



Journal of
**Software
Engineering**

ISSN 1819-4311



Academic
Journals Inc.

www.academicjournals.com



Research Article

Human Activity Recognition Based-on Conditional Random Fields with Human Body Parts

¹Chen Yehui, ²Liu Hao and ²Yi Bo

¹Electronic Communications Engineering College, Anhui Xinhua University, 230088 Hefei, China

²Department of Physics, University of Science and Technology of China, 230026 Hefei, China

Abstract

Background: The RGBD-based human activity recognition has captured extensive concerns of researchers from the domains of entertainment, surveillance, robotics and a variety of systems that involve interactions between persons and electronic devices. However, it is a non-trivial task due to the spatial and temporal variations in the activity data. **Materials and Methods:** This study propose a Conditional Random Fields (CRF) with star structure to model the variations and accurately recognize activity patterns. The human body is partitioned into five parts, the torso, the left arm, the right arm, the left leg and the right leg. Each vertex in this CRF model corresponds to one part of the human body in an activity sequence. Joint angle features are extracted to support this model. **Results:** This method not only takes advantage of multiple features and temporal context but also captures the spatial context among the human body parts. **Conclusion:** Experimental results show that this method achieved a higher recognition rate and it is still effective when self-occlusion happened.

Key words: Human activity recognition, star CRF, joint angle, human body parts, CLC index TP391.4

Received: February 25, 2016

Accepted: June 16, 2016

Published: September 15, 2016

Citation: Chen Yehui, Liu Hao and Yi Bo, 2016. Human activity recognition based-on conditional random fields with human body parts. J. Software Eng., 10: 408-415.

Corresponding Author: Liu Hao, Department of Physics, University of Science and Technology of China, 230026 Hefei, China

Copyright: © 2016 Chen Yehui *et al.* This is an open access article distributed under the terms of the creative commons attribution License, which permits unrestricted use, distribution and reproduction in any medium, provided the original author and source are credited.

Competing Interest: The authors have declared that no competing interest exists.

Data Availability: All relevant data are within the paper and its supporting information files.

INTRODUCTION

Human activity recognition has become a hot topic in the field of computer vision and can be widely used in human-computer interaction, military surveillance, robotics and virtual reality games. The goal of human activity recognition is to automatically detect and analyze human activities from the information acquired from sensors, e.g., a sequence of images, either captured by RGB cameras, range sensors or other sensing modalities.

Recent advances in sensing technology have enabled us to capture the depth information in real-time, which inspired the study on activity recognition from 3D data. In 2011, Shotton *et al.*¹ proposed to extract 3D body joint locations from a single depth image using an object recognition scheme. The human body is labeled as body parts based on the per-pixel classification results. The parts include LU/RU/LW/RW head, neck, L/R shoulder, LU/RU/LW/RW arm, L/R elbow, L/R wrist, L/R hand, LU/RU/ LW/RW torso, LU/RU/LW/RW leg, L/R knee, L/R ankle and L/R foot (left, right, upper, lower). Skeletal joints tracking algorithms were built into the Kinect device, which offers easy access to the skeletal joint locations. This excited considerable interest in the computer vision community. Many algorithms have been proposed recognizing activities using skeletal joint information.

This study addresses the issue of human activity recognition, which is useful in many practical applications. However, human activity recognition is a non-trivial task. The factors that contribute to the challenges include the followings: (1) The activity is of high dimensionality, which is difficult to extract features, (2) Activities are often similar. For example between the activities "Run" and "Walk", there are several similar frames and (3) Because of the sensor' direction, there is always some of the human body parts which are invisible.

To recognize human activities, several approaches have been proposed past few years. The Dynamic Time Warping (DTW)², a method for measuring similarity between two temporal sequences, is one of the most popular temporal classification algorithms. However, DWT is not appropriate for a large number of classes. Another popular classifier for

human activity recognition is Hidden Markov Models (HMM)³, which is proved successful in speech recognition. Uddin *et al.*⁴ proposed a detection method with joint angle feature and HMM model. The HMM model has a set of states that changes with the time, whose process is more in line with people's activity. However, because of the output independence assumption of HMM, it can't get access to the context, resulting in the feature selection limitation. To solve this problem, Phillips *et al.*⁵ proposed maximum entropy method but the maximum entropy model can only find a local optimum value, it also brings a label bias problem. Recently, Conditional Random Fields (CRFs)⁶ has been applied to the activity detection⁷⁻⁹. The CRF model is an undirected probabilistic graph which can model the posterior probability directly. Also, it does not rely on conditional independence assumptions, has the ability to get use of multi-feature and the context. Long *et al.*¹⁰ had used CRF models in the abnormal activity detection task and achieved better detection results.

