Depth Induced Feature Representation for 4D Human Activity Recognition

Runlin Zhao[1]*, Yang Zhao[2]

[1]*Department of Computer Science and Technology, Yuncheng University, Yuncheng, China*

[2]*School of Automation, University of Electronic Science and Technology of China, Chengdu, China*

Received 23 November 2014, www.cmnt.lv

Abstract

Human activity recognition based on RGBD data has drawn considerable attention due to recent emergence of low-cost depth cameras. Essentially, human activities are composed by human bodies moving in four-dimensional space, (x,y,z,t). The traditional human activity recognition approaches usually ignore depth information thus degrading its discriminative performance. In this paper, our contributions are two-fold. First of all, we learn an Activity Depth Mapping (ADM) over each activity from training samples, where Activity Depth Maps are represented by Gaussian Mixture of Models (GMM) and encode depth distributions of activities. Second, we propose a novel feature representation, called Depth-Induced Multiple Channel STIPs (DIMC-STIPs), for activity representation with RGB-D data where both color and depth channels are available. The proposed feature representation is evaluated on the public dataset RGBD-HuDaAct and it remarkably improves the classification accuracy over state-of-the-art approaches.

Keywords: Kinect; human activity recognition; STIPs; GMM

Introduction

Much effort has been made in human activity understanding since human activities play important roles on smart healthcare and wellbeing [9], human-computer interfaces [30], video surveillance [19], and content-based video indexing. Visual activity recognition has been an active research topic in computer vision community [40]. So far, most of the visual action recognition approaches only considered human body movement in x-y-t subvolumes due to the high cost and low availability of depth cameras. In this case, we usually capture activities using color cameras thus losing depth information. Hence, this simplification definitely leads to discriminative performance degradation. However, both physical bodies and motions are of four dimensions, x-y-z-t, in real world. That is, human activities involve not only spatio-temporal axes but also a depth axis. The recent progress in depth sensors (e.g. Microsoft Kinect [30]) has drawn much attention on human activity recognition from RGBD data [30], [26], [18], [33], [35], [41], [43], [42].

Compared with infinite variation in appearance of human activities, depth information is a straightforward yet useful cue. The depth constraints on the structure of the 3D Scenes and activities can be directly transposed into image/ video content [36]. Consequently, learning and inferring depth from images/videos are widely used in various computer vision tasks [36], [28], [25]. The rationalization behind this is that video features that are based on image gradients have different distributions as the depth changes. In a similar token, human activities can be thought of as parts/points moving in 4D space. The conventional video-based features (e.g. STIPs) will have different distributions for different depth ranges. Therefore, we propose to model depth distributions and augment the codebook so that the codewords are depth dependent.

More specifically, we propose a technique, called activity depth mapping (ADM), which is represented by Gaussian Mixture Models (GMM) whose parameters are learned from training samples. The learned ADMs affect activity distribution over depth layers each corresponding to one of components of the GMM components. Instead of using fixed depth layers, we propose a novel depth induced feature representation, called Depth-Induced Multiple Channel STIPs (DIMC-STIPs). Moreover, we discuss two common issues in RGBD data collected by Kinect devices, Kinect calibration and noise removal. Finally, we validate the proposed approach on RGBD-HuDaAct [26].

This paper is organized as follows. Section II reviews some related work. Section III introduces the framework of the proposed approach. Section IV presents how to model and learn activity depth mappings of human activities. We discuss Depth Induced Multi-Channel STIP (DIMC-STIP) feature representation in Section V. Moreover, we discuss two common issues about RGBD data using Kinect devices in section VI. The experimental results are shown in Section VII. Section VIII concludes this paper.

