Activity Monitoring in Assisted Living

David Megins-Nicholas

Supervised Dr Charith Abhayaratne

Interim Report

Contents

Abstract

This project aims to design, implement and test an activity monitoring system for use in an assisted living setting. It was decided that the system aims to be non invasive such that the individual being monitored need not concern themselves with wearing additional devices. The system should be able to be set up by health care professionals and left running.

The system aims to monitor a video feed from a room and identify the pose of an individual at regular intervals. A range of pose types is proposed to be as follows: sitting, lying, standing and walking. From this time series data an estimate of the individuals activity levels can be gauged. The system should ideally run using a single web cam connected to a Raspberry Pi which runs the system. This can be connected to the outside world via a mobile internet connection thought his level of implementation is beyond the scope of the project.

The methods chosen to implement this are a Flexible Mixture of Part Model (FMM) to estimate the pose, Principle Component Analysis (PCA to reduce the dimensionality) and K-Means Clustering to classify the pose. The time series data of poses can then be represented visually and aggregated to measure an individuals activity.

1 Literature Review

A literature review was undertaken to achieve the following aims. Firstly to understand the meaning and importance of assisted living. Secondly to investigate the importance of activity monitoring. To understand current methods used to extract activity features from visual data. To investigate machine learning concepts that will be used to reason about the extracted features. Finally what data sets are currently available to test the system on.

1.1 Assisted Living

1.1.1 Definition

Assisted living is the name given to the set of systems provided to individuals who would otherwise find it difficult or impossible to live independently whilst preserving the individuals dignity. Typically individuals who require assisted living systems are elderly and infirm. They may have multiple conditions that impair their ability to perform daily tasks. For example they may have difficulty with mobility and therefore have modifications made to their houses to enable them to move more freely e.g. additional railings for the individual to hold to whilst getting up or walking. Another example of assisted living technologies is assisted feeding.

A patient may be unable to feed themselves due to decreased motor function as a result of Parkinson's Disease, companies such as Google have released self stabilizing cutlery to allow a patient to feed themselves with dignity[1]. Other assisted feeding technologies are available when impairment is even worse, where the patient has access to robotic arms which can feed help them feed[2].

Systems need not be as physical as assisted feeding, personal alarms used by elderly people also form part of the assisted living ecosystem. These alarms should be carried with them at all times so that they may activate them to raise help with outside agencies such as the Ambulance service. Assisted living technologies can also record information about the state of the individual at risk which can help inform medical decisions about how well they are coping with living independently.

1.1.2 Importance of Activity Monitoring

Monitoring the physical activity of the elderly or infirm has long been an active research issue[3]. Early methods used cumbersome hardware that the individual had to wear for extended periods of time. This clearly does not fully meet the criteria for assisted living as it requires additional upkeep by a professional and does not preserve the dignity of the individual. In addition the measured behavior my not accurately represent the individual as they are fully aware of being monitored. Even in 2003 these physical devices were still being used in research[4] though the devices were still cumbersome. At this point the systems could identify pose transitions which reflect activities such as changing between sitting and standing. These pose transitions will form the core of the activity monitoring system that will be implemented.

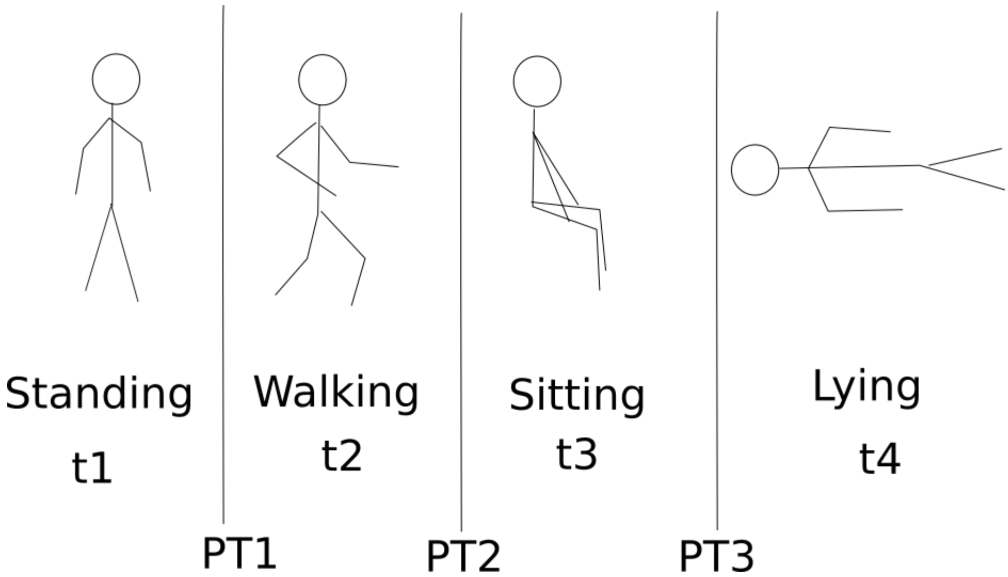

Figure 1: Poses and pose transitions in time series

The figure above shows how a time series data of poses can be generated. By identifying the pose at each time stop it is also possible to identify the pose transitions. These transitions can be weighted to give an overall score of an individuals activity during the day in addition to a breakdown of how long the individual spend standing, sitting, walking or lying down.