In this study, we propose a human activity recognition method (Fig. 1) which can model the human activity class well and recognize them effectively. We observed that: First, even if some different activities have similar frames, these frame's temporal adjacent frames are different. Second there is a class of spatial context. For example, when people walk, two arms will swing alternately on both sides of the body to keep balance. But when people jump, two arms will always swing at the same direction. This spatial context also makes sense to self-occlusion. So the invisible body part's state can be inferred from the visible body parts. For feature extraction, first, the human body is partitioned into five parts, the torso, the left upper limb, the right upper limb, the left lower limb, the right lower limb. Then the joint angle features are extracted which are invariant to the camera location and subject appearance and invariant to human body size. To do recognition, a conditional random fields is designed with star structure called star-CRF which can model human activity with both the temporal context and the spatial context. Also, this method can jointly recognize human activity with five human body parts. The experiment results showed that this method achieved a higher recognition rate and it is still effective when self-occlusion happened.

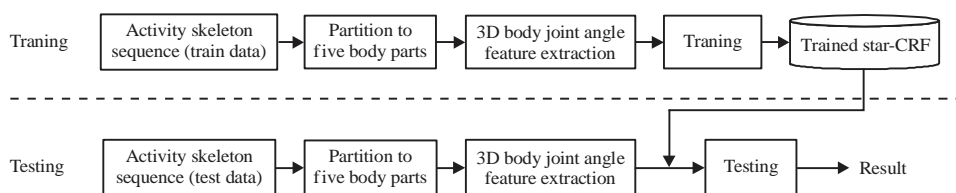


Fig. 1: Overview of proposed framework

MATERIALS AND METHODS

The skeleton structures are different from each other in different motion capture or RGB-D skeleton datasets (Fig. 2). For example, in the CMU Graphics Lab Motion Capture Database and the Motion Capture Database HDM05 the skeleton contain 31 joints, the Prime Sense Natural Interaction Middleware (NiTE) provides 15 joints and the Microsoft Kinect SDK provides 20 joints. Although, the number of joints are different, the essential joints are available in all skeletons. They are head, neck, root, L/R hand, L/R elbow, L/R shoulder, L/R foot, L/R knee, L/R crotch.

These 15 joints are partitioned into 5 parts, torso, right upper limb, left upper limb, right lower limb, left lower limb (Fig. 3).

With a skeleton sequence, to describe the gesture at each frame, the joint angle features are extracted. The skeletal joint angle features are invariant to the camera location and subject appearance and it is invariant to human body size Fig. 4.

Uddin *et al.*⁴ proposed the joint angle feature extraction method. In this study, Uddin's method is used to extract joint angle feature. The 24D feature vector at frame t is:

$$F^t = [\theta_{\text{global_local}}, \theta_{\text{left_shoulder}}, \theta_{\text{right_shoulder}}, \theta_{\text{left_crotch}}, \theta_{\text{right_crotch}}, \theta_{\text{neck}}, \theta_{\text{left_elbow}}, \theta_{\text{right_elbow}}, \theta_{\text{left_knee}}, \theta_{\text{right_knee}}] \quad (1)$$

Then divide it into five parts:

- The torso feature $F_{ts}^t = [\theta_{\text{global_local}}, \theta_{\text{neck}}]$
- The left upper limb feature $F_{lu}^t = [\theta_{\text{left_shoulder}}, \theta_{\text{left_elbow}}]$
- The right upper limb feature $F_{ru}^t = [\theta_{\text{right_shoulder}}, \theta_{\text{right_elbow}}]$
- The left lower limb feature $F_{ll}^t = [\theta_{\text{left_crotch}}, \theta_{\text{left_knee}}]$
- The right lower limb feature $F_{rl}^t = [\theta_{\text{right_crotch}}, \theta_{\text{right_knee}}]$