2 Relate work

Many different approaches have been proposed for human activity recognition. These techniques have been surveyed recently in [40], [27], [24], [35]. Roughly, we divide activity recognition techniques into four categories, Bag-of-Features/SVM (BoF/SVM) approaches [17], Deformable Part Models (DPM) approaches[39], silhouette representation [3], [5], [14], [22], feature trajectories [38], [21], [23], [31]. Most of those activity recognition approaches are

* ***Corresponding author's*** e-mail: yczhaorunlin@163.com

only using $x-y-t$ features. This section mainly presents the related work on activity recognition using $x-y-z-t$ features.

Thanks to the recent emergence of Microsoft Kinect devices, depth based activity recognition has drawn much effort in computer vision community recently [26], [33], [34], [15], [30]. Li el al. [18] proposed a bag-of-3D-points feature representation for activity recognition from depth map sequences, where the 3D points are sampled from the silhouettes of the depth maps. They used an action graph as their classification framework, where each action is encoded in one or multiple paths in the action graph. Each node of the action graph denotes a salient postures. Since activities consist of a sequence of well defined sub-activities, the other category models the dynamics of the activities explicitly using statistical techniques. Sung et al. proposed a hierarchical Maximum Entropy Markov Model (MEMM), where a person's activity is composed of a set of sub-activities and the two-layered graph structure is inferred by using a dynamic programming approach. Sempena et al. proposed exemplar-based sequential single-layered approach using Dynamic Time Warping (DTW) to recognize actions, and performed body part tracking using depth information to recover human body joints in 3D coordinate system [29]. The BoFs/SVM approaches are widely used in activity recognition due to its simplicity and effecttiveness [17], [37], [32]. Ni et al. proposed a Depth-Layered Multi-Channel STIPs (DLMC-STIPs) framework [26], where STIPs were divided into multiple depth layered channels, and afterwards those STIPs within different depth layers are pooled correspondingly. Finally, it yields multiple depth channel histogram representation. Meanwhile, Ni et al. proposed a 3D Motion History Images (3D-MHI) using depth information in the same paper. Zhang et al. proposed a new 4D local spatio-temporal feature that combines both intensity and depth information, which is detected by along the 3D dimensions and the 1D temporal dimension to detect a feature point [41]. Here, Latent Dirichlet Allocation with Gibbs sampling is used as the classifier.

For better evaluating depth based activity recognition approaches, several activity datasets are collected by using Kinect devices in very recent years [26][34][41]. The RGBD-HuDaAct collected by Singapore Advanced Digital Science Center aims at home daily activities [26]. This database includes 12 categories, make a phone call, mop the floor, enter the room, exit the room, go to bed, get up, eat meal, drink water, sit down, stand up, take off the jacket and put on the jacket. Also, there is a background activity that contains different types of random activities. There are 30 subjects to perform these daily activities, which are organized into 14 video capture sessions. Each subject repeats 2-4 times and each video sample spans about 30-150 seconds. Therefore, there are 1189 labeled video samples in total. As the authors mentioned that the size of the database is still growing. The Robot Learning Laboratory at Cornell University collected an unstructured human activities dataset in unstructured environment for smart homes and personal assistive robotics [33], [34]. This dataset were collected by the Kinect sensor in five different environments: office, kitchen, bedroom, bathroom, and living room. Moreover, there are twelve daily activities, brushing teeth, cooking, writing on whiteboard, working on computer, talking on phone, wearing contact lenses, relaxing on a chair, opening a pill container, drinking water, cooking, talking on a chair, and rinsing mouth with water. This dataset not only provides RGBD images, but also provides skeleton motion data. The LIRIS human activities dataset contains RGBD videos showing people performing ten activities taken from daily life, discussion between two or more people, giving an object to another person, putting /taking an object into/from a box/desk, entering/leaving a room without unlocking, trying to enter a room, unlocking and entering a room, leaving baggage unattended, handshaking, typing on a keyboard, telephone conversation [1]. To evaluate the 4D local spatio-temporal features for activity recognition, a dataset is built with a Kinect installed on a Pioneer mobile robot [41]. This dataset currently contains six human activities, lifting, removing, pushing, waving, walking, and signaling. Each activity has 33 samples where each sample lasts 2 to 5 seconds.

