Human Activity Recognition Based on ℜ Transform

Ying Wang Kaiqi Huang Tieniu Tan
National Laboratory of Pattern Recognition,
Institute of Automation, Chinese Academy of Sciences, Beijing, P.R.China
{wangying, kqhuang, tnt}@nlpr.ia.ac.cn

Abstract

This paper addresses human activity recognition based on a new feature descriptor. For a binary human silhouette, an extended radon transform, ℜ transform, is employed to represent low-level features. The advantage of the ℜ transform lies in its low computational complexity and geometric invariance. Then a set of HMMs based on the extracted features are trained to recognize activities. Compared with other commonly-used feature descriptors, ℜ transform is robust to frame loss in video, disjoint silhouettes and holes in the shape, and thus achieves better performance in recognizing similar activities. Rich experiments have proved the efficiency of the proposed method.

1. Introduction

Recognizing human activities from videos is a hot topic of research in computer vision [1, 2], which has a wide range of applications such as intelligent surveillance, analysis of the physical condition of people and caring of aged people [3, 4, 5]. In general, human behavior analysis includes moving object tracking, low-level dynamic information extraction and representation, activity model learning and high-level semantic understanding.

Shape-based features are commonly used in activity recognition because they can be extracted robustly from videos, and they are robust to appearance variations such as color and texture. Two types of shape-based features are used, i.e. silhouette and contour. The silhouette method takes into account all the pixels within a shape, and the contour method only extracts the boundary of a shape. Feature description is a key bridge between low level image feature and high level activity understanding [6, 7]. General contour-based descriptors include wavelets, Fourier descriptors and Hough transform [8, 9, 10]. Since contour descriptors are based on the boundary of a shape, they cannot capture the internal structure information. Consequently, they are limited to certain applications. Common silhouette-based shape descriptors include invariant moment, Zernike moment and wavelet moment[11, 12, 13]. The moments are computationally intensive and sensitive for disjoint shapes or shapes with noise where the silhouette information is not correct. In surveillance, shapes with noise are common because of complex background, and the size of moving object varies with its distance to camera. Therefore we need features invariant to geometry transformation and robust to noise, which is in accordance with the performance of ℜ transform [14].

ℜ transform, a new feature representation, has low computational cost and is effective to recognize similar activity even in the case of disjoint silhouette, silhouette with holes or frame loss data. Moreover, rich experiments prove that it outperforms common shape descriptors in activity sequence recognition.

In this paper, activity models based on features extracted by ℜ transform are trained by a standard learning tool, HMM. The overall system architecture is illustrated in Figure 1.

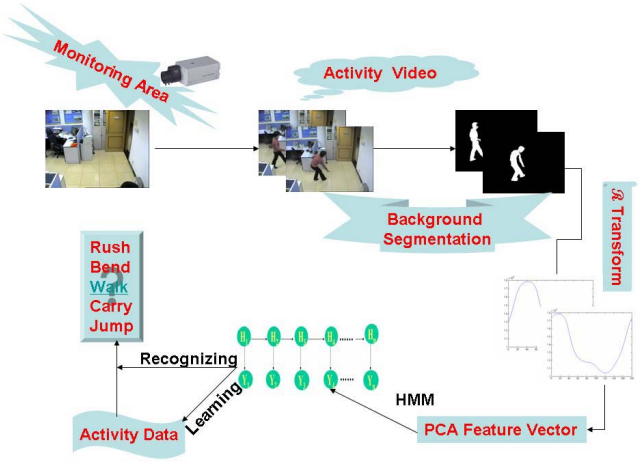

Figure 1. The flowchart of activity recognition based on ℜ Transform.

This paper is organized as follows. Basic theory of ℜ transform is introduced in Section 2. Section 3 demonstrates the effectiveness of the proposed method by com-

1-4244-1180-7/07/$25.00 ©2007 IEEE

parison with other feature descriptors. Finally Section 4 are conclusions and future research directions.

2. Low level feature extraction

2.1. Radon transform

In mathematics, two dimensional Radon transform is the transform consisting of the integral of a function over the set of lines in all directions, which is roughly equivalent to finding the projection of a shape on any given line. For a discrete binary image, each image point is projected to a Radon matrix. Let $f(x, y)$ be an image, its Radon transform is defined by [15]:

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

where $\theta \in [0,\pi]$, $\rho \in [-\infty,\infty]$ and $\delta(.)$ is the Dirac delta-function,

$$\delta(x) = \begin{cases} 1 & \text{if } x = 0 \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

For geometry transformation such as scaling, translation and rotation, Radon transform has the following properties:

For a scaling factor α,

$$R_{adon}\{f(\frac{x}{\alpha},\frac{y}{\alpha})\} = \frac{1}{\alpha}T_{Rf}(\alpha\rho,\theta) \quad [scaling](3)$$

For translation of vector $\overrightarrow{\mu} = (x_0, y_0)$,

$$R_{adon}\{f(x-x_0,y-y_0)\} = T_{Rf}(\rho - x_0\cos\theta - y_0\sin\theta,\theta) \quad [translation](4)$$

For rotation of θ_0

$$R_{adon}\{f_{\theta_0}(x,y)\} = T_{Rf}(\rho,\theta+\theta_0) \quad [rotation](5)$$

From equations (3)-(5), one can see that Radon transform is sensitive to scaling, translation and rotation. Some adaptations are proposed to solve these problems.

