *MHI*. The benefit of it is the low time complexity, but it is influenced by the noise and the intermission time of the motional object easily. *Hidden markov model (HMM)* [10,11] has varied applications in identification territory such as human posture recognition, gesture recognition and facial expression recognition, etc. This model is a double random process system, it employs the *gaussian mixture model (GMM)* to compute the state machine, and outputs the results by highest and probable path. The *neural network (NN)* methodology has noise resistant and automatic learning ability. It is also generally used as a kernel on identification technique. The network has sizable neural cells which are connected with one another. Every cell has an *active function* to resolve the condition of data which will be conveyed to next layer of cells. These data will go through *weight machine* to determine flow rate of data. It will find the global minimum value as a target by using *gradient decent* method, *quasi-Newton* method or *Levenberg-Marquardt* method when it is going to train the network. The network will possess a very fast computing ability after training and its framework is consistent with *distributed computing environment (DCE)*.

2. Motion History Image

The t*imed motion history image (tMHI)* [12-15] is adopting the timestamp of inter-frame difference pixels to record the motion history at that point. It can use a simple replacement and decay operator:

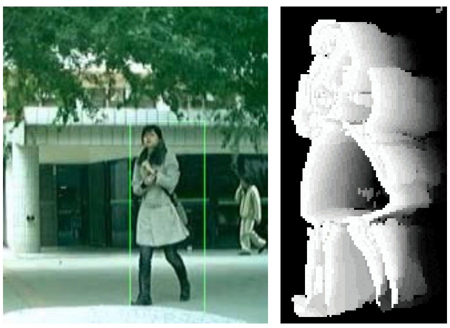
$$tMHI^i(x,y)=\begin{cases}\tau & \text{if } D^i(x,y)=1\\ 0 & \text{if } tMHI^i(x,y)<(\tau-\delta)\\ tMHI^i(x,y) & \text{otherwise}\end{cases} \quad (1)$$

where $D^i(x,y)$ is the inter-frame difference image of frame *i*, τ is the present time and δ is the maximum continuous time of *tMHI* which can adjust all sorts of system speed. The larger value of δ can retain more of the motions of object, it will be set within few seconds ordinarily. Since it has the exceptional characteristics, that is why we can readily get the history information of successive actions of object and further analyze it.

The results of $tMHI^i(x,y)$ is the timestamp but we can't use it directly so that we transform it into 255 scale grey image. The motion history image is demonstrated in Figure 1.

$$Gray^i_{MHI}(x,y)=\frac{tMHI^i(x,y)-(\tau-\delta)}{\delta}\times 255 \quad (2)$$

For the purpose of lessening the dispensable reckoning in the following processes, we compute the bounding box B^i_{MHI} of $Gray^i_{MHI}$ and set it as a region of interest (*ROI*).



(a) (b)

(c) (d)

Figure 1 The motion history image. (a)(c) The original image. (b)(d) The motion history image of moving object.

3. Motion Feature Extraction

Although it is very elaborate on human behavior, it still has its humanistic fixity. We make the *motion gradient magnitude histogram (MGMH)* of human as the major features, and make the global and local motion orientation as the minor features.

3.1 Motion gradient & motion gradient magnitude histogram

We employ the 3x3 *Sobel* matrices to get the gradient and magnitude of *MHI*, the equation is

$$S_x=\begin{bmatrix}1 & 0 & -1\\ 2 & 0 & -2\\ 1 & 0 & -1\end{bmatrix} \quad S_y=\begin{bmatrix}1 & 2 & 1\\ 0 & 0 & 0\\ -1 & -2 & -1\end{bmatrix} \quad (3)$$

$$\theta^i(x,y)=\arctan\left(\frac{G^i_y(x,y)}{G^i_x(x,y)}\right) \quad (4)$$

$$E^i(x,y)=\sqrt{G^i_x(x,y)^2+G^i_y(x,y)^2} \quad (5)$$

where G^i_x and G^i_y are the results of $Gray^i_{MHI}$'s convolution of 3x3 *Sobel* matrices S_x and S_y, $\theta^i(x,y)$ is the direction of gradient and $E^i(x,y)$ is the magnitude of gradient.

