Grading Tai Chi Performance in Competition with RGBD Sensors

Hui Zhang(✉), Haipeng Guo, Chaoyun Liang, Ximin Yan, Jun Liu, and Jie Weng

Department of Computer Science, United International College, 28, Jinfeng Road, Tangjiawan, Zhuhai, Guangdong, China
{amyzhang,hpguog}@uic.edu.hk
{f030300021,f030300052,f030300030,f030300047}@mail.uic.edu.hk

Abstract. In order to grade objectively, referees of Tai Chi practices always have to be very concentrated on every posture of the performer. This makes the referees easy to be fatigue and thus grade with occasional mistakes. In this paper, we propose using Kinect sensors to grade automatically. Firstly, we record the joint movement of the performer skeleton. Then we adopt the joint differences both temporally and spatially to model the joint dynamics and configuration. We apply Principal Component Analysis (PCA) to the joint differences in order to reduce redundancy and noise. We then employ non-parametric Nave-Bayes-Nearest-Neighbor (NBNN) as a classifier to recognize the multiple categories of Tai Chi forms. To give grade of each form, we study the grading criteria and convert them into decision on angles or distances between vectors. Experiments on several Tai Chi forms show the feasibility of our method.

Keywords: Tai Chi · RGBD sensor · Kinect

1 Introduction

Tai Chi, as shortened to Tai Chi Chuan, is a traditional Chinese martial art, which is practiced for both its defense training and its health benefits. Because of its soft and continuously flowing movements, Tai Chi is able to cultivate both peoples mind and physical body into a balance system [11]. Tai Chi has become popular internationally and many Tai Chi schools have been opened around the world. Trainees can follow the coach in order to learn different forms in Tai Chi. Meanwhile, there are a lot of Tai Chi national or international competitions for the performers to improve their skills, such as London Competition for Traditional Tai Chi Chuan or Tai Chi Competition in New York, etc.

In a national Tai Chi competition, there are generally eight referees sitting in six position around the playground (see figure 1 for details). The five referees on the edge of the playground will first manually record the scores from their own view points and show them to the three chief referees after the performer finishing

L. Agapito et al. (Eds.): ECCV 2014 Workshops, Part II, LNCS 8926, pp. 3–13, 2015.
DOI: 10.1007/978-3-319-16181-5_1

his performance. The chief referees will finally give out the final score according to all of the scores collected. Such grading s largely based on manual works. There are also electronic systems for Tai Chi grading utilized in national competitions. Referees press keys on a joystick to deduce a score when he found that the performer makes a mistake. This system, along with the manual procedures, requires the referees concentrating on observing every posture of the performers movement in order to give a justice grade. The referees are easy to get tired and thus subject errors are inevitable during grading. Therefore, an automatic and objective method is urgently needed to solve these problems.

Fig. 1. The position of the referees

To facilitate the manual works, the first task is to work on recognizing different forms of Tai Chi performance. For automatic human action recognition, traditional methods may work on video sequences captured by a single camera. In this case, a video is a sequence of 2D RGB frames in time series. In [1,3,8–10], the spatio-temporal volume-based method have been proposed to compute and the similarity between two action volumes are compared to recognize the action. Another trend of methods is based on motion trajectory for recognizing human activities [13,14]. Human actions were interpreted by the movement of a set of body joints. In [18], naive Bayes mutual information maximization (NBMIM) is introduced as a discriminative pattern matching criterion for action classification.

However, it is not easy to extract and track skeleton joints from 2D video sequences quickly and accurately until the launch of Microsoft Kinect sensors. The Kinect sensor is able to capture RGB sequences as well as depth maps of human action in real time. With its associated SDK or OpenNI, we could model human actions by the motion of a set of key joints [6] with reasonable accuracy. There are applications or research with Kinect supporting martial art practices, such as the Kinect Sports game, the posture classification of Muay

Thai [7], etc. Human action and activity recognition with Kinect become popular research topics recently [5,12,16]. In order to have a fast, simple yet powerful recognition,[15] proposes an actionlet ensemble model to characterize the human actions, which represents the interaction of a subset of human joints. Zanfir et. al. [19] introduce a non-parametric Moving Pose (MP) descriptor considering both pose information as well as differential quantities (speed and acceleration) of the human body joints.