3 The overview of the proposed approach

In this paper, we propose a novel depth induced local feature representation. First, we use Gaussian Mixture Models to capture important depth structures instead of fixed and unified layer dividing strategy. By doing so, we not only discard STIPs from clustered backgrounds due to both camera motion and noises, but also improve the distinctive ability of STIP feature representation by using depth components learned from the depth information of training STIPs. Furthermore, the learned GMM is used in the feature pooling step. Compared to DLMC-STIPs, we pool features along Gaussian depth components encoded in the learned GMM. Second, we use noise removing and Kinect alignment between RGB data and depth data to improve activity recognition performance.

Figure 1 illustrates the system framework of the proposed approach, which consists of two modules: training and testing. The training module learns the depth mapping and the classifier. The test module performs classification for any given video clip. The learning module is different from the conventional bag-of-words based classification framework in two aspects. First, It learns the GMM parameters of the activity depth mappings. Second, it generates the depth induced multi-channel (DIMC) feature representation. Given a set of training video clips, we first extract the STIPs from the RGB channel*. For each STIP, we store both its HOG-HOF feature descriptor and the corresponding depth value. The extracted STIPs are clustered to form a dictionary. The depth values of the extracted STIPs are used to learn the GMM parameters of the activity depth mappings. For each training video, we can then obtain its depth induced multi-channel (DIMC) histogram vector. The DIMC histogram vectors are used to train a SVM. We have developed an efficient spatio-temporal bilateral filter to remove depth noises, and a Kinect calibration for RGBD alignment. Both the depth noise removal and the RGB-D

* http://www.di.ens.fr/laptev/download.html

alignment are important for the performance of the activity depth mappings.

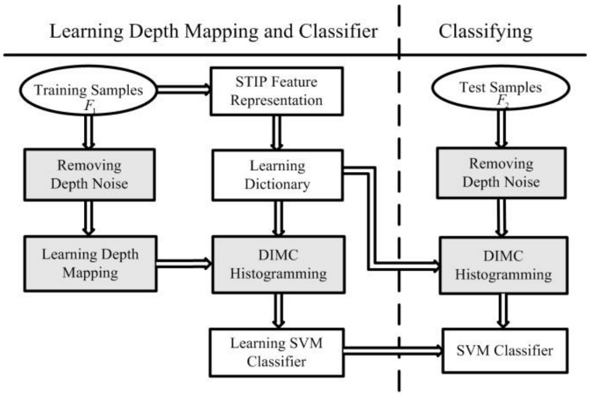

FIGURE 1 The framework of the proposed approach for human activity recognition. Note that the gray rectangles denotes the novel parts in this paper.

4 Learning activity depth mapping

Let $X=\{X^1,X^2,.......,X^C\}$ be all training samples, where $X^C=\{X_i^c\}_{i=1}^{N_c}$ is the set of STIPs extracted from training samples of the c^{th} activity,

where $X^C=\{x_i,y_i,z_i,HOG,HOF\}^T$.

Here, x_i,y_i,z_i,t_i,σ_i are the 3D coordinate *(x, y, z)*, temporal index, and the scale of the detected feature point, respectively. HOG and HOF are the feature vectors of histogram of gradient and optical flow, respectively. Given $Z=\{Z^1,Z^2,.......,Z^C\}$, where $Z^C=\{z_i\}_{i=1}^{N_c}$, we can model class-independent Activity Depth Mappings (ADM) over all activities using GMM

$$f(z)=\sum_{k=1}^{K}w_kN(z\,|\,u_k,\sigma_k)\,, \tag{1}$$

where $N(z\,|\,u_k,\sigma_k)$ is the k^{th} component of the mixture; u_k is depth mean, σ_k is depth variance, and w_K is the mixing weight. Here, each component could correspond to one important depth structure of activities. Moreover, u_k denotes the depth position, and w_K measures the relative importance degree of this component compared with the rest of the mixture. Instead of using class dependent ADM, we use class independent way to model depth distribution. By doing so, we can simply the feature pooling step without falling into the issue of class-dependent depth layer selection in the histogramming step.