Activity monitoring is useful tool medical fields as information gathered on an individuals activity can be used to aid diagnosis and treatment of disease. Studies conducted show that physical activity can effect survival during and after the treatment of certain cancers[5]. If medical professionals can use activity monitoring systems to gain an objective measure of a patients activity they can then assess whether the patient requires more physical activity in their daily routine to improve recovery rates.

In addition activity monitoring can also identify behavioral traits or changes that may indicate the onset of diseases such as Alzheimer's disease. One of the characteristics of this is the phenomenon known as "night wandering". Individuals suffering from the condition experience wakefulness during the night time and as a result "wander" around their room or house[6]. These periods of activities could be discovered in the pose time series data and aid in the diagnosis of such diseases.

1.2 Methods for Monitoring Activity

1.2.1 Wearable Technology

With the boom in wearable technology come wristbands aimed at measuring physical activity. Examples include the Fitbit[7] and the Fuel Band[8]. These products have been used in studies to measure the activity of individuals in an assisted living environment[9]. The study conducted by Tan-hsu et al showed they could measure the activity of an individual with an accuracy reaching almost 99%.

However that study also used RFID tags on the individuals with readers positioned around the room. This improves the measure of activity of the Fitbit but increases installation cost and is more invasive. Individuals may not want the hardware scattered around their house.

Fitbits can also be used to monitor sleep quality, duration and periods of wakefulness during the night. A study by Perez-macias et al showed[10] that devices such as the Fitbit can be successfully used to monitor an individuals sleep. However there are drawbacks to this, devices such as the Fitbit are not sufficiently aware of their context to determine when sleep has started and as such need to be put into a "sleep mode". During the tests the participants had to make certain that the devices were put in this mode. In an assisted living scenario and individual who may lack experience with such devices or not have the the presence of mind to remember to put the device into sleep mode.

Although generally affordable there are also underlying problems with such devices. They are battery powered and as such require charging, individuals who need assistance caring for themselves may not be best positioned to keep these devices charged. The devices need to be removed before washing which means they need to be put back on afterwards, this requires the individual to remember to put the device back on and have the manual dexterity to do so (even individuals in their 20s may sometimes have difficulties Fitbits). The devices also require a smart phone or computer to connect to to upload the data. Smart phone technology has not currently penetrated the elderly market sufficiently for smart phones to be viable. The USB wireless dongle for devices such as the Fitbit do not currently have support for headless Linux computers which would be the obvious low area and cost choice for synchronizing the devices.

After researching hardware methods of tracking activity it was decided that they would not be suitable for this project.

1.2.2 Estimating Activity via Motion Tracking

Activity could be primitively modeled by simple image subtraction. Using a static camera the background of the image can be identified and subtracted from subsequent frames. This will identify regions that have changed which could be assumed to be due to motion of the individual. The center of these regions of difference can be identified in each frame and therefore the motion of the individual can be tracked[11].

This method is not robust. If the camera moves at all then the entire frame is moves. There needs to be careful fine tuning of parameters e.g. thresholding the difference results to identify the important regions of motion, this tweaking means that the system will not be robust to changes in lighting which will definitely happen in an assisted living environment. Changes in lighting will also affect the background therefore the background must be constantly remodeled.

1.2.3 Tracking and Particle Filters

Particle filters use Baysiean statistics and Monte Carlo Simulation methods to track objects in image sequences. It estimates the state of a system by treating a particular state of a system as a particle. These particles fill a state space and their density is related to a probability density function which is the prior probability. The system is then evolved using equations of state why are typically based on Newtonian Mechanics. The system is then sampled stochastically via Monte Carlo methods. This estimates the posterior distribution of states[12].

Chen and Schonfeld incorporated Particle Filters with the SIFT (scale invariant feature transform) algorithm to identify key points of a pose and then track them over the a series of frames[13]. The system can track movement and rotation between multiple frames using only 2D information. This method was not considered further due to its mathematical complexity (particle filters are beyond the scope of EEE6219) and due to the need for initialization which involves starting with a known point on the subject. Knowing a starting point requires the individual to be marked which goes against the aim of being non invasive.

1.2.4 Pose Estimation with Models

Rather than than purely tracking movement in an image sequence pose estimations attempt to estimate the underlying pose of the human body by using knowledge of human anatomy. Models that aim to use prior knowledge of the human anatomy attempt to over come the shortcomings of more holistic methods that purely use the shape of the person. Holistic methods suffer greatly from occlusion and noise and therefore require lots of training data to reach good results. The theory is that that by having an understanding of what parts of the person are displayed the system can deal better with occlusion.

