Hindawi Publishing Corporation
The Scientific World Journal
Volume 2014, Article ID 495071, 10 pages
http://dx.doi.org/10.1155/2014/495071

Research Article

The Approach for Action Recognition Based on the Reconstructed Phase Spaces

Hong-bin Tu and Li-min Xia

School of Information Science and Engineering, Central South University, Hunan 410075, China

Correspondence should be addressed to Hong-bin Tu; tuhongbin310@163.com

Received 1 July 2014; Accepted 11 September 2014; Published 10 November 2014

Academic Editor: Shifei Ding

This paper presents a novel method of human action recognition, which is based on the reconstructed phase space. Firstly, the human body is divided into 15 key points, whose trajectory represents the human body behavior, and the modified particle filter is used to track these key points for self-occlusion. Secondly, we reconstruct the phase spaces for extracting more useful information from human action trajectories. Finally, we apply the semisupervised probability model and Bayes classified method for classification. Experiments are performed on the Weizmann, KTH, UCF sports, and our action dataset to test and evaluate the proposed method. The compare experiment results showed that the proposed method can achieve was more effective than compare methods.

1. Introduction

Automatic recognition of human actions from image sequences is a challenging problem that has attracted the attention of researchers in the past decades. This has been motivated by the desire for application of entertainment, virtual reality, motion capture, sport training [1–3], medical biomechanical analysis, and so on.

In a simple case, where a video is segmented to contain only one execution of a human activity, the objective of the system is to correctly classify the video into its activity category. More generally, the continuous recognition of human activities must be performed by detecting the starting and ending times of all occurring activities from an input video. Aggarwal and Ryoo [4] summarized the general method as single-layered approaches, hierarchical approaches, and so forth. Single-layered approaches represent and recognize human activities directly based on sequences of images. So, they are suitable for the recognition of gestures and actions with sequential characteristics. Single-layered approaches are again classified into two types: space-time approaches and sequential approaches. Space-time approaches are further divided into three categories: space-time volume, trajectories, and space-time features. Hierarchical approaches represent high-level human activities by describing them in terms of simpler activities. Hierarchical approaches usually can be divided into 3 classes: the statistical, the syntactic, and the description-based classes. Recognition systems composed of multiple layers are constructed, which are suitable for the analysis of complex activities. Among all these methods, the space-time approaches are the most widely used ones to recognize simple periodic actions such as "walking" and "waving," and periodic actions will generate feature patterns repeatedly and the local features are scale, rotation, and translation-invariant in most cases. However, the space-time volume approach is difficult in recognizing the actions when multiple persons are present in the scene and it requires a large amount of computations for the accurate localization of actions. Besides, it is difficult in recognizing actions which cannot be spatially segmented. The major disadvantage of the space-time feature is that it is not suitable for modeling more complex activities. In contrast, the trajectory-based approaches have the ability to analyze detailed levels of human movements. Furthermore, most of these methods are view-invariant. Therefore, the trajectory-based approaches have been the most extensively studied approaches.

Several approaches used the trajectories themselves to represent and recognize actions directly. Sheikh et al. [5]

applied a set of 13 joint trajectories in a 4D XYZT space to describe the human action. Yilmaz and Shah [6] presented a methodology to compare action videos by the set of 4D XYZT joint trajectories. Anjum and Cavallaro [7] proposed algorithm based on the extraction of a set of representative trajectory features. Jung et al. [8] designed the novel method to detect event by trajectory clustering of objects and 4D histograms. Hervieu et al. [9] used Hidden Markov models to capture the temporal causality of object trajectories for the unexpected event detection. Wang et al. [10] proposed a nonparametric Bayesian model to analysis trajectory and model semantic region in surveillance. Wang et al. [11] presented a video representation based on dense trajectories and motion boundary descriptors for recognizing human actions. Yu et al. [12] used the novel approach based on weighted feature trajectories and concatenated bag-of-features (BOF) to recognize action. Pao et al. [13] proposed a general user verification approach based on user trajectories, which include on-line game traces, mouse traces, and handwritten characters. Yi and Lin [14] introduced the salient trajectories to recognize. Du et al. [15] proposed an intuitive approach on videos based on the feature trajectories. Psarrou et al. [16] designed the model of the statistical dynamic to recognize human actions by learning prior and continuous propagation of trajectories models.

These approaches approximated the true motion state by setting constraints on the type of the dynamical model [1]. Above all, they required the detailed mathematical and statistical modeling. To solve these problems, we present the approach for action recognition based on the reconstructed phase spaces.

The remainder of this paper is organized as follows. Section 2 presents the modified particle filter that is used to track human key points. In Section 3, we reconstruct the phase space of the total data. Section 4 explains the probability generation model. Section 5 explains action classification. Section 6 explains the results and analysis of the proposed approach. Finally, we conclude the paper in Section 7.

2. Human Key Joints Track

The human body [2] is divided into 15 key points, which are named 15 key joint points for representing the human body structure (torso, pelvis, left upper leg, left lower leg, left foot, right upper leg, right lower leg, right foot, left upper arm, left lower arm, left hand, right upper arm, right lower arm, right hand, and head) [17], which the 15 joints trajectory represents the human body behavior (blue dot represents pelvis, which is the origin of coordinate). Another consideration was that these joints were relatively easy to automatically detect and track in real videos, as opposed to the inner body joints which were more difficult to track. Each key joint had a trajectory as the time was going on and 15 trajectories were used to represent different actions. Therefore, we must track accurately the human body 15 nodes for indicating the human behavior. These are illustrated in Figure 1.

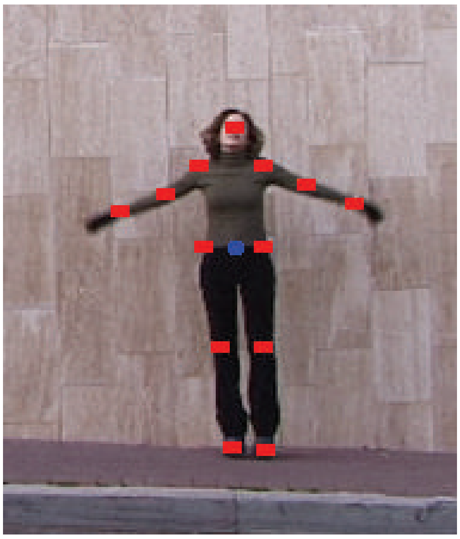


Figure 1: The human joints model. The original photo stems from Weizmann dataset [21].

However, it is difficult to track some key points for occlusion. In this paper, we use the modified particle filters to track these key points. Particle filters are very efficient methods for tracking multiple objects, which they can cope with no-linear and multimodality induced by occlusions and background clutter. But it has been proved that the number of samples increases exponentially with the size of the state vector to be explored. The reason is that one sample dominates the weight distribution and the rest of the samples are not in statistically significant regions. In order to solve the above problem, we adopt the integrated algorithm based on both particle filters and Markov chain Monte Carlo models [18, 19], which is based on drift homotopy for stochastic differential equations and the existing particle filter methodology for multitarget tracking by appending an MCMC step after the particle filter resampling step. The MCMC step is integrated to the particle filter algorithm to bring the samples closer to the observation at the same time respecting the target dynamics.