Now, we can use Expectation Maximization to estimate the parameters of the GMM (refer to [4]), $\Omega=\{w_1,w_2,......,w_K\}$, $u=\{u_1,u_2,......,u_k\}$, and $\sum=\{\sigma_1,\sigma_2,......,\sigma_K\}$. Usually, we can take $K=8\sim 10$ in our experiments. After obtaining GMM distribution, we will remove the component $\{u_k,\sigma_k\}$ with very small weights w_K. To simplify descriptions, we still assume there are K components of GMM in the rest of paper.

Given the GMM of the activity depth mappings, we can classify each STIP into one of the GMM components according to its depth value *z* by

$$k=\arg\max_{l\in\{1,2,........,K\}}\exp\{-\frac{(z-ul)^2}{2\sigma_l^2}\}. \tag{2}$$

We would like to point out that there are two ways to model depth information, whole depth images and just depth of STIPs. The simple yet effective way to use depth information is to just model depth distribution of STIPs. Alternatively, we can first model whole depth distribution of both activities and background. However, most of STIPs focus on human activity not on background. Hence, we use the first way to model activity depth mappings.

5 Depth induced STIP feature representation

Local interest points along with both space and time contain significant local variation of video intensities and motions. Spatio-Temporal Interest Points are one of the most popular action representations, and Laptev et al. proposed the most representative STIPs, 3D Harris detector [16], which is a natural extension of 2D Harris detector [13]. 3D Harris interest points are local extremes of second-moment matrix, a 3-by-3 matrix composed of first order spatial and temporal derivatives. Upon the localization of STIPs, Histogram of Gradient (HOG) and Histogram of Flow (HOF) are important yet popular video features in videos [10], [20], [16], [11], [12]. Hence, STIP features consisting of HOG and HOF are widely used in activity recognition. Bag-of-Feature representation of activities can be obtained by the two steps, STIP feature coding and feature pooling. First, the STIP features are coded by being quantized into visual codewords. Second, each video clip is represented as a histogram vector over a visual codebook by pooling functions. Furthermore, an nonlinear trained SVM can be usually used in activity recognition.

In this paper, we propose a depth induced STIP feature representation as follows.

Given training samples$\{X^1,X^2,.....,X^C\}$, we learn dictionary by unsupervised clustering on HOG and HOF thus resulting in visual codeword set of size M for all activities. Define each video clip $V=\{X_1,X_2,.......,X_N\}$ as a set of N STIPs, where $X_n=\{x_n,y_n,z_n,HOG,HOF\}^T$. In the coding step, we assign feature vector X_n as v_n^k by calculating the distance between z_n and component $\{u_l,\sigma_l\}$ using Eqn. (2). Here, v_n^k is a $K\times M$ assignment vector with one of the element within the range of *k(M-1)* to *(k+1)(M-1)* as 1 and the others as 0s. We use feature pooling to generate the global histogram vector of video clip *V* by aggregating local STIPs

$$h=\frac{1}{N}\sum_{n=1}^{N}v_n^k\ , \tag{3}$$

Similar to depth-layered multi-channel representation in [26], we use depth gaussianization multi-channel representation to form h. In the feature pooling step, the histogram vectors of each video clip over visual codewords are depth dependent. The proposed feature representation is called Depth Induced Multi-Channel STIPs (DIMU-STIPs) feature representation.

