Capturing Global and Local Dynamics for Human Action Recognition

Siqi Nie
Department of Electrical, Computer
and System Engineering
Rensselaer Polytechnic Institute
Troy, New York 12180
nies@rpi.edu

Qiang Ji
Department of Electrical, Computer
and System Engineering
Rensselaer Polytechnic Institute
Troy, New York 12180
qji@ecse.rpi.edu

***Abstract*—Human action analysis has achieved great success especially with the recent development of advanced sensors and algorithms that can effectively track the body joints. Temporal motion of body joints carries crucial information about human actions. However, current dynamic models typically assume stationary local transition and therefore are limited to local dynamics. In contrast, we propose a novel human action recognition algorithm that is able to capture both global and local dynamics of joint trajectories by combining a Gaussian-Binary restricted Boltzmann machine (GB-RBM) with a hidden Markov model (HMM). We present a method to use RBM as a generative model for multi-class classification. Experimental results on benchmark datasets demonstrate the capability of the proposed method in exploiting the dynamic information at different levels.**

I. Introduction

Human action is the combination of the movements of body joints over a time interval. Understanding a complex action requires studying not only the spatial configurations among the body joints, but also how they move at different time scales in the time domain. Capturing the movements of the body joints used to be a difficult task, which significantly limited the performance of previous video-based human action recognition, until the recent emergence of low-cost and reliable depth sensors such as Kinect and efficient pose tracking systems [18] that can provide well-estimated joint positions in real time. Joint trajectories present a more explicit representation of the action dynamics. However, these temporal characteristics of human actions have not yet been thoroughly exploited, partially due to the limitations of current models. In this work, we interpret a human action as a set of 3D trajectories of dominant body joints. We comprehensively investigate the underlying temporal dynamics of these trajectories for action recognition.

Modeling the temporal patterns of body joints of a complex human action is generally addressed by extracting bottom-level spatio-temporal features from the image sequences or designing top-level dynamic models such as hidden Markov model (HMM), dynamic Bayesian network (DBN) or conditional random field (CRF). Time-sliced dynamic models generally assume n^{th} order Markov property and stationary transition. They, hence, can only capture local stationary transitions but cannot represent global moving pattern. Moreover, these assumptions may not hold for many real-world applications. Spatio-temporal features are typically based on local interest points and therefore are also not able to describe the movement pattern throughout the whole action process.

Compared to time-sliced dynamic models, restricted Boltzmann machine (RBM) has been demonstrated to have strong power to capture the joint distribution of the inputs and therefore can be used to model the global patterns when the input is a time sequence of joint positions. To the best of our knowledge, RBM has not yet been applied to analyze the global dynamics of trajectories for action recognition, although it has been widely used in many other applications such as image and document analysis.

To comprehensively model the temporal dynamics of human actions at different levels, we propose a hybrid approach that combines a Gaussian-Binary restricted Boltzmann machine (GB-RBM) to capture the global movement patterns with an HMM to capture the local dynamics. As GB-RBM is a variation of the standard RBM, we use the term RBM in the following sections to represent our model.

The local and global models capture complementary dynamic information at different time scales and are combined through a fusion approach for action classification. A detailed illustration of the framework is given in Figure 1.

The remainder of the paper is organized as follows. Section II presents an overview of the related work. Section III introduces the learning process of RBM for action representation. Section IV demonstrates the fusion method for global and local dynamic models. Experimental results are given in Section V. The paper is concluded in Section VI.

