Automated Classification of Two Persons' Interactive Activities

Hao Zhang*, Zhijing Liu and Haiyong Zhao
School of Computer Science and Technology, Xidian University, Xi'an, P.R.China
Email: zhanghao@mail.xidian.edu.cn, liuzhijing@vip.163.com,
zhaohaiyongym@163.com

***Abstract*—This paper presents a novel classification method for interactive activities, which is represented by $\Re$ transform descriptor and classified by Hidden Markov Models (HMMs). It solves the problem that trades off activity recognition rate and computational complexity, rather than highlight the former exclusively. We extract binary silhouette images after the background model is created. Then the low-level features are described by $\Re$ transform and principal vectors are determined by Principal Component Analysis (PCA). We utilize HMMs to train and classify video sequences, and demonstrate the usability with many sequences. Compared with others, our method is applicable to intelligent surveillance, as its advantage of $\Re$ transform descriptor lying in robustness, computational complexity, geometric invariance and classification performance, and HMM in medium computational cost. So the video surveillance based on these is practicable in (but not limited to) many scenarios where the background is known.**

***Index Terms*—feature extraction, activity classification, $\Re$ transform descriptor, Hidden Markov Models (HMMs), video surveillance**

I. Introduction

Recognizing human activities from videos is a hot topic of research in the computer vision [1], which has a variety of applications, such as behavioral biometrics [2], contend-based video analysis [3], security and surveillance [4,5]. Human behavior analysis includes moving object tracking, low-level dynamic information extraction and representation, activity model learning and high-level semantic understanding.

Building a general activity recognition and classification system is a challenging task, because of variations in the environment, objects and actions. Variations in the environment can be caused by cluttered or moving background, camera motion, occlusion, weather and illumination. Variations in the objects are due to differences in appearance, size or posture of the objects or due to self-motion which is not itself part of the activity. Variations in the action can make it difficult to recognize semantically equivalent actions.

In this paper, we propose a generative method that takes into account silhouette-based shape feature of human motion and Hidden Markov Models (HMMs) in activity recognition. Our aim is to offer a generic solution to human motion categorization via flexible yet highly descriptive Hidden Markov Models. Its advantage is not only helpful in utilizing internal and external information of images, but also easy to distinguish similar shape sequences. Furthermore, it suffers from many factors rarely, i.e. video alignment, noise and images segmenting. Our contribution is that the features of interactive activities are summarized and applied in recognition, and it balances the recognition rate and computational cost. So it is suitable for intelligent surveillance in practicability.

This paper is organized as follows. Section 2 introduces the pioneers' work and summary their advantages and disadvantages, then proposes our method. In Section 3, it describes details of interactive activities outdoors. Followed by Section 4, it introduces $\Re$ transform descriptor and discusses the experiment procedure and results. Finally, we summary the advantages of $\Re$ transform, state the deployment of video surveillance and present the future work in Section 5.

II. Related Works

Using good shape-based features to describe pose and motion has been widely researched in the past few years. As their robustness from videos, they rarely suffer from appearance variations such as color and texture. Generally speaking, there are two popular types of shape-based features, silhouette and contour. The silhouette method takes into account all the pixels within a shape, while the contour only extracts its boundary. Feature description is a key bridge between low level image features and high level activity understanding [6,7]. General contour-based descriptors include wavelets, Fourier descriptors and Hough transform [8-10]. Since contour descriptors are based on the boundary, 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

Manuscript received October 15, 2009; revised December 27, 2009; accepted February 8, 2010.

Based on "Automatic Classification on Human Activities", by Hao Zhang, Zhijing Liu, and Haiyong Zhao, which appeared in the Proceedings of 2009 WASE Global Congress on Science Engineering, Taiyuan, China, December 25-27, 2009. © 2009 WASE.

This research was supported by funds from the Ministry of Education of the People's Republic of China under Grant 2006D90704017.

*corresponding author.

doi:10.4304/jsw.5.8.810-817

[11-13]. The moments are computationally intensive and sensitive to disjoint shapes and its noise where the silhouette information is not accordant. In multi-direction gait recognition [14], the success of Radon descriptor [15] is satisfying, but it is not convenient for the representation of the statistical feature. In surveillance, shapes with noise are common because of the complex background, and the size of moving object varies with the distance. Therefore, what we need is a feature descriptor which highlights invariant to geometry transformation and robust to noise, and $\Re$ transform descriptor [16] is deserved.

