Motions Detection of Senior Citizen Using Machine Learning

V. C. Bagal[1], Pratik Hanpode[2]

Department of MCA, K. K. Wagh Institute of Engineering Education and Research, Pune University, Maharashtra, India

Student, Department of MCA, K. K. Wagh Institute of Engineering Education and Research, Pune University, Maharashtra, India

Abstract: ***Recognizing human motions is important from security point of view at any level and scenario. As there are plenty of human motions in a fraction of second, so classification of each motion is challenging task in real world. A Human activity Recognition System recognizes the Shapes and or orientation depending on implementation to task the system into per forming some job. Movement is a form of nonverbal information. A person can make numerous movements at a time. The proposed work aims to detect the movement and actions of a person using image detection methodology. Human activity recognition (HAR) aims to recognize activities from a series of observations on the actions of subjects and the environmental conditions. The vision-based HAR research is the basis of many applications including video surveillance, healthcare, and human-computer interaction (HCI). The proposed work is suitable to identify objectionable human motions of senior citizen who live alone at home.***

Keywords: Human activity, Image detection, sensing technology, feature extraction, moment feature.

1. Introduction

Human motion detection is an ability to identify human body gesture via sensors and determine human activity or action. Most of the human daily tasks can be simplified or automated if they can be recognized via HAR system. Human activities have an inherent hierarchical structure that indicates the different level so fit, which can be considered as a three-level categorization. First, for the bottom level, there is anatomic element and these action primitives constitute more complex human activities. Afterthe action primitive level, the action/activity comes as the second level. Finally, the complex interactions form the top level, which refers to the human activities that involve more than two persons and objects. In this paper, we follow this three-level categorization namely action primitives, actions/activities, and interactions. This three-level categorization varies a littlefrom previous surveys and maintains a consistent theme. Action primitives are those atomic actions at the limb level, such as “stretching the left arm,” and “raising the right leg.” Atomic actions are performed by a specific part of the human body, such as the hands, arms, or upperbody part .With the upgrades of camera devices, especially the launch of RGBD cameras inthe year 2010, depth image-based representations have been a new research topic and have drawn growing concern years.

2. Literature Survey

There are several surveys in the human activity recognition literature. Gavrila (1999) separated the research in 2D (with and without explicit shape models) and 3D approaches. In Aggarwal and Cai (1999), a new taxonomy was presented focusing on human motion analysis, tracking from single view and multiview cameras, and recognition of human activities. Similar in spirit to the previous taxonomy, Wang et al. (2003) proposed a hierarchical action categorization hierarchy. The survey of Moeslund et al. (2006) mainly focused on pose-based action recognition methods and proposed fourfold taxonomy, including initialization of human motion, tracking, pose estimation, and recognition methods.

A fine separation between the meanings of “action” and “activity” was proposed by Turaga et al. (2008), where the activity recognition methods were categorized according to their degree of activity complexity. Poppe (2010) characterized human activity recognition methods into two main categories, describing them as “top-down” and “bottom-up.” On the other hand, Aggarwal and Ryoo (2011) presented a tree-structured taxonomy, where the human activity recognition methods were categorized into two big sub-categories, the “single layer” approaches and the “hierarchical” approaches, each of which have several layers of categorization.

Modeling 3D data is also a new trend, and it was extensively studied by Chen et al. (2013b) and Ye et al. (2013). As the human body consists of limbs connected with joints, one can model these parts using stronger features, which are obtained from depth cameras, and create a 3D representation of the human body, which is more informative than the analysis of 2D activities carried out in the image plane. Aggarwal and Xia (2014) recently presented a categorization of human activity recognition methods from 3D stereo and motion capture systems with the main focus on methods that exploit 3D depth data. To this end, Microsoft Kinect has played a significant role in motion capture of articulated body skeletons using depth sensors.

Although much research has been focused on human activity recognition systems from video sequences, human activity recognition from static images remains an open and very challenging task. Most of the studies of human activity recognition are associated with facial expression recognition and/or pose estimation techniques. Guo and Lai (2014) summarized all the methods for human activity recognition from still images and categorized them into two big categories according to the level of abstraction and the type of features each method uses.