2.2. $\Re$ Transform

An improved representation of Radon transform, $\Re$ Transform, is introduced [14]:

$$\Re_f(\theta) = \int_{-\infty}^{\infty} T^2_{Rf}(\rho,\theta)d\rho \quad (6)$$

$\Re$ transform has several useful properties. Some of them are relevant to shape representation [14]:

For a scaling factor α,

$$\frac{1}{\alpha^2}\int_{-\infty}^{\infty} T^2_{Rf}(\alpha\rho,\theta)d\rho = \frac{1}{\alpha^3}\int_{-\infty}^{\infty} T^2_{Rf}(\nu,\theta)d\nu = \frac{1}{\alpha^3}\Re_f(\theta) \quad [scaling](7)$$

For translation of vector $\overrightarrow{\mu} = (x_0, y_0)$,

$$\int_{-\infty}^{\infty} T^2_{Rf}((\rho - x_0\cos(\theta) - y_0\sin(\theta)),\theta)d\rho$$
$$= \int_{-\infty}^{\infty} T^2_{Rf}(\nu,\theta)d\nu = \Re_f(\theta) \quad [translation](8)$$

For rotation of θ_0

$$\int_{-\infty}^{\infty} T^2_{Rf}(\rho,(\theta+\theta_0))d\rho = \Re_f(\theta+\theta_0) \quad [rotation](9)$$

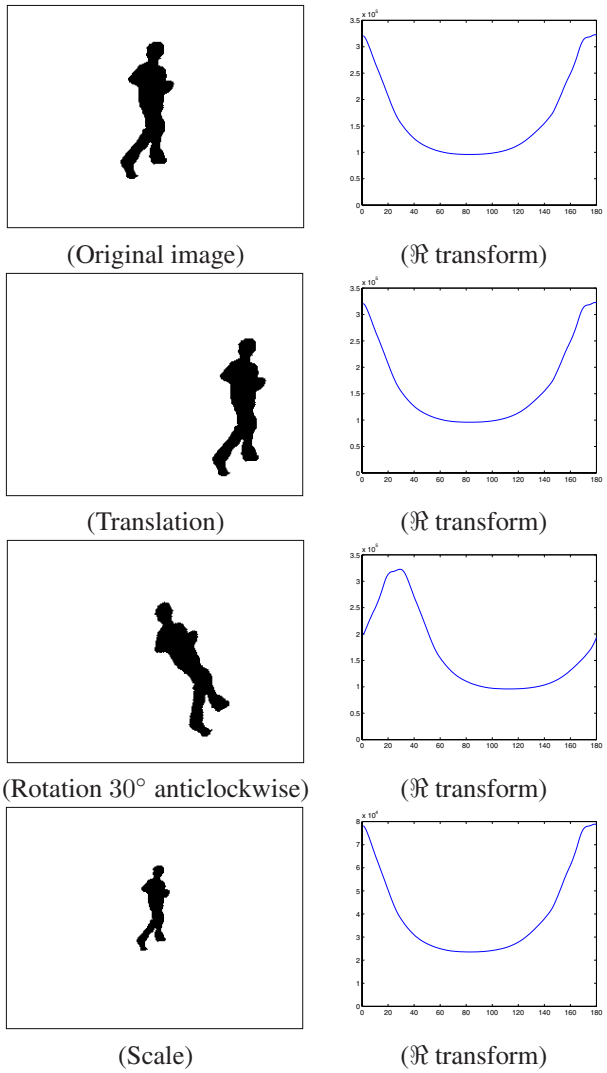

Figure 2. $\Re$-transforms of the same human rushing silhouette which has been scaled, translated and rotated by 30°.

Figure 2 shows the $\Re$ transform of the same object after geometry transformation.

1. Translation in the plane does not change the $\Re$ transform.

2. A rotation of θ_0 in the original image leads to the phase shift of θ_0 in $\Re$ transform. This rarely happens in human activities.

3. Scaling the original image would not change the shape of the $\Re$ transform, but the amplitude would change accordingly.