A problem with strict parts models is that if they are rigid the image they project on the image plane changes due to orientation changes. Therefore many orientations of parts need to be represented and tested against to determine which part is a location in the image. The state of the art model for representing the human body is the flexible mixture of parts described by Yang et all in 2011[14] and it aims to overcome the limitation of rigid parts. Instead they use flexible parts. They are flexible because the parts are decomposed into smaller parts that are less variant to change in view and these are connected by “springs” to allow for joints and deformation.

A score is calculated for each parts location at a point in the image, the score is assigned based on the sum of two values. First is a weighted sum of the HOG[1] descriptor at all points and the second is a term which describes how the placement is connected to parent nodes. The second term for example ensures that legs do not get connected to arms.This method is best represented as a tree or graph. Initially the root node is placed in the optimum location, this is the torso. Then the algorithm traverses the graph until an arbitrary depth is reached (this is specified by the total number of parts in the model). Att each stage it attempts to place each type part at each location in the image each pair is scored. The part and position combination with the highest score at each node on the graph is chosen. The method described by Yang et al reached an accuracy of 74.9% over the entire body using the PARSE[16] data set. This method selects better positions for the head and torso than that of the limb extremities. The figure below shows a schematic result from this algorithm.

[1] the HOG descriptor or “Histogram of Orientated Gradients” is a transform that calculated the local gradients in an image patch and quantizes it into a number of directions then chooses the gradient with the highest bin count[15]. This descriptor is available in OpenCV.

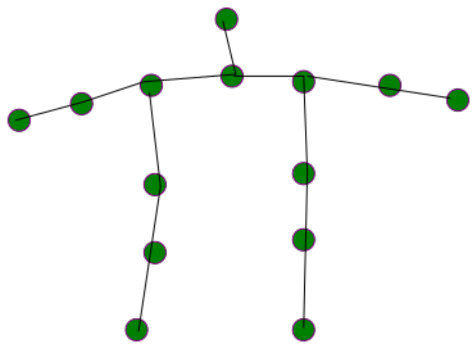

Figure 2: A schematic of the graph that represents the "skeleton" of the pose

Results and implementations based on the flexible mixture of parts are the current state of the art and provide the best placement of the skeleton available. FMM can reach up to 100% accuracy on placing the torso and 99.8% on placing the head when using the Buffy data set[17].

This method was further improved by Cho et al[18] by aggregating different hypothetical poses. Their method does not simply accept that the initial placement of the root node (torso) will provide the optimum results for the entire skeleton. This is because the optimum placement of the root node may not provide the overall optimum pose score. The method chooses the top m root nodes then generates the top poses associated with each root node position and then chooses the optimal pose. This clearly still may not be the global optimum pose but has a better chance of finding it compared the first implementation.

Cho et al's implementation outperforms the Yang and Ramanan's results by reaching 83% accuracy over the whole body on the PARSE data set.

1.3 Machine Learning

The output of the flexible mixture of parts model will be an array of numbers representing the locations of the various joints of the model in image space. Before this is processed the the coordinates will centered so that two identical poses in different regions of the image are classified as the same pose. Once the data has been preprocessed the data will be used in the following machine learning algorithms.

The results of the machine learning algorithms can be tested by using validation methods. Two methods that were investigated were k-folds validation and leave one out validation. K-folds validation[19] hold back a portion of the data set as test data and uses the remaining data to train on, the fit is then tested using the test data. The test is then performed again using a different section of the data. This is repeated until all the data has been held out. Leave one out validation is an extension to the k-folds in that each data point is held out in turn and then tested.

1.3.1 Principle Component Analysis

The skeletal data is clearly highly dimensional as it has the x and y components of many joints. The dimensionality of this should be reduced to investigate how the system can be parameterised. In the case of 2D data PCA attempts to find the orthogonal eigenvalues that describe the semi-major and semi-minor axis of an ellipse that fits the data. The figure below shows a schematic of the principle.

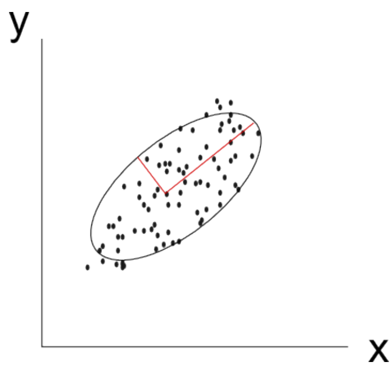

Figure 3: The red vectors are the eigenvalues that can parameterise this data set.

The more dimensions the data has the higher dimension ellipse may be required to fit the data and therefore more eigenvalues are required. The aim is to find parameters that can describe the pose accurately enough to allow clusters of poses to be identified but few enough to make decisions on[19]. The number of principle components used to represent that data will be derived empirically based on the performance of the k-means clustering.

In this project a new implementation of PCA will not be used, the package Scikit-Learn's[20] implementation will be used.

1.3.2 K-Means Clustering

K-means clustering is a machine learning technique that aims to fit n observations into k different clusters[19]. In this case the k clusters are the different poses that need to be identified. The figure below demonstrates the principle.

