HUMAN ACTIVITY RECOGNITION

SENSE MAKING & INSIGHT DISCOVERY

PROJECT REPORT

Team Members –

ANAND RAJAN
A0163200B
GU JIAHAO GAELAN
A0163252M
SUMA MULPURU
A0163231U
SUNIL PRAKASH
A0163449X

AIM

To make sense of data from modality sensors for robust human action recognition. Then develop an algorithm that would classify the different activities into categories using either data from single sensor or from fusion of multiple sensors.

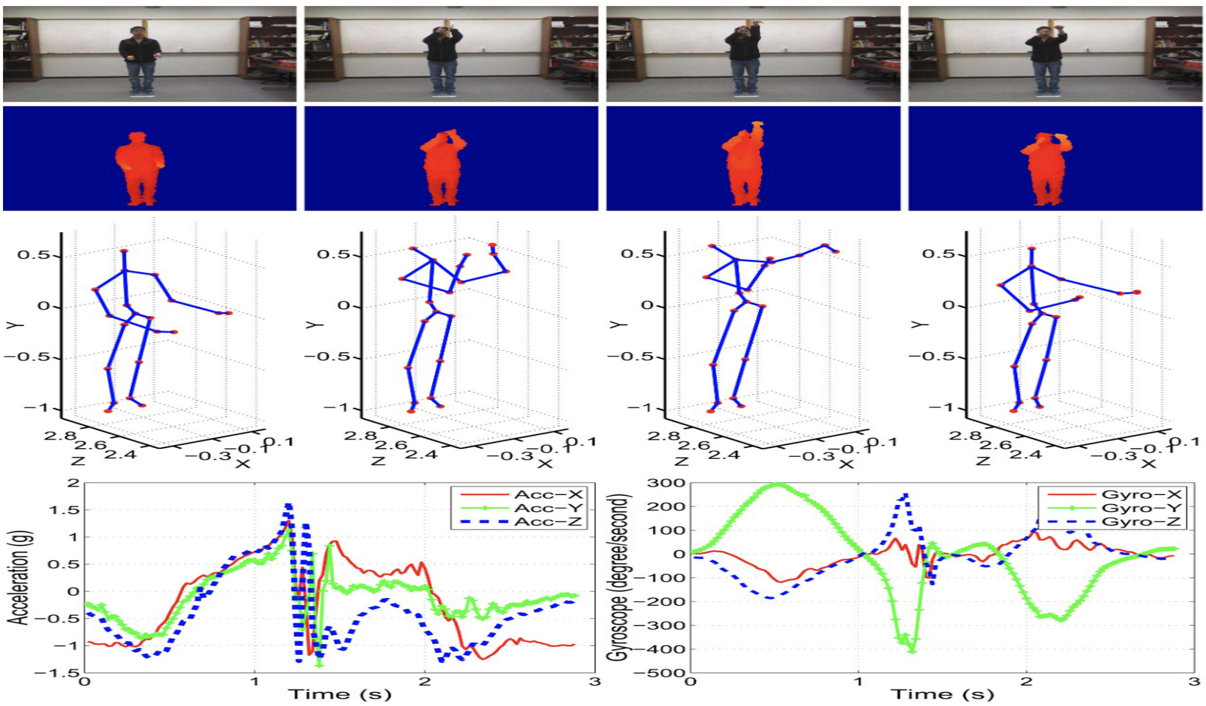
DATA EXPLORATION

Problem Description

The human activity detection has many applications in several fields like biometrics, surveillance, help monitor aged people in home environments etc . Different types of sensors can be used to address this task. The use of multi modal sensors for human activity recognition is increasing these days. Data from four different temporarily synchronised modalities is available. Using this data of either all the types or any single type the human activities must be classified.