In [17], learning algorithms for space-dependent switched dynamical models (SDSDM) recognizes human activities, where switching probabilities should be space dependent in order to be able to represent the interaction between a person and the scene. Though it describes motion tendency globally, it characterizes the person's activities less accurately. A dynamic Bayesian network (DBN) model structure with state duration presented in [18] models and recognizes interacting activities with global and local features. This model, consisting of 3 layers, i.e. image data, object feature and activity recognition, highlights feature analysis and recognition in intelligent surveillance, but it is much more complex than SDSDM. Hierarchical durational-state dynamic Bayesian network (HDS-DBN) presented in [19] contains multiple levels of states and represents multiple scales of motion details. It extracts features in large, middle and small scales, and constructs HDS-DBN to analyze them. Though it has good performance in recognition rate, the computational complexity is much higher.

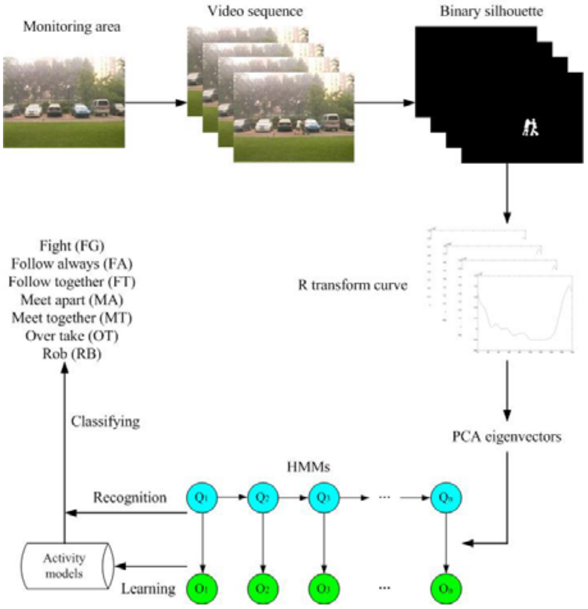


Figure 1. The flowchart of interactive activities recognition.

$\Re$ transform, a new feature representation, has low computational cost and is effective to recognize similar activity even in many cases, i.e. disjoint silhouette, holes and frame loss. There has been some success in activities recognition for single person. Moreover, rich experiments prove that it outperforms common shape descriptors in activity sequence recognition. We adopt $\Re$ transform descriptor to classify interactive activities for 2 persons. Compared with the single, it proposes a more common and practical method for the interaction.

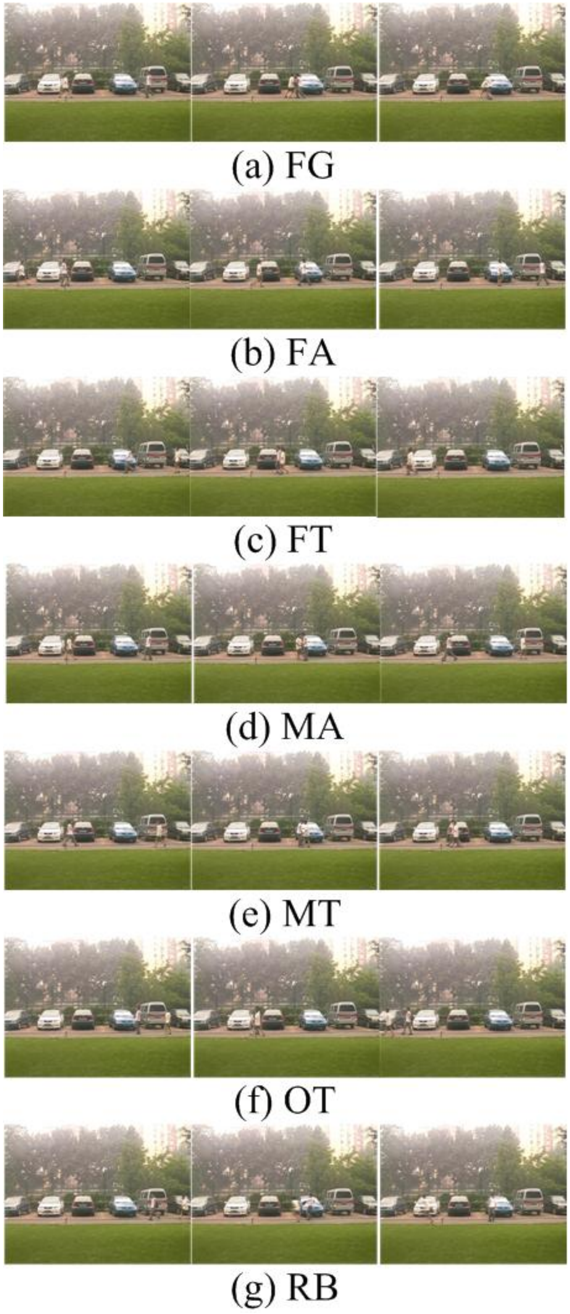
In our method, we input video sequences and extract binary image of silhouettes with background subtract. Then every image is transformed to 180 dimensionalities with $\Re$ descriptor simultaneously. Thus we can obtain primary vectors with PCA before determining parts of them to represent its features as one observation of HMM. So different models with HMM is trained and designed for classification. Therefore, we can feed the test and determine which category they belong to respectively. The overall system architecture is illustrated in Fig. 1. Henceforth, we refer that the training sequences in HMM are termed "gallery" and testing ones are "probe".