Inspired by [17], this paper first record the joint movement of the performers skeleton. Then we adopt the joint differences both temporally and spatially to model the joint dynamics and configuration. We apply Principal Component Analysis (PCA) to the joint differences by reducing redundancy and noise. We then employ non-parametric Nave-Bayes-Nearest-Neighbor (NBNN) as a classifier to recognize the multiple categories of Tai Chi actions.

After the system has recognized the action of the performer, the next task is to mark the quality of the performers action. We convert the text description of the criteria into the grading decisions on angles or distances between vectors. Experiments on several sample Tai Chi Chuan actions show the feasibility of our method. Note that we have used only one Kinect sensor for grading. We plan to use six Kinect later similar to the position configuration of the referees in figure 1 so that the grading results will be comparable to those by the referee.

This paper is organized as follows. Section 2 introduces the feature extraction and dimension reduction. Section 3 provides our classifier for action recognition. Section 4 studies the grading rules and converts them into programmable decisions. Then Section 5 summarized the implementation steps. The experimental results are shown in section 6. Finally, section 7 gives the conclusions.

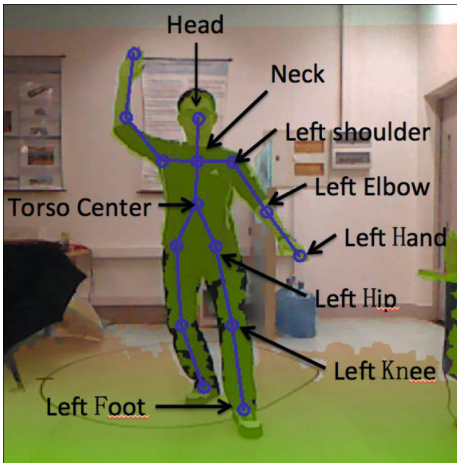

Fig. 2. The joints on the skeleton in OpenNI.

2 Feature Extraction

The human skeleton captured by Kinect sensor could have n joints and their respective 3D positions are $\mathbf{X}_k$ $(k = 1, ..., n)$. In OpenNI, $n = 15$. The joint 3D positions $\mathbf{X}_k = \{x_k, y_k, d_k\}$ are indicated by head, neck, torso center, left shoulder, left elbow, left hand, left hip, left knee, left foot, etc. as shown in figure 2. These joints are defined by the Kinect skeletal tracking system. The joints have hierarchy that the torso center joint as the root and extends to the head, feet and hands. Note that the three coordinate of a joint $\mathbf{X}_k = \{x_k, y_k, d_k\}$ are of inconsistent coordinates, e.g. $\{x_k, y_k\}$ are in screen coordinates and d_k is in world coordinate. Therefore the data normalization has to be first applied to $\mathbf{X}_k$ to avoid bias attributes in greater numeric ranges dominating those in smaller numeric ranges.

An action $\mathbf{A}_i$ could be represented by a sequence of frames f_{ij} $(j = 1, ..., N_i)$, where f_{ij} is a vector containing n coordinates of skeleton joints,

$$\mathbf{A}_i = \{f_{i1}, f_{i2}, ..., f_{iN_i}\}, \tag{1}$$

$$f_{ij} = \begin{pmatrix} \mathbf{X}_{head} \\ \mathbf{X}_{neck} \\ \mathbf{X}_{leftshoulder} \\ \mathbf{X}_{leftelbow} \\ \cdots \\ \mathbf{X}_{rightfoot} \end{pmatrix} = \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \mathbf{X}_3 \\ \mathbf{X}_4 \\ \cdots \\ \mathbf{X}_n \end{pmatrix}. \tag{2}$$

To characterize the action features, we first set the initial frame to approximate the neutral posture. Then we form the preliminary feature representation for each frame by the combination of three feature channels as $f_c = [f_{cc}, f_{cp}, f_{ci}]$ (see figure 3 in detail).

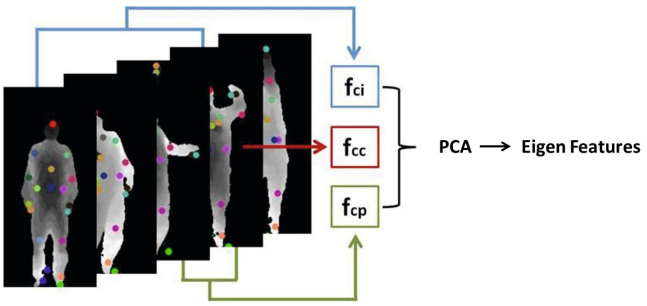

Fig. 3. The framework of representing Eigen features. In each frame, we obtain three feature sets, f_{cc}, f_{cp} and f_{ci} to capture the information of offset, posture, and motion. The normalization and PCA are then applied to obtain Eigen features descriptor for each frame.

