ISSN No: 2348-4845
International Journal & Magazine of Engineering,
Technology, Management and Research
A Peer Reviewed Open Access International Journal

Human Activity Recognition Process Using 3-D Posture Data

G.Nandini
PG Scholar,
Dept of ECE,
PVKK Institute of Technology,
Anantapur, AP, India.

K.Naveen Kumar
Associate Professor,
Dept of ECE,
PVKK Institute of Technology,
Anantapur, AP, India.

S.Ravi Kumar
Associate Professor,
Dept of ECE,
PVKK Institute of Technology,
Anantapur, AP, India.

Abstract:

In this paper, we present a method for recognizing human activities using information sensed by an RGB-D camera, namely the Microsoft Kinect. Our approach is based on the estimation of some relevant joints of the human body by means of the Kinect; three different machine learning techniques, i.e., K-means clustering, support vector machines, and hidden Markov models, are combined to detect the postures involved while performing an activity, to classify them, and to model each activity as a spatiotemporal evolution of known postures. Experiments were performed on Kinect Activity Recognition Dataset, a new dataset, and on CAD-60, a public dataset. Experimental results show that our solution outperforms four relevant works based on RGB-D image fusion, hierarchical Maximum Entropy Markov Model, Markov Random Fields, and Eigenjoints, respectively. The performance we achieved, i.e., precision/recall of 77.3% and 76.7%, and the ability to recognize the activities in real time show promise for applied use.

I. INTRODUCTION:

HERE, we present a novel technique to perform user activity recognition by means of an unobtrusive motion sensor device. In particular, we adopt the Microsoft Kinect as a motion sensor mainly due to its reliability, competitive cost, and its usage for user tracking. The output of the framework proposed here (i.e., the probability of the recognized activity) represents one of the inputs of a more general activity recognition system, which reasons about different information coming from the sensing infrastructure. Human activities can be described as spatiotemporal evolutions of different body postures. We model the human body as a set of *joints* connecting some relevant body parts (e.g., arms or legs), and then, the most significant configurations of joint positions are used to define recurrent *postures*. Our solution uses three different machine learning techniques. First, a set of body joints is detected by means of the Kinect. Then, such a set is clustered by applying the K-means algorithm in order to discover the postures involved in each activity. The obtained postures are validated by support vector machines (SVMs) and hidden Markov models (HMMs) are finally applied to model each activity as a sequence of known postures. For more widespread applicability, we chose to connect the Kinect to a miniature fanless computer, which is able to process the scene with minimum levels of obtrusiveness and low power consumptions. Our current work includes three contributions. Our first contribution is to design an activity recognition method able to guarantee an acceptable accuracy, real-time processing, low power consumption. The second contribution is the release of the public Kinect Activity Recognition Dataset (KARD), which contains 18 Activities, divided into ten gestures and eight actions, each performed three times by ten different subjects. The third contribution is the validation of the proposed method against a well-known public dataset. This paper is organized as follows. Related work is outlined in Section II.

The system architecture is described in Section III. Section IV presents the experimental scenario and the results for two different datasets. Conclusions are presented in Section V.

II. RELATED WORK:

Researchers have explored different compact representations of human actions in the past few decades. In 1975, Johansson's experiment shows that humans can recognize activity with extremely compact observers [8]. Johansson demonstrated his statement using a movie of a person walking in a dark room with lights attached to the person's major joints. Even though only light spots could be observed, there was a strong identification of the 3D motion in these movies. In recent studies, Fuijiyoshi and Lipton [9] proposed to use "star" skeleton extracted from silhouettes for motion analysis. Yu and Aggarwal [10] use extremities as semantic posture representation in their application for the detection of fence climbing. Zia et al. [12] present an action recognition algorithm using body joint-angle features extracted from the RGB images from stereo cameras. Their dataset contains 8 simple actions (e.g., left hand up), and they were all taken from frontal views. Inspired by natural language processing and information retrieval, bag-of-words approaches are also applied to recognize actions as a form of descriptive action unites. In these approaches, actions are represented as a collection of visual words, which is the codebook of spatio-temporal features. Schuldt et al. [21] integrate space-time interest point's representation with SVM classification scheme. Dollar et al. [22] employ histogram of video cuboids for action representation. Wang et al. [23] represent the frames using the motion descriptor computed from optical flow vectors and represent actions as a bag of coded frames. However, all these features are computed from RGB images and are view dependent. Researchers also explored free viewpoint action recognition algorithms from RGB images. Due to the large variations in motion induced by camera perspective, it is extremely challenging to generalize them to other views even for very simple actions.