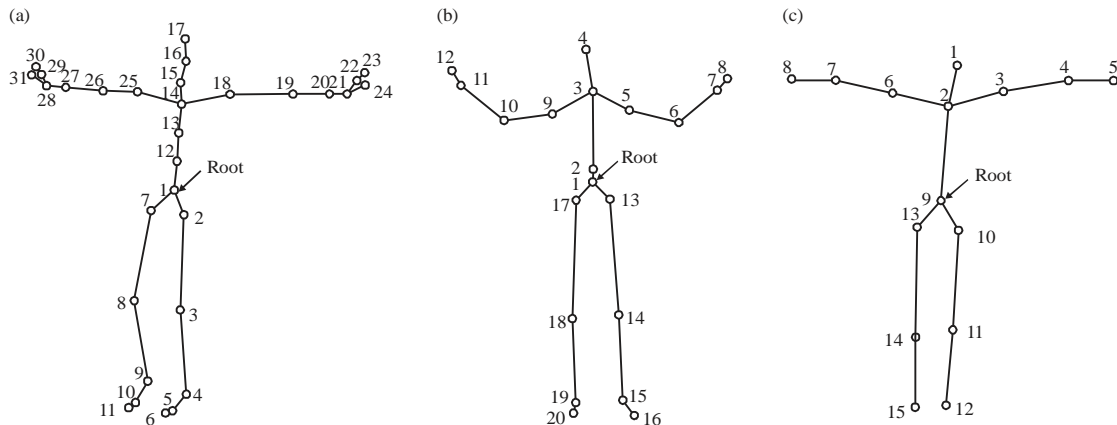


Fig. 2(a-c): Skeleton models in different sources, (a) HDM05, (b) Kinect SDK and (c) Kinect NiTE

In this study, torso feature is used to present the global gesture, left upper limb feature, right upper limb feature, left lower limb feature and right lower limb feature are used to present the left upper gesture, the right upper limb gesture, the left lower limb gesture and the right lower limb gesture respectively.

Star CRF model: The X is a random variable over data sequences to be labeled, Y is a random variable over corresponding label sequences. In this study, X is features and Y is activity classes.

Let $G = \{V, E\}$ be an undirected graph model. Each element in X and Y is a vertex of G . Edges in G represent the relationships among these vertices.

The structure of traditional CRF is show in Fig. 5a. $Y = \{Y^t, t = 1, 2, \dots, N\}$ is an observed sequence, $X = \{X^t, t = 1, 2, \dots, N\}$ is a label sequence represents the states of X . $X^t \in \{1, 2, \dots, S\}$ where, S is the number of activity classes. This structure use one vertex X^t at each frame to represent the state of the frame. There are two sets of edges, one connect X^t and Y^t representing the relationship between the state vertex and the observed vertex, another connect X^t and X^{t+1} , representing the temporal context. The disadvantage of the traditional CRF is it can't get use of the spatial context.

Therefore, a CRF model with star structure showed in Fig. 5 is designed. At each frame, there are five observed vertices and five state vertices and each vertex corresponds to one human body part. Although, the number of vertices increased, the dimension of data at each vertex is reduced. $Y_1^t, Y_2^t, Y_3^t, Y_4^t, Y_5^t$ denote the features of the human torso F_{ts}^t , left upper limb F_{lu}^t , right upper limb F_{ru}^t , left lower limb F_{ll}^t , right lower limb F_{rl}^t , while $X_1^t, X_2^t, X_3^t, X_4^t, X_5^t$ represent the states of the human torso, left upper limb, right upper limb, left lower limb, right lower limb at time t .

In Star CRF, there are five time sequences standing for five parts of human body. Also, there are three sets of edges standing for different relational meanings. The edge set E_a , by which each state is connect with the observation, represents the relationship between the observation sequence and the state sequence. The edge set E_b , by which each state is connect with its neighboring states in the same time sequence, represents the relationship between the states of the same human body part at different times. The edge set E_c , by which each state is connect with its neighboring

states across time sequences, represents the relationship between the states of different parts of human body at the same time.