We can assume [18] the parameters as follows: $Z_{K_1}, \ldots, Z_{K_N}$: the noisy observations, $K_1, \ldots, K_N$: the status of the system particular time, $Z_{K_n} = G(X_{K_n}, \eta_n)$ $(\eta_n, n = 1, \ldots, N)$: the observations functions, $g(X_{K_n}, Z_{K_n})$: the distribution of the observations, and $E[f(X_{K_n} \mid \{Z_{K_n}\}_{i=1}^{N}]$: the conditional expectation.

Given a video sequence and labeled samples of object or background pixels on the first frame [20], we have access to noisy observations of the status of the system particular time.

The filtering problem consists of computing estimates of the conditional expectation. Therefore, we can compute the conditional density of the state of the system $p(X_{T_k} \mid \{Z_{Z_k}\}_{i=1}^{k})$ and define a reference density: $q(X_{K_n} \mid \{Z_{K_n}\}_{n=1}^{N})$. At last, we obtain the weighted sample [18]:

$$E\left[f\left(X_{K_n} \mid \{Z_{K_n}\}_{i=1}^{N}\right)\right] \propto \frac{1}{N}\sum_{n=1}^{N} f\left(X_{K_n}\right) \frac{p\left(X_{K_n} \mid \{Z_{K_n}\}_{i=1}^{N}\right)}{\sum_{n=1}^{N} q\left(X_{K_n} \mid \{Z_{K_n}\}_{i=1}^{N}\right)}, \quad (1)$$

$$p\left(X_{K_n} \mid \left\{Z_{K_n}\right\}_{n=1}^{N}\right) \propto \frac{g\left(X_{K_n}, Z_{K_n}\right) p\left(X_{K_n} \mid \left\{Z_{K_i}\right\}_{i=1}^{N}\right)}{q\left(X_{K_n} \mid \left\{Z_{K_i}\right\}_{i=1}^{N}\right)}, \tag{2}$$

$$p\left(X_{K_n} \mid \left\{Z_{K_i}\right\}_{i=1}^{N}\right) = \int p\left(X_{K_n} \mid X_{K_{n-1}}\right) p\left(X_{K_{n-1}} \mid \left\{Z_{K_i}\right\}_{i=1}^{N-1}\right) dX_{K_{i-1}}. \tag{3}$$

We assume that

$$q\left(X_{K_n} \mid \left\{Z_{K_i}\right\}_{i=1}^{N}\right) \propto p\left(X_{K_n} \mid \left\{Z_{K_i}\right\}_{i=1}^{N-1}\right), \tag{4}$$

and, from (2), we can obtain the formula

$$\frac{p\left(X_{T_k} \mid \left\{Z_{T_i}\right\}_{i=1}^{k-1}\right)}{q\left(X_{T_k} \mid \left\{Z_{T_i}\right\}_{i=1}^{k}\right)} \propto g\left(X_{K_n}, Z_{K_i}\right). \tag{5}$$

The approximation in expression (1) becomes

$$E\left[f\left(X_{K_n} \mid \left\{Z_{K_n}\right\}_{n=1}^{N}\right)\right] \approx \frac{\sum_{n=1}^{N} f\left(X_{K_n}^{n}\right) g\left(X_{K_n}^{n}, Z_{K_n}\right)}{\sum_{n=1}^{N} q\left(X_{K_n} \mid \left\{Z_{K_n}\right\}_{n=1}^{N}\right) g\left(X_{K_n}^{n}, Z_{K_n}\right)}. \tag{6}$$

Thus, we can define the (normalized) weights

$$w_{K_n}^{n} = \frac{q\left(X_{K_n} \mid \left\{Z_{K_i}\right\}_{i=1}^{N}\right) g\left(X_{K_n}^{n}, Z_{K_i}\right)}{\sum_{n=1}^{N} g\left(X_{K_n}^{n}, Z_{K_i}\right)}. \tag{7}$$

The tracking algorithm is described as follows.

(1) Sampling N particles in accordance with the unified weights randomly generated particles form unweighted samples $X_{K_{n-1}}^{n}$ and determination $p(X_{K_n} \mid \{Z_{K_{n-1}}\}_{n=1}^{N-1})$, as follows:

$$p\left(X_{K_n} \mid \left\{Z_{K_{n-1}}\right\}_{n=1}^{N-1}\right) = \prod_{\lambda=1}^{\wedge} p\left(X_{K_\lambda, K_n} \mid \left\{Z_{\lambda, K_n - 1}\right\}_{n=1}^{N-1}\right). \tag{8}$$

(2) Predict by sampling $X_{K_n}^{N}$ from

$$p\left(X_{K_n} \mid X_{K_{n-1}}\right) = \prod_{\lambda=1}^{\wedge} p\left(X_{\lambda, K_n} \mid X_{\lambda, K_{n-1}}\right). \tag{9}$$

(3) Target observation association.

(4) Update and evaluate the weights:

$$w_{T_k}^{n} = \frac{\prod_{\lambda=1}^{\wedge} q\left(X_{K_n} \mid \left\{Z_{K_i}\right\}_{i=1}^{N}\right) g_\lambda\left(X_{\lambda, T_k}^{m}, Z_{\lambda, T_i}\right)}{\sum_{n=1}^{N} \prod_{\lambda=1}^{\wedge} g_\lambda\left(X_{\lambda, T_k}^{m}, Z_{\lambda, T_i}\right)}. \tag{10}$$

(5) By resampling, through the above steps, we can generate independent uniform random variables $\{\theta^i\}_{i=1}^{N}$ $(0 < \{\theta^i\}_{i=1}^{N} < 1)$. Therefore, we can obtain the following equation:

$$\left(X_{K_{n-1}}^{n}, X_{K_n}^{n}\right) = \left(X_{K_{n-1}}^{\prime j}, X_{K_n}^{\prime j}\right), \tag{11}$$

where $\sum_{n=1}^{j-1} w_{T_k}^{n} \leq \theta^j \leq \sum_{n=1}^{j} w_{T_k}^{n}$.

(6) By Markov chain Monte Carlo tracking, we choose a modified drift for $n = 1, \ldots, N$ and $k = 1, \ldots, K$. Construct a Markov chain [18–20] for $Y_{K_n}^{N}$ with initial value $X_{K_n}^{N}$ (the global state of the system is defined by $X_{K_n}^{N}$) and obtain the stationary distribution

$$\prod_{\lambda=1}^{\wedge} g_\lambda\left(Y_\lambda^{n}, Z_{K_i}\right) p_\lambda\left(Y_\lambda^{n} \mid X_{K_n}^{n}\right). \tag{12}$$

(7) Set $X_{K_n}^{N} = Y_{K_n}^{N}$.

(8) Set $n = n + 1$ and go to Step 1.

