Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 100 (2016) 78 – 84

Procedia Computer Science

Conference on ENTERprise Information Systems / International Conference on Project MANagement / Conference on Health and Social Care Information Systems and Technologies, CENTERIS / ProjMAN / HCist 2016, October 5-7, 2016

A Depth Camera-based Human Activity Recognition via Deep Learning Recurrent Neural Network for Health and Social Care Services

S. U. Park[a], J. H. Park[a], M. A. Al-masni[a], M. A. Al-antari[a], Md. Z. Uddin[b], T. -S. Kim[a]*

[a]*Department of Biomedical Engineering, Kyung Hee University, Yongin, Republic of Korea*
[b]*Department of Computer Education, Sungkyunkwan University, Seoul, Republic of Korea*

Abstract

Human activity recognition (HAR) has become an active research topic in the fields of health and social care, since this technology offers automatic monitoring and understanding of activities of patients or residents. Depth camera-based HAR recognizes human activities using features from depth human silhouettes via conventional classifiers such as Hidden Markov Model (HMM), Conditional Random Fields etc. In this paper, we propose a new HAR system via Recurrent Neural Network (RNN) which is one of deep learning algorithms. We utilize joint angles from multiple body joints changing in time which are represented a spatiotemporal feature matrix (i.e., multiple body joint angles in time). With these derived features, we train and test our RNN for HAR. In order to evaluate our system, we have compared the performance of our RNN-based HAR against the conventional HMM- and Deep Belief Network (DBN)-based HAR using a database of Microsoft Research Cambridge-12 (MSRC-12). Our test results show that the proposed RNN-based HAR is able to recognize twelve human activities reliably and outperforms the HMM- and DBN-based HAR. We have achieved the average recognition accuracy of 99.55% for the activities. The results are 7.06% more accurate than that of the HMM-based HAR and 2.01% more accurate than that of the DBN-based HAR.

Peer-review under responsibility of the organizing committee of CENTERIS 2016

* Corresponding author. Tel.: +82-31-201-3731; fax: +82-31-201-3731;
E-mail address: tskim@khu.ac.kr

1877-0509
Peer-review under responsibility of the organizing committee of CENTERIS 2016
doi:10.1016/j.procs.2016.09.126

Keywords: Human Activity Recognition, Depth Imaging Sensor, Deep Learning, Recurrent Neural Network, Health Care

1. Introduction

Human activity recognition (HAR) is a technique to recognize various human activities via external sensors such as inertial or video sensors. In recent years, HAR has evoked significant interest among researchers in the areas of health care, social care, and life care services[1], since it allows automatic monitoring and understanding of activities of patients or residents in smart environments such as smart hospitals and smart homes. For instance, at smart home, a HAR system can automatically recognize residents' activities and create daily, monthly, and yearly activity logs. These life logs can provide residents' habitual patterns which medical doctors evaluate for further health care suggestions. Especially for elderly people, a HAR system can recognize their falls or unusual activity patterns and alert or inform their caregivers.

The basic methodology of activity recognition involves activity feature extraction, modeling, and recognition techniques. Video-based HAR is a challenging task as it has to consider whole body movement of a human and does not follow rigid syntax like hand gestures or sign languages. Hence, a complete representation of a full human body is essential to characterize human movements properly in this regard. Though many researchers have been exploring video-based HAR systems due to their practical applications, accurate recognition of human activities still remains as a major challenge.

Generally, video-based HAR can be divided into two categories according to motion features: namely, marker-based and vision-based. The former is based on a wearable optical marker-based motion capture (MoCap) system that is widely used as it offers an advantage of accurately capturing complex human motions. However, it has a disadvantage that the optical sensors must be attached to the body and requires multiple camera settings. The latter is based on depth video cameras and it is marker-free[2]. This approach is getting more attention these days due to the absence of tracking wearable markers, hence making the HAR system easy to be deployed in daily applications.

As for the recognition techniques, until now HMMs have been widely used in many HAR systems, as HMMs are capable of temporal pattern decoding[3-5]. However, recently recognition via deep learning is getting considerable attentions due to its power to learn deep structures of patterns[6-12]. Basically, deep learning refers to neural networks that exploit layers of non-linear data processing for feature classification. These layers are hierarchically organized and process the outputs of the previous layers. Deep learning techniques have outperformed many traditional methods in computer vision[8-12]. Deep learning techniques are very promising to address the requirements of HAR in two ways. First, performance can be significantly improved over conventional recognition techniques. Second, deep learning approaches have the potential to uncover features that are tied to the dynamics of human motion (i.e., from simple motion encoding in lower layers to more complex motion dynamics in upper layers of the network). This may be useful to scaling up HAR to activities that are more complex.

Recently, there has been a HAR work via Deep Belief Network (DBN)[6] which is one of Deep Neural Networks (DNNs) proposed by Hilton in 2004[7]. DBN uses Restricted Boltzmann Machines (RBMs) in learning and it avoids local minimum problem with less training time. However, Recurrent Neural Networks (RNNs) is a better choice than DBN, since it could offer more discriminative power over DBN as time sequential information can be encoded or learned through RNNs. Although HMM can handle time sequential information, now researchers prefer RNN over HMM for its improved discriminant capability.

In this paper, we present a RNN-based HAR system. We have performed HAR with the features of body joint angles. The performance of RNN for HAR has been compared to other conventional recognizers such as HMM and DBN.

2. Materials and Methods

In this section, we introduce our RNN-based HAR system. Our HAR system proceeds to the following steps. First, we create an input feature matrix of joint angles computed from the MSRC-12 activity dataset[13]. Second, we train RNN with the training feature matrix. Third, we evaluate the trained RNN with test data sets by recognizing