Potential functions $\phi_a(X_i^t)$ is defined for each edge (X_i^t, Y_i^t) in E_a , $\phi_b(X_i^t, X_i^{t+1})$ is defined for each edge (X_i^t, X_i^{t+1}) and $\phi_c(X_i^t, X_j^t)$ is defined for each edge (X_i^t, X_j^t) . Star CRFs are determined as:

$$p(X|Y, \theta) = \frac{1}{Z(Y, \theta)} \exp \left\{ \sum_{t=1}^N \sum_{i=1}^5 \phi_a(X_i^t) + \sum_{t=1}^{N-1} \sum_{i=1}^5 \phi_b(X_i^t, X_i^{t+1}) + \sum_{t=1}^{N-1} \sum_{i=1, j=1, i \neq j}^5 \phi_c(X_i^t, X_j^t) \right\} \quad (2)$$

where, $z(Y, \theta)$ is the normalization constant.

$$Z(Y, \theta) = \sum_X \exp \left\{ \sum_{t=1}^N \sum_{i=1}^5 \phi_a(X_i^t) + \sum_{t=1}^{N-1} \sum_{i=1}^5 \phi_b(X_i^t, X_i^{t+1}) + \sum_{t=1}^{N-1} \sum_{i=1, j=1, i \neq j}^5 \phi_c(X_i^t, X_j^t) \right\} \quad (3)$$

For Multinomial Logistic Regression (MLR)¹¹, observational data does not need to meet the normal distribution restrictions, so that MLR can be adapted to more types of observational data. The potential function $\phi_a(X_i^t)$ is defined by MLR as:

$$\phi_a(X_i^t) = \sum_{s=1}^S \delta(X_i^t = s) \log p(X_i^t = s | Y, \lambda) \quad (4)$$

The potential functions $\phi_b(X_i^t, X_i^{t+1})$ and $\phi_c(X_i^t, X_j^t)$ are defined by the promotion of ising/potts model as:

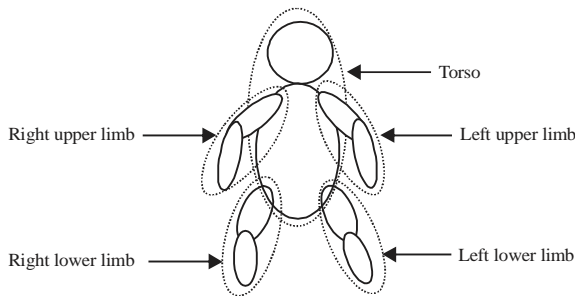


Fig. 3: Five human body parts

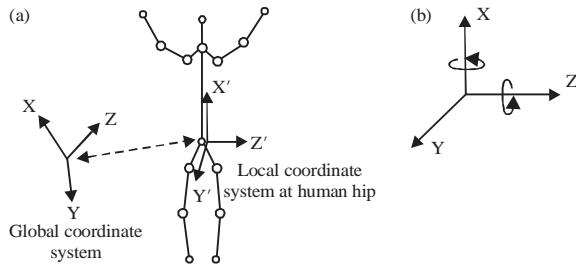


Fig.4(a-b): (a) Articulated skeletal human body model and (b) two rotational euler angles around the X-axis and Z-axis (two DOF at each joint) controlling the movement of each segment

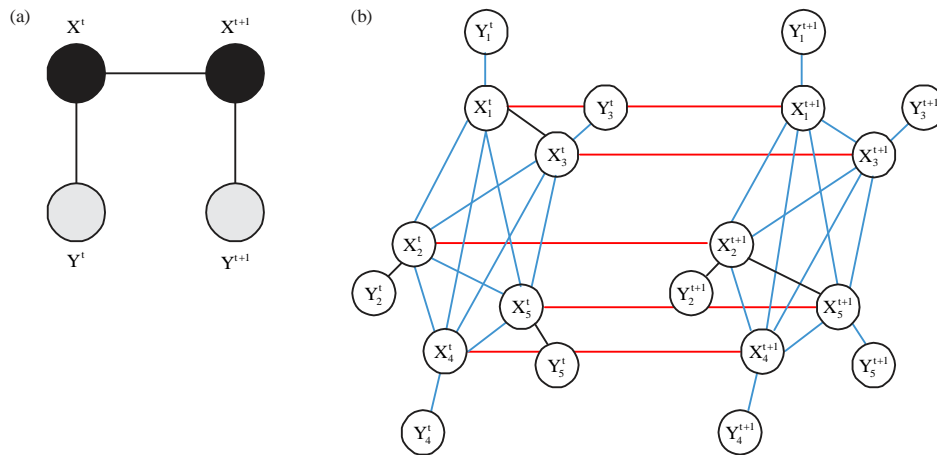


Fig. 5(a-b): The structure of conditional random fields, (a) Traditional CRF and (b) Star CRF