Here f_{cc} is the pair-wise joints differences within the current frame, i.e.,

$$f_{cc} = \{\mathbf{X}_s^c - \mathbf{X}_t^c | s, t = 1, 2, ..., n; s \neq t\}, \tag{3}$$

which is used to characterize the joints' static posture information of current frame-c. f_{cp} is the pair-wise joints differences between the current frame-c and its preceding frame-p, i.e.,

$$f_{cp} = \{\mathbf{X}_s^c - \mathbf{X}_t^p | s, t = 1, 2, ..., n\}. \tag{4}$$

f_{cp} is used to capture the dynamic property of current frame-c. Finally, to represent the overall dynamics of the current frame-c with respect to the initial frame-i, the pair-wise joints differences f_{ci} are computed between frame-c and frame-i, i.e.,

$$f_{ci} = \{\mathbf{X}_s^c - \mathbf{X}_t^i | s, t = 1, 2, ..., n\}. \tag{5}$$

By making use of PCA, we could then reduce redundancy and noise in f_c. As a result, we obtain the Eigen features $\mathbf{E}_j$ representation for each frame f_{ij}. Most energy could be covered in the first few leading eigenvectors and 95% redundant data could be removed.

3 Action Recognition with NBNN Classifier

The Naive-Bayes-Nearest-Neighbor (NBNN) [2] is used here as the classifier for Tai Chi action recognition. The Nearest-Neighbor (NN) has several advantages over most learning-based classifiers. First, it doesn't require the time-consuming learning process. Second, the Nearest-Neighbor naturally deals with a large number of classes. Third, it avoids the over fitting problem. Instead of using NBNN-based image classification [3], we use NBNN-based video classification for Tai Chi action recognition. We directly compute Video-to-Class distance rather than Video-to-Video distance. Therefore the action recognition is performed by

$$C^* = arg\ min \sum_{j=1}^{N_i} ||\mathbf{E}_j - NN_c(\mathbf{E}_j)||^2, \tag{6}$$

where $NN_c(\mathbf{E}_j)$ is the nearest neighbor of $\mathbf{E}_j$ in class-C.

4 Converting Grading Criteria to Angles or Distances between Vectors

From the methods of previous sections, each Tai Chi form could be recognized correctly. Now the next task is to convert the Tai Chi grading criteria of each action into programmable rules. We first need to study the details of the Tai Chi grading criteria [4].

Let's look at Tai Chi Chuan 24 forms. We analyze each form and its grading criteria and found that some of the criteria are related with angles between two

bones. Here is an example that a straight arm is forbidden in Tai Chi Chuan competitions. As the competition rules, the arms should always in bending (see figure 4 for detail).

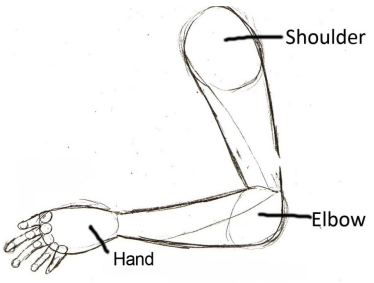

Fig. 4. Only a bend instead of a straight arm is allowed in Tai Chi Chuan competitions

Therefore, we first get the joint points of the shoulder, the elbow and the hand $\mathbf{X}_{shoulder}$, $\mathbf{X}_{elbow}$, $\mathbf{X}_{hand}$. Then we can get the upper arm bone as the vector $\mathbf{B}_{ua} = \mathbf{X}_{shoulder} - \mathbf{X}_{elbow}$ and the lower arm bone as the vector $\mathbf{B}_{la} = \mathbf{X}_{elbow} - \mathbf{X}_{hand}$. Then the angle θ between $\mathbf{B}_{ua}$ and $\mathbf{B}_{la}$ can be calculated as

$$\theta = arc\cos\left(\frac{\mathbf{B}_{ua} \cdot \mathbf{B}_{la}}{|\mathbf{B}_{ua}||\mathbf{B}_{la}|}\right). \tag{7}$$

Therefore according to the creteria, if the angle θ is close to 180^o, corresponding marks will be deducted.