III. Activity Description

The test videos are parts in activities database shot by Institute of Automation, Chinese Academy of Sciences (CASIA). In this database, there are two types of activities, including single person and two interactive persons. Each type is screened in three visual angles, i.e. horizontal, vertical and angle. The interactive contains 7 groups of videos per angle, i.e. fight, follow always, follow together, meet apart, meet together, overtake and

TABLE I. The Descriptions of Two Persons' Activities in Videos

No.	Activity term in database	Abbreviation	Activity description
1	Fight	FG	Two persons walking opposite in a straight line and meeting fight each other.
2	Followalways	FA	Two persons at a distance walk in a straight line in the same direction.
3	Followtogether	FT	Two persons at a distance walk in a straight line in the same direction. The follower catches up and walks in stride with the first one.
4	Meetapart	MA	Two persons walk in a straight line in opposite direction. After meeting, they go on walking respectively.
5	Meettogether	MT	Two persons walk opposite in a straight line. After meeting, they walk side by side in the same direction.
6	Overtake	OT	Two persons at a distance walk in a straight line in the same direction, and then the follower overtakes the first one with speed.
7	Rob	RB	Two persons walk in a straight line in the same direction. The first one goes with bag, and the follower robs the bag while running.



(a) FG

(b) FA

(c) FT

(d) MA

(e) MT

(f) OT

(g) RB

Figure 2. Activities description.

rob, as shown in Table 1 and Fig. 2. Every group includes 4 flips (320×240, 25fps) with avi format, which are between 5 and 15 seconds. All flips are shot outdoors by stationary camera, including 14 persons in 7 types of activities. In our system, horizontal interactive flips are used for learning and classifying.

IV. Experiment and Discussion

A. $\Re$ Transform Descriptor

Feature representation is the key step of human activity recognition because it is an abstraction of original data to a compact and reliable format for latter processing. In this paper, we adopt a novel feature descriptor, $\Re$ transform, which is an extended Radon transform [16].

Two dimensional Radon transform is 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 *f*(*x*,*y*), its Radon transform is defined by [20]:

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

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

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

However, Radon transform is sensitive to the operation of scaling, translation and rotation, and hence an improved representation, called $\Re$ Transform, is introduced [16,21]:

$$\Re_f(\theta) = \int_{-\infty}^{\infty} T_{R^f}^2(\rho,\theta)d\rho \qquad (3)$$

$\Re$ transform has several useful properties in shape representation for activity recognition [16,21]:

(1) Translate the image by a vector $\vec{\mu} = (x_0, y_0)$,

$$\begin{aligned} &\int_{-\infty}^{\infty} T_{R^f}^2((\rho - x_0\cos\theta - y_0\sin\theta),\theta)d\rho \\ &= \int_{-\infty}^{\infty} T_{R^f}^2(\nu,\theta)d\rho = \Re_f(\theta) \end{aligned} \qquad (4)$$

(2) Scale the image by a factor *α*,

$$\begin{aligned} &\frac{1}{\alpha^2}\int_{-\infty}^{\infty} T_{R^f}^2(\alpha\rho,\theta)d\rho \\ &= \frac{1}{\alpha^3}\int_{-\infty}^{\infty} T_{R^f}^2(\nu,\theta)d\theta = \frac{1}{\alpha^3}\Re_f(\theta) \end{aligned} \qquad (5)$$