3. Methodology

Activity detection is related to the position of a human at a given time in a stiff image or sequence of images i.e. moving images. In case of a moving sequence, itcan be followed by tracking of the movement in the scene, but this is morerelevant to the applications such as sign language. The underlying concept ofactivity detection is that human eyes can detect objects, which machines cannot, with that much accuracy as that of humans. From a machine point of view it is just like a man fumble with his senses to find an object. The factors, which make the activity detection task difficult to solve are: The human pose in the image varies due to its changing position whether it be sitting, standing, bending or sleeping .The rotation can be both in and out of theplane. **Movement Recognition** means interpreting human actions via mathematical algorithms using images and camera samples. However, the identification and recognition of posture, gait, proxemics, and human behaviors is also the subject of gesture recognition techniques. However, the typical approach of a recognition system has been shown in the below figure:

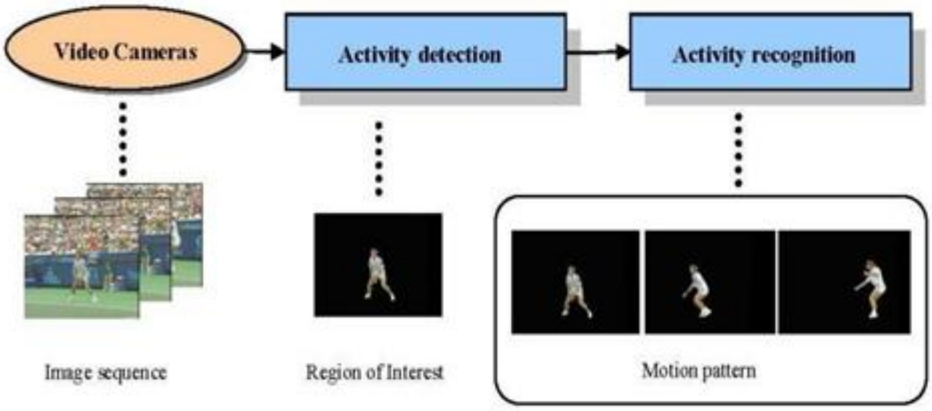

Figure 1.1: Hand Gesture Recognition Flow Chart

The proposed work is real time activity recognition which ultimately controls the image with a jpg extension and the camera samples of real-time web cam method. During the project, four gestures were chosen to represent four navigational commands that are sitting, standing, bending and sleeping. A simple computer vision application was written for the detection and recognition of the four gestures and their translation into thecorresponding commands for the actions and tracking.. Thereafter, the program was tested on a web cam with actual movement of the person in real-time and the results were observed.

The sense of sight is arguably the most important of man's five senses. It provides a huge amount of information about the world that is rich in detail and delivered at the speed of light. However, human vision is not without its limitations, bothphysical and psychological. Through digital imaging technology and computers,man has transcended many visual limitations. While computers have been central to this success, for the most part man is the sole interpreter fall the digital data. For along time, the central question has been whether computers can be design to analyze and acquire information from images autonomously in the same natural way humans can. The main difficulty for computer vision as a relatively young discipline is thecurrent lack of a final scientific paradigm or model for human intelligence and human vision itself on which to build a infrastructure for computer or machine learning. The use of images has an obvious drawback. Humans perceive the world in 3D, but current visual sensors like cameras capture the world in 2D images. The result is the natural loss of a good deal of information in the captured images. Without a proper paradigm to explain the mystery of human vision and perception, the recovery of lost information (reconstruction of the world) from 2D images represents a difficult hurdle for machine vision. However, despite this limitation, computer vision has progressed, riding mainly on the remarkable advancement of decade old digital image processing techniques, using the science and methods contributed by other disciplines such as optics, neurobiology, psychology, physics, mathematics, electronics, computer science, artificial intelligence and others. Computer vision techniques and digital image processing methods both draw the proverbial water from the same pool, which is the digital image, and therefore necessarily over lap. Image processing takes a digital image and subjects it to processes, such as noise reduction, detail enhancement, or filtering, for producing another desired image as the result. For example, the blurred image of a car registration plate might be enhanced by imaging techniques to produce a clear photo of the same so the police might identify the owner of the car. On the other hand, computer vision takes a digital image and subjects it to the same digital imaging techniques but for the purpose of analysing and understanding what the image depicts. For example, the image of a building can be fed to a computer and thereafter be identified by the computer as a residential house, a stadium, high-rise office tower, shopping mall, or a farm barn. The proposed work has considered geometrical shape of human pose and based on defined thresholds and real time parametric variation, the segmentation for human position is accomplished. Based on retrieved specific shape, certain application-oriented commands have to be generated. The predominant uniqueness of the proposed scheme is that it does not employ any kind of prior training and it is functional in real time without having any databases or training datasets. Unlike tradition approaches of images, datasets-based recognition system; this approach achieves human activity recognition in real time, and responds correspondingly. This developed mechanism neither introduces any computational complexity nor does it cause any user interferences to achieve tracing of human gesture. The primary goal of the proposed work in this paper is to study emerging CNN architectures and compare the impact on the recognition rate of the human activity due to the different architectures. Also, a modified architecture is developed and also tested. The architectures are described in the following subsections. The word “single modal” is used to indicate that the inputs are achieved from one type of sensor while “multi-modal” is used to indicate that inputs from multiple sensors are used as input to the CNN.