Other criteria may relate to the distance between two joint points or the distance between a joint point and the ground plane. For example, if the performer performs the lunge motion (see figure 5 for detail), he is not allowed to drag his step on the ground when he moves his left foot. So we need to calculate the distance between the left foot and the ground plane. With the depth of the points on ground captured by Kinect, we can easily calculate the ground plane $\mathbf{P}_g : ax+by+cz+d = 0$. The normal of the ground plane can be directly obtained as $\mathbf{N}_g = (a, b, c)$. Thus the distance $\mathbf{D}_{xp}$ between the joint point $\mathbf{X} = (x, y, z)$ and the plane $\mathbf{P}_g$ is

$$\mathbf{D}_{xp} = \frac{|ax + by + cz + d|}{\sqrt{a^2 + b^2 + c^2}}. \tag{8}$$

Here that $\mathbf{D}_{xp}$ is the least distance that performers left foot should raise from the ground. Note that different people has different height, so that the distance would be varied. The body size should be scaled to a reference size first before we measure $\mathbf{D}_{xp}$.

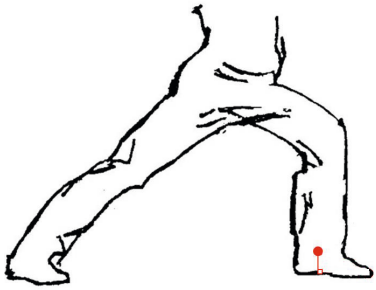

Fig. 5. In a lunge motion arm, the performer is not allowed to drag his step

5 Implementation

To implement this Tai Chi grading system, first we have to prepare a database for storing the joint positions of standard expert's actions captured by a single Kinect sensor. For each action, we normalize the data and form the feature matrix. Then we apply PCA to the feature data to reduce the data dimension.

During testing, when the system detect a new video input from Kinect, the referee has to indicate the start and end frame for different actions of the performer. Then for each action, the joint positions are stored and then normalized. We now use them to form the feature matrix. PCA will be applied to the performer feature data to reduce the data dimension. Thereafter, we can decide which category the performer belongs to by using NBNN classifier.

Within each action category, corresponding grading criteria are applied to postures such that the postures are graded objectively. Finally, the overall grade for the performer is provided automatically by the system.

The detailed procedures can be described in the following algorithm.

6 Experimental Results

Since we use OpenNI for developing our Kinect system, there are 15 joints in each frame. After normalization, we will have a huge feature dimension. f_{cc}, f_{cp} and f_{ci} contains 105, 255 and 255 pair-wise comparisons, respectively. Since each comparison generates three values $(\Delta x, \Delta y, \Delta d)$, this results in a dimension of $3 \times (105 + 255 + 255) = 1845$. Then PCA is applied to reduce redundancy and noise to obtain the Eigen features representation for each frame. From our experiments, the 95% energy is covered in the first 13 leading eigenvectors.

We take Tai Chi Chuan 24 forms as the example. Table 1 lists the details of the 24 forms and totally there are 33 postures.

Here we use the commencing position and its corresponding grading criteria for example. We will describe how to qualify criteria with conditional decisions. In figure 6, it is clear to see the postures for the commencing position.

Algorithm 1. Tai Chi Grading Procedure with a Kinect Sensor.

1: prepare a database for storing the joint positions of standard expert's actions captured by a Kinect sensor;
2: normalize the data stored;
3: for each action, form the feature matrix f_c with equation (3), (4), (5) (see section 2);
4: apply PCA to the feature data to reduce the data dimension;
5: during testing, the referee indicates the start and end frame for each action of the performer;
6: then for each action, the joint positions are stored and also normalized;
7: form the feature matrix;
8: apply PCA to the performer feature data to reduce the data dimension;
9: decide which category the performer belongs to by using NBNN classifier with equation (6);
10: within each action category, apply corresponding grading criteria to each postures by making use of equation (7) and (8);
11: the overall grade for each action is summed automatically in order to get the total mark.