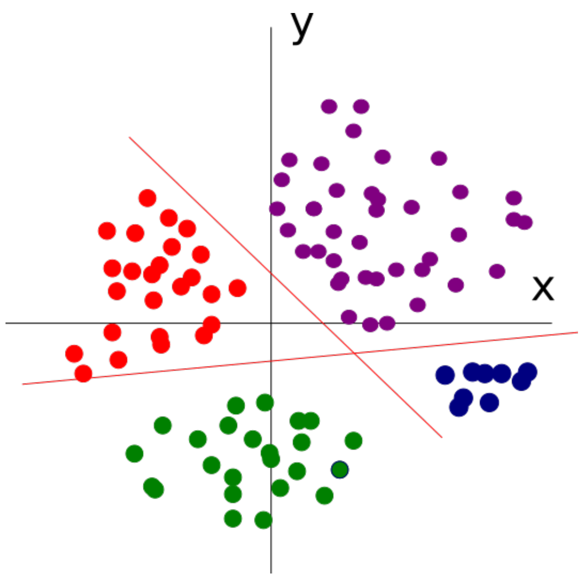

Figure 4: A simplified illustration of K-means clustering

Each of the different coloured groups should represent a different type of pose. The algorithm will be trained using training data whilst holding back test data or new observations. During the training the equations of the lines separating these clusters will be estimated to find a local optima. Clearly this is a simplification as the actual data will be of higher dimensions and the lines will be planes or even higher dimensional objects.

Much like PCA the K-means clustering method in the Scikit-Learn[20] software package will be used rather than implementing and testing a new solution.

1.4 Data Sets

Data sets will be needed to train and test the performance of the system. These data sets will consist of still images or video recordings of typical activities undertaken in an assisted living scenario. In addition data sets that are used in the flexible mixture of parts method will also be used to test the performance of the implementation against the original paper[18] even though the images in these sets is not representative of the activities in an assisted living environment e.g. playing Baseball.

1.4.1 Existing Data Sets

The data sets used compare the performance of the implementation of the FMM model and Cho et al's implementation are the PARSE[16] and LSP[21] data sets. Both of these data sets were created to test the performance of pose estimation using sports persons as the targets. The figure below shows a typical pose from the two data sets. The PARSE data set consists of 305 images each of which has had the persons joints manually labeled with x and y positions. The Leeds Sports Pose (LSP) data set consists of 2000 images of sports persons each of which has had 14 different joints manually annotated. Both sets annotations are stored in a MATLAB array, these will need to be converted into a more usable format.

Figure 5: A baseball player post swing, available via the Wikimedia Foundation released under a Creative Commons License

Other data sets were investigated as the PARSE and LSP data sets will clearly not be satisfactory to identify the poses required by the system.

The Kitchen Data Set from Technische Universität München[22] provides data on 17 different kitchen activities including setting the table and preparing meals. Each activity is recorded from 4 different view points and contains joint annotations. The data is available in both video and still image format.

Figure 6: A still frame from the TUM Kitchen Data Set

Serre Labs at Brown University provide another data set called Breakfast Actions data set[23]. The data set included over 77 hours of video from multiple cameras preparing ten different meals. Prepared by 52 different individuals. The Serre Labs and TUM data sets provide a far more realistic examples of activities performed in the house than the PARSE and LSP data sets.

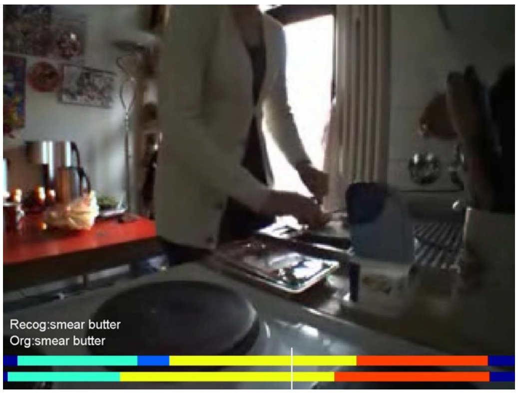

Figure 7: Example frame from the Breakfast Actions data set

An additional data set that was used when testing the FMM in Cho et al's implementation[18] was the Buffy data set[17] this uses still images from the TV series Buffy the Vampire slayer, this database could be used as additional validation of the FMM implementation but would not be useful in testing the overall system in an assisted living setting.

This uses The SUN data set[24] features images of people sitting or lying down but there may not be enough example of these to properly train and test. Therefore a new data set must be created.

1.4.2 Data Set Creation

It is likely that additional data will be required for training and testing of the algorithms. This is due to the lack of data sets with large amounts of people sitting or lying down especially in a household environment. The data set should have the following features:

- have imagery taken from multiple different view angles, this will be use useful to test the robustness of the system with respect to changes in view,
- imagery taken with different levels of illumination in the scene, this will be used to test the robustness with respect to changes in lighting,

- have a variety of individuals be recorded performing the actions, this will reduce the bias towards young white men,
- have participants wear different clothes, this is clearly important as individuals to not wear the same clothes every day.