Therefore, $\Re$ transform is invariant under translation and scaling if we resize the image into a normalized scale, which is feasible in activity recognition. Compared with Radon transform, $\Re$ transform is robust to geometry transformation as shown in Table 1. For each frame in the video, a 180-dimensional feature vector, instead of the 2D Radon matrix, is extracted to represent the motion of the human body. The information that the initial silhouette sequences carry is transformed in a more compact way and invariant to geometry transformation.

Table 1. The difference between Radon transform and $\Re$ transform.

	Radon Transform	$\Re$ Transform
Scaling α	A scaling of both ρ and the transform amplitude.	Only an amplitude scaling of the $\Re$ transform
Translation x_0, y_0	A shift of its transform in ρ	Invariant
Rotation θ_0	A shift of the transform in θ_0	A shift of the transform in θ_0

3. Experimental analysis

3.1. $\Re$ transform of different activities

Our recognition system uses a stationary camera which works in an office environment. The experiments are based on 150 low resolution video sequences (320×240, 25fps) of thirty different people, each performing five natural activities including rushing, carrying a bag, suddenly bending down when walking , walking normally and jumping. One hundred of them are used in learning while the others 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× 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 as shown in the fourth part of Figure 3.

Figure 4 shows the key frame silhouettes of different activities, and their respective $\Re$ transform. The figures in the second column shows the difference of each activity. Moreover, the difference will keep for several continuous frames

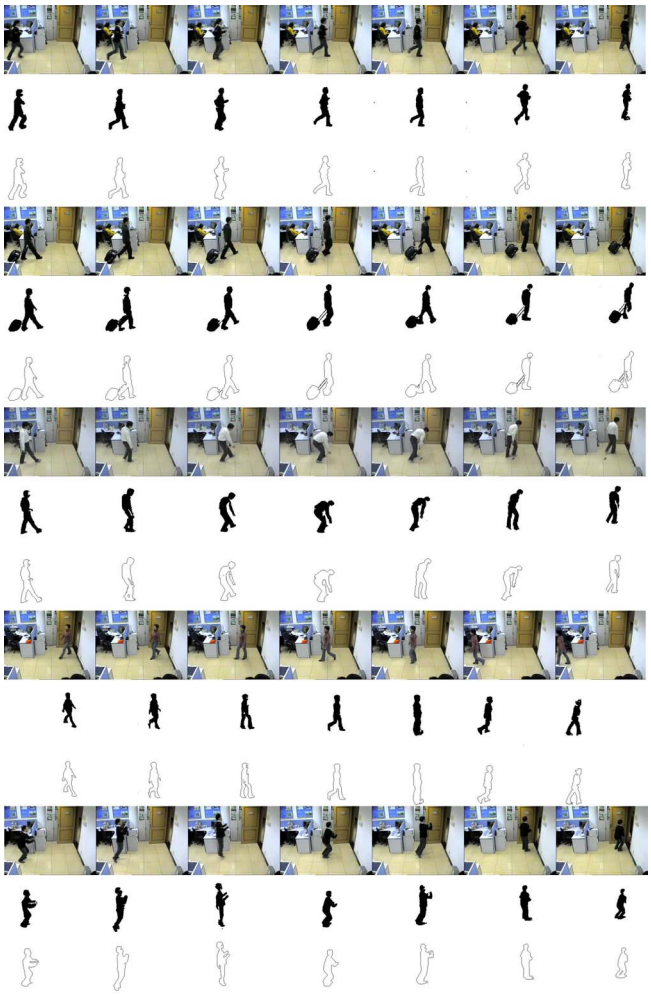

Figure 3. Examples of video sequences and extracted silhouettes and contours.

in the activity sequence, as shown in the last column in Figure 4. This figure illustrates that $\Re$ transform could represent the characteristics of different activities:

1. Compared with other activities, the $\Re$ transform curve of carrying a bag fluctuates in the range from $140°$ to $170°$. Obviously, this is caused by the bag.

2. The shape representation of bending down varies slightly. And also, it has a peak close to $160°$.

3. Rushing, walking and jumping activities have similar shape transform representations. Both rushing and jumping have fluctuations while walking is smooth. This is caused by the motion of limbs. There is intense movement in rushing, so it has two peaks while jumping has one peak and walking in general has no peak. However, the amplitude of rushing is always higher than that of walking.

This figure shows that $\Re$ transform can describe the spatial information sufficiently and characterize the different activity shape effectively.