Table 1. Tai Chi Chuan 24 forms

1. Commencing position	2. Part the wild horses mane to both sides (3)
3. White crane spreads its wings	4. Brush knee and twist hip on both sides (3)
5. Hand strums the lute	6. Repulse the monkey both sides (4)
7. Grasp the birds tail, left side	8. Grasp the birds tail, right side
9. Single whip	10. Wave hands like clouds (3)
11. Single whip	12. High pat on horse
13. Kick with the right heel	14. Strike opponents temple with fists
15. Turn body and kick left heel	16. Squatting and standing on one leg left side
17. Squatting and standing on one leg right side	18. A fair maiden threads the shuttle both sides
19. Pluck needle from the sea bottom	20. Open fan through the back
21. Turn body wrench, parry, punch	22. Apparent close-up
23. Cross-hands	24. Closing form

(a) (b) (c) (d) (e)

Fig. 6. Commencing position.

The following shows the grading rules and how to translate it into conditional decisions.

– Open two feet (see figure 6 (b) for detail). If the feet do not have the same width with that of the shoulders, 0.1 point will be deducted. To convert the criteria into qualified rules, we first connect the two foot joints and also connect the two shoulder joints. If the length of the two line segments has apparent difference or they are not perpendicular to the normal of the ground plane, we will deduct 0.1 point.
– Slowly raise the two arms forward horizontally (see figure 6 (c) for detail). If the hand or elbow joints are higher than the shoulders, 0.1 point will be deducted. We calculate the distance from the hand / elbow joints to the ground plane and the distance from the shoulder joints to the ground plane. If the former ones are larger than the later, we will deduct 0.1 point.
– Move the arms up (see figure 6 (d)) and then down (see figure 6 (e) for detail). If one of the elbow joints is above the hand joints, 0.1 point will be deducted. We calculate the distance from the elbow joints to the ground plane and the distance from the hand joints to the ground plane. If the former is larger than or equal to the latter, we will deduct 0.1 point.

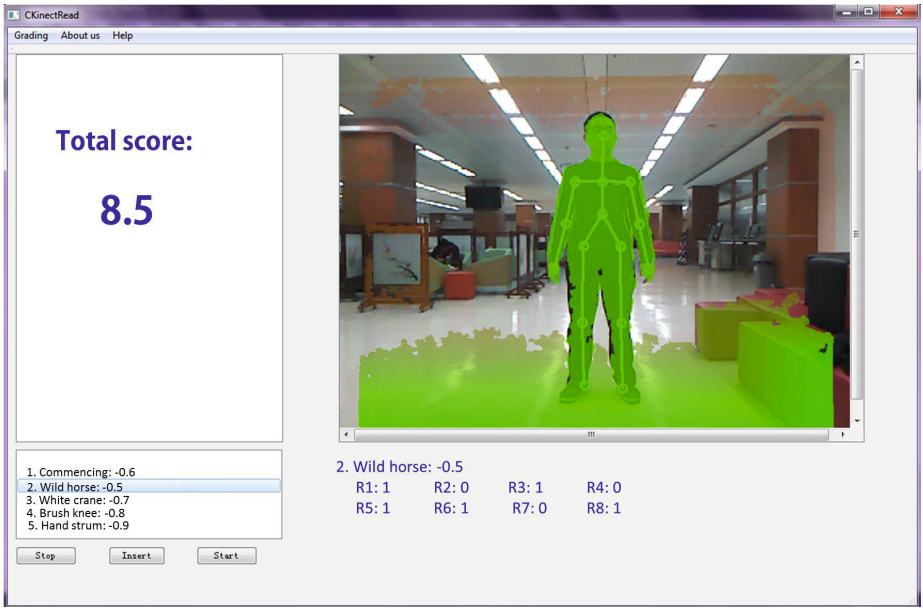

Fig. 7. Tai Chi Grading Interface

Here we only show three rules for grading the commencing position. There are in fact 8 rules in our implementation for grading each form in Tai Chi 24 forms. Through the study and on-the-spot investigation of Tai Chi, Tai Chi grading

criteria are converted into the quantified rules by applying different algorithms introduced in section 4.

Figure 7 illustrates the user interface of our system. The skeleton joints are shown together with the input video. The deducted grade and which rule is broken are illustrated in the bottom. And the total grade is given in the top left panel.

7 Conclusions

In this paper, we introduced a Tai Chi Chuan grading system with the Microsoft Kinect sensor. We first capture the joint movement of the performers skeleton. Then we record the joint differences both temporally and spatially to model the joint dynamics and configuration. Principal Component Analysis is then allied to the joint differences in order to reduce redundancy and noise. Then non-parametric Nave-Bayes-Nearest-Neighbor (NBNN) is employed as a classifier to recognize the multiple categories of Tai Chi forms. To grade the quality of each posture, we convert the competition grading criteria into decision on angles or distances between vectors. Experiments on several sample Tai Chi Chuan forms show the feasibility of our method.