Once the data is recorded the poses must then be annotated. It will be useful to annotate joints on the participants as well as giving and overall pose classification. The database of images should also feature metadata on the illumination and view, this will be used during testing to identify which areas the system performs well in and which areas it fails.

The main draw back to creating a data set is the requirement to submit a proposal to the University Ethical Committee as live human participants are involved. To mitigate the risk of being held up by the Ethical approval the proposal for the experiment will be made as early as possible such that data acquisition can be started on June 29th in accordance with the plan.

2 Plan

2.1 Overview

After undertaking the literature search a plan was made which describes the functionality that the project aims to fulfill.

The complete system will be able to receive a video signal placed in an assisted living environment where one individual operates. The system will undertake any calibration e.g. background subtraction and camera adjustments. It will then take still frames from the video signal and estimate the skeleton of a person present. Through machine learning algorithms PCA and and K-means clustering the skeleton will then be classified into a type. The types of poses intended to be identified are "standing, walking, sitting and lying down". From the pose time series data pose transitions will be identified, these pose transitions will be used to measure the individuals activity.

2.2 Tools

The system will be prototyped using a combination of the Python programming language, the iPython notebook[25], SciPy[26], Scikit-Learn[20] and OpenCV[27]. These are all open source tools and many of which are essentially Python wrappers for C/C++ code. This will provide the ease of use of Python with the speed of C/C++.

Python was chosen as the language due to the familiarity with it in addition to the large number of packages and tools available to improve performance and functionality.

The iPython notebook is a tool that is becoming widely used in data science fields. It allows the user to create rich documents that contain executable python code. In addition code can be separate blocks which can be run individually or sequentially. This allows the user to import the data in one such block and only run that once. Then new features can be prototyped and refined in subsequent blocks. These are can be run separately without having to read the data in again. iPython is also developed with a client/server model in mind, an iPython server can be set up and the notebooks accessed through the web. This allows the code to be run on more powerful machine from anywhere in the world. This means that the local computer will not need to be left running whilst the system is being tested or the machine learning algorithms are being trained. If more compute power is required the iPython server can be run on scale-able Amazon EC2 instances[28].

SciPy is a package for Python that adds more functionality. Such as Numpy which improves Pythons mathematical capabilities, the backend of Numpy is written in C/C++

which allows for faster vector and matrix mathematics than in standard Python. Matplotlib which is a useful package for presenting data in graphs, it has a very similar syntax to MATLAB that can link into existing Python code. Pandas is a package that adds improved data structures for machine learning purposes. Pandas introduces the “data frame” that allows large amounts of data to be sorted and organised for easy use in machine learning problems[29].

Scikit-Learn[20] is a package that adds a wide array of machine learning toots to Python. In this project the specific machine learning methods that are expected to be used are Principal Component Analysis (PCA) and k-means clustering. Scikit-Learn adds these methods to Python which will mean that new implementations of these methods will not have to be designed and tested.

OpenCV[27] is an open source Computer Vision package. It is primarily used in a C/C++ context but more recently a Python wrapper has been developed. The the Python implementation adds improved methods to deal with image and video inputs as well as a multitude of standard computer vision functions.

As all the tools used in the development of the system are open source then it is reasonable to want to make the system open source as well. At the beginning of the project a discussion will take place as to whether this is feasible with respect to University of Sheffield guidelines and the usage of algorithms derived from other groups work can be used in the project whilst remaining open source.

2.3 Gantt Chart

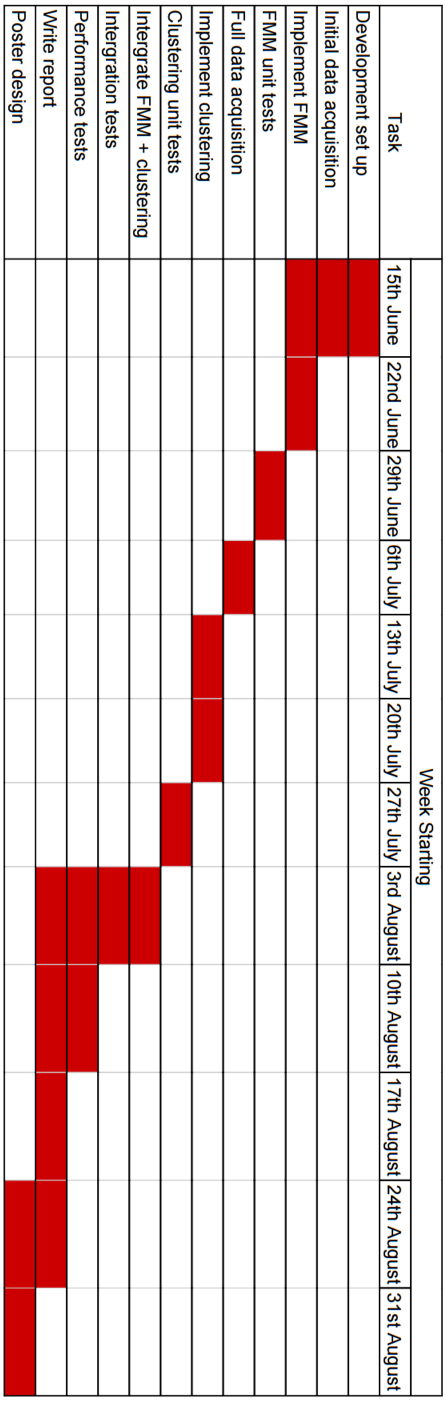

Figure 8: Gantt chart of planned work

The development set up stage of the project will involve setting up the software packages as well as familiarization with their use e.g. OpenCV. This should not take long as many of the packages are already in use. This stage may involve the creation of reproducible virtual machines which can be created and used as a consistent test bench for the software.