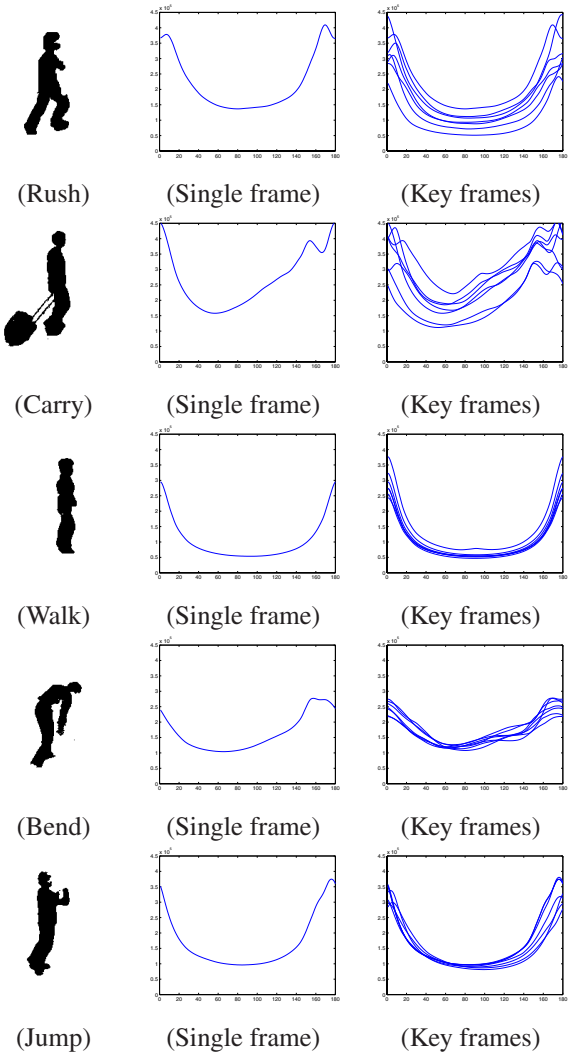

Figure 4. $\Re$ transform of single frame and key frames for different activities.

3.2. Experimental data

The resultant silhouettes contain holes and intrusions due to imperfect subtraction, shadows and color similarities with the background. To train the activity models, holes, shadows and other noise are removed manually. The fine-tuned data are taken as ground truth data.

The raw data include such cases as disjoint silhouettes, silhouettes with holes and silhouettes with missing parts. Compared with the ground truth data, they are incomplete data.

Shadow and other noises may add an extra part to the human silhouette, and thus result in redundant data.

The incomplete data and redundant data are of low quality, and thus they are used for testing the performance of the $\Re$ transform. Figure 5 shows some such examples.

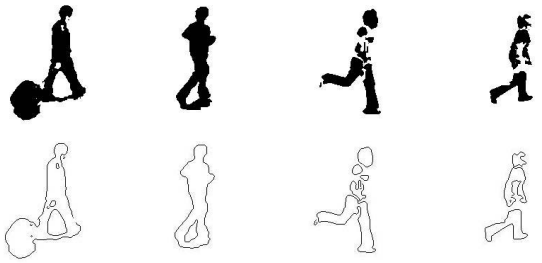

Figure 5. Examples of noisy data in the dataset.

In order to test the robustness of $\Re$ transform , we remove the first 10 frames, the middle 15 frames and the last 15 frames of each testing sequence. These artificially generated data are defined as frame loss data.

In experiments, 100 sequences of ground truth data are used for training. For each data category, 50 sequences are used for testing.

Figure 6 shows the $\Re$ transform of the rushing shape in different data case. For the cases of incomplete data and ground truth data, the $\Re$ transform is similar, but the transform of redundant data varies significantly in the peak of the curve. In fact, $\Re$ transform is sensitive to this type of redundant data which will have negative effect on activity recognition.

3.3. Activity learning and recognition

The proposed feature descriptor contains the spatial information about the pose of the human body. The dynamic information, specifically, the human postures varying with time characterize the difference between different activities. HMM is appropriate to characterize the variation of activity [16], which is trained for each activity class. The number of model states and GMMs are selected according to experience (Table 2). Trained HMM models are then used to compute each model's similarity to a new input sequence.

Table 2. The number of HMM states and GMMs.

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

Because $\Re$ transform is non-orthogonal, the shape vector of 180 dimensionality is redundant. In general, PCA is used to obtain the compact and accurate information in each video sequence. According to primary analysis of each activity, we find that 15 principal components are enough to represent the 98% variance. However, in experiments, the performance is not satisfactory. So we try other methods to reduce the dimensionality of feature vector for each frame.

