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; tuhongbin30@163.com

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

Academic Editor: Shifei Ding

Copyright © 2014 H.-b. Tu and L.-m. Xia. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

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
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
user verification approach based on user trajectories, which
are used to represent different actions. Therefore, we must track
these trajectories 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: $Z_{K_0}, \ldots, Z_{K_N}$: 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 $E[f(X_{K_n} \mid \{Z_{K_i}\}_{i=1}^N]$ 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 $p(X_{T_k} \mid
\{Z_{K_i}\}_{i=1}^k)$ and define a reference density: $q(X_{K_n} \mid \{Z_{K_i}\}_{i=1}^N$.
At last, we obtain the weighted sample [18]:

$$
\sum_{n=1}^{N} f(X_{K_n} \mid \{Z_{K_i}\}_{i=1}^N) \frac{p(X_{K_n} \mid \{Z_{K_i}\}_{i=1}^N)}{\sum_{n=1}^{N} q(X_{K_n} \mid \{Z_{K_i}\}_{i=1}^N)}
$$

(1)
\[
p(X_n | Z_{K_n}) = \frac{g(x_{K_n}, Z_{K_n})}{q(x_{K_n}, Z_{K_n})} p(X_{K_n} | Z_{K_n}^{N}) q(x_{K_n}, Z_{K_n})^{N},
\]
\[
(2)
\]
\[
\begin{align*}
& p(X_n | Z_{K_n}^{N}) \\
& = \int p(X_n | X_{K_n}) \frac{p(X_{K_n} | Z_{K_n}^{N-1})}{q(x_{K_n}, Z_{K_n})} dX_{K_n}.
\end{align*}
\]
\[
(3)
\]
We assume that
\[
q(x_{K_n}, Z_{K_n}^{N}) \propto p(X_{K_n} | Z_{K_n}^{N-1}).
\]
and, from (2), we can obtain the formula
\[
\begin{align*}
& \frac{p(X_n | Z_{K_n}^{N})}{q(x_{K_n}, Z_{K_n})} \propto g(x_{K_n}, Z_{K_n}).
\end{align*}
\]
\[
(4)
\]
The approximation in expression (1) becomes
\[
\begin{align*}
E\left[f(x_{K_n} | Z_{K_n}^{N})\right] & = \frac{\sum_{n=1}^{N} f(x_{K_n}^{n}, Z_{K_n}^{n}) g(x_{K_n}^{n}, Z_{K_n}^{n})}{\sum_{n=1}^{N} g(x_{K_n}^{n}, Z_{K_n}^{n}) g(x_{K_n}^{n}, Z_{K_n}^{n})}.
\end{align*}
\]
\[
(5)
\]
The tracking algorithm is described as follows.

(1) Sampling \( N \) particles in accordance with the unified weights randomly generated particles form unweighted samples \( X_n^{N} \) and determination \( p(X_{K_n} | Z_{K_n}^{N-1}) \), as follows:
\[
\begin{align*}
P(X_n | Z_{K_n}^{N-1}) & = \prod_{\lambda=1}^{N} p(X_{K_n}^{\lambda} | Z_{K_n}^{\lambda-1}^{N-1}) .
\end{align*}
\]
(2) Predict by sampling \( X_n^{N} \) from
\[
\begin{align*}
P(X_n | X_{K_n-1}) & = \prod_{\lambda=1}^{N} p(X_{K_n}^{\lambda} | X_{K_n-1}^{\lambda}) .
\end{align*}
\]
(3) Target observation association.
(4) Update and evaluate the weights:
\[
\begin{align*}
& w_{K_n}^{n} \propto \frac{\prod_{\lambda=1}^{N} q(X_{K_n}^{\lambda} | Z_{K_n}^{\lambda-1}) g_{\lambda}(X_{K_n}^{\lambda}, Z_{\lambda_T})}{\sum_{n=1}^{N} \prod_{\lambda=1}^{N} g_{\lambda}(X_{K_n}^{\lambda}, Z_{\lambda_T})}.
\end{align*}
\]
(5) By resampling, through the above steps, we can generate independent uniform random variables \( \{|\theta_{i}^{N}\}_{i=1}^{N} \) \((0 < |\theta_{i}^{N}| < 1)\). Therefore, we can obtain the following equation:
\[
\begin{align*}
& (X_{K_n}^{\lambda}, X_{K_n}^{\lambda}) = (X_{K_n}^{\lambda}, X_{K_n}^{\lambda}),
\end{align*}
\]
(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
\[
\begin{align*}
& \prod_{\lambda=1}^{N} g_{\lambda}(Y_{\lambda}^{n}, Z_{\lambda_T}) p_{\lambda}(Y_{\lambda}^{n} | X_{K_n}^{n}) .
\end{align*}
\]
(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 \( \tau \) 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 \( x(t) \)
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 a dynamical system as the possibly nonlinear map, which represents the temporal evolution of state variables