The initial data acquisition involves the creating of a limited number of images of a person in a variety of poses. These are the images that will be used during the prototyping of the flexible mixture of parts pose estimation stage of the project. These images will be used to test that the model is correctly estimating the pose. Some images will be held back for unit testing of the FMM.

Implementation of FMM is the stage where the flexible mixture of parts model is implemented such that a the "skeleton" of the poses recorded in the previous stage can be estimated. At this the design should be able to read in single images and output the skeleton.

FMM unit tests will ensure that the FMM system is provides correct results and can handle errors. The images held back during initial data acquisition will now be used to verify that the model can estimate poses. These images are analogous to the test data set where as the images used during development are the training data set.

Once the FMM has been implemented and proved to have worked a larger data set of images and video will be produced and tagged with their pose names and metadata. This data will be processed by the FMM model and used to train and test the clustering algorithm.

Implementing clustering will involve using principle component analysis to reduce the dimensionality of the skeletons and implementing k-means clustering to allow the discrimination of poses. The end product of this will be a system that can receive skeletons as inputs and return a named pose.

Testing the clustering will involve train and testing the clustering using methods such as cross validation e.g. leave one out cross validation or k folds cross validation.

Integrating FMM and clustering will join the two models together such that a video stream can be fed in, samples can be taken and processed and a pose type is returned. The data of these poses can be processed and plotted to form an activity log.

Integration testing will ensure that the final system is functional and can handle errors that may arise e.g. multiple people in the image or no people in the image and disconnection of the image stream. Performance testing will identify the correctness of pose identification and whether or not the system can run in real time.

Report writing will start early to ensure that it is finished on time. The poster design should not take long as it is the representation of the report.

2.4 Milestones

There are four milestones set to determine whether or not the project is progressing as planned. Meetings will be held with the supervisor at these milestones to discuss progress. If progress is less than planned then technical problems will be discussed and any changes to to overall plan will be discussed and put into effect.

The first milestone is at the end of the third week. At this point the flexible mixture of parts model should be fully implemented and tested. At this point the code should be able to take sample images, identify the skeleton and return the data. This is the most challenging part of the project. Being a week behind at this stage is manageable due to built in time later on.

The second milestone is at the end of the seventh week. By this time clustering should be fully implemented and tested. This milestone should be easily achievable due to previous experience with machine learning in COM6509[30]. If this proceeds well it can provide a buffer for milestone one taking longer.

Milestone three is at the end of integration tests at the end of week 8. At this point a video stream should be input with samples taken at time intervals with the identified pose. There should be a method of visualizing this data through graphs or plots. Milestone three allows to project to branch depending on how much time is remaining. If progress to this point is slow then the integration aspect of the project can be dropped and the performance testing can continue using still images rather than a video feed. If necessary the code can be rolled back to the separate models and the data can be passed between them in a more manual fashion. If the progress is on time then the performance testing can continue as planned. If this milestone has been reached before the end of week eight then fall detection can be investigated though this may require audio signals[31].

The final milestone is at the end of week ten. At this point the performance testing should be finished an the final report should be in progress. At this point if fall detection is not finished or nearly finished work on it should be ceased as there will not be enough time to test the correctness and performance of it.

2.5 Risks

2.5.1 Health and Safety

At the time of submission the RACIE form that this section is based on has been submitted and has been approved at the Supervisor level.

Electronics with mains power supply will be used e.g. computers there is a very small risk of electrocution due to damage. The effect of electric shocks include burns and cardiac arrest. To mitigate the risk all equipment will be PAT tested and will be visually inspected for damage before use. This will make the risk negligible.

As the primary activity in the project is programming there is a low risk that repetitive movements and poor posture could cause repetitive strain injury or musculoskeletal injury. To mitigate this proper posture will be maintained whilst sitting at the desk (straight back and forearms parallel to the desk). If symptoms of repetitive strain injury develop then the frequency of breaks can be increased as well as changing mouse and keyboards to more ergonomic models. Changing working between laptops and desktops will also change the movements being made which can reduce the symptoms.

The use of display screens poses a low risk of poor posture causing musculoskeletal pain and sore eyes or headaches. To mitigate poor posture work spaces will have fully adjustable seating and layout to give proper posture. Regular breaks will be taken (5 minutes per hour) to minimise risk of headaches and sore eyes.