II. Related Work

Human activity recognition has been widely investigated in the past few decades. Depending on the action complexity, human actions are categorized into four different levels: gestures, actions, interactions and group activities [1]. In this work, we focus on action recognition, *i.e.*, a single person's activities that may be composed of multiple gestures organized temporally, such as 'waving', 'running', and 'jumping'. Research in action recognition generally follows two paths: single-layered approaches and hierarchical approaches. Single-layered approaches recognize human actions directly from

1051-4651/14 $31.00 © 2014 IEEE
DOI 10.1109/ICPR.2014.340

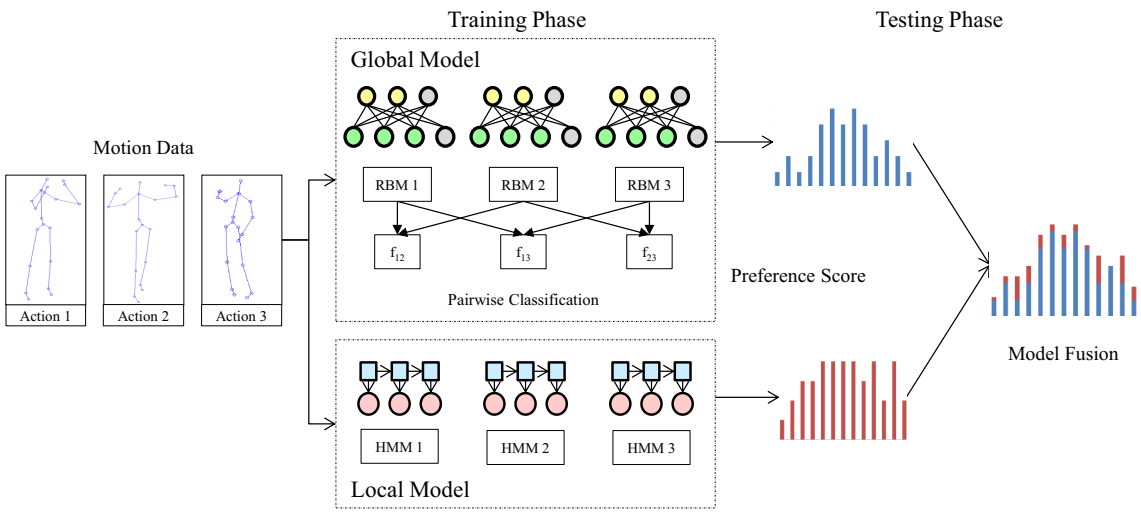

Fig. 1. Framework of the proposed method. For each class of action, one RBM and one HMM are trained to represent the global and local dynamics respectively. Each pair of RBMs $\mathcal{M}_i$ and $\mathcal{M}_j$ forms a pairwise classifier, which gives a preference score toward action class i or j. The preference scores of RBM and HMM are combined to make the final prediction.

sequential images, while hierarchical approaches represent actions with simpler sub-actions.

In single-layered approaches, sequence of images may be considered as 3D volume [16], trajectories [15], or spatio-temporal features. The most widely used spatio-temporal features for visible videos are histogram of gradients (HOG) and histogram of flows (HOF), which capture the local appearance or motion information. Features from depth images and joint trajectories have also been developed recently with the development of inexpensive and reliable depth sensors. For instance, Wang *et al.* [20] propose an LOP features which are the frequency coefficients of Fourier Transform of local features extracted from the depth images around human joints. Given some specifically designed features, template matching [16], neighborhood-based method [22] and other models are typically used to make predictions.

Hierarchical approaches typically include statistical approaches and description-based approaches. Statistical approaches construct statistical state-based models that are concatenated hierarchically. Conditional random fields [7] and hidden Markov models [12] are common examples of statistical models. These dynamic models, either generative or discriminative, assume stationary transition and hence are only able to capture local temporal interactions between several consecutive frames. Description-based approaches divide human actions into sub-events. Prediction is made by modeling the temporal and spatial relationship of sub-events [6].

Restricted Boltzmann machine and its variants are generally used as a tool for feature learning or data pre-processing yet could also be used for modeling the motion data. For example, Wang *et al.* [21] uses RBM to get a prior probability for finger trace. Larochelle and Bengio [10] uses RBM to generate features for character recognition. Our work is inspired by the idea of Taylor *et al.* [19], where a Conditional RBM (CRBM) is proposed to model the temporal transitions between consecutive time slices and generate pseudo movement sequences. Still, CRBM models local dynamics by assuming n^{th} order Markov property. Conditioned on previous slices, it models the information of the current time slice.

