ISSN: 2454-9940

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

E-Mail :
editor.ijasem@gmail.com
editor@ijasem.org

www.ijasem.org

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN2454-9940
www.ijsem.org
Vol 13, Issuse . 3 Sep 2022

Human Action Recognition From Depth Maps And Postures Using Deep Learning

Mr.E.Sunil,

Abstract:
Human Activity Recognition is one of the active research areas in computer vision for various contexts like security surveillance, healthcare and human computer interaction. Over the past years, several methods published for human action recognition using RGB (red, green, and blue), depth, and skeleton datasets. Most of the methods introduced for action classification using skeleton datasets are constrained in some perspectives including features representation, complexity, and performance. However, there is still a challenging problem of providing an effective and efficient method for human action discrimination using a skeleton dataset. The first input is depth images and second input is a proposed moving joints descriptor (MJD) which represents the motion of body joints over time, in order to maximize feature extraction for accurate action classification, CNN channels are trained with different inputs and for Score fusion we are planning to use neural networks. Our proposed method was implementation on public datasets like MSRAction3D.

Keywords: MJD, MSRAction3D, CNN, Human action.

INTRODUCTION

Human action recognition has been a popular research topic for a long time, not only because of their widespread application in a variety of applications, such as intelligent surveillance systems, human-robot interaction, and home care systems, but also because it remains a difficult research problem. Convolutional Neural Networks (CNN) have been widely used in many research areas, particularly computer vision and pattern recognition, thanks to the advancement of deep learning over the last few years, and have achieved remarkable performance on classification, detection, segmentation, and tracking tasks. In the subject of action recognition, there are a few things to keep in mind. As human-machine interaction becomes one of the most researched topics in multimedia processing, traditional communication techniques are being developed in order to address technological advancements and enable disabled people to communicate with machines and understand their activities using computer computing. Many research works have attempted to model and then recognise people's behavior using motion analysis. In this paper, we focus on human behaviour analysis from video scenes, and it is worth noting that many information is hid behind gesture, rapid motion, and walking pace[1],[2].Due to the expressive features provided by the two types of data, recent human action recognition research has been

Assistant Professor, Dept. of CSE,
Malla Reddy Engineering College (Autonomous), Secunderabad, Telangana State

directed toward using depth maps or body postures to represent the action[3]. A strong representation that gives distinct aspects of each action for classification is critical to the success of an action recognition algorithm. For some activities, using depth map data for action recognition is still confusing, resulting in incorrect classification because two actions may appear similar from the front, but they appear differently from the side[4].When substantial occlusions occur, the depth maps collected by the depth cameras are quite noisy, and the 3D positions of the tracked joints may be completely incorrect, increasing intraclass variances in the actions[5].They propose ”Skepxels,” a spatio-temporal format for skeletal sequences that uses CNN’s 2D convolution kernels to fully leverage ”local” connections between joints. Thye use Skepxels to convert skeleton movies into images with adjustable dimensions, and then use the generated images to build a CNN-based framework for successful human action recognition.

LITERATURE SURVEY

Zhao-Xuan Yang [7] Two single-view RGB datasets (KTH and TJU), two well-known depth datasets (MSR action 3-D and MSR daily activity 3-D), and one revolutionary multiview multimodal dataset are all used to validate the proposed methodology (MV-TJU). In both RGB and depth modalities, the extensive experimental data show that this method outperform the popular 2-D/3-D component model-based approaches and other competitive techniques for numerous human action recognition. C. Krishna Mohan:[8] They proposed leveraging action bank features to recognise human actions in videos using a deep fully convolutional architecture. Action bank features are linear patterns that describe the resemblance of the video to the action bank films. They are computed against a predefined set of pictures defined as an action ban Miki, Hiroshi:[9] They offer a method for recognising objects that examines the relationship between human behaviour and object functions, with the goal of improving our method by adding human activities into dynamic object segmentation. Ziaeefard, Maryam :[10] This research proposes a revolutionary normalized-polar histogram-based human action recognition algorithm. The action of accumulating skeletonized pictures was described.