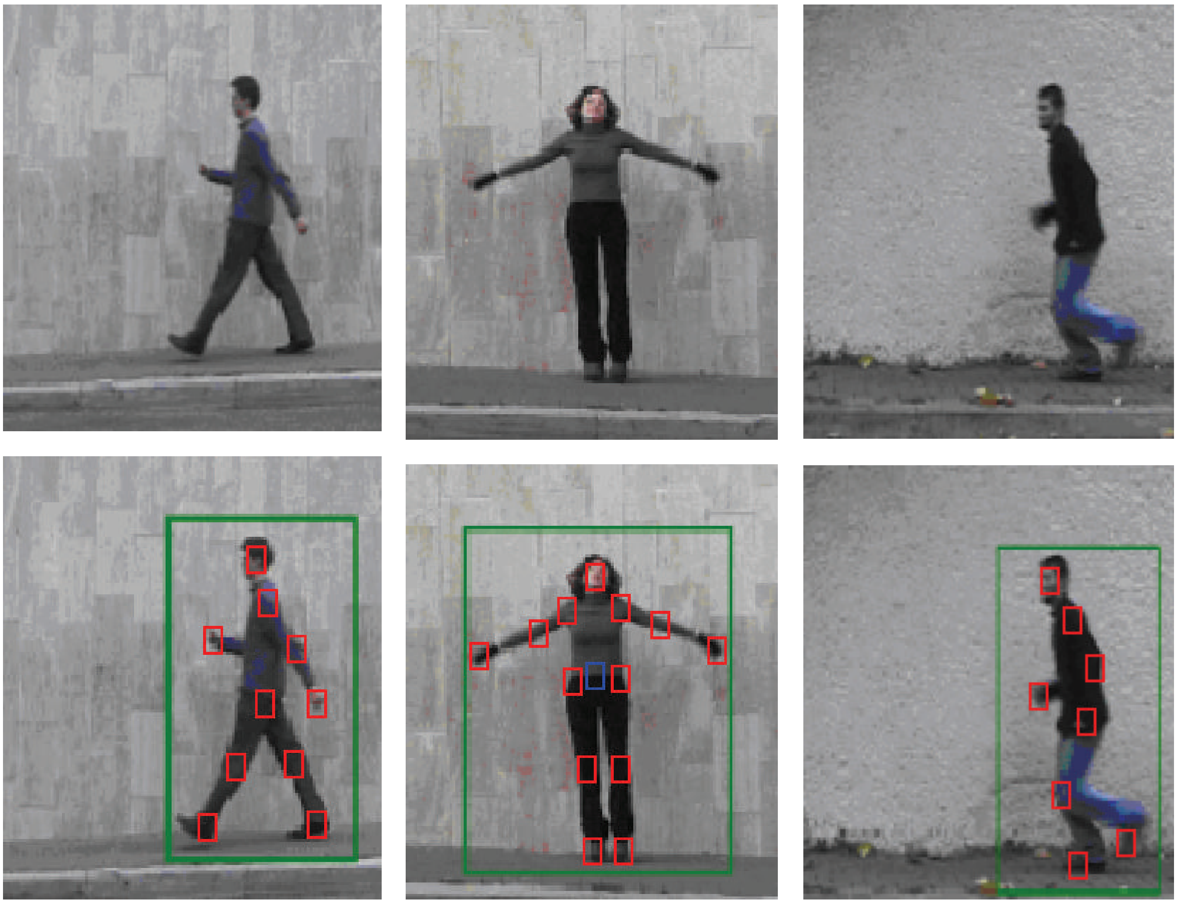
Using the tracking algorithm, we can obtain key points trajectories, which are used to recognize human behavior. Figure 2 depicts the results of human target tracking.

3. Phase Space Reconstruction

At present, the phase space reconstruction has been used in many research fields. de Martino et al. [22] constructed the trajectory space and refer to the phase space in the dynamic system. Paladin and Vulpiani [23] presented the embedding trajectory dimension, which was similar to reconstruct the embedding dimension of the phase spaces of the dynamic system. Fang and Chan [24, 25] present the unsupervised ECG-based identification method based on phase space reconstruction in order to save the picking up characteristic points. Nejadgholi et al. [26] used the phase space reconstruction for recognizing the heart beat types. In this paper, we use the phase space reconstruction for human action recognition.

We use the linear dynamic systems instead of the traditional gradient and optical flow features of interest points to recognize action. The linear dynamic system [27] is suitable to deal with temporally ordered data, which has been used in several applications in computer vision, such as tracking, human recognition from gait, and dynamic texture. The temporal evolution of a measurement vector can be modeled by the dynamic system. In this case, we use the linear dynamic system to model the spatiotemporal model. In this series, it is sometimes necessary to search for patterns not only in the time series itself, but also in a higher-dimensional transformation of the time series. We can estimate the delay time τ and embedding dimensions d in reconstructed phase space in order to extract more useful information from human action trajectories. These parameters can be computed as follows.

The phase portrait of a dynamic system [28] described by a one-dimensional time series of measured scalar values $x(t)$



(a) Walking

(b) Jacking

(c) Running

Figure 2: The target tracking results. The original photo stems from Weizmann datasets [21].

can be reconstructed in a k-dimensional state space. From the time-series signal, we can construct an m-dimensional signal $x(t)$. We define [28] a dynamical system as the possibly nonlinear map, which represents the temporal evolution of state variables

$$x(t) = [x_1(t), x_1(t), \ldots, x_m(t)] \in R^m. \tag{13}$$

de Martino et al. [22] pointed out that the phase space reconstruction based on Taken's theory is equivalent to the original attractor if m is large enough by suitable hypotheses.

Each point in the phase space is calculated according to [26]. Consider

$$x_n = [x_n x_{n-\tau}, \ldots, x_{n-(d-1)\tau}]\, n = (1 + (d-1)\tau), \ldots, N, \tag{14}$$

where x_n is the nth point in the time series, delay times τ is the time lag, N is the number of points in the time series, and d is the dimension of the phase space. From now on, η is used to denote this set of body model variables describing human motion.

The reconstructed phase space is shown by López-Méndez and Casas and Takens [28, 29] for the large enough m, which is a homeomorph m (embedding dimension) of the true dynamical system in the generated time series. We used Takens' theorem to reconstruct state spaces by time-delay embedding. In our case, parameters [26, 28] are defined as follows:

η: the temporal evolution;

$Y^{N}_{K_n}$: time series (scalar), and we want to characterize

$$Y^{\eta N}_{K_n} = [z^{\eta}(t), z^{\eta}(t+\tau), \ldots, z^{\eta}(t+(m-1)\tau)]. \tag{15}$$

$Y^{\eta N}_{K_n}$ is a point in the reconstructed phase space, m is the embedding dimension, and τ is the embedding delay. Therefore, the phase space can be reconstructed by stacking sets of m (the large enough m) temporally spaced samples. The embedding delay τ determines the properties of the reconstructed phase space.

At first, the embedding delay using the mutual information method was determined [26] and the estimated delay was used to obtain the appropriate embedding dimension [30]. Once both the embedding delay and the embedding dimension have been estimated, is performed [26] as follows:

$$\widehat{x}^{\eta} = \begin{bmatrix} \overline{Y^{\eta N}_{K_n\ 1+(d-1)\tau}} \\ \overline{Y^{\eta N}_{K_n\ 2+(d-1)\tau}} \\ \vdots \\ \overline{Y^{\eta N}_{K_n\ n+(d-1)\tau}} \end{bmatrix}$$

$$= \begin{bmatrix} \overline{{z^{\eta}}_{1+(d-1)\tau}(0)}, \dots, \overline{{z^{\eta}}_{1+(d-1)\tau}(\tau)}, \dots, \overline{{z^{\eta}}_{1+(d-1)\tau}((m-1)\tau)} \\ \overline{{z^{\eta}}_{2+(d-1)\tau}(\tau)}, \dots, \overline{{z^{\eta}}_{2+(d-1)}(t+\tau)}, \dots, \overline{{z^{\eta}}_{2+(d-1)}(t+(m-1)\tau)} \\ \vdots \\ \overline{{z^{\eta}}_{N}(N-1-(m-1)\tau)}, \dots, \overline{{z^{\eta}}_{N}(N-1-(m-2)\tau)}, \dots, \overline{{z^{\eta}}_{N}(N-1)} \end{bmatrix}. \tag{16}$$