Feature Extraction

To enhance the feature space, we calculated the pitch, roll and normal value for each of the accelerometer using the following equations,

$$pitch=\mathbf{atan}(\sqrt{}x\mathbf{2}+z\mathbf{2})$$

$$roll=\mathbf{atan}(-x/z)$$

$$norm=\sqrt{}x\mathbf{2}+y\mathbf{2}+z\mathbf{2}$$

This increased the feature space from 12 attributes to 24 attributes. Segmentation For segmentation, our approach was to consider sliding windows, of various sizes, to aggregate the data points within the window. This helps us capture the chronological variation between the data points. Performing this activity, mitigates the risk that the classifier itself isn’t temporal in nature. It is observed that the accuracy for prediction for non-overlapping windows is higher, however the confidence interval is also wide. Here, if the window size is n, then the new data set would be 1/n the original dataset. This increases the risks of over-fitting, due to decrease in training points.

In overlapping windows, though the size of dataset still reduces, it isgreater than 1/n. Also overlapping windows ensure that the transition of time is maintained and the data points are not independent of each other. Mean and standard deviationas aggregation functions initially and performed segmentation on raw dataset. This newdataset was then tested against Naïve Bayes and Random Forest, the top two classifiers fromour preliminary analysis. Non-overlapping windows vs. overlapping window in both the cases, we found the peak to be around window size of 13-14, i.e. 1.95s -2.10s. Further to confirm our findings, we tested the overlapping window for newly generated features. Again, Human Activity Recognition: Group C| 6 this time we considered the mean and variance for each attribute across the window. This time we only considered Random Forest, since it had given us the best results in previous case. Figure 8. Accuracy across different window sizes for Random Forest with new features Finally, we concluded that window size of 14 gave us the best results. Next, we moved on to feature extraction. 7.3. Feature Extraction Having multiple accelerometers increases redundancy in the data being observed. Thus, as the first step of feature extraction we decided to test which accelerometers give us new information and are relevant as opposed to the redundant ones. For carry out this analysis, we took all the 48 new features, and grouped them by the accelerometer number. Then we performed the performance test on the exhaustive combination of sensors using Random Forest. Accuracy comparison for exhaustive combination set of sensors.

4. Results

In this proposed work, it keeps the track of a senior citizen person and its movements using a webcam of the laptop and the image that it uploads. Here we considered dataset for four major actions done by a human on his daily basis whether it be standing, sitting, sleeping and bending and the techniques and algorithms they employ and the success/ failure rate of these systems. Accordingly, we made a detailed comparison of these systems and analysed their efficiency. Following are the sample data points of human motion

class	x1	y1	z1	v1	x2	y2	z2
waving	0.517079	0.344584	0.11697	0.999978	0.523217	0.333297	0.09798
waving	0.517671	0.344379	0.15576	0.999979	0.523309	0.332956	0.13444
waving	0.517681	0.343629	0.16771	0.999981	0.523273	0.331897	0.14694
waving	0.517669	0.343105	0.15622	0.999982	0.523238	0.331445	0.13549
waving	0.517498	0.343236	0.1778	0.999984	0.523212	0.331677	0.15625
waving	0.517377	0.34331	0.18259	0.999985	0.523196	0.331842	0.16106
waving	0.517273	0.343358	0.18565	0.999986	0.523181	0.331978	0.16415
waving	0.517432	0.343356	0.09031	0.999986	0.523284	0.332066	0.06918

class	x1	y1	z1	v1	x2	y2	z2
sitting	0.500846	0.535521	0.12959	0.999991	0.509913	0.522087	0.11012
sitting	0.500061	0.529321	0.11515	0.999992	0.508062	0.517514	0.09252
sitting	0.498902	0.526781	0.10962	0.999992	0.506191	0.514383	0.08739
sitting	0.497237	0.525568	0.09876	0.999992	0.504268	0.512694	0.07619
sitting	0.496669	0.524522	0.08428	0.999993	0.503714	0.511525	0.06123
sitting	0.496044	0.524368	0.08305	0.999993	0.503196	0.511225	0.05888
sitting	0.495476	0.524373	0.08325	0.999993	0.502773	0.511121	-0.0628
sitting	0.494889	0.524444	0.08332	0.999994	0.502344	0.511131	0.06327
sitting	0.495037	0.525066	0.08544	0.999994	0.502454	0.511862	0.06668
sitting	0.495146	0.525554	0.09091	0.999994	0.502598	0.512426	0.07149
sitting	0.495226	0.525846	0.10738	0.999994	0.502723	0.512838	0.09014