There is a low risk of trips and falls due to wiring and cables. When building datasets or testing the system wires and cables may need to be run between webcams and computers. If these trail on the floor there is a low risk that persons may trip and fall over them. To mitigate this risk all cables and wires will be placed away from main thoroughfares and can be taped to the floor if this cannot be avoided.

2.5.2 Completion

There are several risk that may cause the project to be delayed such that it cannot be finished by the deadline. These risks include:

- Data loss
- Equipment failure
- Delays to ethical approval

Data loss provides the largest risk. As the project is entirely software based in theory the entire project could be lost in the event of a disaster e.g. fires, floods, theft or accidental deletion providing that there were insufficient backups.

To avoid data loss daily backups will be made. At the end of the working day all code written will be backed up to several locations. These locations include:

- on device backups, code will be backed up in a separate directory to avoid the deletion of files due to mistyped commands eg "rm -r *". This backup should be performed on a separate internal hard drive
- off device backups, code will be backed up on additional hardware e.g. other computers of external hard drives so that if the primary development machine fails minimal data is lost.
- off site backups, some backups will be stored away from the lab to mitigate the risk of loss due to fires, flood or theft. These backups will be stored at home which is a 30 minute walk from the lab, this should be a large enough separation to ensure that if there is a disaster at the lab then the data is still safe.
- cloud backups, data will also be backed up to cloud services e.g. Dropbox and Copy to provide additional off site redundancy.

Any data sets generated will also need to be backed up. However as these will contain potentially large quantities of video these files could be very large. As such it may be impractical to back them up in as many places as would ideally be necessary e.g. there may not be sufficient space on Dropbox or Copy to store the entire data set.

The version control system "git"[32] will be used so that if changes are made to the code that introduces too many bugs or significantly deteriorates performance any changes made can be rolled back to previous commits. This open source tool also allow branches to be made and checked out so that new features can be developed independently and then merged at a later date.

Equipment failure poses a low risk to the completion of the project. As the data will be heavily backed up development can be continued on any other computer with limited switching cost. The system will also be built with no specialist hardware therefore replacing components e.g. web cams will be low cost.

If the available data sets are not sufficient for testing and training purposes a new data set can be generated by recording activities. If the only participant is myself then there is not requirement for ethical improvement. However basing a data set on only one person introduces biases into the data set. For example the data set will favor the particular way in which the participant performs those activities as well as the physical attributes of the participant. Therefore ideally the data set should be produced using multiple participants. However by using other humans in the experimentation this would require ethical approval from the University Ethics Council. This can be a lengthy process which could delay the production of the data set.

2.6 Metrics and Success

The performance of the system will need to be tested while changing a number of different variables and at different stages.

The flexible mixture of parts system for estimating human poses should perform at least as well as the method described in the paper "Accurate Human Pose Estimation by Aggregating Multiple Pose Hypotheses Using Modified Kernel Density Approximation" where was proposed. There for should reach 83% on the PARSE[16] data set and 63% on the LSP[21] data set. In their paper[18] they outline how well their method performed one each body part.

The system should be robust to at least small changes in viewing angle and to changes in illumination. To test this it will be possible to use the new data set and its additional meta data on view angle and illumination to separate the test data and specifically test it under certain view angles or lighting levels. From this metrics of pose identification accuracy can be produced. It will be possible to say that the system reaches $x\%$ accuracy in poor illumination and at a certain viewing angle but can achieve $y\%$ accuracy over all. From this it will be possible to identify the systems strengths and weaknesses.

Another metric of success will be how fast the system can process a frame. The more estimations of pose that can be produced the more accurate statistics can be drawn for the subjects activity. More accurate metrics are more likely to be useful in diagnosis or patient care.

3 Conclusion

The project aims to to build a system that can estimate the pose of a an individual from a video stream by estimating a skeleton and then using machine learning algorithms to classify these poses into types. These types can then be used as a measure of the individuals activity over time. The system should use a single camera and should be able to perform in near real time (multiple samples per minute). Real time results may be hard to achieve due to the choice to develop in Python, however the system could be ported to a faster language like C or C++ to increase the performance but that is beyond the scope of the project.

The largest issue faced is estimating the skeleton using the flexible mixture of parts. From the literature search this is the state of the art in skeleton estimation using only one camera and no depth feed. However skeleton estimation is a wide spread technique that and therefore should be possible to achieve.

References

[1] C. Norris, “Liftware : Tremor Cancellation Technology Stabilize your Hand Tremor,” 2014.

[2] W. Song and J. Kim, “Novel Assistive Robot for Self-Feeding,” *Robotic Systems - Applications, Control and Programming*, vol. 1, 2009.

[3] R. a. Washburn, K. W. Smith, A. M. Jette, and C. a. Janney, “the Physical Activity (Pase): Development and Evaluation,” *J Clin Epidemiology*, vol. 46, no. 2, pp. 153–162, 1993.

[4] B. Najafi, K. Aminian, A. Paraschiv-Ionescu, F. Loew, C. J. Büla, and P. Robert, “Ambulatory system for human motion analysis using a kinematic sensor: Monitoring of daily physical activity in the elderly,” *IEEE Transactions on Biomedical Engineering*, vol. 50, no. 6, pp. 711–723, 2003.