Unlike these works, RBM is used as a generative model in this research, which models the high dimensional sequential data and returns the likelihood of the input. Moreover, RBM is combined with an HMM to jointly capture the global and local dynamics of human actions. By utilizing an approach to estimate the relative partition functions of RBMs, we are able to compare between different RBMs, and thus make predictions.

III. Modeling Trajectories

In this work, we propose to capture the global patterns of human joint trajectories using the *restricted Boltzmann machine* (RBM). We choose RBM due to its capability to model complex patterns in high dimensional data. One RBM is learned to capture the global moving pattern of one type of action. In this section, we will firstly give a brief introduction of restricted Boltzmann machine. We will then introduce how to use RBM to model a sequence of motion data. An approach to estimate the partition function of RBM is then proposed to perform classification among multiple actions.

A. Restricted Boltzmann Machine

A restricted Boltzmann machine (RBM) is a generative stochastic neural network that can learn a probability distribution over a set of inputs. As shown in Figure 2, all the neurons form a bipartite graph: they have input units, corresponding to data, hidden units that are learned, and each connection in an RBM must connects a visible unit to a hidden unit. In our work, the hidden units are binary and the visible variables are assumed to follow normal distribution.

The energy function $E(\mathbf{v}, \mathbf{h})$ is parameterized in Equation 1. Variable a_i is the bias, σ_i is the standard deviation of the Gaussian distribution for visible unit v_i. If the data is normalized in each dimension, then $\sigma_i = 1, a_i = 0$. b_j is the bias of the hidden unit h_j. The joint distribution of the visible and hidden variables is given in Equation 2. With continuous inputs, the partition function Z can be computed by the integral over all visible nodes and summation over all hidden units.

$$E(\mathbf{v}, \mathbf{h}) = \sum_i \frac{(v_i - a_i)^2}{2\sigma_i^2} - \sum_{ij} \frac{v_i}{\sigma_j} w_{ij} h_j - \sum_i b_j h_j \,, \quad (1)$$

$$p(\mathbf{v}, \mathbf{h}) = \frac{1}{Z} \exp\left(-E(\mathbf{v}, \mathbf{h})\right) \,. \quad (2)$$

The probability of an observation can be calculated by marginalizing over the hidden variables, as shown in Equation 3.

$$p(\mathbf{v}) = \frac{1}{Z} \sum_{\mathbf{h}} \exp\left(-E(\mathbf{v}, \mathbf{h})\right) \,. \quad (3)$$

Hidden units in the RBM have two states: on and off. Given an input vector $\mathbf{v}$, the binary state h_j is set on with probability

$$p(h_j = 1|\mathbf{v}) = \sigma(b_j + \sum_i v_i w_{ij}) \,, \quad (4)$$

where $\sigma(x)$ is the logistic sigmoid function $1/(1+\exp(-x))$.

A hidden unit h_j is connected to all the inputs, so it is activated when there exists some specific pattern in the visible layer through Equation 4. The pattern is captured by the weights connecting each element in $\mathbf{v}$ to h_j. Thus the hidden layer $\mathbf{h}$ represent important patterns of $\mathbf{v}$.

Parameters of RBM include the weights of the connections between the hidden units and visible units as well as their biases. They are usually learned using the Contrastive Divergence (CD) [3] method to get an approximate Maximum-Likelihood solution.

B. Modeling Actions using RBM

Typically, the input to the RBM is the limited to a single image. As we interpret an action as a combination of the 3D trajectories of human joints, we propose to use RBM to model the whole sequence of an action. The basic idea is to feed the joint positions along the temporal trajectory as the inputs to RBM, as shown in Figure 3, where the tth visible variable corresponds to the joint positions at time slice t. Given a total of N actions, N RBM's $\{\mathcal{M}_1, \mathcal{M}_2, \cdots, \mathcal{M}_N\}$ are learned, with each model $\mathcal{M}_i$ learning the temporal dynamics for action $\mathcal{A}_i$.