For each frame in activity sequence, the shape vector of 180-D is divided into 3 parts and thus a feature matrix of 60×3 is formed. Then PCA is used to reduce the dimen-

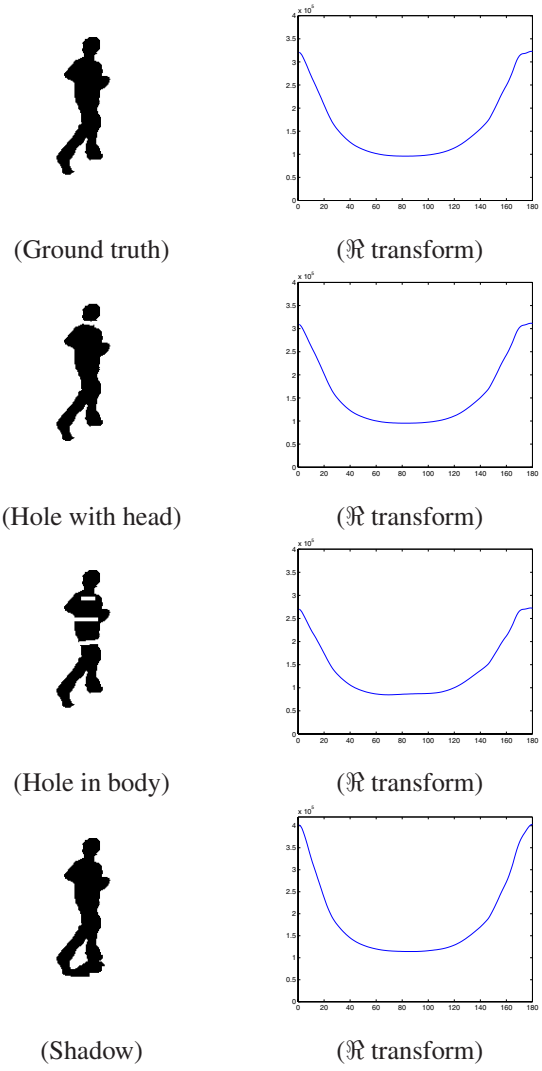

Figure 6. ℜ transform for data of different qualities.

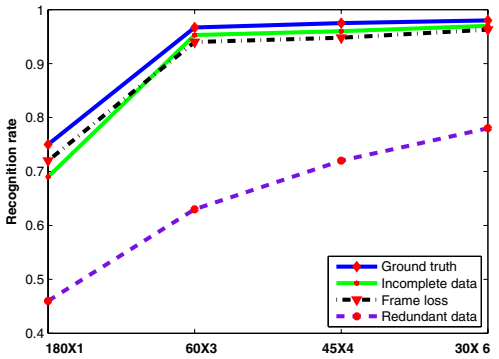

Figure 7. The curve of activity recognition rates with different PCA methods for different data.

sionality of each feature matrix to 2×3. Finally, a vector of 6×1 is concatenated from the 2×3 matrix to train HMM. Similarly, the shape vector is divided to form matrixes of 45×4 and 30×6. After using PCA, the dimension of these matrixes are reduced to 2×3 as well. The recognition rates in different division schemes, which are represent by red marks in each curve, are shown in Figure 7. Higher recognition rates are achieved when the dividing scheme proceeds to 30×6, but with the subtle improvement, the computation cost gets much higher. To balance computation cost and recognition rates, we adopt the division scheme of 3×60 in our HMM parameter learning and testing experiments.

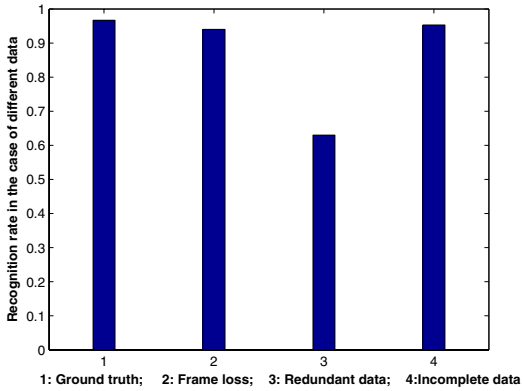

Figure 8. The recognition rates of ℜ transform with different data .

Table 3. Recognition results with different data using ℜ transform.

	Ground truth	Frame loss	Redundant data	Incomplete data
Bend	√	√	3Carry, 1Jump	√
Carry	√	1Bend	2Rush	√
Jump	1Walk, 1Rush	1Rush, 2Walk	2Rush, 2Walk	2Rush, 1Walk
Rush	√	√	1Jump, 1Walk	√
Walk	√	√	2Rush, 2Jump	√

Figure 8 illustrates the recognition results with HMMs in accordance with Table 3 (where √ means recognizing all activities correctly, 2Walk means two activities are recognized as "Walk" and so on). Note that there are on 50 sequences in each testing class, which are different with training data. In spite of the high similarities between rushing and walking, misclassifications never occur in the case of frame loss data and incomplete data. Our shape descriptor captures both boundary and internal content of the shape, so they are more robust to noise, such as internal holes and separated shape. While in the case of silhouette with shadow,

the performance of $\Re$ transform is slightly worse than other cases. This shows that $\Re$ transform is suitable for the background segmentation methods with low false positive rate but keeping some false negative rate [17]. Generally speaking, low level features based on $\Re$ transform are effective for recognizing similar activity even in the case of noisy data.

