ABNORMAL ACTIVITY RECOGNITION IN OFFICE BASED ON ℜ TRANSFORM

Ying Wang, Kaiqi Huang and Tieniu Tan

National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences
(wangying, kqhuang, tnt@nlpr.ia.ac.cn)

ABSTRACT

This paper introduces an abnormal activity recognition method based on a new feature descriptor for human silhouette. For a binary human silhouette, an extended radon transform, ℜ transform, is employed to represent low-level features. The information that the initial silhouette carries is transformed in a compact way preserving important spatial information of the activities. Then a set of HMMs based on the features extracted by our method are trained to recognize abnormal activities. Experiments have proved the accuracy and efficiency of the proposed method, and the comparison with Fourier descriptor illustrates its robustness to disjoint shapes and shapes with holes.

Index Terms— ℜ Transform, HMM, Abnormality Recognition, Feature Descriptor, Surveillance.

1. INTRODUCTION

Human activity analysis plays a significant role in computer vision and pattern recognition [1]. As a research focus, abnormal activity recognition has a wide range of applications such as intelligent surveillance, analysis of the physical condition of people and caring of aged people [2, 3, 4]. In general, human behavior analysis works on topics including background segmentation, object detection, tracking, feature extraction, learning and recognition of human activities. Feature description, as a bridge between low level image processing and high level activity understanding, plays a key role in human activity analysis. In this paper a feature descriptor with good distinguishing capability but low dimensionality is proposed.

Two kinds of features, silhouette and contour, are commonly used. The silhouette method takes into account all the pixels within a shape, and the contour method only extracts the boundary of a shape. General contour-based descriptors include Wavelets, Fourier descriptors and Hough transform [5, 6, 7]. Since contour descriptors are based on the boundary of a shape, they cannot capture the internal structure information. Furthermore, these methods are not suitable for disjoint shapes or shapes with holes where the boundary information is not available. Consequently, they are limited to certain applications. [8] proves that silhouette-based approaches outperform the contour-based methods in many cases. Examples of silhouette-based shape representation are geometric moments, Hu moments and some other varieties based on moment theory [9, 10]. But the moments are computationally intensive and most transforms are not centroid independent and scale free. However, in surveillance, the size of moving object varies with its distance to camera. Therefore we need features invariant to translation and scaling of moving people, and ℜ transform is adopted.

Our research focuses on recognizing some abnormal activities in office such as rushing in, carrying a package out of the office and suddenly bending down when walking. Our algorithm comprises two major parts: motion feature representation and abnormal activity recognition. An extended Radon transform is used to extract low dimensional feature vector for every frame. The shape sequences is trained by a standard learning tool, HMM. The overall system architecture is illustrated in Figure 1.

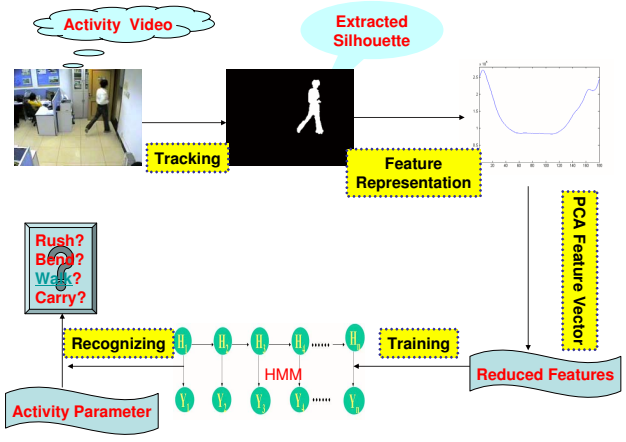

Fig. 1. The flowchart of abnormality activity recognition based on ℜ Transform.

This paper is organized as follows: basic theory of ℜ transform is introduced in Section 2. Section 3 defines the abnormal activity, and the effectiveness of the proposed method is proved with experiments in Section 4. Finally Section 5 are conclusions and future research directions.