Data Set Explanation

The dataset is a freely available dataset named UTD-MHAD, that has four temporarily synchronized data modalities. These modalities include RGB videos, depth videos, skeleton positions, and inertial signals from a Kinect camera and a wearable inertial sensor for a comprehensive set of 27 human actions performed by 8 subjects (4 males, 4 females). Each subject repeated each action 4 times. After removing three corrupted sequences, the dataset includes 861 data sequences.



LITERATURE SURVEY

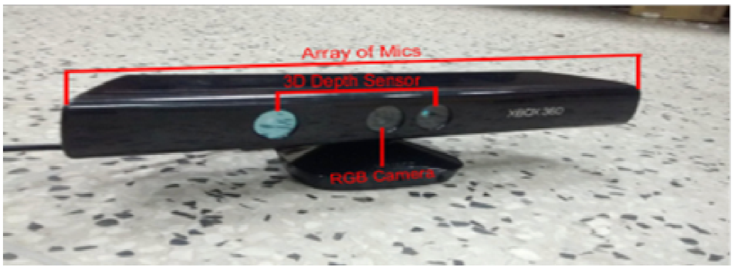
Human Activity Recognition is a hot research topic as it may enable different applications from most commercial (gaming) to most assistive (help monitor aged people in home environments) ones. Basic human activities such as sitting, walking, lying down, standing etc. are recognized in real time using simple features, to accomplish a bigger goal of developing an elderly people health monitoring system using Kinect. Human activity is complex and dynamic and therefore the algorithms should model nuances in human activities. Each activity may in turn comprise of sub activities. Example the activity of brushing teeth consists of sub activities such as squeezing toothpaste, bringing brush close to face, brushing and so forth. Hands play an especially important role in carrying out many activities. Motion information is also important for classifying a person's activities. This task can be accomplished by using different sensors like wearable sensors or even vision based devices.

<u>Wearable Sensors</u>

Wearable technology (also called wearable gadgets) is a category of technology devices that can be worn by a consumer and often include tracking information related to health and fitness. Other wearable tech gadgets include devices that have small motion sensors to take photos and sync with mobile devices. Wearable sensors have diagnostic, as well as monitoring applications. Their current capabilities include physiological and biochemical sensing, as well as motion sensing. The wearable inertial sensor used here was the low-cost wireless inertial sensor. The placement of wearable sensors is related to the locations where the sensors are placed and how they are attached to those locations. Wearable inertial tracking is well accepted due to its convenience for free-style motion tracking with high accuracy. With its character of low power consumption, wearable sensor consequently provides long-term and accurate monitoring of daily activities in free living environment. Wearable inertial tracking allows for unlimited estimation of limb orientations under fast motions, which could improve the performance of motion capture evidently. Human soft tissue artifact is a main source of errors, no matter the wearable sensors are mounted on a garment or directly attached to skin.

<u>Kinect Sensors</u>

Kinect is a Motion sensor that provides a natural user interface (NUI) which allows users to interact intuitively and without any intermediary device, such as a controller. By the use of this depth sensors i.e Kinect it is possible to design activity recognition systems exploiting depth maps, that are a good source of information as they are not affected by environmental light variations and can provide the body shape, thus simplifying the problem of human detection and segmentation. Kinect is an unobtrusive motion sensor device which is reliable, of competitive cost, powerful, easy to set up, use and accessible. With affordable devices like Kinect" depth information has also come into play". This type of device facilitates extraction of 3D points of body joints. The Kinect hardware contains a depth sensor, a color (RGB) camera, a four-microphone array and motorized tilt. The Kinect technology has revolutionized the way humans interact with machines. Action recognition has a wide range of application areas such as computer vision, robotics, machine learning, ambient intelligence, medical and many more commercial uses. In medical purpose doctors can operate a patient from a remote location using Kinect.