$$x(t) = [x_1(t), x_2(t), \ldots, x_m(t)] \in \mathbb{R}^m.$$   \hspace{2cm} (13)

de Martino et al. \cite{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 \cite{26}. Consider

$$x_n = [x_n, x_{n-\tau}, \ldots, x_{n-(d-1)\tau}] n = (1 + (d - 1)\tau), \ldots, N,$$   \hspace{2cm} (14)

where $x_n$ is the $n$th point in the time series, delay times $\tau$ 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, $\eta$ 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 \cite{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 \cite{26, 28} are defined as follows:

$\eta$: the temporal evolution;

$Y_{K_n}^{\eta}$: time series (scalar), and we want to characterize

$$Y_{K_n}^{\eta} = [z^\eta(t), z^\eta(t + \tau), \ldots, z^\eta(t + (m - 1)\tau)].$$   \hspace{2cm} (15)

$Y_{K_n}^{\eta}$ is a point in the reconstructed phase space, $m$ is the embedding dimension, and $\tau$ 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 $\tau$ determines the properties of the reconstructed phase space.

At first, the embedding delay using the mutual information method was determined \cite{26} and the estimated delay was used to obtain the appropriate embedding dimension \cite{30}. Once both the embedding delay and the embedding dimension have been estimated, is performed \cite{26} as follows:
$$\hat{X}^n = \begin{bmatrix}
Y_{K_n}^n 1+(d-1)\tau \\
Y_{K_n}^n 2+(d-1)\tau \\
\vdots \\
Y_{K_n}^n m(d-1)\tau
\end{bmatrix}$$

$$= \begin{bmatrix}
\frac{z_{ij}^n 1+(d-1)\tau}{\tau}, \ldots, \frac{z_{ij}^n (m-1)\tau}{\tau}, \ldots, \frac{z_{ij}^n (m-1)\tau}{\tau} \\
\frac{z_{ij}^n (t+\tau)}{\tau}, \ldots, \frac{z_{ij}^n (t+(m-1)\tau)}{\tau}, \ldots, \frac{z_{ij}^n (t+(m-1)\tau)}{\tau} \\
\vdots \\
z_{ij}^n (N-1-(m-1)\tau), \ldots, z_{ij}^n (N-1-(m-2)\tau), \ldots, z_{ij}^n (N-1)
\end{bmatrix}$$

We use the phase space $\hat{X}^n$ 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]:

$$H_k = \lim_{\tau \to 0} \lim_{t \to -\infty} \frac{1}{t} \left[ \sum_{i=1}^{N_k} P_i \sum \log \left( \frac{1}{P_i} \right) \right].$$

(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:

$$p(x \theta_i) = \sum_{j=1}^{C} P_i \ p_i(x \theta_i).$$

(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, $p_i (\theta_i)$ is the probability of the sample belonging to the $i$th class, $\theta_i$ represents the object classes and the covariance matrix of pixel. Therefore, likelihood functions [34] were defined as follows:

$$\log p(\theta_i \mid X) = \log \left( \prod_{i=1}^{M} p(\theta_i) p(x_i \mid \theta_i) \prod_{i=M+1}^{2M} p(x_i \mid \theta) \right)$$

$$= \sum_{i=1}^{M} \log p(\theta_i) p(x_i \mid \theta_i) + \sum_{i=M+1}^{2M} \log p(x_i \mid \theta).$$