One way to address the problem is to store templates from several canonical views and interpolate across the stored views [29, 30]. Scalability is a hard problem for this approach. Another way is to map an example from an arbitrary view to a stored model by applying homography. The model is usually captured using multiple cameras [31]. Weinland et al. [32] model action as a sequence of exemplars which are represented in 3D as visual hulls that have been computed using a system of 5 calibrated cameras. Parameswaran et al. [33] define a view-invariant representation of actions based on the theory of 2D and 3D invariants. They assume that there exists at least one key pose in the sequence in which 5 points are aligned on a plane in the 3D world coordinates. Weinland et al. [34] extend the notion of motion-history [35, 29] to 3D. They combine views from multiple cameras to build a 3D binary occupancy volume. Motion history is computed over these 3D volumes and view-invariant features are extracted by computing the circular FFT of the volume. The release of the low-cost RGBD sensor Kinect has brought excitement to the research in computer vision, gaming, gesture-based control, and virtual reality. Shotton et al. [6] proposed a method to predict 3D positions of body joints from a single depth image from Kinect. Xia et al. [24] proposed a model based algorithm to detect humans using depth maps generated by Kinect. There are a few works on the recognition of human actions from depth data in the past two years. Li et al. [25] employ an action graph to model the dynamics of the actions and sample a bag of 3D points from the depth map to characterize a set of salient postures that correspond to the nodes in the action graph. However, the sampling scheme is view dependent. Lalal et al. [27] utilize the Radon transformation on depth silhouettes to recognize human home activities. The depth images were captured by a ZCAM [28]. This method is also view dependent. Sung et al. [26] extract features from the skeleton data provided by Prime Sense from RGBD data from Kinect and use a supervised learning approach to infer activities from RGB and depth images from Kinect.

Considering they extract features from both types of imageries, the result is interesting but at the same time not as good as one would expect. In this work, we present an action recognition algorithm using a HOJ3D representation of postures constructed from the skeletal joints' locations extracted from depth images. Taking advantage of the Kinect devise and J. Shotton et al.'s algorithm [6], this method improves on the previous ones in that it achieves excellent recognition rates and is also view invariant and real time.

III. ACTIVITY RECOGNITION SYSTEM:

The system proposed here (see Fig. 1) aims at automatically inferring the activity performed by the user according to a set of known postures. The system can be decomposed into three components addressing three different aspects. The first is responsible for *features detection*, that is for the extractionof a set of points to be used for distinguishing different body postures. The detection and classification of such Postures is accomplished by the *posture analysis* techniques, based on K-means and SVM, and, finally, *activity recognition* is performed by means of HMMs built on the set of known postures.

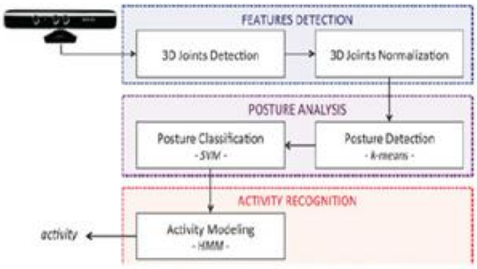

Fig. 1. Images captured by the Kinect are processed to detect a set of joints, which are subsequently normalized with respect to scale and position. These joints represent the features used to define a set of postures, which are detected by applying a K-means clustering and classified by means of SVMs. HMMs are finally used to model an activity in terms of postures and classify new sequences coming from the Kinect.

A. Features Detection:

The first processing step consists in identifying the features of interest. Since our goal is to understand what activity the user is performing at a given time, we need to track movements focusing on those body parts, which are mostly involved while executing a particular activity. The human body consists of many interacting systems, none of which can work in isolation. In particular, we started from the musculoskeletal system, which is responsible for supporting the human body and enabling its movements in accordance with the stimuli provided by the nervous system. To describe the user's movements, we chose to track the human skeleton focusing on significant parts such as head, neck, torso, arms, legs, hands, and feet. The different parts of the human skeleton can be modeled as segments connected to each other by nodes, called *joints*, which limit the movement of each body Since we are interested in a more general representation suitabl for a dynamic environment, we performed some preliminary tests to measure the relevance of the set of joints provided by the Kinect.part in the 3-D space.

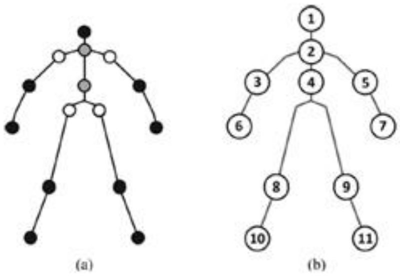

Fig. 2. (a) Fifteen joints detected by means of the Kinect. Reference joints (gray): *neck, torso*. Selected joints (black): *head, elbows, hands, knees, feet*. Discarded joints (white): *shoulders, hips*. (b) Eleven joints of the feature set.