It can get the *motion gradient magnitude histogram* $MGMH^i$ processed from the gradient $\theta^i(x,y)$ and magnitude $E^i(x,y)$ of $Gray^i_{MHI}$ (Eq. (6)), and is described in Figure 2. This can express the orientation and gradient magnitude of the moving object clearly. To decrease the dimension of features, we scale the gradient by 5 degrees, which can gain the 72 sets of motion gradient magnitude. Then the normalization (Eq. (7)) is applied since the speed and distance of the object will influence the magnitude value of motion, and we can achieve a good results after this process.

$$\hat{E}^i(\phi,x,y)=\begin{cases}E^i(x,y) & \text{if } \theta^i(x,y)=\phi \text{ and } Gray^i_{MHI}(x,y)\neq 0\\ 0 & \text{otherwise}\end{cases} \quad (6)$$

$$MGMH^i(\phi)=\sum_{x,y}\hat{E}^i(\phi,x,y)$$

where $x,y\in B^i_{MHI}, \phi\in Z$ and $0\le\phi<360$.

$$MGMH^{i}_{normal}(\phi) = \frac{MGMH^{i}(\phi) - \left(\sum_{n=1}^{\#\phi} MGMH^{i}(\phi) \Big/ \#\phi\right)}{\sqrt{\sum_{m=1}^{\#\phi}\left(MGMH^{i}(m) - \left(\sum_{n=1}^{\#\phi} MGMH^{i}(n) \Big/ \#\phi\right)\right)^{2} \Big/ \#\phi}} \quad (7)$$

where $MGMH^{i}$ is the motion gradient magnitude histogram of moving object *i*, and the $\#\phi$ is the amount of sets of degree of gradient after quantizing process.

The radar maps of human behavior are obtained from the motion gradient magnitude histogram, it has 11 kinds of human behavior representing walking, running, bending, squatting, sitting (all with leftward/rightward) and standing up. Figure 2 is the motion gradient magnitude histogram of walking action with leftward and rightward direction.

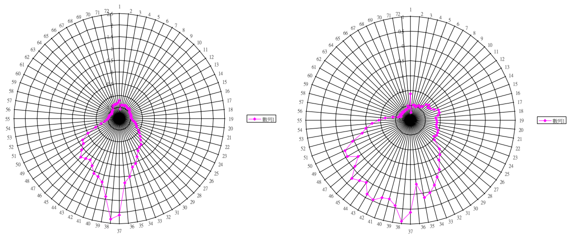


Figure 2 The motion gradient magnitude histogram of working action.

3.2 Motion orientation of object

Only applying the motion gradient magnitude histogram feature is insufficient to analyze the human actions due to there are similar gradient magnitude in minority actions. So we add the motion orientation of object feature to enhance the features. The motion orientation includes both global and local orientation which is

$$O \in O_g, \left[O_l^t, O_m^t, O_r^t, O_l^m, O_r^m, O_l^b, O_m^b, O_r^b\right] \quad (8)$$

where O_g is the global motion orientation and the set of $[O_l^t, O_m^t, O_r^t, O_l^m, O_r^m, O_l^b, O_m^b, O_r^b]$ represent the top-left, top-middle, top-right, middle-left, middle-right, bottom-left, bottom-middle and bottom-right of all eight motion orientations separately.

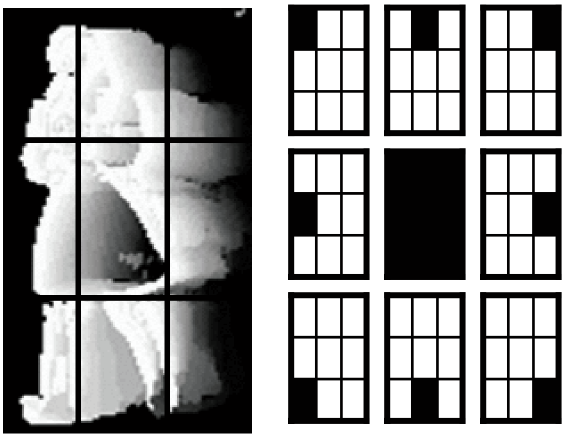
The global motion orientation can represent the object's moving direction. It can carry off that from the following equation