6 Experiment results and analysis

6.1. EXPERIMENT SETUP

Local In this paper, we use RGBD-HuDaAct database [26] for validating the proposed algorithm. For better comparison with DLMC-STIPs, we follow the experiment setup in [26]. Hence, we use 18 subjects with 9 capture sessions, and 702 video samples belonging to 13 activity categories for evaluating the proposed approach. In our experiments, the dimensions of HOG and HOF are 72 and 90, respectively. We use Ivan's STIP implementation to extract interest point[*], and classifying accuracy and class confusion matrix are used as evaluation approach. Moreover, we use LibSVM [8] to classify human activities as multi-class classification and a leave-one-out strategy is used to evaluate the generalization capability of the proposed approach.

In the following sections, we use two sets of experiments to validate different issues of the proposed approach: (1) How many GMM components are the best for activity recognition? (2) Is a general dictionary better than a depth induced dictionary?

6.2. THE NUMBER OF GMM COMPONENTS

In our experiments, we compare activity recognition among DLMC-STIPs, the traditional STIPs, and Depth Induced Mutli-channel STIPs (DIMC-STIPs). We perform K-means clustering to the set of STIP features thus resulting in codebooks with size M. In our experiment, we take M = 256 due to only slight performance difference among different M . The comprehensive evaluation over different values of M, 128, 256, and 512, can refer to [26](Table 2). Instead of fixed equally spaced layers dividing strategy in [26], we divide all STIPs into different depth layers, where each depth layer corresponds to one of components of the GMM. Hence, we generate the histogram vector of each video clip over each depth layer according to depth distribution of all STIPs. Finally, we concatenate K channel histogram vectors into an M×K dimensional feature vectors.

Table 1 shows the activity recognizing accuracies using different number of GMM components, K = 2, 3, 4, 5. The parameters of learned GMM shows in Table 2. From this table, we can see that: (1) The accuracy of activity recognition using K = 3 is better than those of the others. (2) The general Activity Depth Models contain two classes, human posture depth distribution and background depth distribution. According to the components of learned GMM, it's easy to see that one component with much bigger mean since the background is much further away from Kinect cameras.

TABLE 1 Experimental comparison on the RGBD-HuDaAct dataset among different number of GMM components

K	K=2	K=3	K=4	K=5
Acc.	86.37%	87.71%	85.14%	83.49%

TABLE 2 The parameters of learned GMM, K=3

Para.	Comp. 1	Comp. 2	Comp. 3
μ	0.62912	117.81	195.33
σ	1.07	29.24	20
ω	4.5%	82.57%	12%

Table 3 shows the experimental comparison between DLMC-STIPs and DIMC-STIPs. The average accuracy for DLMC-STIPs is 85.50% and for DIMC-STIPs is 87.71%. We can see that the DIMC-STIPs approach significantly outperforms DLMC-STIPs.

TABLE 3 Experimental comparison between DLMC-STIPs and DIMC-STIPs

Types	DLMC-STIPs	DIMC-STIPs
Accu.	85.5%	87.71%

6.3. DEPTH LAYERS-DEPENDENT&DEPTH LAYERS-INDEPENDENT DICTIONARY LEARNING

In this section, we validate that depth induced dictionary learning is not benefit for dictionary learning. There are two potential cases, general dictionary and depth induced dictionary. In depth induced dictionary learning, we can generate different codewords over different depth layers. Table 4 shows that the performance of general dictionary is better than that of depth induced dictionary.

7 Conclusion

In this paper, we have proposed an depth induced multiple channel feature representation for activity recognition. We have introduced Activity Depth Mappings to model depth distribution of all activities by learning the parameters of GMMs. The approach has been evaluated on the popular public activity dataset, RGBD-HuDaAct. The experiments show that the proposed approach is better than existing approaches.

TABLE 4 Experimental comparison on the RGBD-HuDaAct dataset using general dictionary and depth induced dictionary, respectively, K=3

Types	General dictionary	Depth-induced dictionary
Accu.	87.71%	86.97%

Acknowledgements

This research was supported by the Yuncheng University *Production-Study-Research Cooperation Research Project* NO.CY-2014011. We also thank the anonymous reviewers for their valuable suggestions.