As a motion pattern, it incorporates distance and angle. The most important aspect of this action is highlighted in cyan. It is represented by convolving all skeleton model frames round their centre. To categorize people’s behavior, a multi-class SVM with two levels was utilised, first with generic features and subsequently with salient features. Mejdi DALLEL :[11] O”Industrial Human Action Recognition Dataset (InHARD)” is a large-scale RGB+Skeleton action recognition dataset that they introduced. Our database contains 4804 distinct action samples split over 38 movies from 14 different types of industrial activity. They complete the development of an end-to-end regression classification LSTM network in order to assess our dataset using the metrics suggested. Yang Si :[12] They created a novel neural network by combining an autoencoder with a pattern recognition neural network. human activities deep neural network recognition. The model they suggested was confirmed by Experiments were conducted, and the model’s benefits were discovered. by comparing performance They presented a model that was available. numerous significant contributions:

They revealed a unique approach of combining multiframe data into a single picture in their study. This strategy is flexible. a method for mechanically extracting human action traits based on the deep neural network Lei Zong ,Chen Xu ,HongLin Yuan :[12] Currently, typical RF fingerprint recognition relies on a preset determination formula, which has drawbacks such as a high prior knowledge demand and a limited application range. They should employ an RF fingerprint identification method based on convolutional neural networks to tackle these issues (CNN). The study focuses on three aspects: RF fingerprint extraction, convolutional neural network architecture, and identification and verification of wireless transmitters. [13] fingerprint information is not utilised fraudulently. The deep convolutional neural network (DCNN) outperforms hand-crafted feature techniques. Most CNN models have the drawback of fixed scale images, however they have a new FLD method termed an improved DCNN with image scaling. For the first time, the confusion matrix is used as a performance indicator in FLD. The amounts of the experimental results based on the LivDet 2011 and LivDet 2013 data sets further confirm that our method outperforms others in terms of detection performance. [14] To match latent fingerprints collected at crime scenes to a huge collection of reference prints and provide a candidate list of prospective mates, an automated latent fingerprint recognition system with high accuracy is required. They present an automated latent fingerprint recognition technique that uses Convolutional Neural Networks (ConvNets), and their results against a reference database of 100K rolled prints are 64.7 percent for the NIST SD27 and 75.3 percent for the WVU latent databases.

EXISTING SYSTEM

Sometimes might be finding products is easy than waiting in the billing queue because it consumes more time of the customer. So now by taking the motivation of this scenario which was regularly done in all the Shoppe we are designing this system which can be benefited for the customer in all the means and also it was benefited for the Shoppe owner also. So, we design a system by this, the customer can know their bill while adding the items in the cart. The best and most useful example of this Supermarket Basket is that if a customer purchases can easily billed.

PROPOSED SYSTEM

This system brings new innovation than existing shopping system. The main purpose of this project is to provide centralized and automated billing system using web. Along with the automatic billing some special features incorporated are along .We use new term that is Supermarket Basket.

MODULES:

1. **Data Collection:** The first step is to collect multivariate time series data from the phone’s and the watch’s sensors. The sensors are sampled with a constant frequency of 30 Hz. After that, the sliding window approach is utilized for segmentation, where the time series is divided into subsequent windows of fixed duration without interwindow gaps (Banos et al., 2014). The sliding window approach does not require preprocessing of the time series, and is therefore ideally suited to real-time applications.

2. **Preprocessing:** Filtering is performed afterwards to remove noisy values and outliers from the accelerometer time series data, so that it will be appropriate for the feature extraction stage. There are two basic

types of filters that are usually used in this step: average filter (Sharma et al., 2008) or median filter (Thiemjarus, 2010). Since the type of noise dealt with here is similar to the salt and pepper noise found in images, that is, extreme acceleration values that occur in single snapshots scattered throughout the time series. Therefore, a median filter of order 3 (window size) is applied to remove this kind of noise.