2. FEATURE EXTRACTION BY $\Re$ TRANSFORM

The goal of shape transform is to use the fewest possible measures to characterize the object adequately so that it may represent the original information unambiguously. Radon transform of an image $f(x, y)$ is defined by [11]:

$$T_{R^f}(\rho, \theta) = \int_{-\infty}^{\infty} f(x, y)\delta(x\cos\theta + y\sin\theta - \rho)dxdy \quad (1)$$

where $\delta(.)$ is the Dirac delta-function, $\theta \in [0, \pi]$ and $\rho \in [-\infty, \infty]$. However, Radon transform is sensitive to the operation of scaling, translation and rotation. and hence an improved representation, called $\Re$ Transform, is introduced [12]:

$$\Re_f(\theta) = \int_{-\infty}^{\infty} T^2_{\Re^f}(\rho, \theta)d\rho \quad (2)$$

The $\Re$ transform has several useful properties. some of them are relevant of shape representation [12, 13]:

$$\frac{1}{\alpha^2}\int_{-\infty}^{\infty} T^2_{\Re^f}(\alpha\rho, \theta)d\rho = \frac{1}{\alpha^3}\Re_f(\theta) \qquad [scaling](3)$$

$$\int_{-\infty}^{\infty} T^2_{\Re^f}((\rho - x_0\cos(\theta) - y_0\sin(\theta)), \theta)d\rho = \Re_f(\theta) \qquad [translation](4)$$

$$\int_{-\infty}^{\infty} T^2_{\Re^f}(\rho, (\theta + \theta_0))d\rho = \Re_f(\theta + \theta_0) \qquad [rotation](5)$$

Note that only rotation modifies the function, and $\Re$ transform is invariant under translation and scaling if we resize the image into a normalized scale. Figure 2 shows two $\Re$ transform results of the same object after geometric transformations. In this manner, only the rotation leads to a 25° phase shift of $\Re$ function. The information that the initial silhouette sequences carry is transformed in a compact way preserving important description of the activities. Finally, a feature vector with 180 dimensions, instead of the 2D shape matrix is extracted.

The computation of shape descriptor is linear. The processing time in the case of 2D $\Re$ transform is about 0.047s for a shape of 320 $\times$240 pixels. The results are obtained by MATLAB on a Pentium 4, 3.2 GHz running under windows XP.

3. ABNORMAL ACTIVITY LEARNING AND RECOGNITION

In office, the normal activity is defined as walking into and out of the room in normal speed. Rushing, walking out with a bag or suddenly bending down when walking are defined as abnormal activities.

The proposed feature descriptor just contains the spatial information about the pose of the human body. The dynamic information, to be specific, the human postures varying with time characterize the discriminative difference between different activities. Obviously, HMM is appropriate to characterize the variation of activity [14]. Because $\Re$ transform is non-orthogonal, the shape vector of 180 dimensionality is redundant. So PCA is first used to obtain the compact and accurate information in each video sequence. According to primary analysis of each activity, we find that 15 principal component are enough to represent the 98% variance. Then, a HMM model is trained for each activity class. Table1 gives the number of model states and GMMs for each activity in our experiments. Trained HMM models are then used to compute each model's similarity to a new input sequence.

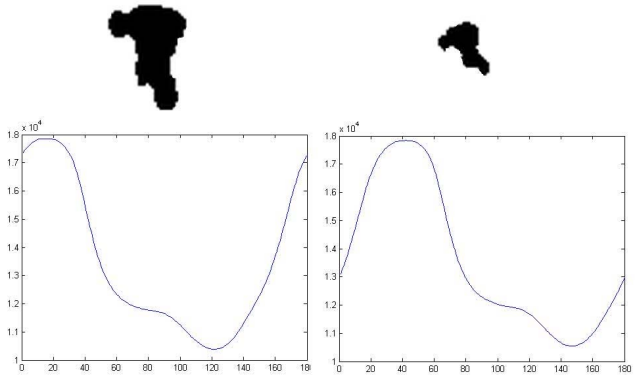

Fig. 2. $\Re$-transforms of the same human bending silhouette which has been scaled, translated and rotated by 25°.