(3) Rotate the image by an angle θ_0,

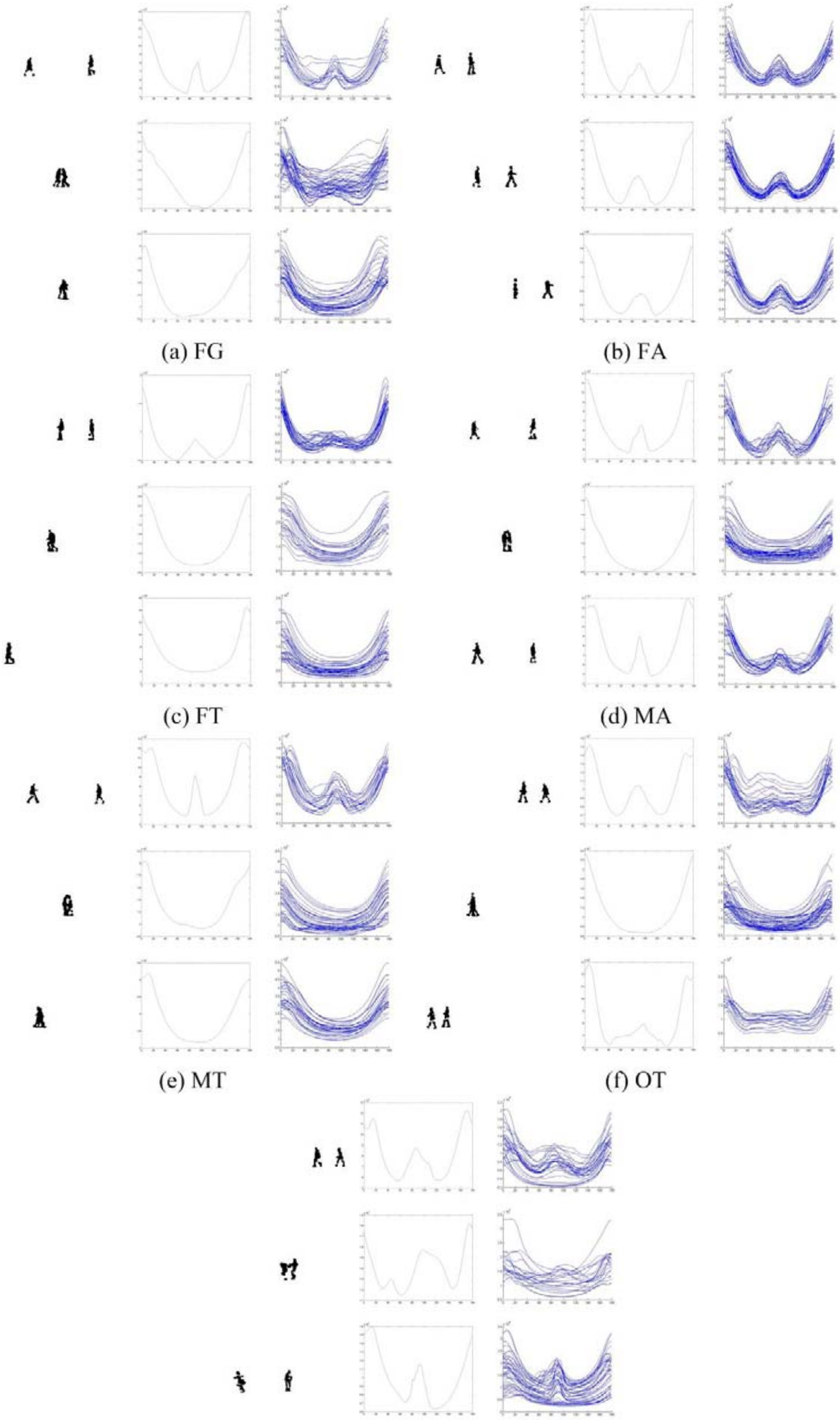
$$\int_{-\infty}^{\infty} T_{R^f}^2(\rho,\theta+\theta_0)d\rho = \Re_f(\theta+\theta_0) \qquad (6)$$

According to the symmetric property of Radon transform, and let $\nu = -\rho$,

$$\begin{aligned} &\int_{-\infty}^{\infty} T_{R^f}^2(-\rho,\theta\pm\pi)d\rho \\ &= -\int_{\infty}^{-\infty} T_{R^f}^2(\nu,\theta\pm\pi)d\nu \\ &= \int_{-\infty}^{\infty} T_{R^f}^2(\nu,\theta\pm\pi)d\nu = \Re_f(\theta\pm\pi) \end{aligned} \qquad (7)$$

From (4)-(7), one can see that:

(1) Translation in the plane does not change the result of $\Re$ transform.