(19)

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

Unsupervised part should be written as

$$\sum_{i=M+1}^{2M} \log p(x_i \mid \theta) = \sum_{i=M+1}^{2M} \sum_{j=1}^{C} p(x) p(x_i \mid \theta_j).$$

(20)

Finally, we can obtain the log-likelihood function

$$\log p(\theta_i \mid X) = \log \left( \prod_{i=1}^{M} p(\theta_i) p(x_i \mid \theta_i) \prod_{i=M+1}^{2M} p(x_i \mid \theta) \right)$$

$$+ \sum_{i=M+1}^{2M} \sum_{j=1}^{C} p(x) p(x_i \mid \theta_j).$$

(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)

$$\hat{p}_j^{M} = \frac{\theta_j^M p(x_i \mid \theta_j^M-1)}{\sum_{j=1}^{M} \theta_j^M p(x_i \mid \theta_j^M-1)}.$$ 

(22)
The phase space reconstruction of walking

The phase space reconstruction of jogging

The phase space reconstruction of running

Table 1: Confusion matrix for KTH dataset.

<table>
<thead>
<tr>
<th></th>
<th>a1</th>
<th>a2</th>
<th>a3</th>
<th>a4</th>
<th>a5</th>
</tr>
</thead>
<tbody>
<tr>
<td>a1</td>
<td>0.95</td>
<td>0.01</td>
<td>0.02</td>
<td>0.00</td>
<td>0.01</td>
</tr>
<tr>
<td>a2</td>
<td>0.01</td>
<td>0.93</td>
<td>0.02</td>
<td>0.10</td>
<td>0.00</td>
</tr>
<tr>
<td>a3</td>
<td>0.00</td>
<td>0.02</td>
<td>0.90</td>
<td>0.00</td>
<td>0.01</td>
</tr>
<tr>
<td>a4</td>
<td>0.01</td>
<td>0.00</td>
<td>0.00</td>
<td>0.92</td>
<td>0.30</td>
</tr>
<tr>
<td>a5</td>
<td>0.03</td>
<td>0.00</td>
<td>0.02</td>
<td>0.00</td>
<td>0.82</td>
</tr>
</tbody>
</table>

$M^{-1}$ 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:

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

where $p^M(\theta_j)$ is the posterior distribution of the $k$ category, $\text{COV}_j(\bullet)$ is the covariance matrix, $\mu$ 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.
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 hand clapping (a5).

Algorithm 1

5. Action Classification

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

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

\[
p(X_j | Y_j) = \frac{1}{\sqrt{2\pi} \sqrt{\sum p(Y_j | X_i)}} \times \exp \left( \frac{1}{2} (X - \mu_j)^T \sum^{-1} (Y - \mu_j) \right),
\]
where $\mu_Y$ is mean vector and $\sum_p (Y_j | X_i)$ is the covariance matrix. The operation of the classifier is shown in Algorithm 1.

### 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

<table>
<thead>
<tr>
<th>Method</th>
<th>Average recognition rate (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>The proposed method</td>
<td>91.10</td>
</tr>
<tr>
<td>Martinez-Contreras et al. [38]</td>
<td>85.20</td>
</tr>
<tr>
<td>Chaaraoui et al. [39]</td>
<td>87.30</td>
</tr>
<tr>
<td>Zhang and Gong [40]</td>
<td>88.60</td>
</tr>
</tbody>
</table>

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

$$Y = \arg\max_i \max_j p(Y, X_j).$$

(26)
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 hand clapping (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: Martinez-Contreras et al. [38], Chaarou 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 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


Submit your manuscripts at http://www.hindawi.com