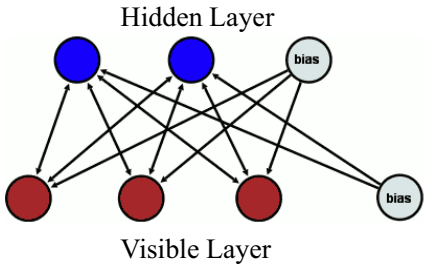

Fig. 2. Graphical Illustration of RBM

RBM can be efficiently learned using the contrastive divergence algorithm (CD) [3]. However parameter estimation of an RBM still faces one or more of the following challenges. Due to the non-convexity property of RBM, only local optimal solutions can be achieved. Different initializations could end up with different estimated parameters. Moreover, parameters are estimated in a generative manner and therefore it does not necessarily benefit action classification. We propose a model selection approach to simultaneously address all the above issues.

Model selection is performed for every action in turn. Consider selecting a model for the ith action, we first generate K candidate RBMs $\{M_i^k : k = 1, \cdots, K\}$ from different initializations. These RBM candidates are then evaluated on the training set $\{V_i^j\}$, with V_i^j representing the jth sample of the ith action. The score of each model M is defined in Equation 5, where $E(V_i^j|M)$ corresponds to the energy of V_i^j on model M. The basic idea is that we hope the selected model can maximally differentiate the samples of the ith action from other actions in terms of their likelihood. Finally the model that produces the highest score is selected as the model for the ith action. The procedure is repeated for N times until all the models are selected.

$$Score(M) = \frac{\sum_j \exp(-E(V_i^j|M))}{\sum_i \sum_j \exp(-E(V_i^j|M))} \,. \quad (5)$$

The difference between using energy function and likelihood function is the partition function. Since we are computing the likelihood based on one single model, the partition function is a constant, which can be omitted in Equation 5.

With the RBM learned for each action, it is infeasible to compare between models, because calculating the likelihood requires calculating the partition functions, which is intractable for RBM with large number of hidden units. Nevertheless, there still exists a method to estimate the relative partition function between different RBM's. For binary classification, Schmah *et al.* [17] propose a method to discriminatively estimate the difference of log-partition functions of two RBMs.

$$t_{ij} = \log Z_i - \log Z_j \,. \quad (6)$$

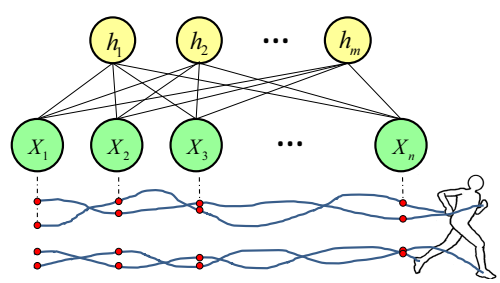

Fig. 3. Modeling Actions with RBM

We extend this approach to multi-class classification with a label ranking procedure [9] (Section IV).

C. Local Dynamic Model

Local dynamic models capture the local interactions between consecutive frames. In this work we implement hidden Markov model as a local dynamic model.

An HMM is defined by the prior of the hidden states, the transition probabilities and the emission probabilities. The well-known Expectation-Maximization algorithm (EM) [14] can be employed to estimate the parameters. The hidden states in HMM are generalization of the input sequence. For instance, if the inputs are actual joint positions, then the hidden states represent some specific joint positions which are crucial in the sequence. In this way a sequence can be transformed into a sequence of states. To recognize actions, we follow the same procedure as RBM and learn a group of HMM's, each of which corresponds to one action. Given the query sample, its likelihood for each HMM is calculated using the Forward-Backward procedure.