We use the phase space $\widehat{x}^{\eta}$ as signatures, where each one of the model variables constitutes a time series from the reconstructed phase space. The time series [28] model provides a better performance to recognize the action model based on independent scalar time series, which are based on action recognition method. Therefore, we get the phase space corresponding to each point trajectory, which contained the joint point of occlusion and nonocclusion. Besides, we choose Kolmogorov-Sinai entropy [31, 32] as another feature for analyzing the dynamics human action. K-S entropy (HKS) is the average entropy per unit time. We define it as the following [32]:

$$H_k = \frac{\lim_{\kappa \to 0} \lim_{t \to \infty} \left[\sum_{i=1}^{N_t} P_\kappa \sum \log (1/P_k) \right]}{t}. \tag{17}$$

Therefore, each trajectory of the human action can be described as the 3-dimensional feature vector according to the 9-dimensional feature vector of each key joint and 90-dimensional feature vector of each action.

Figure 3 shows the reconstructed phase space of the total joint point.

4. Probability Generation Model

These are a few labeled actions; however, a large number of unlabeled actions need be recognized. Therefore, we use the semisupervised probability model.

It is assumed that [34] the action is generated by a mixture generative model of distribution function $p(x \mid \theta_i)$. Then, we can obtain the generative model [34] as follows:

$$p(x\theta_i) = \sum_{i=1}^{c} p(\theta_i)\, p_i(x\theta_i). \tag{18}$$

It is generally assumed that the distribution of the feature space is almost consistent with a Gaussian distribution or a multinomial distribution for human action images. x is the feature vector of the training sample, $p(\theta_i)$ is the probability of the sample belonging to the ith class, θ_i represents the object classes and the covariance matrix of pixel. Therefore, likelihood functions [34] were defined as follows:

$$\log p(\theta_i \mid X) = \log \left(\prod_{i=1}^{M} p(\theta_i)\, p(x_i \mid \theta_i) \prod_{i=M+1}^{2M} p(x_i \mid \theta) \right)$$
$$= \sum_{i=1}^{M} \log p(\theta_i)\, p(x_i \mid \theta_i) + \sum_{i=M+1}^{2M} \log p(X_i \mid \theta). \tag{19}$$

The first part is supervised classification and the second is called unsupervised part.

Unsupervised part should be written as

$$\sum_{i=l+1}^{2M} \log p(x_i \mid \theta) = \sum_{i=l+1}^{2M} \sum_{j=1}^{c} p(x)\, p(x_i \mid \theta_i). \tag{20}$$

Finally, we can obtain the log-likelihood function

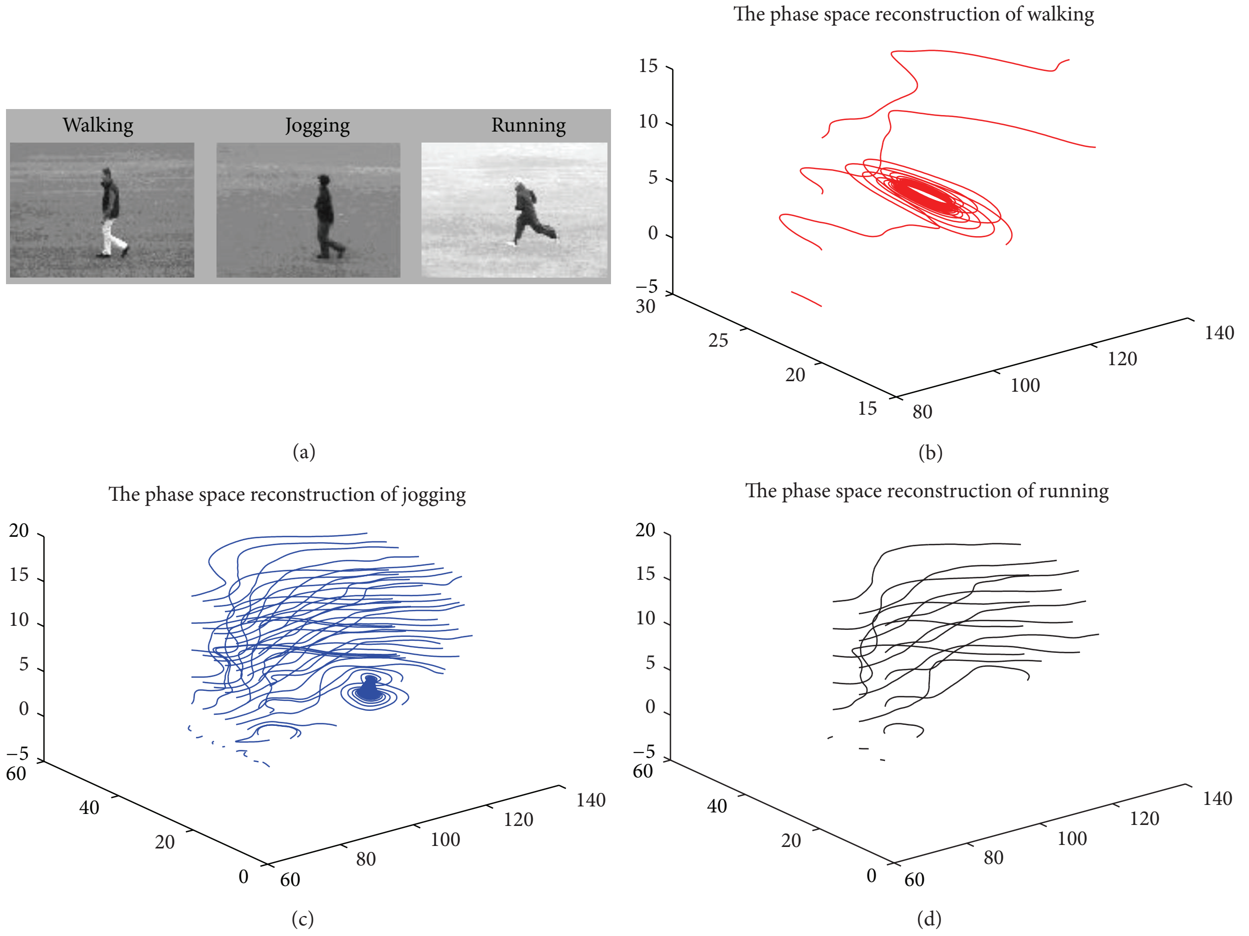
$$\log p(\theta_i \mid X) = \log \left(\prod_{i=1}^{M} p(\theta_i)\, p(x_i \mid \theta_i) \prod_{i=M+1}^{2M} p(x_i \mid \theta) \right)$$
$$+ \sum_{i=l+1}^{2M} \sum_{j=1}^{c} p(x)\, p(x_i \mid \theta_i). \tag{21}$$

In this case, we build the relationship between the unlabeled samples and the learning sample. EM is also an iterative algorithm which has two main steps: expectation and maximization.