$$\phi_b(X_i^t, X_i^{t+1}) = \sum_{s, h \in \{1, 2, \dots, S\}} (\sigma_{sh}^b)^T g_{t, t+1}(Y) \delta(X_i^t = s) \delta(X_i^{t+1} = h) \quad (5)$$

$$\phi_c(X_i^t, X_j^t) = \sum_{s, h \in \{1, 2, \dots, S\}} (\sigma_{sh}^c)^T g_{i, j}(Y) \delta(X_i^t = s) \delta(X_j^t = h) \quad (6)$$

where, $g_{t, t+1}$ and $g_{i, j}$ are feature vectors extracted from the pair $(t, t+1)$ and the pair (i, j) of the observational data. The λ is the parameter vector of the potential function $\phi_a(X_i^t)$. The σ_{sh}^b and σ_{sh}^c are the parameter vectors of the potential functions $\phi_b(X_i^t, X_i^{t+1})$ and $\phi_c(X_i^t, X_j^t)$. $\theta = \{\lambda, \sigma^b, \sigma^c\}$ is the set of parameters in the model, where $\theta = \{\lambda, \sigma^b, \sigma^c\}$, $\sigma^c = \{\sigma_{sh}^c, s, h = 1, 2, \dots, S\}$.

Parameter estimation: For estimating the parameters $\theta = \{\lambda, \sigma^b, \sigma^c\}$, it can be assumed that M independent identically distributed labeled training samples $\{X^m, Y^m, m = 1, 2, \dots, M\}$. The parameters can be estimated by the standard maximum-likelihood (ML) approach. The ML training chooses parameter values such that the logarithm of the likelihood:

$$\tilde{\theta} = \arg \max_{\theta} \{\log(p(X|Y, \theta))\} = \arg \max_{\theta} \{\log(\prod_{m=1}^M p(X^m | Y^m, \theta))\} \quad (7)$$

For the Star CRFs used in this study, likelihood maximization can be performed using a gradient ascent (BFGS) method¹².

Inference: To recognize human activity, the most possible label sequence of observed sequence need to be inferred. There are many inference algorithms for CRFs, including forward-backward algorithm, Viterbi algorithm, junction tree¹³ and Loopy Belief Propagation (LBP)^{14,15}. In this study, the popular inference algorithm LBP is used to release the restrictive condition.

The Star CRF model can be rewritten as the following form:

$$p(X|Y, \theta) \propto \prod_{t=1}^N \prod_{i=1}^S \xi_a(X_i^t) \prod_{t=1}^{N-1} \prod_{i=1}^S \xi_b(X_i^t, X_i^{t+1}) \prod_{t=1}^N \prod_{i=1, j=1, i \neq j}^S \xi_c(X_i^t, X_j^t) \quad (8)$$

Where:

$$\xi_k(X_k) = \exp\{\phi_a(X_k)\}$$

$$\xi_b(X_i^t, X_i^{t+1}) = \exp\{\phi_b(X_i^t, X_i^{t+1})\}$$

$$\xi_c(X_i^t, X_j^t) = \exp\{\phi_c(X_i^t, X_j^t)\}$$

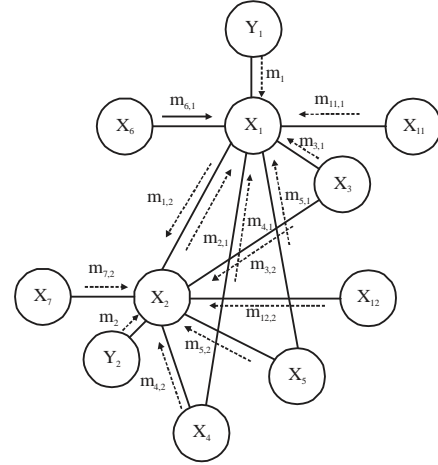


Fig. 6: Local message passing in star CRF network

Let $m_{kl}(X_i)$ be the message that node X_k sends to X_l , $m_k(X_k)$ be the message that observed node Y_k send to node X_k and $b_k(X_k)$ be the belief at node X_k (Fig. 6). The standard "Max-product" algorithm is shown below:

- Initialize all messages $m_{kl}(X_i)$ as uniform distributions and messages $m_{kl}(X_k) = \xi_k(X_k)$
- Update messages $m_{kl}(X_i)$ iteratively for $d = 1, 2, \dots, D$:

$$m_{kl}^{d+1}(X_i) \leftarrow \gamma \max_{X_k} \left\{ \xi_{kl}(X_k, X_i) m_k(X_k) \prod_{g \in \eta_{k,l}} m_{gk}^d(X_k) \right\} \quad (9)$$

- Compute beliefs:

$$b_k(X_k) \leftarrow m_k(X_k) \prod_{g \in \eta_k} m_{gk}(X_k) \quad (10)$$

$$X_k^{MAP} = \arg \max_{X_k} b_k(X_k) \quad (11)$$

RESULTS AND DISCUSSION

The proposed method is used to recognize human activities and compare the performance to existing methods.

Microsoft Research Cambridge-12 Kinect (MRC12K)¹⁶ gesture database is used in our experiments. Microsoft Research Cambridge-12 Kinect gesture database is captured by the depth sensor Kinect. The dataset is comprised of 119 sequences collected from 30 people performing 12 gestures (Table 1). In total, there are 1223 gesture instances. The motion files contain tracks of 20 joints estimated using the Kinect Pose Estimation pipeline.

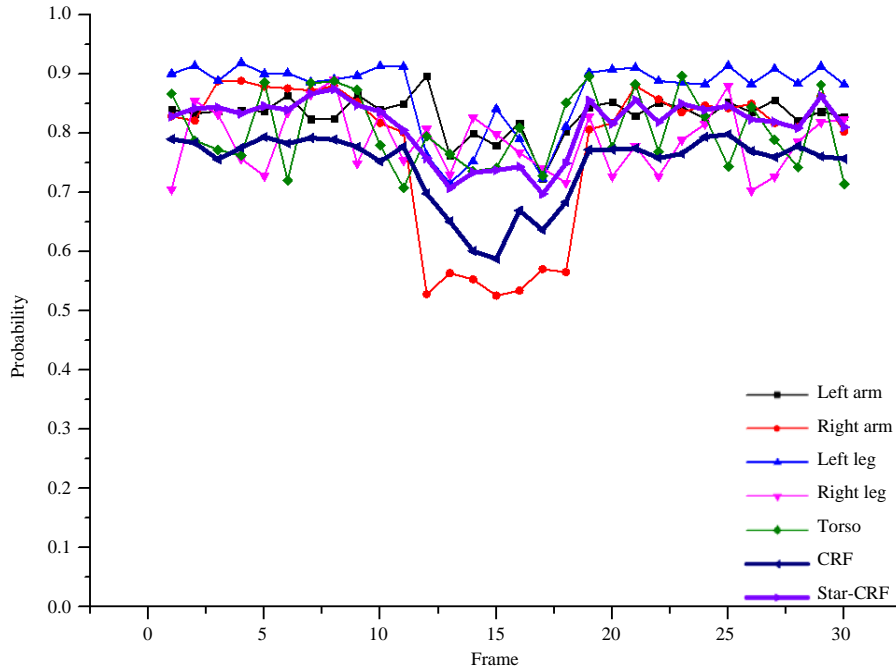


Fig. 7: Temporal evolution of probabilities when self-occlusion happened. The right arm is invisible between frame 11 and 19, the probability of recognition with CRF is reduced to 0.6. With Star-CRF, the probability is reduced to 0.7, because the Star-CRF can get use of the spatial context of body parts

Experimental platform in this experiment is the personal computer with Inter CPU E3-1230v3, 16.0 GB memory and Windows 8.1×64 operating system. The main function of algorithm coded in C++ were called in Matlab. The parameters $\theta = \{\lambda, \sigma^b, \sigma^c\}$ of the Star CRF components were learned by the standard ML approach. Each parameter vector λ was initialized to be 0, while the elements of σ^b and σ^c were initialized to be random numbers between 0 and 0.5.

In order to evaluate the performance of the proposed method. The correct rate and error rate should be calculated. The correct rate is calculated by:

$$R = \frac{C}{N} \times 100\%$$

where, C and N represent the number of correct recognition and test activities.