3. Feature Extraction: Here, each resulting segment will be summarized by a fixed number of features, i.e., one feature vector per segment. The used features are extracted from both time and frequency domains. Since, many activities have a repetitive nature, i.e., they consist of a set of movements that are done periodically like walking and running. This frequency of repetition, also known as dominant frequency, is a descriptive feature and thus, it has been taken into consideration.

4. Standardization: Since, the time domain features are measured in (m/s 2), while the frequency ones in (Hz), therefore, all features should have the same scale for a fair comparison between them, as some classification algorithms use distance metrics. In this step, Z-Score standardization is used, which will transform the attributes to have zero mean and unit variance, and is defined as

$$xnew = (x-\mu)/ \sigma$$

where μ and σ are the attribute's mean and standard deviation respectively (Gyllensten, 2010).

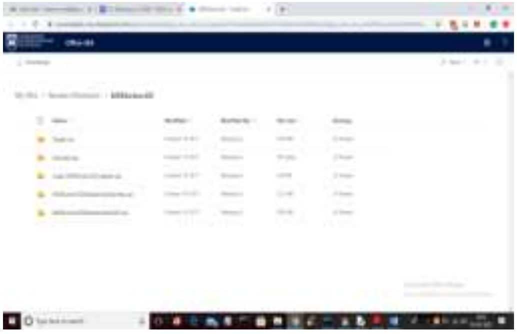
Human Action Recognition from depth maps and Postures using Deep Learning

In this paper author is using CNN (Convolution Neural Networks) algorithm to recognize human action as this algorithm will extract important features by filtering same data multiple times in order to maximize chances of accurate action classification, CNN channels are trained with different inputs features which will not happen in existing RGB Depth algorithm which will get train on two features such as images and skeleton data.

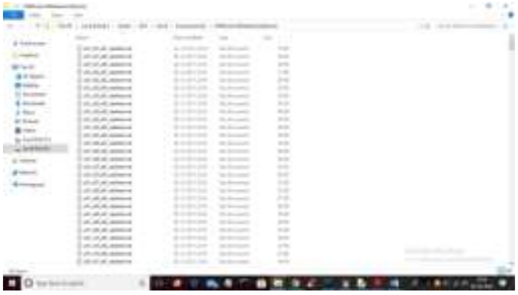
As existing algorithm are not efficient so author using CNN algorithm which already proves its success in various fields such as image classification, weather and stock prediction etc.

To train CNN algorithm author is using MSRAction3D skeleton dataset which contains 20 different actions such as 'high arm wave', 'horizontal arm wave', 'hammer', 'hand catch', 'forward punch', 'high throw', 'draw x', 'draw tick', 'draw circle', 'hand clap','two hand wave', 'side-boxing', 'bend', 'forward kick', 'side kick', 'jogging', 'tennis swing', 'tennis serve', 'golf swing', 'pick up & throw'.

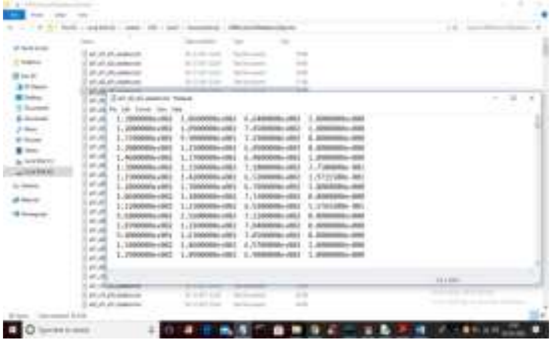
All this actions data are taken from MSRAction3D dataset and below is the screen shots of that dataset



From above page we downloaded 'MSRAction3DSkeleton(20joints)' dataset and this dataset is captured using DEPTH cameras so it will record only skeleton values and below are the dataset files screen shot



In above dataset screen each file contains skeleton data and in file name a01 refers to action 1 (actions are from 1 to 20) and s01 is the subject id and e01 is the instance ID. From above files we are training CNN with action details only so after training when we upload any file then CNN will predict action from that file. Each file will contains skeleton values as below screen