E-step: this step predicts the labels of each unlabeled sample by calculating from the last iteration parameters in (21)

$$p_{ji}^{M} = p\left(\theta_i^M \mid X_i\right) = \frac{\theta_i^M p\left(x_i \mid \theta_i^{M-1}\right)}{\sum_{j=1}^{M} \theta_i^{M-1} p(x_i)\, p\left(x_i, \theta_i^{M-1}\right)}, \tag{22}$$

where p_{ji}^{M} is the current prediction of model i unlabeled samples conditioned on the current distributed parameter,



(a) (b)

(c) (d)

Figure 3: Examples of the reconstructed phase space of the missing data. Hip rotations for walking, jogging, and running actions in the KTH dataset [33]. (a) shows original images. (b) shows the result of reconstructing the reconstructed phase space of the missing data. (b1) shows the phase space reconstruction of right foot motion. (b2) shows the phase space reconstruction of right elbow motion. (b3) shows the phase space reconstruction of right elbow motion. (c) shows the reconstructed phase space of the total occlude joint point. (c1) shows the phase space reconstruction of walking. (c2) shows the phase space reconstruction of jogging. (c3) shows the phase space reconstruction of running.

Table 1: Confusion matrix for KTH dataset.

	a1	a2	a3	a4	a5
a1	**0.95**	0.01	0.02	0.00	0.01
a2	0.01	**0.93**	0.02	0.10	0.00
a3	0.00	0.02	**0.90**	0.00	0.01
a4	0.01	0.00	0.00	**0.92**	0.30
a5	0.03	0.00	0.02	0.00	**0.82**

$M{-}1$ is the previous state value, and M is the current state value.

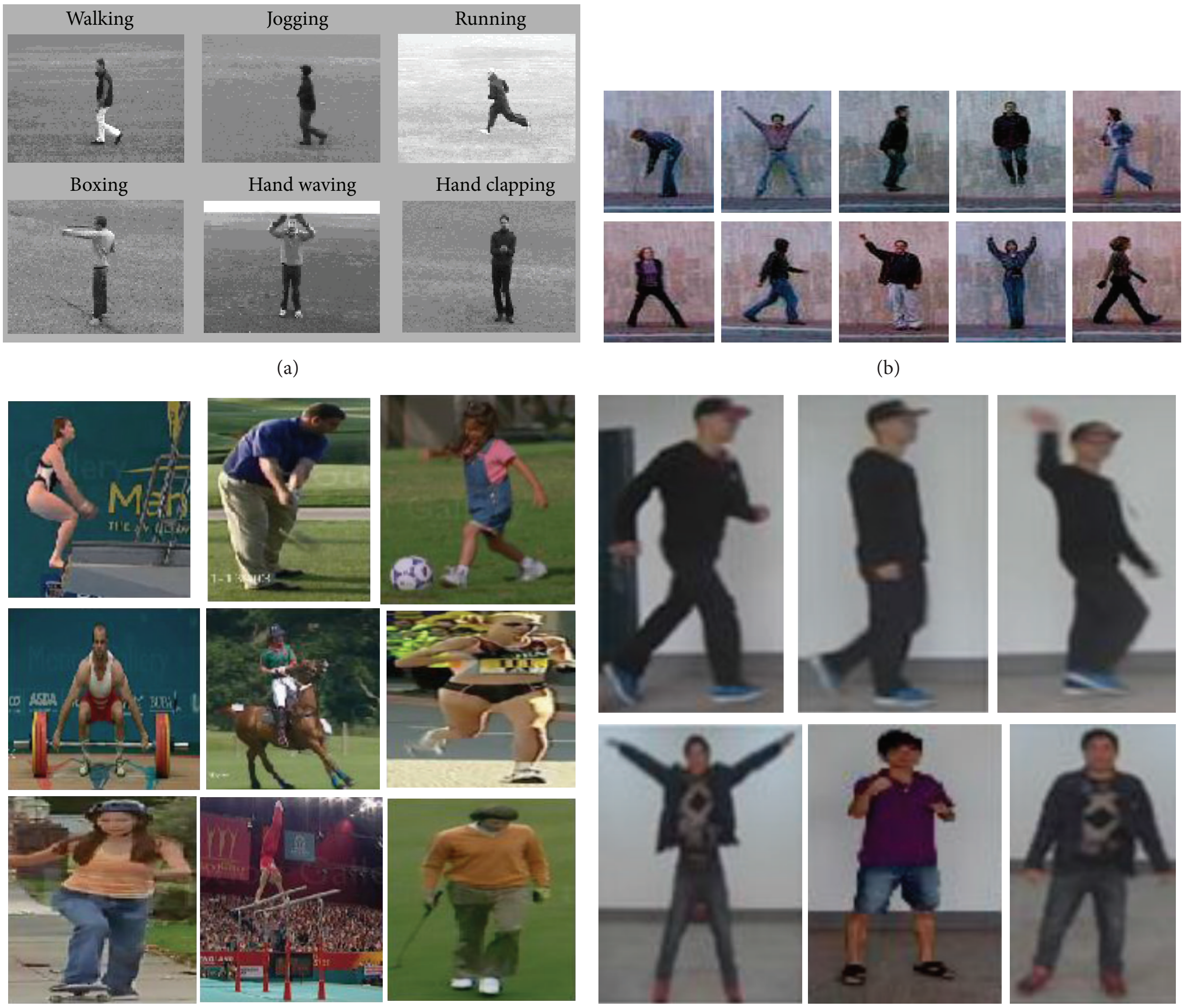
M-step: we calculate the current parameters by maximizing the likelihood function as follows:

$$p_{ji}^{M}(i,j)=\frac{\sum_{k=k+1}^{M}p_{ji}^{M-1}+M_j}{M+l},$$

$$\mu_j^{M}=\frac{\sum_{k=1}^{M}p_{ji}^{M-1}x_k+\sum_{k=1}^{u_j}X'_{jk}}{\sum_{k=1}^{M_j}P_{jk}+up^{M-1}(\theta_i)+l_M},$$

$$\sum_{j}^{(M)}=\frac{\sum_{k=1}^{2M}p_{jk}^{M-1}\mathrm{COV}_j(x_k)+\sum_{k=1}^{u_j}\mathrm{COV}_j(x'_{jk})}{\sum_{k=1}^{M}P_{jk}+up^{M-1}(\theta_i)+l_j}, \tag{23}$$

where $p^{M}(\theta_i)$ is the posterior distribution of the k category, $\mathrm{COV}_j(\bullet)$ is the covariance matrix, u is the number of unlabeled sample, j is the number of the label sample, l_j is the number of the label sample within the j class, and x'_{jk} is the k label sample within the j class. When the change of the likelihood function between two iterations goes below the threshold, we stop the iteration and export the parameters. Threshold is determined empirically as 0.06.