The error rate is calculated by:

$$ER = \frac{Y+D}{N} \times 100\%$$

where, D represent the number of detection errors, Y represent the number of non-recognize gestures. Star-CRF jointly model five body parts, so it get five labels after

inference respective to the five body parts. In the five labels, if there are no more than 3 labels are the same label, it can be considered as a non-recognized activity. The activity recognition results are showed in Table 2.

In order to evaluate the performance of Star-CRF, experiments on MRC12K database with Hidden Markov Model (HMM) and Conditional Random Fields (CRF) are also taken. The experiments results are shown in Table 3.

Compare with the results of HMM, CRF, T-CRF¹⁷ and Star-CRF, the correct recognition rate of the Star-CRF is higher than those of the HMM, CRF and T-CRF. Because of the output independence assumption of HMM, it can only get access to the above information in temporal context, the correct recognition rate of HMM is 78.3% lower than CRF, T-CRF and Star-CRF. The CRF can get access of temporal context but it ignore the spatial context. Hyunsook Chung proposed a Threshold-CRF to do activity recognition and get a correct recognition rate of 91.9%, however the threshold need to be set manually. Star-CRF has the ability to model human activity with both the temporal context and the spatial context, get a higher correct recognition rate of 93.8%.

The performance of Star-CRF is also tested when self-occlusion happened. Figure 7 shows a recognition result with CRF and Star-CRF for a sequence contain the gestures "Lift outstretched arms". The time evolution of probabilities

Table 1: Microsoft Research Cambridge-12 Kinect (MRC12K) gesture database

ID	Description	Tap name
1	Start music/raise volume	Lift outstretched arms
2	Crouch or hide	Duck
3	Navigate to next menu	Push right
4	Put on night vision goggles	Goggles
5	Wind up the music	Wind it up
6	Shoot with a pistol	Shoot
7	Take a bow to end the session	Bow
8	Throw an object	Throw
9	Protest the music	Had enough
10	Change weapon	Change weapon
11	Lay down the tempo of a song	Beat both
12	Kick to attack an enemy	Kick

Table 2: Recognition results with Star-CRF on MRC12K database

ID	N	C	Y	R (%)	ER (%)
1	101	96	3	95.0	5.0
2	102	95	5	93.1	6.9
3	101	99	2	98.0	2.0
4	101	92	6	91.1	8.9
5	101	99	1	98.0	2.0
6	92	83	6	90.2	9.8
7	103	99	2	96.1	3.9
8	103	98	4	95.1	4.9
9	105	96	3	91.4	8.6
10	103	94	4	91.2	8.8
11	106	96	2	90.6	9.4
12	105	100	3	95.2	4.8

C: No. of correct recognition, N: No. of test activities, R: Rate, Y: No. of Non-recognize gesture and ER: Error rate

Table 3: Human abnormal activity recognition results

ID	HMM (%)	CRF (%)	T-CRF (%)	Star-CRF (%)
1	78.2	81.2	93.1	95.0
2	74.5	75.5	95.1	93.1
3	85.1	87.1	93.1	98.0
4	83.1	83.2	89.1	91.1
5	79.2	87.1	97.0	98.0
6	73.9	80.4	82.4	90.2
7	86.4	85.4	95.1	96.1
8	83.5	85.4	95.1	95.1
9	73.3	77.1	89.5	91.4
10	72.8	77.7	93.2	91.2
11	65.1	71.7	84.9	90.6
12	84.7	84.8	94.3	95.2
Average	78.3	81.4	91.9	93.8

are illustrated by curves. In Fig. 7, the right upper limb part of the actor is invisible between frame 11 and frame 19. The probability of CRF is reduced to 0.6, while the probability of CRF is reduced to 0.7. The reason is the CRF only get use of temporal context and the spatial context makes Star-CRF to get a better performance.

CONCLUSION

In this study, a novel method of human activity recognition was proposed. In order to do feature extraction, depth information from Microsoft' Kinect was used. First, the

temporal context and the spatial context in human activity sequence were discussed. Second, human body were partitioned into five body parts and extract five feature vectors for the five body parts. Third, a Star-CRF model was proposed to model the five body parts with the temporal context and the spatial context.