(a) FG

(b) FA

(c) FT

(d) MA

(e) MT

(f) OT

(g) RB

Figure 3. The images of features with $\mathfrak{R}$ transforms

(2) A scaling of the original image only induces the change of amplitude. Here in order to remove the influence of body size, the result of $\Re$ transform is normalized to the range of [0, 1].

(3) A rotation of θ_0 in the original image leads to the phase shift of θ_0 in $\Re$ transform. In this paper, recognized activities rarely have such rotation.

(4) Considering (7), the period of $\Re$ transform is π. Thus a shape vector with 180D is sufficient to represent the spatial information of silhouette.

Therefore, $\Re$ transform is robust to geometry transformation, which is appropriate for activity representation. According to [21], $\Re$ transform outperforms other moment based descriptors, such as Invariant moment, Wavelet moment and Zernike moment, on similar but actually different shape sequences, and even in the case of noisy data.

B. *Feature Representation of $\Re$ Transform*

All interactive videos in CASIA activities database begin with moving person entering the camera view and end in leaving the camera view. Background model is created by Gaussian Mixture Models (GMMs). The foreground images are subtracted by background subtraction and noise is removed with a median filter by a 3×3 template. Finally, a predetermined threshold is used to obtain binary images, as shown in Fig. 2.

Any interactive activity is divided into three phases, including prologue, climax and epilogue, each of which consists of two parts.

Fig. 3 shows the key frame silhouettes of different activities and respective $\Re$ transform of their own. The figures in the first column display the key frame silhouettes. The second column describes the curves of $\Re$ transform. The last column illustrates respective $\Re$ transform in their phase. There are distinct differences in the same activity and the same phase. To illuminate the curve distinctions in different parts naturally, we represent their shapes as follow:

(1)The curve with 3 maximums and 2 minimums is defined as "W", as shown in Figure 3(b);

(2)The curve with 2 maximums and 1 minimum is defined as "U", as shown in Figure 3(c);

(3)The curve fluctuating frequent with remarkable amplitude is labeled with "+" as superior character after previous capital letter.

In conclusion, Fig. 3 illustrates that $\Re$ transform could represent the different characteristics of the interactions. Table 2 shows the characteristic letters in different parts and phases.

TABLE II. THE CHARACTERISTIC LETTERS OF INTERACTIVE ACTIVITIES BETWEEN TWO PERSONS

Activity	Prologue	Climax	Epilogue
FG	WW	WW	UU
FA	WW	WW	WW
FT	WW	UU	UU
MA	WW	UU	WW
MT	WW	UU	UU
OT	WW	UU	UW
RB	UW	W^+W^+	WU

Compared with the other 4 types of activities, Table 2 shows that the curves in FG, FA and RB with distinct characters could be distinguished significantly. The other types of activities are divided into 2 groups, i.e. FT and MT, MA and OT. In contrast with FT, the second maximum in MT curve is more significant. The average of the second maximums is higher than the total amplitude. On the contrary, the second maximum in FT is less significant. Its maximum is approximate to the minimum of amplitude. Similarly, for MA and OT curve, there are not only amplitude differences above but also variation with tendency in phase of epilogue. In details, there is continuous "W" in MA, while there is from "U" to "W" in OT. This figure shows that $\Re$ transform can describe the spatial information sufficiently and characterize the different activity shape effectively.

C. *Experiment Data*

Experiment data consists of 7 classes, each of which includes 4 groups of videos. Binary silhouette is extracted from every video, as shown in Fig. 2. Then we select continuous 100 frames for experiment. As the silhouettes are not deal with manually, there are disjoint, noise and incomplete parts in images. It concludes that the data are common and suitable for surveillance system.

D. *Activity Learning and Classification*

The $\Re$ transform 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. HMMs are appropriate to characterize the variation of interactive activity [22], which is trained for each activity class. The number of model states and GMMs are determined according to experience, as shown in Table 3. Trained HMM models are then used to compute the similarity for a testing sequence.

Because $\Re$ transform is non-orthogonal, the shape vector with 180 dimensionalities 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, because of higher computational complexity, the performance in experiments is not satisfactory. So we try other methods to reduce the dimensionality of feature vector for each frame.