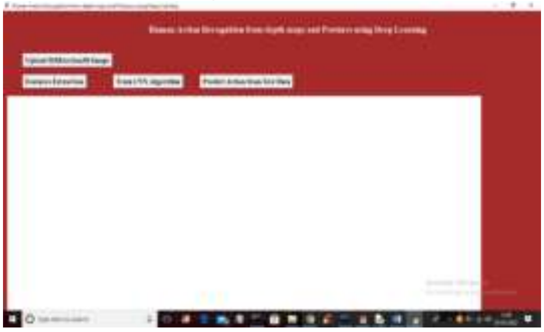


To implement this project we have designed following modules

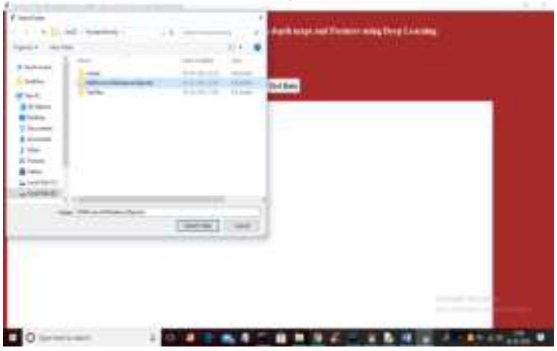
1) Upload MSRAction3D Image: using this module we will upload action dataset to application
2) Features Extraction: using this module we will read each file and then extract features (dataset values) values and action value will be consider as class label and then we will visualize movement in graph format
3) Train CNN Algorithm: using this module we will input extracted features to CNN and then CNN will get trained and then will apply TEST data on trained model to calculate accuracy and confusion matrix graph
4) Predict Action from Test Data: using this module will upload test file and then CNN will read features from that test file and then identify action from that test file

SCREEN SHOTS

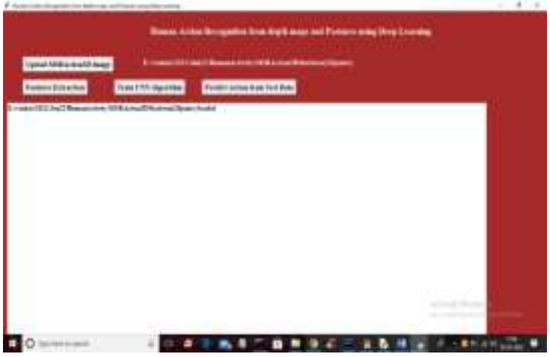
To run project double click on 'run.bat' file to get below screen



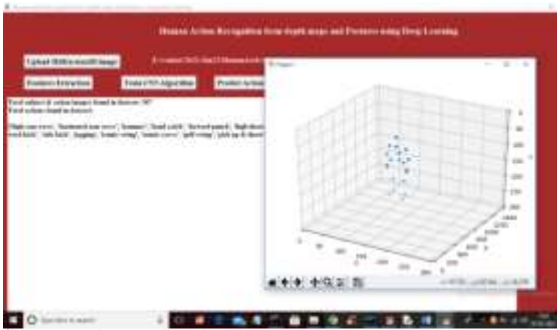
In above screen click on 'Upload MSRAction3D Image' button to upload dataset and to get below screen

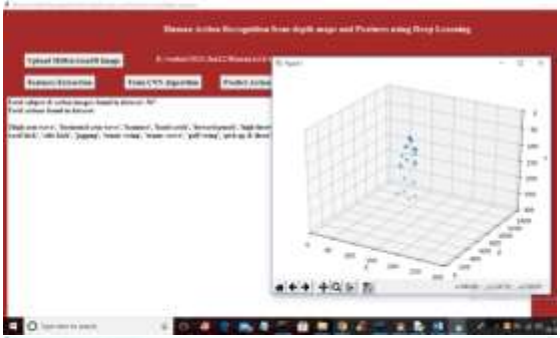


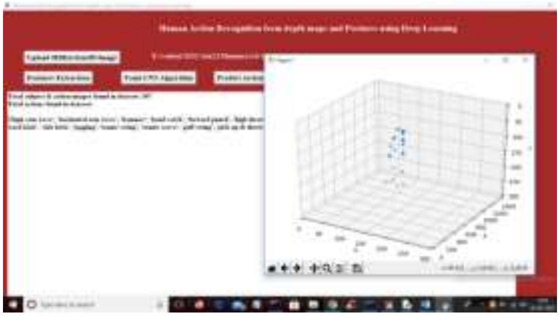
In above screen selecting and uploading MSRACTION dataset and then click on 'Select folder' button to load dataset and to get below screen