In this study, a public data set, the Microsoft Research Cambridge-12 Kinect gesture database is used. The experimental results demonstrated that the proposed method can be efficient and effective for human activity recognition. Near-term future study includes extending the proposed method to real-world applications.

ACKNOWLEDGMENT

We thank the National Natural Science Foundation of China (No. 61071173).

REFERENCES

1. Shotton, J., T. Sharp, A. Kipman, A. Fitzgibbon and M. Finocchio *et al.*, 2013. Real-time human pose recognition in parts from single depth images. *Commun. ACM*, 56: 116-124.
2. Pham, C.H., Q.K. Le and T.H. Le, 2014. Human action recognition using dynamic time warping and voting algorithm. *VNU J. Sci. Comput. Sci. Commun. Eng.*, 30: 22-30.
3. Lehrmann, A.M., P.V. Gehler and S. Nowozin, 2014. Efficient nonlinear markov models for human motion. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, June 23-28, 2014, Columbus, OH., pp: 1314-1321.
4. Uddin, M.Z., N.D. Thang, J.T. Kim and T.S. Kim, 2011. Human activity recognition using body joint-angle features and hidden Markov model. *ETRI J.*, 33: 569-579.
5. Phillips, S.J., R.P. Anderson and R.E. Schapire, 2006. Maximum entropy modeling of species geographic distributions. *Ecol. Modell.*, 190: 231-259.
6. Lafferty, J.D., A. McCallum and F.C.N. Pereira, 2001. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. *Proceedings of the 18th International Conference on Machine Learning*, June 28-July 1, 2001, Williamstown, MA., USA., pp: 282-289.
7. Song, Y., D. Demirdjian and R. Davis, 2012. Continuous body and hand gesture recognition for natural human-computer interaction. *ACM Trans. Interact. Intell. Syst.*, Vol. 2, No. 1. 10.1145/2133366.2133371
8. Siddiquie, B., S. Khan, A. Divakaran and H. Sawhney, 2013. Affect analysis in natural human interaction using joint hidden conditional random fields. *Proceedings of the IEEE International Conference on Multimedia and Expo*, July 15-19, 2013, San Jose, CA., pp: 1-6.

9. Liu, A.A., W.Z. Nie, Y.T. Su, L. Ma, T. Hao and Z.X. Yang, 2015. Coupled hidden conditional random fields for RGB-D human action recognition. *Sig. Process.*, 112: 74-82.
10. Long, Z., G. Li, X. Jinsheng and L. Hao, 2012. Video abnormal target description based on CRF model. *Proceedings of the International Conference on Audio, Language and Image Processing*, July 16-18, 2012, Shanghai, pp: 519-524.
11. Li, J., J.M. Bioucas-Dias and A. Plaza, 2012. Spectral-spatial hyperspectral image segmentation using subspace multinomial logistic regression and Markov random fields. *IEEE Trans. Geosci. Remote Sens.*, 50: 809-823.
12. Wallach, H., 2003. Efficient training of conditional random fields. *Proceedings of the 6th Annual CLUK Research Colloquium*, January 6-7, 2003, Edinburgh, UK., pp: 1-8.
13. Vats, D. and R.D. Nowak, 2014. A junction tree framework for undirected graphical model selection. *J. Mach. Learn. Res.*, 15: 147-191.
14. Yasuda, M., S. Kataoka and K. Tanaka, 2014. Erratum: Inverse problem in pairwise Markov random fields using loopy belief propagation [*J. Phys. Soc. Jpn.* 81, 044801 (2012)]. *J. Phys. Soc. Jpn.*, Vol. 83. 10.7566/JPSJ.83.048001
15. Smith, D. and V. Gogate, 2014. Loopy belief propagation in the presence of determinism. *Proceedings of the 17th International Conference on Artificial Intelligence and Statistics*, Volume 33, April 22-25, 2014, Reykjavik, Iceland, pp: 895-903.
16. Fothergill, S., H. Mentis, P. Kohli and S. Nowozin, 2012. Instructing people for training gestural interactive systems. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, May 5-10, 2012, Austin, TX., USA., pp: 1737-1746.
17. Chung, H. and H.D. Yang, 2013. Conditional random field-based gesture recognition with depth information. *Opt. Eng.*, Vol. 52, No. 1. 10.1117/1.OE.52.1.017201.