It is used for retrieving 3D information of a scene analyzing the depth map and skeletal joint information of the human body. This helps the Kinect sensor to identify the type of action being performed by the person such as standing, walking, punching, sitting, waving etc. The Kinect sensor senses the environment and generates a depth map for it. The human body is tracked using skeletal tracking by using the mean shift algorithm. In skeletal tracking, the Kinect sensor recognizes 24 joints in the human body which represent different body parts. Using the 3D joint information, the Kinect identifies the gestures and actions being performed by the human body and then the machine responds according to the action input. If the human body is taken as a model set of joints connecting the relevant body parts, then the most significant configurations of its positions are used to define recurrent postures. The activity recognition method should be able to guarantee an acceptable accuracy, real time processing, low power consumption which Kinect provides. Kinect follows the below steps to perform image acquisition:

1. Colour image frames extraction: Out of various resolutions available we have to obtained only 640x480 resolution images.

2. Depth image frame extraction: It is also obtained at the resolutions of 640x480 images.

3. Skelton data used to track and extract the activity of user.

4. Background subtraction

Skeleton streams are the most important features of a Kinect. It provides the position and location of the persons whether they tracked or not. The skeletons which are not tracked are given to zero value returned.

TECHNOLOGICAL APPROACH

The approach employed is the Hidden Markov Model (HMM), which works very well with sequential data and is essentially a Markov process with hidden and unobservable states. The actions (or output) are visible to the viewer, but the sequences that lead to the output are hidden. The model was trained on the skeleton, inertial and depth data for the first 3 actions - swipe left, swipe right and wave.

EXPERIMENTAL RESULTS & PERFORMANCE EVALUATION

The results of the testing were favourable in general. The skeleton model achieved an accuracy of 67%, inertial achieved 94% and depth, 67%. The inertial model was the most accurate due to the nature of the data since the actions were only recorded in a few seconds and there was insignificant data loss, the model would recognize the actions very well.

As for the depth model, it depends on visual details and requires some effort to extract features based on edge detection (performed with the Sobel-Feldman algorithm). Initially the features were extracted and convoluted(CNN) before applying the HMM classification.Since the depth sequences may be sensitive to occlusions and the textures of the images are not as good as that of coloured images,the model turned out to be inaccurate. After applying the edge detection algorithm and the corresponding feature extraction,the accuracy improved considerably to 67%.The results could possibly be improved with a larger dataset acquired through additional sensors.

FUTURE RESEARCH CONSIDERATIONS

In order to realize the full potential in HAR (Human Activity Recognition) systems, some topics need further investigation. A list of those topics is mentioned below:

Activity Recognition Data Set: The quantitative comparison of Human Activity Recognition approaches has been hindered by the fact that each system works with a different dataset. In that direction, various datasets publicly open to the research community can be included which can be used as benchmarks to evaluate new approaches.

Concurrent and overlapping activities: The assumption that an individual only performs one activity at a time is true for basic ambulation activities. In general, human activities are rather overlapping and concurrent. Since only few works have been reported in this area, we foresee great research opportunities in this field.

Crowd Human Activity Recognition: The recognition of human activities has been somehow individualized, i.e., the majority of the systems predict activities in a single user. If we could gather activity patterns from a significant sample of people in certain area (e.g., a city, a state, or a country), that information could be used to estimate levels of sedentarism , exercise habits, and even health conditions in a target population.

REFERENCES

1. https://www.hindawi.com/journals/cin/2016/4351435/
2. http://www.utdallas.edu/~cxc123730/ICIP2015-Chen-Final.pdf
3. http://research.ijcaonline.org/rtfem2016/number1/rtfem45111.pdf
4. https://www.microsoft.com/en-us/research/wp-content/uploads/2016/02/06247813.pdf
5. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3354997/
6. https://www.webopedia.com/TERM/W/wearable_technology.html
7. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4721778/
8. http://ieeexplore.ieee.org/document/6990523/
9. https://link.springer.com/chapter/10.1007/978-3-319-04114-8_40
10. https://link.springer.com/chapter/10.1007/978-3-319-11973-1_3