HMM is treated as a local dynamic model because it assumes stationary transition and Markov property of the states. We only consider the transition between two consecutive frames.

The score of HMM is simply the likelihood of the observation, which is easy to compute using the Forward-Backward algorithm.

IV. Fusion of Global and Local models

In this section, we introduce how we transform unnormalized likelihood of RBM into confidence score, and together with likelihood of HMM for action recognition.

A standard procedure to classify a query sample $\mathbf{v}$ is to compute its likelihood for all the models, and choose the model with the greatest likelihood, as shown in Equation 7.

$$y = \arg\max_i p(\mathbf{v}|\mathcal{M}_i)\,, \tag{7}$$

where y is the predicted result for instance $\mathbf{v}$.

Let $p'(\mathbf{v})$ denote the unnormalized likelihood in RBM with $\log p(\mathbf{v}) = \log p'(\mathbf{v}) - \log Z$. A confidence score for a sequence is defined as:

$$\mathcal{F}_{ij}(\mathbf{v}) = \frac{1}{1+\exp(-\alpha(\log p'(\mathbf{v}|\mathcal{M}_i) - \log p'(\mathbf{v}|\mathcal{M}_j) - t_{ij}))}\,, \tag{8}$$

where parameter α modifies the distribution of the score, in case all the scores are too close to 0 or 1.

The output of such "soft" binary classifier can be interpreted as a confidence value in the classification: the closer the output $\mathcal{F}_{ij}$ to 1, the stronger the decision of choosing action $\mathcal{A}_i$ is supported. A valued preference relation $\mathcal{R}_{\mathbf{v}}$ is defined for any query instance $\mathbf{v}$:

$$\mathcal{R}_{\mathbf{v}}(i,j) = \begin{cases} \mathcal{F}_{ij}(\mathbf{v}) & if \quad i<j \\ 1-\mathcal{F}_{ij}(\mathbf{v}) & if \quad i>j \end{cases}\,. \tag{9}$$

In our approach, we evaluate the score as sum all the confidence value

$$\mathcal{S}_{\mathbf{v}}(i) = \sum_{j\neq i} \mathcal{R}_{\mathbf{v}}(i,j)\,. \tag{10}$$

The global and local temporal information can be integrated at different levels of the learning process. In this paper we propose to combine them in the prediction phase. The score of RBM and HMM models are linearly combined (Equation 11) with a tuned weight ω, which maximize the recognition accuracy on a validation set, and the label with the highest score is proposed as the final decision.

$$S(\mathbf{v}) = S_{\text{RBM}}(\mathbf{v}) + \omega S_{\text{HMM}}(\mathbf{v})\,. \tag{11}$$

V. Experiments

We evaluate our algorithm on three datasets: MSRC-12 Kinect gesture dataset [5], G3D dataset [2], and MSR Action3D dataset [11]. Models that will be compared in the experiments include a global model RBM, a local model HMM, and the combined model. We will also compare our proposed approach with other related works.

A. MSRC-12 Dataset

The Microsoft Research Cambridge-12 Kinect gesture dataset consists of sequences of human movements, represented as body-part locations, and the associated gesture to be recognized by the system. The data set includes 594 sequences and 719,359 frames collected from 30 people performing 12 gestures. The motion files contain tracks of 20 joints estimated using the Kinect Pose Estimation pipeline. The body poses are captured at a sample rate of 30Hz with an accuracy of about two centimeters in joint positions.

To deal with the tracking noise, anisotropic diffusion [13] is employed to smooth the trajectories, correcting noise, yet preserving meaningful changes in motion. Figure 4 illustrates an example of anisotropic diffusion.