In [8], due to the sensitiveness of the IR sensor, some joints are misdetected if two segments overlap (e.g., hands touching other body parts), or not detected at all due to the presence of objects between the sensor and the user. For this reason, we evaluated the system by measuring the recognition rates achieved on a limited number of selected subset of joints. Some noisy joints that are redundant (i.e., wrists, ankles) due to their closeness to other joints (i.e., hands, feet) or not relevant at all for activity recognition (i.e., spine, neck, hip and shoulders) have been discarded. The final set of joints we chose as features is shown in black in Fig. 2(a), while the joints we discarded are white. Since the appearance of the skeleton depends on several factors, as, for example, the distance between the user and the sensor, the detected features need to be normalized for scale. For doing that, we moved the detected joints to a new coordinate system fixed at the torso (considering as the *x*-direction the left-right hip axis) and all features have been scaled according to a reference distance, *h*, between the neck and the torso joints.

B. Posture Analysis:

As already mentioned, our idea is that each activity can be considered as a sequence of different configurations of joints. In order to identify those configurations that are effectively related to meaningful users postures, a classification procedure is needed. SVMs [3] are supervised learning models used for binary classification and regression, which aim to find the optimal separating hyperplane between two classes according to some labeled training samples. Unfortunately, building the training set non large-scale data is a costly operation, which may also lead to worse performance because of the presence of noise. Thus, a more effective way of building the training set could be to select the most informative samples, that is, in our case, the most recurrent joint configurations. we overcome the problem of recognizing the same activities performed at different speeds. Moreover, the posture-based representation does not affect the capacity of the system to distinguish among different activities with different durations.

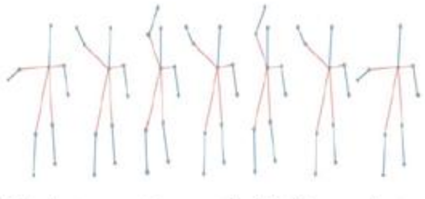

Fig. 3. Posture sequence from one repetition of the "high arm wave" gesture.

That we in those cases, a greater number of postures would be involved making longer activities intrinsically different from the shorter ones. In Fig. 3, an example of the posture sequence extracted from one repetition of the "high arm wave" gesture is shown.

C. Activity Recognition:

In order to fully satisfy the design requirements, the system should also correctly classify multiple instances of the same activity, which may generally involve different sequences of postures.

The activity recognition process is based on HMMs similarly to what is described in [11] and [29]. We modeled each activity using a discrete HMM, whose observed symbols are the postures we have previously extracted. In a system whose instantaneous condition may be represented as belonging to one of *N* distinct states, we denote the different states as *S* = {*S*1, *S*2, . . . , *SN* }, and the state at time *t* as *qt*.

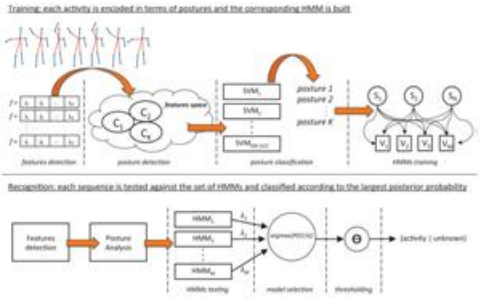

Fig. 4. Activity Recognition Process.

During the training, each activity is analyzed to extract a set of postures, which are used to build an HMM. A new activity is recognized by testing the corresponding posture sequence against the set of HMMs and selecting the model with the largest posterior probability. The activity recognition process is described in Fig. 4. The training phase consists of four steps: 1) for each activity, the features of interest are detected; 2) the features space is organized into *k* clusters, which represent the most significative postures; 3) the detected postures are refined by means of SVMs classification; and 4) an HMM which models the activity is built. To recognize an activity, we need to 1) detect the features, 2) detect and classify the postures involved in the activity, 3) test the posture sequence against all HMMs; 4) select the model which maximizes the posterior probability, and 5) compare such probability against a threshold to classify an activity as known or unknown.

V. CONCLUSION:

In this study, we presented a framework for human activity recognition using 3-D posture data. In particular, we referred to a scenario where the whole environment is equipped with a number of sensory nodes capable of unobtrusive monitoring of some raw measures such as temperature, humidity, and light level. In this context, the Kinect is responsible for gathering high-level information about what the user is doing. In order to obtain a suitable representation of the human body, we detected 11 relevant joints and encoded a relevant set of joints into *postures*. Thus, since each posture represents a recurrent pattern of joints positions, an activity can be described as a sequence of known postures. To support a real office environment, we mainly focused on a solution made of simple processing blocks, which are functional in the scenario we considered. Other approaches could perform better on single tasks, e.g., providing more reliable posture representation mechanisms or more complex activity models, but we aimed to develop a framework which can be easily integrated in a more general AmI system. To this end, we evaluated the effectiveness of our technique using two different datasets. The first is KARD, a new public dataset we collected to overcome the unreliability of some other existing data collections. The second is CAD-60, which allowed comparison with some state-of-the-art techniques. The experiments showed that our method is able to capture a general model of the activity regardless of the user.