We assume A^s is the image of frame s defined by

TABLE III. THE NUMBER OF HMM STATES AND GMMS

	FG	FA	FT	MA	MT	OT	RB
States	2	2	2	2	2	2	2
GMMs	2	2	2	3	3	3	3

$$A^s_{m\times n} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \qquad (8)$$

After $\Re$ transform, we obtain the shape vector as follow:

$$B^s_{180\times 1} = \begin{pmatrix} b_1 & b_2 & \cdots & b_{180} \end{pmatrix}^T$$

Then it is divided into j parts with the same length, each of which is given by

$$B'^s_l = \begin{pmatrix} b_{i\cdot(l-1)+1} & b_{i\cdot(l-1)+2} & \cdots & b_{i\cdot(l-1)+k} \end{pmatrix}^T$$

where $i,j\in Z^+$, $i\cdot j = 180$, $1\le k\le i$, $l\in L$ and $L=\{l\,|\,l\in Z^+, 1\le l\le j\}$. We construct the matrix B'^s as the column vector B'^s_l in sequence, and reduce its dimensionalities with PCA, as shown in (9)

$$C^s_{j\times j} = \begin{pmatrix} c_{11} & c_{12} & \cdots & c_{1j} \\ c_{21} & c_{22} & \cdots & c_{2j} \\ \vdots & \vdots & \ddots & \vdots \\ c_{j1} & c_{j2} & \cdots & c_{jj} \end{pmatrix} = \begin{pmatrix} C'^s_1 & C'^s_2 & \cdots & C'^s_j \end{pmatrix} \qquad (9)$$

where $C'^s_l = \begin{pmatrix} c_{1l} & c_{2l} & \cdots & c_{jl} \end{pmatrix}^T$.

We assume the set of eigenvalues as $E'^s = \{e_l\}$. We denote $f(C'^s_l) = e_l$ and obtain corresponding eigenvector as (10)

$$C''^s_t = \underbrace{\arg\max_{r_t\in L-\{r_1,r_2,\cdots,r_{t-1}\}} \cdots \arg\max_{r_2\in L-\{r_1\}} \arg\max_{r_1\in L}}_{t} \{E'^s\} \qquad (10)$$

where $t\in L$. Then eigenvector, $D^s = \begin{pmatrix} C''^s_1 & C''^s_2 & \cdots & C''^s_t \end{pmatrix}^T$, is formed to train and test HMMs.

For the sake of balance between classification accuracy and computational complexity, we conclude that j is 5 and t is 3 according to [21]. Since the video sequences in the database are limited in quantity, random testing is used for classification. In each class of interactive activity, one sequence is denoted randomly as probe, while the others are galleries. According to (8)-(10), 7 probes are tested individually in the HMMs. After we implement them 20 times, the results are shown in Table 4.

We can draw several conclusions from Table 4. At first, in all phases of FA, the curves have rarely significant changes in contrast with the others, thus the CCR is the highest. Next, the FG, MA and RB are much similar in the corresponding phase, so the CCR is higher than the other 3 classes as well. Finally, compared with previous 4 classes, though there are distinct characteristics in FT, MT and OT, less significant changes induces lower CCR.

$\Re$ transform descriptor captures both boundary and internal content of the shape, so they are more robust to noise [21], such as internal holes and disjoint parts. While

TABLE IV. THE CORRECT CLASSIFICATION RATE (CCR) OF 7 ACTIVITIES (%)

	FG	FA	FT	MA	MT	OT	RB
CCR	90	100	80	95	80	85	90

TABLE V. THE COMPARISON OF OUR METHOD TO [21]

	Condition	Type of activities	Extracted feature		performance		
			Silhouette	*Frame*	*Phase*	*Stabilization*	*Practicability*
Wang et al. [21]	Indoor	Single person	Complete	Key frame	Medium	Medium	Medium
Our method	outdoor	Two persons' interaction	Incomplete	Continuous frames	better	better	better

TABLE VI. THE COMPARISON OF OUR METHOD TO OTHERS

Method	Characteristics	Model	Destination	Computational complexity
Nascimento et al. [17]	Centroid, trajectory	Space-dependent switched dynamical model	Analyzing the person's position and track to describe its movement and inducing his activity	Low
Du et al. [18]	Speed, direction, distance, contour, moment, ratio of height to width	Dynamic Bayesian Network (DBN)	Analyzing two persons' features and their relations synthetically to depict interactive activities	High
Du et al. [19]	Speed, distance, angle, silhouette, center of mass	Hierarchical durational-state dynamic Bayesian network (HDS-DBN)	Analyzing two persons' features in different dimensions to describe interactive activities	High
Our method	Silhouette, contour	Hidden Markov Models (HMMs)	Analyzing two persons' silhouette and contour synthetically to depict interactive activities	Medium

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 [23]. Generally speaking, low level features based on $\Re$ transform are effective for recognizing similar activity even in the case of noisy data.