Table 1. The number of HMM states and GMMs.

	Rush	Carry	Bend	Walk
States	2	2	3	2
GMMs	1	1	2	1

4. EXPERIMENTAL ANALYSIS

Our recognition system uses a stationary camera which works in an office environment. The experiments are based on 150 low resolution video sequences (320$\times$240, 25fps) of thirty different people, each performing four natural activities including "rushing", "carrying a bag", "suddenly bending down when walking" and "walking normally". Half of them are used in learning while the others are used for recognition.

All of these videos begin with the moving object entering the monitor domain and end in leaving the camera view. The median background of each video sequence is subtracted and noise is removed with a median filter by a 3$\times$ 3 template. A predetermined threshold is used to obtain the binary images. Figure 3 shows some examples of extracted silhouettes and contours for each activity sequences. In experiments, we normalize moving direction based on body symmetry in order to

remove the influence of rotation.

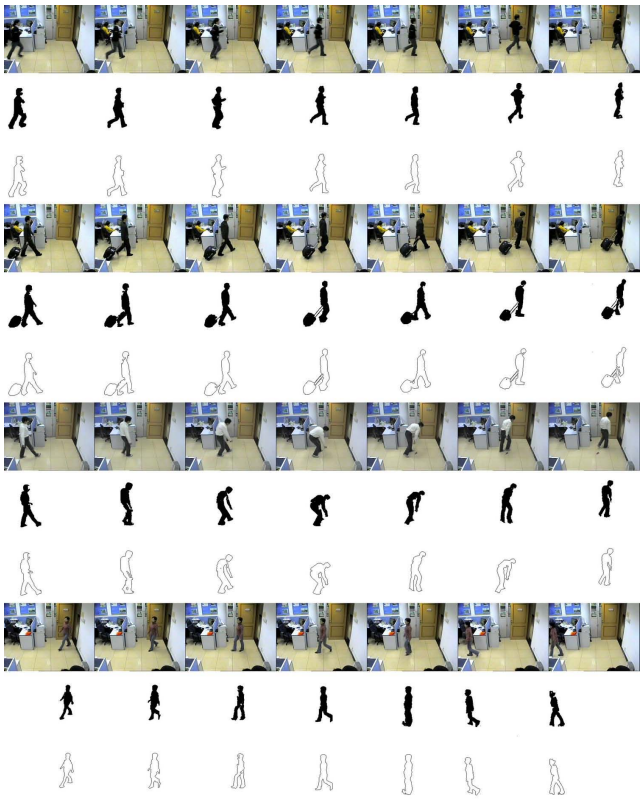

Fig. 3. Example of video sequences and extracted silhouettes and contours.

Figure 4 shows the key frame silhouette of different activities, and their respective $\Re$ transform results. The transformation results in the second row shows the difference of each activity. Moreover, the difference will keep for some time in the activity sequence, as shown in the last row in Figure 4. This figure illustrates that $\Re$ transform could represent the characteristics of different activities:

(Fig. 4.a and 4.d). Referring to the "Rush" and "Walk" activities, we can see that they have similar shape transform representations. Both of them are smooth, however, the amplitude of "Rush" is always higher than that of "Walk".

(Fig. 4.b). Compared with "Rush" and "Walk", the $\Re$ transform curve of "Carry" fluctuates in the range from $140°$ to $170°$. Obviously, this is caused by the bag.

(Fig. 4.c). We note that the shape representation of "Bend" varies slightly. And also, it has a peak close to $160°$.

This figure shows that $\Re$ transform can describe the spatial information sufficiently and characterize the different activity shape effectively. Table 2 shows the recognition results with trained HMMs. We can see that "Carry" and "Bend" get higher recognition rates. Because of the similarities between "Rush" and "Walk", misclassifications occur (15% "Rush" activities are recognized as "Walk", and 10% "Walk" as "Rush").