Due to the slow and smooth motion of Tai Chi Quan, our method works well in the good indoor environment. In the future, we need to extend our work so that the method could be used to grade Tai Chi performance in real playground environment. Furthermore, separate forms are evaluated but not the motion coherence which is very important in Tai Chi performance. We would next focus on the motion coherence. Another future work is to use multiple Kinect sensors to capture skeleton joints and provide grading. Six Kinect sensors are required as their positions can be located as those of the referees in figure 1. The individual grading will be collected and a statistical result is expected to give the final grade. This could also solve the self-occlusion problem caused by the performer rotation.

Acknowledgments. This work is supported by the National Natural Science Foundation of China (Project no. 61005038) and an internal funding from United International College (Project no. R201312).

References

1. Bobick, A., Davis, J.: The recognition of human movement using temporal templates. IEEE Trans. on Pattern Analysis and Machine Intelligence **23**(3), 257–267 (2001)
2. Boiman, O., Shechtman, E., Irani, M.: In defense of nearest-neighbor based image classification. In: Proceedings of the International Conference on Computer Vision and Pattern Recognition, pp. 1–8 (2008)
3. Dollar, P., Rabaud, V., Cottrell, G., Belongie, S.: Behavior recognition via sparse spatio-temporal features

4. Federation, I.W.: International wushu competition rules. International Wushu Federation (2005)
5. Han, J., Shao, L., X, D., Shotton, J.: Enhanced computer vision with microsoft kinect sensor: A review. IEEE Transactions on Cybernetics, **43**(5) 1317–1333
6. Johansson, G.: Visual perception of biological motion and a model for its analysis. Journal of Attention Perception and Psychophysics **14**(2), 201–211 (1973)
7. Kaewplee, K., Khamsemanan, N., Nattee, C.: Muay thai posture classification using skeletal data from kinect and k-nearest neighbors. In: Proceedings of the International Conference on Information and Communication Technology for Embedded Systems (ICICTES 2014) (2014)
8. Klaser, A., Marszalek, M., Schmid, C.: A spatio-temporal descriptor based on 3d gradients. In: Proceedings of British Machine Vision Conference (2008)
9. Laptev, I.: On space-time interest points. International Journal of Computer Vision, **64**(2)
10. Laptev, I., Marszalek, M., Schmid, C., Rozenfeld, B.: Learning realistic human actions from movies. In: Proceedings of the International Conference on Computer Vision and Pattern Recognition, pp. 1–8 (2008)
11. Lee, M.S., Ernst, E.: Systematic reviews of tai chi: An overview. British Journal of Sports Medicine **46**(10), 713–718 (2011)
12. Liu, L., Shao, L.: Learning discriminative representations from rgb-d video data. In: Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)
13. Parameswaran, V., Chellappa, R.: View invariance for human action recognition. Journal of Attention Perception and Psychophysics **66**(1), 83–101 (2001)
14. Sun, J., Wu, X., Yan, S., Cheong, L., Chua, T., Li, J.: Hierarchical spatio-temporal context modeling for action recognition. In: Proceedings of the International Conference on Computer Vision and Pattern Recognition, pp. 2004–2011 (2009)
15. Wang, J., Liu, Z., Wu, Y., Yuan, J.: Learning actionlet ensemble for 3d human action recognition. IEEE Trans. Pattern Anal. Mach. Intell. **36**(5), 914–927 (2014)
16. Wu, D., Shao, L.: Leveraging hierarchical parametric networks for skeletal joints based action segmentation and recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Columbus, USA (2014)
17. Yang, X., Tian, Y.: Eigenjoints-based action recognition using nave-bayes-nearest-neighbor. In: Proc. Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 14–19 (2012)
18. Yuan, J., Liu, Z., Wu, Y.: Discriminative video pattern search for efficient action detection. IEEE Trans. Pattern Anal. Mach. Intell. **33**(9), 1728–1743 (2011)
19. Zanfir, M., Leordeanu, M., Sminchisescu, C.: The moving pose: An efficient 3d kinematics descriptor for low-latency action recognition and detection