E. Result Comparison

As shown in Table 5, there are several advantages of ours in contrast with [21] as follow. Condition is more complex so that it is approximate to practice. Activity type is more in quantity. Extracted silhouette is less complete, and it can meet different condition. Continuous frames are more suitable for analysis in interactive activities than key frames. In general, our method has significant advantages in classification, i.e. significant in phase, stabilization and practicability, thus it highlights intelligent surveillance in complex condition.

As shown in Table 6, the characteristics and model in [17] is comparable simplicity in contrast with the others, so its computational complexity is low. However, it could not afford to describe interactive activities substantially, thus it hardly satisfy intelligent surveillance. In [18] and [19], the methods are more complex in model and precise in representation of interactive activities. Simultaneously, its computational cost is so a bit high that it goes against to surveillance application. Though it is not perfect in precise representation of interaction, it is medium in computational complexity and higher in correct classification rate, thus it is more practicable for surveillance system.

V. Conclusions

In this paper, we have proposed a method using $\Re$ transform as a shape descriptor to represent 2 persons' interaction in each frame and employed HMMs for recognition.

According to the comparison with common methods, this method representing and classifying the interactions based on $\Re$ transform has several advantages. Firstly, it does not require video alignment and is applicable to many scenarios where the background is known, because $\Re$ transform is invariant to scale and translation. Secondly, $\Re$ transform gets the high recognition rate for similar but actually different shape sequences, and even in the case of incomplete data. Thirdly, our shape descriptor captures both boundary and internal content of the shape. For this reason, it is more robust than noise, internal holes and disjoint parts. Fourthly, the computation of shape descriptor is linear, so the computation cost of 2D $\Re$ transform is low. Finally, our method solves the problems in images denoising and segmenting with $\Re$ transform and HMMs. We construct a common and practicable model for the interactions by experiment and improve the correct classification rate. Generally speaking, the video surveillance with this method can be deployed indoors and outdoors and be applicable to many cases, i.e. hall, corridor, road and park etc. It can assist the operators to detect and recognize different activities in many scenarios simultaneously.

In contrast with single person's recognition, our method accomplishes classification of the interactions from incomplete data in complex condition, which is steady in performance and good in practicability. Compared with related works, it has the advantages of medium computational complexity and high correct classification rate in video surveillance, though they are more precise in recognition of activities.

In the future, we focus on 3 parts as follow. First, we shall implement it on large dataset with more types of activities. Next, the $\Re$ transform descriptor and HMMs will be improved in details furthermore. Finally, classifying method in different views will be under study. In general, we improve the surveillance performance by increasing correct classification rate and reducing computational complexity.

Acknowledgements

The authors would like to thank CASIA to provide activity database and the anonymous reviewers for their constructive comments.

References

[1] Pavan Turaga, Rama Chellappa, V. S. Subrahmanian, and Octavian Udrea, "Machine Recognition of Human Activities: A Survey," IEEE Trans. Circuits Syst. Video Technol., vol. 18, no. 11, pp. 1473-1488, November 2008.

[2] S. Sarkar, P. J. Phillips, Z. Liu, I. R. Vega, P. Grother, and K. W. Bowyer, "The Human ID gait challenge problem: Data sets, performance, and analysis," IEEE Trans. Pattern Anal. Mach. Intell., vol. 27, no. 2, pp. 162–177, February 2005.

[3] Y. Rui, T. S. Huang, and S. F. Chang, "Image retrieval: Current techniques, promising directions and open issues," J. Visual Commun. Image Represent., vol. 10, no. 1, pp. 39–62, March 1999.

[4] N. Vaswani, A. K. Roy-Chowdhury, and R. Chellappa, "Shape activity: A continuous-state HMM for moving/deforming shapes with application to abnormal activity detection," IEEE Trans. Image Process., vol. 14, no. 10, pp. 1603–1616, October 2005.

[5] H. Zhong, J. Shi, and M. Visontai, "Detecting unusual activity in video," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2004, pp. 819–826.

[6] PLA Filiberto. etc., "Extracting Motion Features for Visual Human Activity Representation," Lecture Note in Computer Science, vol. 3522, no. 1, pp. 537-544, 2005.

[7] S. Hongeng, R. Nevatia, and F. Bremond, "Video-based event recognition: activity representation and probabilistic recognition methods", Comput. Vision Image Understanding, vol. 96, no. 2, pp. 129-162, November 2004.

[8] Chuang C.-H., Kuo C.-C.J., "Wavelet Descriptor of Planar Curves: Theory and Applications", IEEE Trans. Image Processing, vol. 5, no. 1, pp. 56-70, January 1996.