(a)

(b)

(c)

(d)

Figure 4: Sample frames from our datasets. The action labels in each dataset are as follows: (a) KTH dataset [33]: walking (a1), jogging (a2), running (a3), boxing (a4), and hand clapping (a5); (b) Weizmann dataset [21]: bending (a1), jumping jack (a2), jumping forward on two legs (a3), jumping in place on two legs (a4), running (a5), galloping sideways (a6), walking (a7), waving one hand (a8), and waving two hands (a9); (c)UCF sports action dataset [37]: diving(a1), golf swinging (a2), kicking (a3), lifting (a4), horseback riding (a5), running (a6), skating (a7), swinging (a8), and walking (a9); (d) our action dataset: walking (a1), jogging (a2), running (a3), boxing (a4), and handclapping (a5).

```
(1) Estimate a priori probabilities for each category by training set $D$.
(2) Calculate the mean and covariance matrix of trained samples for each category.
(3) Put the nonclassified samples into various categories of Bayesian discriminant.
```

Algorithm 1

5. Action Classification

We can recognize the human action by trained classified samples by the Bayes classified method [35, 36]:

$$p\left(Y_i \mid X_j\right) = \frac{p\left(X_j \mid Y_i\right)}{\sum p\left(X_j \mid Y_i\right) p\left(Y_i\right)}. \qquad (24)$$

Because our generation model is based on the assumption of a Gaussian mixture distribution, we can obtain the following equation:

$$p\left(X_i \mid Y_j\right) = \frac{1}{\sqrt{2\pi}\sqrt{\left|\sum p\left(Y_j \mid X_i\right)\right|}} \times \exp\left(\frac{1}{2}\left(X-\mu_Y\right)^T \sum^{-1}\left(Y-\mu_Y\right)\right), \qquad (25)$$

Table 2: Confusion matrix for the Weizmann dataset.

	a1	a2	a3	a4	a5	a6	a7	a8	a9
a1	**1.00**	0.01	0.02	0.00	0.20	0.00	0.10	0.05	0.02
a2	0.01	**0.96**	0.02	0.03	0.00	0.00	0.00	0.04	0.00
a3	0.00	0.00	**0.80**	0.10	0.13	0.00	0.02	0.01	0.00
a4	0.00	0.01	0.00	**0.95**	0.00	0.20	0.04	0.00	0.00
a5	0.00	0.01	0.00	0.00	**0.85**	0.00	0.00	0.30	0.02
a6	0.01	0.00	0.03	0.00	0.05	**0.91**	0.02	0.00	0.01
a7	0.00	0.03	0.00	0.00	0.01	0.00	**0.94**	0.00	0.02
a8	0.00	0.03	0.04	0.10	0.00	0.00	0.00	**0.98**	0.00
a9	0.00	0.00	0.20	0.00	0.10	0.00	0.00	0.03	**1.00**

Table 3: Confusion matrix for the UCF sports dataset.

	a1	a2	a3	a4	a5	a6	a7	a8	a9
a1	**0.97**	0.02	0.01	0.00	0.15	0.00	0.10	0.05	0.02
a2	0.00	**0.95**	0.01	0.00	0.00	0.02	0.00	0.03	0.00
a3	0.01	0.00	**0.82**	0.15	0.10	0.00	0.02	0.02	0.00
a4	0.00	0.00	0.00	**0.92**	0.10	0.10	0.00	0.00	0.00
a5	0.00	0.01	0.20	0.00	**0.88**	0.00	0.00	0.10	0.02
a6	0.01	0.00	0.02	0.00	0.05	**0.93**	0.05	0.01	0.02
a7	0.00	0.04	0.00	0.00	0.00	0.00	**0.92**	0.00	0.02
a8	0.00	0.02	0.03	0.10	0.00	0.00	0.00	**0.97**	0.00
a9	0.00	0.10	0.30	0.04	0.10	0.00	0.00	0.00	**1.00**

Table 4: Confusion matrix for our dataset.

	a1	a2	a3	a4	a5
a1	**0.98**	0.00	0.00	0.01	0.02
a2	0.00	**0.96**	0.01	0.00	0.00
a3	0.00	0.02	**0.87**	0.01	0.00
a4	0.00	0.20	0.00	**0.88**	0.02
a5	0.02	0.10	0.00	0.00	**0.86**

Table 5: Comparison with other approaches on KTH action dataset.

Method	Average recognition rate (%)
The proposed method	92.30
Martínez-Contreras et al. [38]	89.20
Chaaraoui et al. [39]	91.20
Zhang and Gong [40]	90.60

Table 6: Comparison with other approaches on the Weizmann action dataset.

Method	Average recognition rate (%)
The proposed method	89.10
Martínez-Contreras et al. [38]	85.10
Chaaraoui et al. [39]	87.20
Zhang and Gong [40]	85.40

where μ_Y is mean vector and $\sum p\left(Y_j \mid X_i\right)$ is the covariance matrix. The operation of the classifier is shown in Algorithm 1.

Table 7: Comparison with other approaches on UCF sportsaction dataset.

Method	Average recognition rate (%)
The proposed method	91.10
Martínez-Contreras et al. [38]	85.20
Chaaraoui et al. [39]	87.30
Zhang and Gong [40]	88.60

Table 8: Comparison with other approaches on our action dataset.

Method	Average recognition rate (%)
The proposed method	90.30
Martínez-Contreras et al. [38]	88.80
Chaaraoui et al. [39]	89.60
Zhang and Gong [40]	87.10

Therefore, we obtain the result of human recognition as follows:

$$Y = \arg_Y \max p\left(Y_i X_j\right). \tag{26}$$

6. Experimental Result

In this section, firstly, four action datasets are used for evaluating the proposed approach: Weizmann human motion dataset [21], the KTH human action dataset [33], the UCF sports action dataset [37], and our action dataset (Table 8). Secondly, we compare our method with some other popular methods under these action datasets. We use a Pentium

4 machine with 2 GB of RAM, and the implementation on MATLAB to experiment, similar to [3]. Representative frames of this dataset are shown in Figure 4.

6.1. Evaluation on KTH Dataset. The KTH dataset is provided by Schuldt which contains 2391 video sequences with 25 actors showing six actions. Each action is performed in 4 different scenarios, which contain some human actions (walking (a1), jogging (a2), running (a3), boxing (a4), and hand waving (a5)).

Representative frames of this dataset are shown in Figure 4(a). The classified results are shown in Table 1.

6.2. Evaluation on Weizmann Dataset. The Weizmann dataset is established by Blank, which contains 83 video sequences, showing nine different people, with each performing nine different actions including bending (a1), jumping jack (a2), jumping forward on two legs (a3), jumping in place on two legs (a4), running (a5), galloping sideways (a6), walking (a7), waving one hand (a8), and waving two hands (a9). Representative frames of this dataset are shown in Figure 4(b). The classified results are shown in Table 2.

6.3. Evaluation on UCF Sports Action Dataset. The UCF sports action dataset is as follows. This dataset consists of several actions from various sporting events from the broadcast television channels. The actions in this dataset include diving (a1), golf swinging (a2), kicking (a3), lifting (a4), horse-back riding (a5), running (a6), skating (a7), swinging (a8), and walking (a9). Representative frames of this dataset are shown in Figure 4(c). The classified results are shown in Table 3.

6.4. Evaluation on Our Action Dataset. Our action dataset is as follows.

We capture the behavior video in the laboratory. It contains five types of human actions (walking (a1), jogging (a2), running (a3), boxing (a4), and handclapping (a5)). Some sample frames are shown in Figure 4(d). The classified results achieved by this approach are shown in Table 4.

6.5. Algorithm Comparison. In this case, we compare the proposed method with the three methods: Martínez-Contreras et al. [38], Chaaraoui et al. [39], and Zhang and Gong [40] in four datasets. In Tables 5, 6, and 7, it is obvious that the low recognition accuracy existed in these methods for the complex occlusion situation and the complex beat, motion, and other group actions. The average accuracy in our method is higher than that in the comparative method.

The experimental results show that the proposed approach can get satisfactory results and overcome these problems by comparing the average accuracy with that in [38–40].