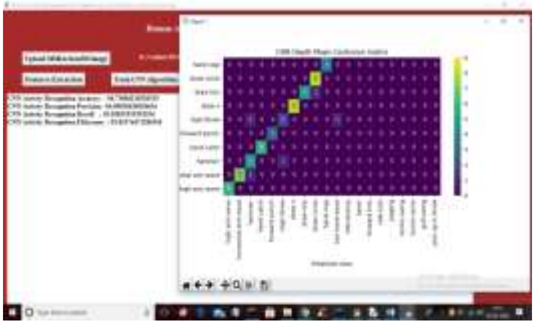
In above screen dataset loaded and now click on 'Features Extraction' button o read all files then build a features array and then visualize one skeleton image





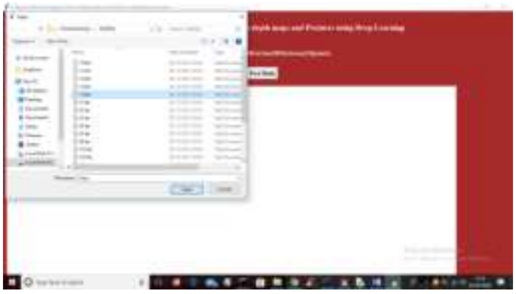


In above screen I am displaying total files found in dataset as 567 and displaying available 20 various actions such as 'high arm wave', 'horizontal arm wave' etc. In above graph you can see skeleton is moving which indicate action of person in dataset. When you upload then you will see skeleton movement in graph. After seeing close that graph and then click on 'Train CNN Algorithm' button to train CNN and to get below output

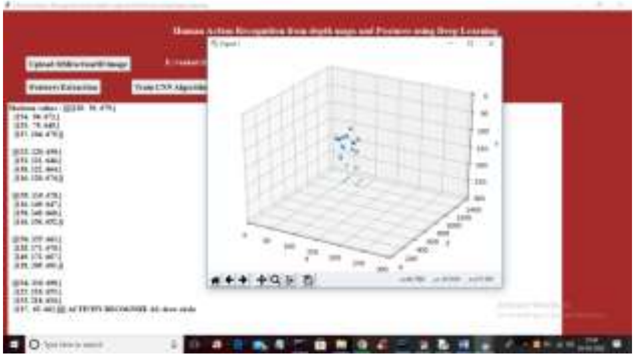


In above screen with CNN we got action recognition accuracy as 94% and in confusion matrix graph x-axis represents predicted action classes and y-axis represents original classes and we can all classes prediction values showing in diagnol boxes are the correct prediction and out of diagnol are the wrong prediction and very few value are there out of diagnol so CNN performance is good and it got 94% accuracy also. Now close above graph

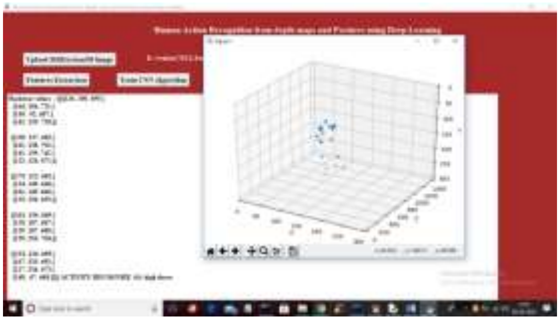
Now click on 'Predict Action from Test Data' button to upload test data file and then CNN will recognize action from that test file data.