[9] D. Zhang, G. Lu, "Shape-based image retrieval using generic Fourier descriptor," Signal Process. Image Commun., vol. 17, no. 10, pp. 825-848, November 2002.

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

[11] D. Zhang, G. Lu, "Study and evaluation of different Fourier methods for image retrieval," Image Vision Comput., vol. 23, no. 1, pp. 33-49, January 2005.

[12] C. H. Teh, R. T. Chin, "On image analysis by the methods of moments," IEEE Trans. Pattern Anal. Mach. Intell., vol.10, no. 4, pp. 496-513, July 1988.

[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, no. 5, pp. 438-460, 1992.

[14] Nikolaos V. Boulgouris, Zhiwei X. Chi, "Gait Recognition Using Radon Transform and Linear Discriminant Analysis," IEEE Trans. Image Process., vol. 16, no. 3, pp. 731-740, March 2007.

[15] P. Toft, "The Radon transform—Theory and implementation," Ph.D. dissertation, Denmark Tech. Univ., Lyngby, 1996.

[16] S. Tabbone, L. Wendling, and J.-P. Salmon, "A new shape descriptor defined on the Radon transform", Comput. Vision Image Understanding, vol. 102, no. 1, pp. 42-51, April 2006.

[17] Jacinto C. Nascimento, Mario A. T. Figueiredo, and Jorge S. Marques, "Recognition of human activities using space dependent switched dynamical models," Proc. Int. Conf. Image Process., vol. 3, pp. 852-855, 2005.

[18] Youtian Du, Feng Chen, Wenli Xu, and Yongbin Li, "Recognizing Interaction Activities using Dynamic Bayesian Network," Proc. Int. Conf. Pattern Recognit., pp. 618-621, 2006.

[19] Youtian Du, Feng Chen, and Wenli Xu, "Approach to human activity multi-scale analysis and recognition based on multi-layer dynamic bayesian network," Acta Automatica Sinica, vol. 35, no. 3, pp. 225-232, March 2009.

[20] Deans, S. R.: Applications of the Radon Transform. Wiley Interscience Publications, Chichester, 1983.

[21] Ying Wang, Kaiqi Huang, and Tieniu Tan, "Human Activity Recognition Based on $\Re$ Transform," Proc. IEEE Comput. Soc. Conf. Comput. Vision Pattern Recognit., pp. 3722-3729, 2007.

[22] Yamato, J., Ohya, J., Ishii, K., "Recognizing Human Action in Time Sequential Images Using a Hidden Markov Model," Proc. IEEE Comput. Soc. Conf. Comput. Vision Pattern Recognit., 1992.

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

Hao Zhang was born in Qingdao, Shandong, P.R.China, 1980. He received the B.Eng. degree in computer science and technology from Qingdao Institute of Architecture and Engineering, Qingdao, Shandong, P.R.China, in 2003. Then he received the M.Eng. degree in computer application from Xi'an Shiyou University, Xi'an, Shaanxi, P.R.China, in 2007. He is currently working toward the Ph.D. degree in the School of Computer Science and Technology, Xidian University, Xi'an, Shaanxi, China. His current research interests include computer vision, pattern recognition, image processing, and their application in activity recognition and classification.

Zhijing Liu was born in Xi'an, Shaanxi, Province, P.R.China, in 1957. He received the B.Eng. degree in computer engineering from Institute of Northwestern Telecommunications Engineering, Xi'an, Shaanxi, P.R.China, in 1982. His major field of study are computer vision and data mining.

He has been teaching and researching at Xidian University since 1982. He is a professor and a supervisor for Ph.D. students in Xidian University. He has published more than 100 papers.

Prof. Liu has acted as member of the Expert Advisory Committee of the leading group of the informatization of Shaanxi province, and is a fellow of the Committee of Experts of manufacturing informatization of Shaanxi province, committeeman of the city of Xi'an manufacturing informatization Expert Committee, and appraisal expert of the Committee of Awards of enterprise technology innovation of Shaanxi province.

Haiyong Zhao was born in Liaocheng, Shandong, P.R.China, in 1981. He received the B.Eng. degree in computer science and technology from Shandong University of Technology, Zibo, Shandong, P.R.China, in 2004. Then he received the M.Eng. degree in software engineering from Shandong University, Shandong, P.R.China, in 2007. He is currently working toward the Ph.D. degree in the School of Computer Science and Technology, Xidian University, Xi'an, Shaanxi, P.R.China. His current research interests include computer vision, object detection and image processing, and their application.