$$O_\theta = \theta_{ref} + \frac{\sum_{x,y} angDiff\,(\theta(x,y), \theta_{ref}) \times norm(\tau, \delta, tMHI^{i}(x,y))}{\sum_{x,y} norm(\tau, \delta, tMHI^{i}(x,y))}) \quad (9)$$

where $x, y \in B^{i}_{MHI}$ and $Gray^{i}_{MHI}(x,y) \neq 0$

θ_{ref} is the peak value θ in the gradient histogram, *angDiff*() is the minimum deviation value of θ_{ref} and $\theta(x,y)$, $tMHI^i(x,y)$ will be normalized from 0 to 1 using the current timestamp τ and duration δ by *norm*() function, and B^{i}_{MHI} is the set of coordinate of bounding box of moving object *i*.

The local motion orientations are segmenting the $Gray^{i}_{MHI}$ image into 9 grids which grant the number from 1 to 9, and compute each grille's motion orientation but excluding number 5th grid. The segment ivision table is shown as follows



(a) (b)

Figure 3 Local motion orientations segmenting. (a) The local motion orientations segmenting table. (b) The black area is the segmented region which will be computed by the motion orientation process.

4. The Behavior Identification

We applied the *error back-propagation neural network* [16] to identify the behavior of object after obtaining the features of object's gradient histogram, global motion orientation and local motion orientation. Figure 4 shows the framework of neural network.

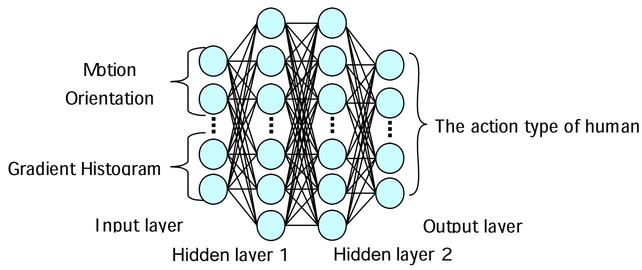


Figure 4 The framework of neural network for identifying the human behavior.

The *active function* of the neural network is

$$a_j = f(\sum_i w_{ji} a_i) \quad (10)$$

where the w_{ji} is the weight of the j_{th} node in current layer to i_{th} node in previous layer, and α_i is the output of the i_{th} node in previous layer. The *f(x)* is the sigmoid function $f(x) = \frac{1}{1+e^{-x}}$ (11)

and the *learning rule* is $\nabla w_{ji} = \eta \delta_j a_i \quad , 0 < \eta \leq 1$ (12)

where
$$\delta_j = \begin{cases} (d_j - a_i) f'(S_j) & \text{if } j \text{ is an output unit} \\ f'(S_j) \sum_{k=1}^{n} \delta_k w_{kj} & \text{if } j \text{ is an hidden unit}, n \in h+1 \end{cases}$$

∇w_{ji} is the rectifying value of weight w_{ji}, η is the learning rate which will be set to 0~1, δ_j is the error value of the output and target in j_{th} node of current layer *h*, δ_k is the error value of the output and target in k_{th} node of next layer *h*+1, and S_j is the income sum.

There are 11 kinds of actions shown in Table 1, which will be given a symbol to represent the human behavior.

Table 1. The symbol of human behavior.

Human behavior	Symbol	
	Rightward	Leftward
Walking	RW	LW
Running	RR	LR
Bending	RB	LB
Squatting	RSQ	LSQ
Sitting	RST	LST
Standing up	UP	

5. Experimental Results

Several test image sequences with single and multiple object are used in the experiments. The average correct identification rate is about 98% from 299 human actions. Figure 5 shows an example for behavior identification of running.