class	x1	y1	z1	v1	x2	y2	z2
jumping	0.489901	0.336113	0.11866	0.999985	0.496924	0.323576	0.09493
jumping	0.48996	0.335983	0.11903	0.999985	0.496851	0.323422	0.09615
jumping	0.489949	0.335823	0.11774	0.999985	0.496646	0.323257	0.09523
jumping	0.489822	0.335822	0.13131	0.999985	0.49633	0.323255	0.10883
jumping	0.48979	0.335788	0.13668	0.999985	0.496183	0.323197	0.11444
jumping	0.489651	0.336064	0.14286	0.999985	0.495892	0.323485	0.12109
jumping	0.48944	0.336148	0.14048	0.999985	0.495584	0.323573	0.1188
jumping	0.489459	0.335884	0.13884	0.999985	0.495579	0.323406	0.11734
jumping	0.489476	0.335198	0.13765	0.999984	0.495575	0.322803	0.11611
jumping	0.489477	0.334328	0.13159	0.999984	0.495578	0.321898	0.11003

class	x1	y1	z1	v1	x2	y2	z2
walking	0.503558	0.136574	0.21581	0.999967	0.513273	0.122703	0.18293
walking	0.506055	0.11526	0.11591	0.99997	0.516097	0.099409	0.07268
walking	0.508093	0.110003	0.09917	0.999973	0.518832	0.095974	0.05232
walking	0.508528	0.09168	0.09124	0.999975	0.519557	0.074842	0.05149
walking	0.508109	0.088123	0.08296	0.999977	0.51855	0.071301	0.04342
walking	0.506784	0.077834	0.06204	0.999978	0.516523	0.061591	0.02838
walking	0.506242	0.05809	0.08993	0.999977	0.51497	0.03972	0.05069
walking	0.505513	0.052671	0.10658	0.999974	0.513447	0.03486	0.07119
walking	0.499904	0.034398	0.11297	0.999973	0.506966	0.020089	0.07926
walking	0.495544	0.017124	0.13227	0.999973	0.503043	0.005077	0.09887
walking	0.4922	0.019491	0.14421	0.999974	0.500848	0.002537	0.11398

5. Conclusion

We carried out a comprehensive study of human motion classification. The accuracy of the machine to detect the current action of the human at that particular moment of time. This accuracy is the success of the machine that it learns from the detection techniques the accuracy can thus vary based on the loss and success ratios. The a maximum accuracy achieved is of 98%. Background of the pictures should be plain to get accurate analysis of recognition of gestures and poses.

References

[1] Aggarwal, J. K., and Cai, Q. (1999). Human motion analysis: a review. Comput. Vis. Image Understand. 73, 428–440. doi:10.1006/cviu.1998.0744

[2] Akata, Z., Perronnin, F., Harchaoui, Z., and Schmid, C. (2013). “Label-embedding for attribute-based classification,” in Proc. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (Portland, OR), 819–826.

[3] Alahi, A., Ramanathan, V., and Fei-Fei, L. (2014). “Socially-aware large-scale crowd forecasting,” in *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (Columbus, OH), 2211–2218.

[4] AlZoubi, O., Fossati, D., D’Mello, S. K., and Calvo, R. A. (2013). “Affect detection and classification from the non-stationary physiological data,” in *Proc. International Conference on Machine Learning and Applications* (Portland, OR), 240–245.

[5] Amer, M. R., and Todorovic, S. (2012). “Sum-product networks for modeling activities with stochastic structure,” in Proc. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (Providence, RI), 1314–1321.

[6] Andriluka, M., Pishchulin, L., Gehler, P. V., and Schiele, B. (2014). “2D human pose estimation: new benchmark and state of the art analysis,” in Proc. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (Columbus, OH), 3686–3693.

[7] Anirudh, R., Turaga, P., Su, J., and Srivastava, A. (2015). “Elastic functional coding of human actions: from vector-fields to latent variables,” in Proc. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (Boston, MA), 3147–3155.

[8] Bandla, S., and Grauman, K. (2013). “Active learning of an action detector from untrimmed videos,” in *Proc. IEEE International Conference on Computer Vision* (Sydney, NSW), 1833–1840.