3.4. Comparison with contour-based Fourier descriptor

To demonstrate the superiority of the feature representation ability based on $\Re$ transform extracted 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 may be divided into several parts because of noise, occlusion and color similarity with background. Here the major part is selected as the required contour after erosion and dilatation with rectangle template (10×4).

Each point on the contour can be represented by $z_k = x_k + iy_k, (k = 1, \ldots, 512)$. We unwrap each moving contour counterclockwise from the top point, and use point sequence to express each contour as a complex vector $[z_1, z_2 \ldots, z_{512}]$. 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 low frequencies symbolize the general contour, and the high frequencies represent the details of the contour. In our experiments, the coefficients from 1 to 15 are selected as features. HMMs with the same structure are also used in activity training and recognition.

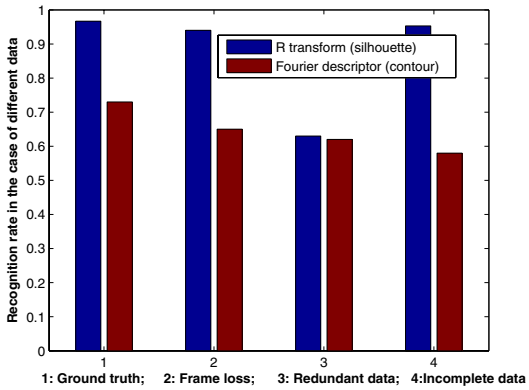

Figure 9. Recognition rate of $\Re$ transform and Fourier descriptor.

From Figure 9, $\Re$-transform method generally performs better than the Fourier-based method. In experiments, an object could be several disconnected parts or contains holes. Because the Fourier descriptors could not extract the interior information in the contour, the spatial information could not represent the original shapes accurately, while $\Re$ transform takes all the information within the shape into account, so it is capable of capturing the intrinsic characteristics of the shapes.

3.5. Comparison with silhouette-based moment descriptor

Moments are used to describe silhouette, which have many desirable properties such as rotation invariance and multi-level representation. Moreover, different moments have their own advantages; Zernike Moment (ZM) and pseudo-Zernike (p-ZM) are robust to noise, Wavelet moment (WM) has efficient feature extraction ability, and Invariant moment (IM) has low computation cost. In this paper, the feature expression capability, robustness and computation cost of these moments are compared with $\Re$ transform.

3.5.1 Invariant Moments

A set of 7 invariant moments (IM) which are invariant to rotation, scaling and translation are given by Hu in 1961 [18]. Four of them are used in our activity recognition.

$$\phi_1 = \eta_{20} + \eta_{02} \tag{10}$$

$$\phi_2 = (\eta_{20} - \eta_{02})^2 + 4\eta_{11}^2 \tag{11}$$

$$\phi_3 = (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2 \tag{12}$$

$$\phi_4 = (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2 \tag{13}$$

where $\eta_{pq} = \frac{\mu_{pq}}{\mu_{00}^{\gamma}}$, μ_{pq} is central moment of a digital image $f(x, y)$ and $\gamma = 1 + \frac{(p+q)}{2}$ for $p + q = 2, 3...$ [18].

3.5.2 Zernike Moments

The Zernike function of order p with repetition q is defined in the polar coordinate system (r, θ) as

$$V_{n,m}(x, y) = V_{n,m}(r\cos\theta, r\sin\theta) = R_{n,m}(r)e^{jm\theta}, x^2+y^2 \leq 1 \tag{14}$$

where $R_{n,m}(r)$ is the orthogonal radial polynomials, which are orthogonal over the interior of the unit circle $x^2 + y^2 = 1$.

$$R_{n,m}(r) = \sum_{s=0}^{\frac{(n-|m|)}{2}} (-1)^s \frac{(n-s)!}{s! \times (\frac{n-2s+|m|}{2})!(\frac{n-2s-|m|}{2})!} r^{n-2s} \tag{15}$$

For a digital image $f(x, y)$ in the polar coordinate, ZM is given by [19]:

$$Z_{nm} = \frac{n+1}{\pi} \sum_r \sum_\theta f(r, \theta) V_{n,m}^*(r, \theta) \ \triangle r \triangle \theta \tag{16}$$

where $n = 0, 1, 2...; 0 \leq |m| \leq n$; and $n - |m|$ is even.