As the size of the visible layer of an RBM is fixed, linear interpolation is performed to convert all sequences into the same length (20 frames for each sequence). In this work, we only use the 3D location information of four dominant joints (*i.e.*, two hands and two feet) due to the limited number of samples. However the proposed approach can be applied to model more joints if training data are adequate. The 3D positions of the body joints along all three dimensions (x, y and z) are concatenated as the 240-dimension input vectors for the RBM model. The size of hidden layer is set to be 150 according to the suggestion in Hinton [8].

The dataset is constructed both to measure the performance of recognition systems and evaluate various methods of teaching human subjects how to perform different actions. So it is partitioned along different methods of instruction, such as text-only or text and video. In our work, different instructions are

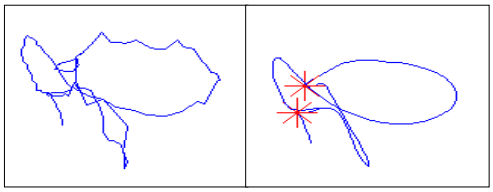

Fig. 4. Illustration of the pre-processing of trajectories. The left figure shows the original tracking result of a joint and the right figure shows the processed trajectory after anisotropic diffusion smoothing. It is clear that the high frequency noise can be effectively removed while the turning points in a trajectory are reserved.

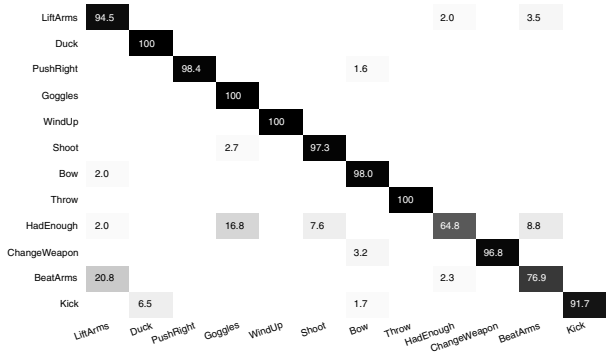

Fig. 5. The confusion matrix of the proposed method on MSRC-12 dataset

ignored and only video-based actions are selected to evaluate the performance of the proposed algorithm. The action markers provided with the dataset are used to segment the actions from long sequences.

4-fold cross-subject validation configuration is used in our experiment. Detailed results are shown in Table I, and the confusion matrix is shown in Figure 5. The local dynamic model achieves a recognition accuracy of 85.2%, while the global dynamic model reaches 89.8%. This demonstrates the importance of incorporating global dynamics. Combining global and local dynamic models, the proposed method can achieve an even better recognition accuracy of 93.1%. In particular our method outperforms the state-of-the-art method as reported in Ellis *et al.* [4]. According to the confusion matrix, our algorithm performs pretty well on most of the actions, and only fails on a small portion of actions such as *Had Enough* and *Lift Arms*.

B. G3D Dataset

G3D dataset is an action dataset containing a range of gaming actions captured by Microsoft Kinect. The dataset contains 10 subjects performing 20 gaming actions. Synchronized video, depth and skeleton data are available in this dataset. We only use the skeleton information in the experiment.

The action segmentation is manually labeled. The input vectors are extracted following the same procedure in Section V-A. Half of the samples are used as testing data, 5 samples from each action as validation data, and the other samples as training data.

To compare with the baseline method [2], we compute the F1 score for each category of actions. The result is shown in Table II. The proposed method outperforms baseline model for

TABLE I. PERFORMANCE COMPARISON OF DIFFERENT METHODS ON MSRC-12 DATASET

Method	Accuracy
Hidden Markov Model	85.2%
Ellis *et al.* [4]	88.7%
RBM	89.8%
Proposed Method	**93.1%**

most of the actions, but also encounters some failures in the actions of *Tennis* and *ThrowBowlingBall*. The reason is that when there is occlusion of the body parts, the Kinect tracker may fail occasionally, and gives the inferred results that will affect the accuracy, which is the case of *TennisSwingBackhand, Golf* and *ThrowBowlingBall*. Especially in *Golf* action, only one side of the subject can be seen by the camera, so the position of one leg is inferred using the tracking procedure, which brings in much trouble. Also, the movement range of the action *Walk* and *Jump* is relatively small, and may be confused with each other. However, the overall accuracy of our algorithm is acceptable. From Table III, the combined model outperforms the global and local models, as expected.