[*] http://www.di.ens.fr/laptev/download.html

References

[1] http://liris.cnrs.fr/voir/activities-dataset/.
[2] http://nicolas.burrus.name/index.php/research/kinectcalibration.
[3] Ben-Arie J, Wang Z, Pandit P, and Rajaram S. Human activity recognition using multidimensional indexing. IEEE Transactions on Pattern Analysis and Machine Intelligence, 24(8):1091–1104, 2002.
[4] Bishop C. Pattern recognition and machine learning. springer New York, 2006.
[5] Bobick A and Davis J. The recognition of human movement using temporal templates. IEEE Transactions on Pattern Analysis and Machine Intelligence, 23(3):257–267, 2001.
[6] Bruder H, Raupach R, Klotz E, Stierstorfer K, and Flohr T. Spatiotemporal filtration of dynamic ct data using diffusion filters. In Proceedings of SPIE, 2009 : 725857-725810.
[7] Camplani M, Salgado L, and Im´agenes G de. Efficient spatio-temporal hole filling strategy for kinect depth maps. In Proceedings of SPIE, 2012, 8920
[8] Chang C and Lin C. LIBSVM: a library for support vector machines. ACM Transactions on Intelligent Systems and Technology, 2(3):27, 2011.
[9] Cheng H, Liu Z, Zhao Y, and Ye G. Real world activity summary for senior home monitoring. In IEEE International Conference on Multimedia and Expo, 2011.
[10] Dalal N and B. Triggs. Histograms of oriented gradients for human detection. In IEEE Conference on Computer Vision and Pattern Recognition, 2005.
[11] Danafar S and Gheissari N. Action recognition for surveillance applications using optic flow and SVM. In Asian Conference on Computer Vision, 2007.
[12] Efros A, Berg A, Mori G, and Malik J. Recognizing action at a distance. In IEEE International Conference on Computer Vision, 2003.
[13] Harris C and Stephens M. A combined corner and edge detector. In Alvey vision conference, 1998.
[14] Hu M. Visual pattern recognition by moment invariants. IEEE Transactions on Information Theory, 8(2):179–187, 1962.
[15] Lai K, Bo L, Ren X, and Fox D. A large-scale hierarchical multi-view RGB-D object dataset. In IEEE International Conference on Robotics and Automation, 2011.
[16] Laptev I. On space-time interest points. International Journal of Computer Vision, 64(2):107–123, 2005.
[17] Laptev I, Marszalek M, Schmid C, and Rozenfeld B. Learning realistic human actions from movies. In IEEE Conference on Computer Vision and Pattern Recognition, 2008.
[18] Li W, Zhang Z, and Liu Z. Action recognition based on a bag of 3d points. In IEEE Conference on Computer Vision and Pattern Recognition Workshop, 2010.
[19] Liu J, Luo J, and Shah M. Recognizing realistic actions from videos in the 'wild'. In IEEE Conference on Computer Vision and Pattern Recognition, 2009.
[20] Lowe D. Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision, 60(2):91–110, 2004.
[21] Matikainen P, Hebert M, and Sukthankar R. Trajectons: Action recognition through the motion analysis of tracked features. In IEEE International Conference on Computer Vision Workshops, 2009.
[22] Meng H, Pears N, and Bailey C. A human action recognition system for embedded computer vision application. In IEEE Conference on Computer Vision and Pattern Recognition, 2007.
[23] Messing R, Pal C, and Kautz H. Activity recognition using the velocity histories of tracked keypoints. In IEEE International Conference on Computer Vision, 2009.
[24] Moeslund T, Hilton A, and Kr¨uger V. A survey of advances in vision-based human motion capture and analysis. Computer Vision and Image Understanding, 104(2):90–126, 2006.
[25] Nedovic V, Smeulders A, Redert A, and Geusebroek J. Depth information by stage classification. In IEEE International Conference on Computer Vision, 2007.