7. Conclusions and Future Work

In this paper, we present a novel method of human action recognition, which is based on the reconstructed phase space. Firstly, the human body is divided into 15 key points, whose trajectory represents the human body behavior, and the modified particle filter is used to track these key points for self-occlusion. Secondly, we reconstruct the phase space for extracting more useful information from human action trajectories. Finally, we can construct use the semisupervised probability model and Bayes classified method to classify. Experiments were performed on the Weizmann, KTH, UCF sports, and our action dataset to test and evaluate the proposed method. The compare experiment results showed that the proposed method can achieve was more effective than compare methods.

Our future work will deal with adding complex event detection by the phase space-based action representation and action learning and theoretical analysis of their relationship, involving more complex problems, such as dealing with more variable motion and interpersonal occlusions.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper (such as financial gain).

Acknowledgments

This research work was supported by the Grants from the Natural Science Foundation of China (no. 50808025) and the Doctoral Fund of China Ministry of Education (Grant no. 20090162110057).

References

[1] L. Tan, L. Xia, J. Huang, and S. Xia, “Human action recognition based on pLSA model,” *Journal of National University of Defense Technology*, vol. 35, no. 5, pp. 102–108, 2013.

[2] H.-B. Tu, L.-M. Xia, and L.-Z. Tan, “Adaptive self-occlusion behavior recognition based on pLSA,” *Journal of Applied Mathematics*, vol. 2013, Article ID 506752, 9 pages, 2013.

[3] H.-B. Tu, L.-M. Xia, and Z.-W. Wang, “The complex action recognition via the correlated topic model,” *The Scientific World Journal*, vol. 2014, Article ID 810185, 10 pages, 2014.

[4] J. K. Aggarwal and M. S. Ryoo, “Human activity analysis: a review,” *ACM Computing Surveys*, vol. 43, no. 3, pp. 1–42, 2011.

[5] Y. Sheikh, M. Sheikh, and M. Shah, “Exploring the space of a human action,” in *Proceedings of the 10th IEEE International Conference on Computer Vision (ICCV ’05)*, pp. 144–149, Beijing, China, October 2005.

[6] A. Yilmaz and M. Shah, “Recognizing human actions in videos acquired by uncalibrated moving cameras,” in *Proceedings of the 10th IEEE International Conference on Computer Vision (ICCV ’05)*, vol. 1, pp. 150–157, October 2005.

[7] N. Anjum and A. Cavallaro, “Multifeature object trajectory clustering for video analysis,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 18, no. 11, pp. 1555–1564, 2008.

[8] C. R. Jung, L. Hennemann, and S. R. Musse, “Event detection using trajectory clustering and 4-D histograms,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 18, no. 11, pp. 1565–1575, 2008.

[9] A. Hervieu, P. Bouthemy, and J.-P. Le Cadre, “A statistical video content recognition method using invariant features on object trajectories,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 18, no. 11, pp. 1533–1543, 2008.

[10] X. Wang, K. T. Ma, G.-W. Ng, and W. E. L. Grimson, "Trajectory analysis and semantic region modeling using a nonparametric bayesian model," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '08)*, pp. 1–8, Anchorage, Alaska, USA, June 2008.

[11] H. Wang, A. Kläser, C. Schmid, and C.-L. Liu, "Dense trajectories and motion boundary descriptors for action recognition," *International Journal of Computer Vision*, vol. 103, no. 1, pp. 60–79, 2013.

[12] J. Yu, M. Jeon, and W. Pedrycz, "Weighted feature trajectories and concatenated bag-of-features for action recognition," *Neurocomputing*, vol. 131, pp. 200–207, 2014.

[13] H.-K. Pao, J. Fadlil, H.-Y. Lin, and K.-T. Chen, "Trajectory analysis for user verification and recognition," *Knowledge-Based Systems*, vol. 34, pp. 81–90, 2012.

[14] Y. Yi and Y. Lin, "Human action recognition with salient trajectories," *Signal Processing*, vol. 93, no. 11, pp. 2932–2941, 2013.

[15] J.-X. Du, K. Yang, and C.-M. Zhai, "Action recognition based on the feature trajectories," in *Intelligent Computing Theories and Applications*, vol. 7390 of *Lecture Notes in Computer Science*, pp. 250–257, Springer, Berlin, Germany, 2012.

[16] A. Psarrou, S. Gong, and M. Walter, "Recognition of human gestures and behaviour based on motion trajectories," *Image and Vision Computing*, vol. 20, no. 5-6, pp. 349–358, 2002.

[17] N.-G. Cho, A. L. Yuille, and S.-W. Lee, "Adaptive occlusion state estimation for human pose tracking under self-occlusions," *Pattern Recognition*, vol. 46, no. 3, pp. 649–661, 2013.

[18] V. Maroulas and P. Stinis, "Improved particle filters for multi-target tracking," *Journal of Computational Physics*, vol. 231, no. 2, pp. 602–611, 2012.

[19] T. Penne, C. Tilmant, T. Chateau, and V. Barra, "Markov chain monte carlo modular ensemble tracking," *Image and Vision Computing*, vol. 31, no. 6-7, pp. 434–447, 2013.

[20] M. K. Pitt, R. dos Santos Silva, P. Giordani, and R. Kohn, "On some properties of Markov chain Monte Carlo simulation methods based on the particle filter," *Journal of Econometrics*, vol. 171, no. 2, pp. 134–151, 2012.

[21] L. Gorelick, M. Blank, E. Shechtman, M. Irani, and R. Basri, "Actions as space-time shapes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 12, pp. 2247–2253, 2007.

[22] S. de Martino, M. Falanga, and C. Godano, "Dynamical similarity of explosions at Stromboli volcano," *Geophysical Journal International*, vol. 157, no. 3, pp. 1247–1254, 2004.

[23] G. Paladin and A. Vulpiani, "Anomalous scaling laws in multifractal objects," *Physics Reports*, vol. 156, no. 4, pp. 147–225, 1987.

[24] S.-C. Fang and H.-L. Chan, "Human identification by quantifying similarity and dissimilarity in electrocardiogram phase space," *Pattern Recognition*, vol. 42, no. 9, pp. 1824–1831, 2009.

[25] S.-C. Fang and H.-L. Chan, "QRS detection-free electrocardiogram biometrics in the reconstructed phase space," *Pattern Recognition Letters*, vol. 34, no. 5, pp. 595–602, 2013.

[26] I. Nejadgholi, M. H. Moradi, and F. Abdolali, "Using phase space reconstruction for patient independent heartbeat classification in comparison with some benchmark methods," *Computers in Biology and Medicine*, vol. 41, no. 6, pp. 411–419, 2011.

[27] H. Wang, C. Yuan, G. Luo, W. Hu, and C. Sun, "Action recognition using linear dynamic systems," *Pattern Recognition*, vol. 46, no. 6, pp. 1710–1718, 2013.

[28] A. López-Méndez and J. R. Casas, "Model-based recognition of human actions by trajectory matching in phase spaces," *Image and Vision Computing*, vol. 30, no. 11, pp. 808–816, 2012.

[29] F. Takens, "Detecting strange attractors in turbulence," in *Dynamical Systems and Turbulence*, Lecture Notes in Mathematics, pp. 366–381, Springer, Berlin, Germany, 1981.