C. MSR Action3D Dataset

MSR Action3D Dataset is dataset of 20 actions including both depth images and skeleton tracking results. The dataset reasonably cover the various movements of arms, legs, torso and their combinations. Each action is performed by 10 subjects, repeated 2 or 3 times. There are 567 sequences all together. We use the same four joint positions as our features. Following the same cross-subject setting as [11], 5 subjects for each action are selected for testing. For the remaining 5 subjects, 4 are used for training and 1 is used for validation.

This dataset poses many challenges for recognition: there exist small between-class variations (*e.g., high arm wave* and *horizontal arm wave*), and some actions involve complex interactions among the body parts, thereby leading to large amount of occlusions (one leg in front of the other or part of body is outside the camera range) which significantly decreases the tracking performance.

Table IV illustrates recognition rate of the the local model (HMM), the global model (RBM), the combined approach as well as the results reported in [11], [20]. From the results we can see that the global model outperforms both local models by about 20%. This demonstrates the importance of global dynamics for discriminating actions, and with the proposed classification approach, RBM can successfully capture the global dynamics for action recognition. Meanwhile, by

TABLE II. F1 SCORE OF PROPOSED MODEL AND BASELINE MODEL

Action	Bloom *et al.* [2]	Proposed Method
Fighting	70.46	**96.15**
Golf	**83.37**	76.19
Tennis	56.44	**81.97**
Bowling	**80.78**	61.54
FPS	53.57	**96.45**
Driving	84.24	**90.24**
Misc	78.21	**95.24**
Avg.	72.44	**85.40**

TABLE III. COMPARISON OF DIFFERENT METHODS ON G3D DATASET

Method	Accuracy
Hidden Markov Model	77.4%
RBM	84.0%
Proposed Method	**86.4%**

TABLE IV. COMPARISON OF DIFFERENT METHODS ON ACTION3D DATASET

Method	Accuracy
Hidden Markov Model	55.3%
Li *et al.* [11]	74.7&
RBM	79.6%
Proposed Method	**80.2%**
Wang *et al.* [20]	88.2%

combining the local model and global model, the proposed method can further improve the classification accuracy.

The performance of our method is below the method reported in [20]. Such result is reasonable because our method only uses a subset of the joints, and we do not use any features from the depth images. The feature is much less than other methods that consider both appearance and shape. The proposed method needs more information to classify among similar actions such as *draw X, draw tick, and draw circle*. The model also cannot handle severely noisy or corrupted data like *bend*. However, the proposed algorithm performs quite well on the distinctive actions, especially complicated actions which involve both hands and feet, like *tennis swing*.

VI. CONCLUSION

In this paper we propose a novel approach that captures both local and global dynamical information of the human joint trajectories for action recognition. The contributions of this paper are as follows. First, we introduce the Gaussian-Bernoulli restricted Boltzmann machine to model the motion data and capture the global dynamics of human actions. RBM is used as a generative model for dynamic modeling. A model selection method is introduced to generate discriminative models. We further propose a novel classification approach to apply RBM for action recognition. Second, we combine RBM with hidden Markov model using a fusion procedure to jointly exploit global and local patterns. Finally, experimental results demonstrate the effectiveness of the proposed approach.

ACKNOWLEDGMENT

The work described in this paper is supported in part by the grant N00014-12-1-0868 from the Office of Navy Research.

REFERENCES

[1] J. Aggarwal and M. S. Ryoo. Human activity analysis: A review. *ACM Computing Surveys (CSUR)*, 43(3):16, 2011.

[2] V. Bloom, D. Makris, and V. Argyriou. G3d: A gaming action dataset and real time action recognition evaluation framework. In *Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 7–12. IEEE, 2012.