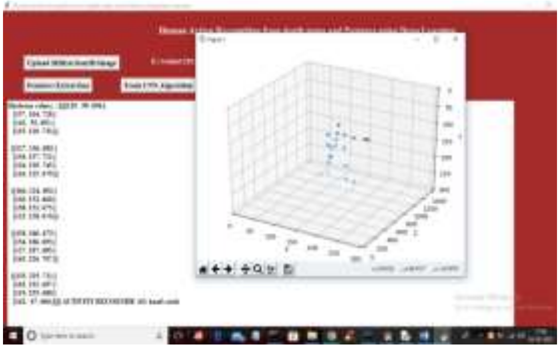
In above screen selecting and uploading '14.txt' file and then click on 'Open' button to load test file and to get below action recognition result



In above screen all values in square bracket are the skeleton values and in last line we got output as 'Activity Recognized as 'draw circle' and in graph we can see the movement of the skeleton



In above screen action recognized as 'high throw'



In above screen action recognized as 'hand catch' and similarly you can upload other files and test them

CONCLUSION

A method for human action recognition from depth map and posture data using deep convolutional neural networks has been proposed. Two action representations and three convolutional neural networks channels were used to maximize feature extraction by fusing the results of the three CNN channels together. The method has been evaluated on three public benchmark datasets. The classification accuracy of the three datasets are better than most existing state of the art methods that are based on either depth data or posture data. This work claims that different action representations provide different cues. One representation carries action features that are absent in the other representation.

REFERANCES

1 C. Schuldt, I. Laptev, and B. Caputo, "Recognizing human actions: A local SVM approach," in Proc. IEEE 17th Int. Conf. Pattern Recognit., vol. 3. Cambridge, U.K., Aug. 2004, pp. 32–36.

2] J. Sun, X. Wu, S. Yan, L.-F. Cheong, T.-S. Chua, and J. Li, "Hierarchical spatio-temporal context modeling for action recognition," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Miami, FL, USA, Jun. 2009, pp. 2004–2011.

3 I. Laptev, M. Marszalek, C. Schmid, and B. Rozenfeld, "Learning realistic human actions from movies," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Anchorage, AK, USA, Jun. 2008, pp. 1–8.

4] W. Li, Z. Zhang, and Z. Liu, "Action recognition based on a bag of 3D points," in Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Workshops, San Francisco, CA, USA, Jun. 2010, pp. 9–14.

5] J. Wang, Z. Liu, Y. Wu, and J. Yuan, "Mining actionlet ensemble for action recognition with depth cameras," in Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on, pp. 1290–1297, IEEE, 2012

6] Q. Ke, M. Bennamoun, S. An, F. Sohel, and F. Boussaid, "Skeleton optical spectra based action recognition using convolutional neural networks," arXiv preprint arXiv:1703.03492, 2016.

7 Liu, An-An; Su, Yu-Ting; Jia, Ping-Ping; Gao, Zan; Hao, Tong; Yang, Zhao-Xuan (2015). Multipe/Single-View Human Action Recognition via Part-Induced Multitask Structural Learning. IEEE Transactions on Cybernetics, 45(6), 1194–1208. doi:10.1109/tcyb.2014.2347057

8 Ijjina, Earnest Paul; Mohan, C. Krishna (2014). [IEEE 2014 13th International Conference on Machine Learning and Applications (ICMLA) - Detroit, MI, USA (2014.12.3-2014.12.6)] 2014 13th International Conference on Machine Learning and

Applications - Human Action Recognition Based on Recognition of Linear Patterns in Action Bank Features Using Convolutional Neural Networks. , (), 178–182. doi:10.1109/icmla.2014.33

9 Miki, Hiroshi; Kojima, Atsuhiro; Kise, Koichi (2008). [IEEE 2008 Second International Conference on Future Generation Communication and Networking (FGCN) - Hainan, China (2008.12.13-2008.12.15)] 2008 Second International Conference on Future Generation Communication and Networking - Environment Recognition Based on Human Actions Using Probability Networks. , (), 441–446. doi:10.1109/fgcn.2008.62