3.5.3 Pseudo-Zernike Moments

Pseudo-Zernike moment differs from Zernike moment in that the real-valued radial polynomial are defined as [20]:

$$R_{n,m}(r) = \sum_{s=0}^{n-|m|} (-1)^s \frac{(2n+1-s)!}{s! \times (n-s+|m|)!(n-s-|m|)!} r^{n-s} \quad (17)$$

where $n = 0, 1, 2...; 0 \leq |m| \leq n$.

3.5.4 Wavelet Moments

According to [21], wavelet transform is particularly suited for extracting local discriminative features. Here we use the cubic B-spline as mother wavelet:

$$\varphi(r) = \frac{4\alpha^{n+1}}{\sqrt{2\pi(n+1)}} \sigma_\omega \times \cos(2\pi f_0(2r-1)) exp(-\frac{(2r-1)^2}{2\sigma_\omega^2(n+1)}) \quad (18)$$

where $n = 3, \alpha = 0.697066, f_0 = 0.409177$, and $\sigma_\omega^2 = 0.561145$. The cubic B-spline wavelet moments for a digital image $f(x, y)$ in the polar coordinate can be defined as [22]:

$$W_{m,n,q} = \sum_r \sum_\theta f(r, \theta) \varphi_{m,n}(r) e^{-iq\theta} \triangle r \triangle \theta \quad (19)$$

where

$$\varphi_{m,n}(r) = 2^{\frac{m}{2}} \varphi(2^m r - 0.5n) \quad (20)$$

$m = 0, 1, 2, 3, ..., n = 0, 1, ...2^{m+1}$ and $q = 0, 1, 2, 3$

It is easy to prove that Hu Moment is invariant to scale, translation and rotation, and other moments are rotation invariance. Compared with $\Re$ transform, silhouette normalization is needed when using the moments in activity recognition.

Table 4. Time cost for different silhouette-based methods.

	IM	ZM	p-ZM	WM	$\Re$
TC	0.031s	0.078s	0.109s	0.234s	0.047s

For an image of 320 × 240 pixels, Table 4 lists the computing time (TC) of each moment and $\Re$ transform. The results are obtained by MATLAB on a Pentium 4, 3.2 GHz running under window XP. Obviously, WM has the highest computational cost while IM has the lowest TC. The processing time in the case of 2D $\Re$ transform is about 0.047s, just higher than IM.

The first 5 columns of Figure 10 demonstrate that the recognition rate of WM is superior to those of other descriptors in the case of ground truth data. However, IM and $\Re$ transform obtain good recognition rates just slightly lower than WM. We may explain this by Table 5. The results show that WM can discriminant similar activity with subtle differences such as walking and rushing while other moments could not. However, $\Re$ transform obtains similar

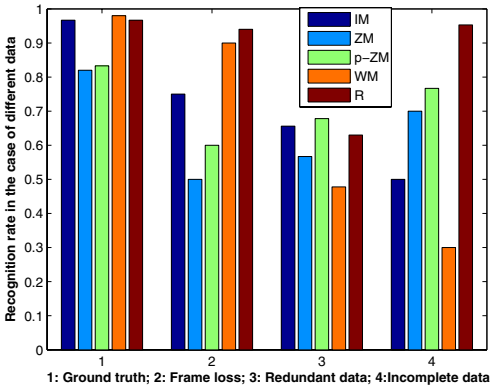

Figure 10. The recognition rates for moments and $\Re$ transform.

Table 5. Recognition results with ground truth.

Activity	IM	ZM	p-ZM	WM	$\Re$
Bend	√	√	√	√	√
Carry	√	2Walk	1Walk	√	1Jump
Jump	1Bend	1Rush, 1Walk	2Rush	1Walk	1Walk
Rush	1Walk	2Walk, 1Carry	3Walk	√	√
Walk	√	1Rush, 1Carry	1Rush, 1Carry	√	√

Table 6. Recognition results with frame loss data.

Data	IM	ZM	p-ZM	WM	$\Re$
First	65%	48%	55.6%	77.8%	93%
Middle	72.7%	52.8%	47.2%	75%	94%
Last	78%	55%	60%	90%	95%

results.

Table 6 illustrates the recognition rates of frame loss data. The lost frames are at different positions of the whole sequence. Compared with the moments, $\Re$ transform achieves the highest recognition rates. From Table 6, the performance of each moment with the data loss of the last 15 frames are better than those with the data loss of the first 10 frames. This demonstrates that the influence of first frames of activity sequence is more significant than that of last frames. However, for the tested cases of frame loss in activity sequence, $\Re$ transform preserves relatively high recognition rates. This proves that $\Re$ transform is more robust to frame loss than the moments.