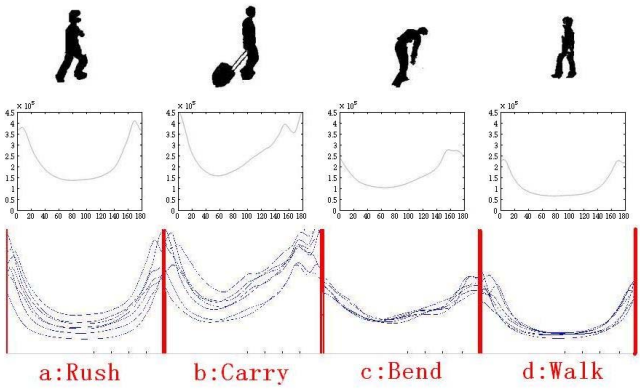

Fig. 4. Key frame and their respective R transform curve for different office activity.

Table 2. Recognition Results with HMMs.

	Rush	Carry	Bend	Walk
Rush	85%	0	0	15%
Carry	0	95%	5%	0
Bend	0	5%	90%	5%
Walk	10%	0	0	90%

Generally speaking, low level features based on $\Re$ transform are effective for recognizing abnormal activity.

4.1. Comparison with contour based feature

To demonstrate the superiority of the spatial information based on $\Re$ transform collected from each frame of a surveillance video, a comparison experiment is conducted with contour-based features. In our experiments, 512 points are sampled to represent the outer contour of each object using a border following algorithm based on connectivity. However, the completed contour is divided into several parts because of noise, occlusion and color similarity with background. The biggest part is selected as the required contour after erosion and dilatation with rectangle template (10×4) like a human body.

Each point on the contour can be represented by $z_k = x_k + iy_k, (k = 1, \ldots, 512)$. Each contour is expressed by a complex vector $[z_1, z_2 \ldots, z_{512}]$, where $z_1 \ldots z_{512}$ are points of activity contour unwrapped in counterclockwise order from the top point. Then Fourier descriptors, a sequence of complex coefficients of Fourier transform for contour vector, represent the shape of an object in the frequency domain where the lower frequencies symbolize the general contour, and the higher frequencies represent the details of the contour. In our experiments, the coefficients from 1 to 15 are selected as our features. The HMM structure of this method is same with that of $\Re$-transform based method as shown in Table 1.

From Figure 6, it should be noted that $\Re$-transform method generally performs better than the Fourier-based method. This

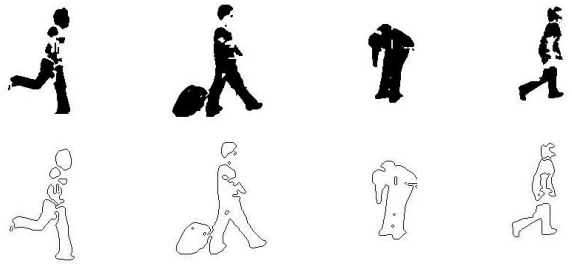

Fig. 5. Examples of noisy data in the dataset.

is because the contour-based shape descriptors make use of only the boundary information, ignoring interior content of a shape. However, an object is often divided into several disconnected parts or contains holes, as shown in Figure 5. This is due to imperfect subtraction caused by color similarities with the background and self-occlusion. Therefore the spatial information carried by the Fourier descriptors could not represent original shapes accurately, while $\Re$ transform is capable of capturing the intrinsic characteristics of the shapes. In summary, as a region-based feature description method, $\Re$ transform is robust to incomplete shape and distortion of body.

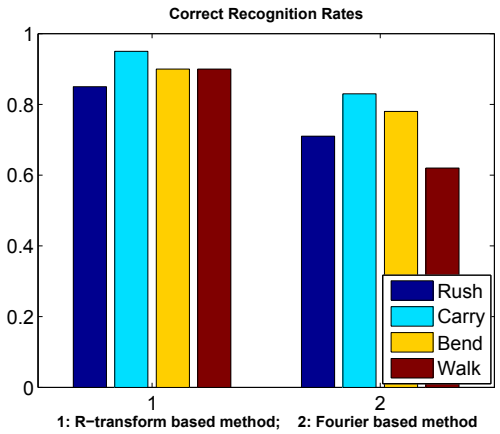

Fig. 6. Correct Recognition Rate between two methods.