[5] J. a. Meyerhardt, E. L. Giovannucci, M. D. Holmes, A. T. Chan, J. a. Chan, G. a. Colditz, and C. S. Fuchs, “Physical activity and survival after colorectal cancer diagnosis,” *Journal of Clinical Oncology*, vol. 24, no. 22, pp. 3527–3534, 2006.

[6] S. Ancoli-Israel, M. R. Klauber, J. C. Gillin, S. S. Campbell, and C. R. Hofstetter, “Sleep in non-institutionalized Alzheimer’s disease patients.,” *Aging (Milan, Italy)*, vol. 6, no. 6, pp. 451–458, 1994.

[7] “Fitbit,” *http://www.fitbit.com/*.

[8] “Nike Fuel,” *http://www.nike.com/us/en_ us/c/nikeplus-fuel*.

[9] T.-h. T.-I. S. Member, M. Gochoo, K.-h. Chen, F.-r. Jean, F.-j. Shih, and C. F. Ho, “Indoor Activity Monitoring System for Elderly Using RFID and FitBit Flex Wristband,” pp. 41–44, 2014.

[10] J. M. P.-m. Ieee, S. Member, H. Jimison, E. Member, I. Korhonen, E. S. Member, M. Pavel, and S. I. Fellow, “Comparative assessment of sleep quality estimates using home monitoring technology,” pp. 4979–4982, 2014.

[11] H. Tao, “Object Tracking and Kalman Filtering,” *CMPE 264: Image Analysis and Computer Vision*.

[12] Z. H. E. Chen, “Bayesian Filtering : From Kalman Filters to Particle Filters , and Beyond VI Sequential Monte Carlo Estimation : Particle Filters,” vol. 182, no. 1, pp. 1–69, 2003.

[13] C. Chen and D. Schonfeld, “A particle filtering framework for joint video tracking and pose estimation.,” *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society*, vol. 19, no. 6, pp. 1625–1634, 2010.

[14] Y. Yang and D. Ramanan, “Articulated Pose Estimation with Flexible Mixtures of Parts,” *Computer Vision and Pattern Recognition, 2011 IEEE Conference on*, 2011.

[15] G. Tsai, “Histogram of Oriented Gradients,” *Lecture Series at The University of Michican*, 2010.

[16] D. Ramanan, “Learning to parse images of articulated bodies,” *Advances in Neural Information Processing Systems*, vol. 19, pp. 1129–1136, 2007.

[17] V. Ferrari, M. Mar, and A. Zisserman, “Pose Search : retrieving people using their pose,” 2009.

[18] E. Cho and D. Kim, "Accurate Human Pose Estimation by Aggregating Multiple Pose Hypotheses Using Modi fi ed Kernel Density Approximation," vol. 22, no. 4, pp. 445–449, 2015.

[19] Chapman and Hall, *A First Course in Machine Learning.* 2011.

[20] F. Pedregosa, G. Varoquax, and A. Gramfort, "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[21] S. Johnson and M. Everingham, "Clustered Pose and Nonlinear Appearance Models for Human Pose Estimation," in *Proceedings of the British Machine Vision Conference*, 2010.

[22] M. Tenorth, J. Bandouch, and M. Beetz, "The TUM kitchen data set of everyday manipulation activities for motion tracking and action recognition," *2009 IEEE 12th International Conference on Computer Vision Workshops, ICCV Workshops 2009*, pp. 1089–1096, 2009.

[23] H. Kuehne, A. Arslan, and T. Serre, "The Language of Actions : Recovering the Syntax and Semantics of Goal-Directed Human Activities," *Cvpr*, 2014.

[24] J. Xiao, J. Hays, and K. Ehinger, "SUN Database: Large-scale Scene Recognition from Abbey to Zoo," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2010.

[25] F. Pérez and B. E. Granger, "IPython: A system for interactive scientific computing," *Computing in Science and Engineering*, vol. 9, no. 3, pp. 21–29, 2007.

[26] E. Jones, T. Oliphant, and P. Peterson, "Open source scientific tools for Python."

[27] G. Bradski, "OpenCV Library Software Package," 2000.

[28] G. Juve, E. Deelman, K. Vahi, G. Mehta, B. Berriman, B. P. Berman, and P. Maechling, "{Data Sharing Options for Scientific Workflows on Amazon EC2}Data Sharing Options for Scientific Workflows on Amazon EC2," *22nd IEEE/ACM Conference on Supercomputing (SC10)*, no. November, 2010.

[29] W. McKinney, "pandas: a Foundational Python Library for Data Analysis and Statistics," *Python for High Performance and Scientific Computing*, pp. 1–9, 2011.

[30] N. Lawerence, "COM6509," 2014.

[31] S. R. Ke, H. Thuc, Y. J. Lee, J. N. Hwang, J. H. Yoo, and K. H. Choi, *A Review on Video-Based Human Activity Recognition*, vol. 2. 2013.

[32] L. Torvalds, "git Software Package," 2005.