The last 10 columns of Figure 10 demonstrate the recognition rates with redundant data and incomplete data respectively. For incomplete data, $\Re$ transform is superior to the moments, which recognizes all walking and rushing correctly. For redundant data, recognition rate of each moment and $\Re$ transform decreases significantly, so they are all sensitive to data redundancy. For $\Re$ transform, it is the integral of lines of all directions, and thus shadows of various shapes

will lead to unstable features extracted from the silhouettes, which will deteriorate the recognition performance.

4. Conclusion

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

According to the comparison experiments with common shape descriptors, the method based on $\Re$ transform 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, $\Re$ transform gets the high recognition rate for similar but actually different shape sequences, and even in the case of frame loss data and incomplete data. Third, our shape descriptor captures both boundary and internal content of the shape. For this reason, it is more robust to noise, internal holes and separated shape. Finally, the computation of shape descriptor is linear, so the computation cost of 2D $\Re$ transform is low. Moreover, this approach can also be applied to other tasks such as gait recognition, content-based image retrieval and face animation.

Our future work will focus on evaluation of the feature descriptor on large dataset, and the scalability of the proposed method.

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.

References

[1] Weiming Hu, Tieniu Tan, Liang Wang, and Steve Maybank, "A survey on visual surveillance of object motion and behaviors", *IEEE Trans. on Systems, Man and Cybernetics, Part C: Applications and Reviews*, Vol. 34, pp. 334-352, 2004. 1

[2] Liang Wang, Weiming Hu, and Tieniu Tan, "Recent developments of human motion analysis", *Pattern Recognition*, Vol. 36, No. 3, pp. 585-601, 2003. 1

[3] Rota M, Thonnat M, "Video Sequence Interpretation for Visual Surveillance", *IEEE Proceedings of Visual Surveillance*, pp. 10-20, 2000. 1

[4] Christian Bauckhage, John K. Tsotsos, Frank E. Bunn, "Detecting Abnormal Gait", *The 2nd Canadian Conference on Computer and Robot Vision*, pp. 282-288, 2005. 1

[5] Gao, J., Wactlar, H., Hauptmann, A., Bharucha, A., " Dining Activity Analysis Using a Hidden Markov Model", *Computer Vision and Image Understanding*, Vol. 2, 2004. 1

[6] PLA Filiberto. etc., "Extracting Motion Features for Visual Human Activity Representation", *Iberian conference on pattern recognition and image analysis*, 2005. 1

[7] S. Hongeng, R. Nevatia and F. Bremond, "Video-based event recognition: activity representation and probabilistic recognition methods", *CVIU*, Vol. 96, 2004. 1

[8] 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. 1

[9] D. Zhang, G. Lu, " Shape-based image retrieval using generic Fourier descriptor", *Signal Process.: Image Commun.*, Vol. 17, pp. 825-848, 2002. 1

[10] V.F. Leavers, " Shape Detection in Computer Vision Using the Hough Transform", *Springer-Verlag*, 1992. 1

[11] D. Zhang, G. Lu, "Study and evaluation of different Fourier methods for image retrieval", *Image Vis. Comput.*, Vol. 23 , pp. 33-49, 2005. 1

[12] 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. 1

[13] 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. 1

[14] S. Tabbone, L. Wendling, J.-P. Salmon, "A new shape descriptor defined on the Radon transform", *Computer Vision and Image Understanding*, Vol. 102, 2006. 1, 2

[15] S.R. Deans, " Applications of the Radon Transform", *Wiley Interscience Publications, New York*, 1983. 2

[16] Yamato, J., Ohya, J., Ishii, K., "Recognizing Human Action in Time Sequential Images Using a Hidden Markov Model", *CVPR*, 1992. 4

[17] M Karaman, L Goldmann, TS Da Yu, - Proceedings of SPIE, 2005 , " Comparison of Static Background Segmentation Methods", *Proceedings of SPIE*, 2005. 6

[18] Hu M.K., "Visual Pattern Recognition by Moment Invariants", *IRE Trans. on Info. Theory*, pp. 179-187, 1962. 6

[19] Teague M. R, "Image analysis via the general theory of moments", *Journal of Optimal Society of American*, pp. 920-930, 1980. 6

[20] R Mukundan, K. R. Ramakrishnan, "Moment Functions in Image Analysis: Theory and Applications", *World Scientific Publishing Singapore* 1998. 7

[21] I. Daubechies, "Ten lectures on Wavelets", *CBMS-NSF Regional Conf. Series in Applied Mathematics*, 1992. 7

[22] D. Shen, H.H.S. Ip, "Discriminative wavelet shape descriptors for recognition of 2-D patterns", *Pattern Recognition*, Vol. 32 (2), pp. 151-165, 1999. 7