In particular, the activity models we built are independent of who performs the action, independent of the speed at which the actions are performed, scalable to large number of actions, and expandable with new actions. Moreover, since repeated sequences of the same posture are merged, the proposed method is able to recognize the same class of activities performed with different time durations. Using the public Cornell Activity Dataset, we obtained an overall precision and recall of 77.3% and 76.7%, respectively, demonstrating that our framework outperforms four of the techniques we considered as reference. Due to the requirements of the overall AmI system, we implemented a real prototype of the activity recognition module by connecting the Kinect to a miniature computer getting a realtime processing of the observed scene with minimum levels of obtrusiveness and low power consumptions.

Analogously to other approaches, the main limitations or our system are primarily related to the capacity of the Kinect of providing a stable video stream and, consequently, a reliable joint detection mechanism. In this regard, future work can concern the improvement of the pose estimation process in order to deal with frame loss and body occlusions, which are the main causes nof misclassification.

REFERENCES:

[1] J. Yamato, J. Ohya, and K. Ishii, “Recognizing human action in time-sequential images using hidden Markov model,” in Proc. IEEE Comput. Soc. Conf. Comput. Vision Pattern Recog. Proc., 1992, pp. 379–385.

[2] M. P. V. Kellokumpu and J. Heikkila, “Human activity recognition using sequences of postures,” in Proc. IAPR Conf. Mach. Vision Appl., 2005, pp. 570–573.

[3] B. Scholkopf and A. J. Smola, Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond. Cambridge, MA, USA: MIT Press, 2001.

[4] A. Oikonomopoulos, I. Patras, and M. Pantic, “Spatiotemporal salient points for visual recognition of human actions,” IEEE Trans. Syst., Man, Cybern. B, Cybern., vol. 36, no. 3, pp. 710–719, Jun. 2005.

[5] G. Willems, T. Tuytelaars, and L. Gool, “An efficient dense and scaleinvariant spatio-temporal interest point detector,” in Proc. 10th Eur. Conf. Comput. Vision, 2008, pp. 650–663.

[6] S. J. Preece, J. Y. Goulermas, L. P. J. Kenney, D. Howard, K. Meijer, and R. Crompton, “Activity identification using body-mounted sensors— A review of classification techniques,” Physiol. Meas., vol. 30, no. 4, pp. R1–R33, 2009.

[7] L. Bao and S. S. Intille, “Activity recognition from user-annotated acceleration data,” in Pervasive Computing (ser. Lecture Notes in Computer Science,

vol. 3001), A. Ferscha and F. Mattern, Eds. Berlin, Germany: Springer, 2004, pp. 1–17.

[8] P. Cottone, G. Lo Re, G. Maida, and M. Morana, "Motion sensors for activity recognition in an ambient-intelligence scenario," in Proc. PerCom Workshops, 2013, pp. 646–651.

[9] J. Wang, Z. Liu, Y. Wu, and J. Yuan, "Learning actionlet ensemble for 3D human action recognition," IEEE Trans. Pattern Anal. Mach. Intell., vol. 36, no. 5, pp. 914–927, May 2014.

[10] W. Li, Z. Zhang, and Z. Liu, "Action recognition based on a bag of 3D points," in Proc. IEEE Comput. Soc. Conf. Comput. Vision Pattern Recog. Workshops, 2010, pp. 9–14.

Author's Details:

G.Nandini, Completed B.Tech in ECE Department from BIT Institute of technology, Hindupur. Persuing Masters in Digital Electronics and Communication Sytem (DECS) in PVKK Institute of Technology, Anantapur.
Mail id : nandini.yadav431@gmail.com

Mr.K.Naveen Kumar, Completed B.Tech in ECE Department from SKD Engineering College, Anantapur. Completed Masters in Sri Venkateswara College of engineering and Technology (SVCET), Chittor. Currently working as Associate Professor in Dept of ECE,PVKK Institute of Technology, Anantapur.
Mail.id:sainaveen705@gmail.com

Mr.S.Ravi Kumar, Completed B.Tech in ECE Department from G PULLA REDDY Engineering College, Kurnool. Completed Masters in Digital Systems and Computer Electronics in BITS Engineering College, Warangal. Currently working as Associate Professor in Dept of ECE ,PVKK Institute of Technology, Anantapur.
Mail id : ravik.s4u2020@gmail.com