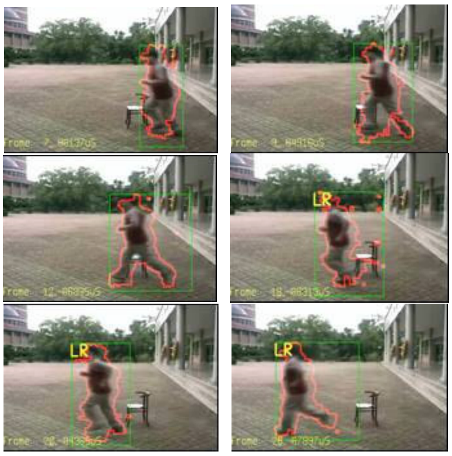


Figure 5 Behavior identification of running

6. Conclusions

In this paper, we develop a real time human behavior identification system based on motion history image theory. We apply the characteristic of *MHI* to preserve the object's motion information, and employ the motion gradient magnitude histogram and motion orientation features to identify the action of object by the error back propagation neural network. It utilizes the object's color information purposely for multiple objects' behavior identification.

When an object enters or leaves the scene, we can't get plentiful motional features to represent the moving object that causes the wrong identification. The ratio of identification depends on the correctness of the moving object detection, these flaws will get improved by a better motion feature selection. We can identify the human's basic behavior just now, but the action features are insufficient for complicated actions. In the future researches, we can add the contour of the object or other features to improve the above-mentioned problems.

References

[1] Fang-Hsuan Cheng & Yu-Liang Chen, "Real time multiple objects tracking and identification based on Discrete Wavelet Transform", *Pattern Recognition*, Vol.39, No.6, pp.1126-1139, 2006.

[2] I. Haritaoglu, D. Harwood, L. S. Davis, "W^4: Who? When? Where? What? A real time system for detecting and tracking people", *Face and Gesture Recognition Workshop,* 1998.

[3] R. Collins, A. Lipton, T. Kanade, H. Fujiyoshi, D. Duggins, Y. Tsin, D. Tolliver, N. Enomoto & O. Hasegawa, "A system for video surveillance and monitoring", *VSAM final report Carnegie Mellon University, Technical Report:* CMU-RI-TR-00-12, 2000.

[4] D. Meyer, J. Denzler & H. Niemann, "Model based extraction of articulated objects in image sequences for gait analysis", *IEEE International Conference on Image Processing,* pp.78-81, Santa Barbara, California 1997.

[5] J. Barron, D. Fleet & S. Beauchemin, "Performance of optical flow techniques", *International Journal of Computer Vision, 12 (1): 42-77, 1994.*

[6] A. Lipton, H. Fujiyoshi & R. S. Patil, "Moving target classification and tracking from real-time video", *IEEE Workshop on Applications of Computer Vision,* pp.8-14, Princeton, NJ, 1998.

[7] S. Ju, M. Black & Y. Yaccob, "Cardboard people: a parameterized model of articulated image motion", *IEEE International Conference on Automatic Face and gesture Recognition*, pp. 38-44, Killington, Vermont USA,1996.

[8] K. Rohr, "Towards model-based recognition of human movements in image sequences", *CVGIP: Image Understanding,* 59(1), pp.94-115, 1994.

[9] A. Bobick & J. Davis, "The representation and Recognition of Action Using Temporal Templates", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 23, No. 3, pp.257-267, 2001

[10] J. Yamato, J. Ohya & K. Ishii, "Recognizing human action in time-sequential images using hidden Markov model", *IEEE Conference on Computer Vision and Pattern Recognition,* pp.379-385, Champaign, Illinois, 1992.

[11] M. Brand, N. Oliver, & A. Pentland, "Coupled hidden Markov models for complex action recognition", *IEEE Conference Computer Vision and Pattern Recognition,* pp.994-999, Puerto Rico, 1997.

[12] J. W. Davis, "Hierarchical motion history images for recognizing human motion", *Proceedings. IEEE Workshop on Detection and Recognition of Events in Video*, pp.39-46, 2001.

[13] G. R. Bradski & J. Davis, "Motion segmentation and pose recognition with motion history gradients", *Fifth IEEE Workshop on Applications of Computer Vision*, pp.238-244, Dec. 2000.

[14] J. Davis & G. Bradski, "Real-time motion template gradients using Intel CVLib", *IEEE ICCV Workshop on Frame-rate Vision,* September 1999.

[15] J. Davis, "Recognizing movement using motion histograms", *MIT Technical Report #487,* March 1999.

[16] N. Baba, "A new approach for finding the global minimum of error function of neural networks", *Neural Networks,* vol. 2, 367, 1989.