[30] L. Cao, "Practical method for determining the minimum embedding dimension of a scalar time series," *Physica D: Nonlinear Phenomena*, vol. 110, no. 1-2, pp. 43–50, 1997.

[31] S. Ali, A. Basharat, and M. Shah, "Chaotic invariants for human action recognition," in *Proceedings of the 11th IEEE International Conference on Computer Vision (ICCV '07)*, pp. 1–8, Rio de Janeiro, Brazil, October 2007.

[32] L.-M. Xia, J.-X. Huang, and L.-Z. Tan, "Human action recognition based on chaotic invariants," *Journal of Central South University*, vol. 20, no. 11, pp. 3171–3179, 2013.

[33] I. Laptev, *Local spatio-temporal image features for motion interpretation [Ph.D. thesis]*, Computational Vision and Active Perception Laboratory (CVAP), NADA, KTH, Stockholm, Sweden, 2004.

[34] R. Guangbo, Z. Jie, M. Yi, and Z. Rong'er, "Generative model based semi-supervised learning method of remote sensing image classification," *Journal of Remote Sensing*, vol. 14, no. 6, pp. 1097–1104, 2010.

[35] N. F. Lepora, M. Evans, C. W. Fox, M. E. Diamond, K. Gurney, and T. J. Prescott, "Naive Bayes texture classification applied to whisker data from a moving robot," in *Proceedings of the International Joint Conference on Neural Networks (IJCNN '10)*, pp. 1–8, July 2010.

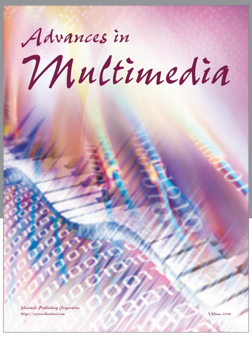
[36] L. Liu, L. Shao, and P. Rockett, "Human action recognition based on boosted feature selection and naive Bayes nearest-neighbor classification," *Signal Processing*, vol. 93, no. 6, pp. 1521–1530, 2013.

[37] M. D. Rodriguez, J. Ahmed, and M. Shah, "Action MACH: a spatio-temporal maximum average correlation height filter for action recognition," in *Proceedings of the 26th IEEE Conference on Computer Vision and Pattern Recognition (CVPR '08)*, Anchorage, Alaska, USA, June 2008.

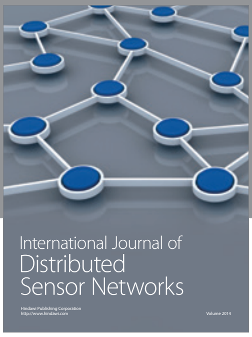
[38] F. Martínez-Contreras, C. Orrite-Uruñuela, E. Herrero-Jaraba, H. Ragheb, and S. A. Velastin, "Recognizing human actions using silhouette-based HMM," in *Proceedings of the 6th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS '09)*, pp. 43–48, Genova, Italy, September 2009.

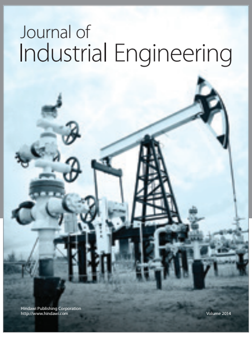
[39] A. A. Chaaraoui, P. Climent-Pérez, and F. Flórez-Revuelta, "Silhouette-based human action recognition using sequences of key poses," *Pattern Recognition Letters*, vol. 34, no. 15, pp. 1799–1807, 2013.

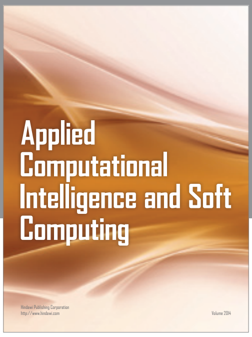
[40] J. Zhang and S. Gong, "Action categorization by structural probabilistic latent semantic analysis," *Computer Vision and Image Understanding*, vol. 114, no. 8, pp. 857–864, 2010.

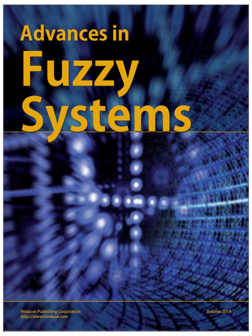

Advances in
Multimedia

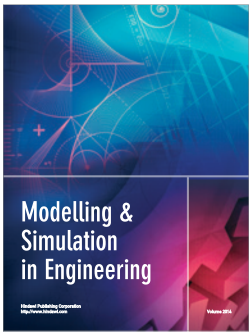
The Scientific
World Journal

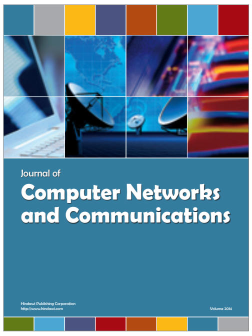

International Journal of
Distributed
Sensor Networks

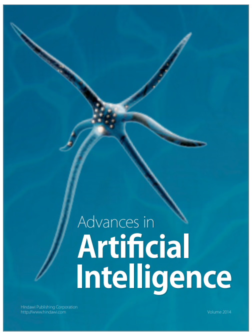

Journal of
Industrial Engineering

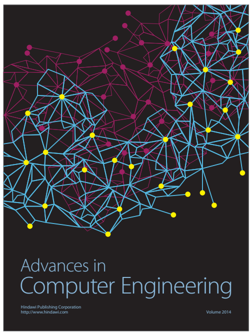

Applied
Computational
Intelligence and Soft
Computing


Advances in
Fuzzy
Systems

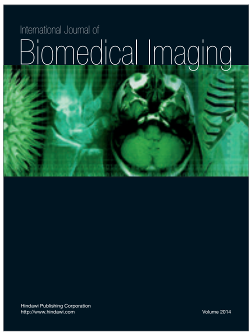

Modelling &
Simulation
in Engineering

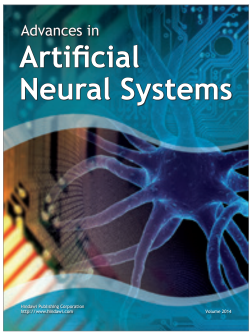

Journal of
Computer Networks
and Communications

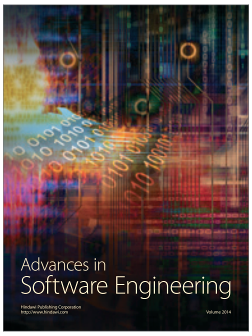

Advances in
Artificial
Intelligence


Advances in
Computer Engineering

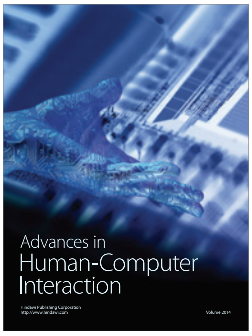
International Journal of
Computer Games
Technology

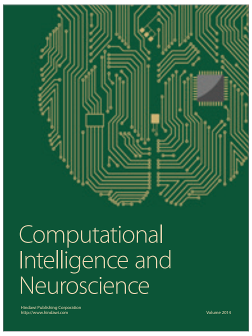

International Journal of
Biomedical Imaging


Advances in
Artificial
Neural Systems


Advances in
Software Engineering

Journal of
Robotics


Advances in
Human-Computer
Interaction


Computational
Intelligence and
Neuroscience

International Journal of
Reconfigurable
Computing

Journal of
Electrical and Computer
Engineering