[3] M. Carreira-Perpinan and G. Hinton. On contrastive divergence learning. In *Artificial Intelligence and Statistics*, volume 2005, page 17, 2005.

[4] C. Ellis, S. Masood, M. Tappen, J. LaViola, and R. Sukthankar. Exploring the trade-off between accuracy and observational latency in action recognition. *International Journal of Computer Vision*, pages 1–17, 2012.

[5] S. Fothergill, H. M. Mentis, P. Kohli, and S. Nowozin. Instructing people for training gestural interactive systems. In J. A. Konstan, E. H. Chi, and K. Höök, editors, *CHI*, pages 1737–1746. ACM, 2012.

[6] A. Gupta and L. S. Davis. Objects in action: An approach for combining action understanding and object perception. In *Computer Vision and Pattern Recognition (CVPR)*, pages 1–8. IEEE, 2007.

[7] L. Han, X. Wu, W. Liang, G. Hou, and Y. Jia. Discriminative human action recognition in the learned hierarchical manifold space. *Image and Vision Computing*, 28(5):836–849, 2010.

[8] G. Hinton. A practical guide to training restricted boltzmann machines. *Momentum*, 9:1, 2010.

[9] E. Hüllermeier, J. Fürnkranz, W. Cheng, and K. Brinker. Label ranking by learning pairwise preferences. *Artificial Intelligence*, 172(16):1897–1916, 2008.

[10] H. Larochelle and Y. Bengio. Classification using discriminative restricted boltzmann machines. In *International Conference on Machine Learning*, pages 536–543. ACM, 2008.

[11] W. Li, Z. Zhang, and Z. Liu. Action recognition based on a bag of 3d points. In *Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 9–14. IEEE, 2010.

[12] F. Lv and R. Nevatia. Recognition and segmentation of 3-d human action using hmm and multi-class adaboost. *European Conference on Computer Vision (ECCV)*, pages 359–372, 2006.

[13] P. Perona and J. Malik. Scale-space and edge detection using anisotropic diffusion. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 12(7):629–639, 1990.

[14] L. Rabiner. A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286, 1989.

[15] C. Rao and M. Shah. View-invariance in action recognition. In *Computer Vision and Pattern Recognition (CVPR)*, volume 2, pages II–316. IEEE, 2001.

[16] M. D. Rodriguez, J. Ahmed, and M. Shah. Action mach a spatio-temporal maximum average correlation height filter for action recognition. In *Computer Vision and Pattern Recognition (CVPR)*, pages 1–8. IEEE, 2008.

[17] T. Schmah, G. E. Hinton, S. L. Small, S. Strother, and R. S. Zemel. Generative versus discriminative training of rbms for classification of fmri images. In *Advances in Neural Information Processing Systems (NIPS)*, pages 1409–1416, 2008.

[18] J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman, and A. Blake. Real-time human pose recognition in parts from single depth images. In *Computer Vision and Pattern Recognition (CVPR)*, pages 1297–1304. IEEE, 2011.

[19] G. Taylor, G. Hinton, and S. Roweis. Modeling human motion using binary latent variables. *Advances in Neural Information Processing Systems (NIPS)*, 19:1345, 2007.

[20] J. Wang, Z. Liu, Y. Wu, and J. Yuan. Mining actionlet ensemble for action recognition with depth cameras. In *Computer Vision and Pattern Recognition (CVPR)*, pages 1290–1297. IEEE, 2012.

[21] Z. Wang, G. Schalk, and Q. Ji. Anatomically constrained decoding of finger flexion from electrocorticographic signals. In *Advances in Neural Information Processing Systems (NIPS)*, 2011.

[22] A. Yilma and M. Shah. Recognizing human actions in videos acquired by uncalibrated moving cameras. In *International Conference on Computer Vision (ICCV)*, volume 1, pages 150–157. IEEE, 2005.