[26] Ni B, Wang G, and Moulin P. RGBD-HuDaAct: A color-depth video database for human daily activity recognition. In IEEE International Conference on Computer Vision Workshops, 2012.
[27] Poppe R. A survey on vision-based human action recognition. Image and Vision Computing, 28(6):976–990, 2010.
[28] Saxena A, Chung S, and Ng A. Learning depth from single monocular images. Advances in Neural Information Processing Systems, 18:1161, 2006.
[29] Sempena S, Maulidevi N, and Aryan P. Human action recognition using dynamic time warping. In IEEE International Conference on Electrical Engineering and Informatics, 2011.
[30] Shotton J, Fitzgibbon A, Cook M, Sharp T, Finocchio M, Moore R, Kipman A, and Blake. A Real-time human pose recognition in parts from single depth images. In IEEE Conference on Computer Vision and Pattern Recognition, 2011.
[31] Sun J, Wu X, Yan S, Cheong L, Chua T, and Li J. Hierarchical spatio-temporal context modeling for action recognition. In IEEE Conference on Computer Vision and Pattern Recognition, 2009.
[32] Sun X, Chen M, and Hauptmann A. Action recognition via local descriptors and holistic features. In IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2009.
[33] Sung J, Ponce C, Selman B, and Saxena A. Human activity detection from RGBD images. In AAAI workshop on Pattern, Activity and Intent Recognition, 2011.
[34] Sung J, Ponce C, Selman B, and Saxena A. Unstructured human activity detection from RGBD images. IEEE International Conference on Robotics and Automation, 2012.
[35] To F. On-body sensing: From gesture-based input to activity-driven interaction. Invisbile Computing, 2010.
[36] Torralba A and Oliva A. Depth estimation from image structure. IEEE Transactions on Pattern Analysis and Machine Intelligence, 24(9):1226–1238, 2002.
[37] Ullah M, Parizi S, and Laptev I. Improving bag-of-features action recognition with non-local cues. In British Machine Vision Conference, 2010.
[38] Wang H, Klaser A, Schmid C, and Liu C. Action recognition by dense trajectories. In IEEE Conference on Computer Vision and Pattern Recognition, 2011.
[39] Wang Y and Mori G. Hidden part models for human action recognition: Probabilistic vs. max-margin. IEEE Transactions on Pattern Analysis and Machine Intelligence, (99):1–1, 2011.
[40] Weinland D, Ronfard R, and Boyer E. A survey of vision-based methods for action representation, segmentation and recognition. Computer Vision and Image Understanding, 115(2):224–241, 2011.
[41] Zhang H and Parker L. 4-dimensional local spatio-temporal features for human activity recognition. In IEEE/RSJ International Conference on Intelligent Robots and Systems, pages 2044–2049, 2011.
[42] Zhao Y, Liu Z and Cheng H. RGB-Depth feature for 3D human activity recognition. IEEE/China Communications, 10(7):93–103, 2013.
[43] Zhao Y, Liu Z and Cheng H. Combining rgb and depth map features for human activity recognition. APSIPA Annual Summit and Conference, December:3–6, 2012.

Authors

Runlin Zhao, 1960.10, Shanxi, China

Current position, grades: Associate professor in the Department of Computer Science and Technology, Yuncheng University.
University studies: Yuncheng University
Scientific interest: Artificial intelligence, intelligent control, image processing, and computer application.
Publications: 5
Experience: B.Eng in Computer Science and Technology from Taiyuan University of Technology, China in 1983.

Yang Zhao, 1988.10, Shanxi, China

Current position, grades: PhD. Candidate in the School of Automation, UESTC.
University studies: University of Electronic and Science Technology of China.
Scientific interest: Pattern Recognition and Human Action Recognition.
Publications: 3
Experience: B.Eng in Automation from University of Electronic and Science Technology of China (UESTC), China in 2010.