5. CONCLUSION

In this paper, we use $\Re$ transform to represent the activity in each frame as a shape descriptor and employ HMM to model the variational trend with time.

Our approach has several advantages. First, it does not require video alignment and is applicable in many scenarios where the background is known, because $\Re$ transform is invariant to scale and translation. Second, our shape descriptor capture both boundary and internal content of the shape. For this reason, they are more robust to noise, internal holes and separated shape. Third, the computation of shape descriptor is linear. So the time cost of 2D $\Re$ transform is low. Finally, this approach can also be applied to other tasks such as gait recognition, content-based image retrieval and face animation.

Acknowledgment

The work reported in this paper was funded by research grants from the National Basic Research Program of China (No. 2004CB318110), the National Natural Science Foundation of China (No. 60605014, No. 60335010 and No. 2004DFA06900) and CASIA Innovation Fund for Young Scientists.

6. REFERENCES

[1] Weiming Hu, Tieniu Tan, Liang Wang, and Steve Maybank, "A survey on visual surveillance of object motion and behaviors", *IEEE Transactions on Systems, Man and Cybernetics, Part C: Applications and Reviews*, Vol. 34, No. 3, pp. 334-352, 2004.

[2] Rota M, Thonnat M, "Video Sequence Interpretation for Visual Surveillance", *IEEE Proceedings of Visual Surveillance*, pp: 10-20, 2000.

[3] Christian Bauckhage, John K. Tsotsos, Frank E. Bunn, "Detecting Abnormal Gait", *Proceedingsof the 2nd Canadian Conference on Computer and Robot Vision*, pp: 282-288, 2005.

[4] Gao, J., Wactlar, H., Hauptmann, A., Bharucha, A., " Dining Activity Analysis Using a Hidden Markov Model", *Computer Vision and Image Understanding*, Vol. 2, pp: 915- 918, 2004.

[5] Chuang C.-H. and Kuo C.-C.J., " Wavelet Descriptor of Planar Curves: Theory and Applications", *IEEE Trans. on Image Processing*, Vol. 5(1), pp. 56-70,1996.

[6] D. Zhang, G. Lu, " Shape-based image retrieval using generic Fourier descriptor", *Signal Process.: Image Commun.*, Vol. 17, pp. 825-848, 2002.

[7] V.F. Leavers, " Shape Detection in Computer Vision Using the Hough Transform', *Springer-Verlag*, 1992.

[8] D. Zhang, G. Lu, "Study and evaluation of different Fourier methods for image retrieval", *Image Vis. Comput.*, Vol. 23 , pp. 33-49, 2005.

[9] C.H. Teh, R.T. Chin, "On image analysis by the methods of moments", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.10, pp. 496-512, 1988.

[10] R.J. Prokop, A.P. Reeves, "A survey of moment-based techniques for unoccluded object representation and recognition", *CVGIP: Graph. Model. Image Process*, Vol. 54 (5) pp. 438-460, 1992.

[11] S.R. Deans, " Applications of the Radon Transform", *Wiley Interscience Publications, New York*, 1983.

[12] S. Tabbone, L. Wendling, J.-P. Salmon, "A new shape descriptor defined on the Radon transform", *Computer Vision and Image Understanding*, Vol. 102, 2006.

[13] Y. Wang, K. Huang and T. Tan, "Human Activity Recognition based on $\Re$ Transform", *The 7th IEEE International Workshop on Visual Surveillance*, 2007.

[14] Yamato, J., Ohya, J., Ishii, K., "Recognizing Human Action in Time Sequential Images Using a Hidden Markov Model", *CVPR*, 1992.