

Cognitive affect detection through application of computer vision on depth

School of Engineering and Informatics
BSc Computer Science
3rd year Project – Final Report
Oscar Trott
132130

Date: 24/04/2017

Word Count: 14,839

Supervisors: Luc Berthouze, Harry Witchel (BSMS)

Declaration of Authorship

This report is submitted as part requirement for the degree of BSc (Hons) in Computer Science at the University of Sussex. It is the product of my own labour except where indicated in the text. The report may be freely copied and distributed provided the source is acknowledged.

Signed:

Date:

Acknowledgements

I would like to take the time to thank my project supervisors, Dr. Luc Berthouze and Dr. Harry Witchel for their support and guidance through the execution of this research, while also providing me with the opportunity to investigate an interesting application of technologies, combining disciplines and greatly expanding my knowledge. This work makes use of other works in select areas, these are cited and sourced as appropriate. I would also like to thank Tom Ranji who was central to the experiments and data collection.

Contents

Abstract.....	i
1 Introduction	1
2 Professional considerations	3
2.1 Code of Conduct.....	3
2.2 Ethical issues	3
2.2.1 Data Security.....	4
2.2.2 Anonymity in film.....	4
2.2.3 Informed consent	4
2.2.4 Legally responsible adults.....	4
3 Problem Background.....	5
4 Requirements Analysis.....	7
4.1 Identify cognitive affective state and subclass	7
4.2 Use movement as the source of features	7
4.3 Identify optimal feature set for classification task	7
4.4 Programming language	8
4.5 Kinect	8
4.6 Balancing quantities of data	8
4.7 Data can be analysed/classified post event.....	9
4.8 Data capture efficiency in terms of computer power	9
6 Implementation.....	10
6.1 Data storage.....	10
6.2 Footage capture	11
6.3 Footage playback.....	11
6.4 Footage trimming	12
6.5 Noise reduction	12
6.6 Background Estimation	13
6.6.1 Applying Background Estimation to Depth.....	13
6.7 Farnebäck Optical Flow	18
6.7.1 Applying Farnëback Optical Flow to Depth.....	20
6.8 Flow Clustering.....	22
6.8.1 DBSCAN (Depth Based Spatial Clustering of Applications with Noise) ..	22
6.8.2 Alpha-Beta Filter	23
6.8.3 Validation.....	24
6.9 Classification.....	27

6.9.1	k-NN (k-Nearest Neighbours).....	27
6.10	Libraries and Software	28
7	Results	29
7.1	Background Estimation	29
7.2	Clustered Optical Flow.....	30
7.3	Results Review	31
8	Conclusion.....	33
8.1	Project Objectives	33
8.2	Original Objectives	33
8.2.1	Identify cognitive affective state and subclass.....	33
8.2.2	Use movement as the source of features	33
8.2.3	Identify optimal feature set for classification task	33
8.2.4	Programming language	34
8.2.5	Kinect.....	34
8.2.6	Balancing quantities of data.....	34
8.2.7	Data can be analysed/classified post event	34
8.2.8	Data capture efficiency in terms of computer power	34
8.3	Extended Objectives.....	34
8.3.1	Quantify movement	34
8.3.2	Evaluate classical luminance-based methods on depth data	34
8.3.3	Evaluate advantages and disadvantages of various systems performing the same functionality	34
8.4	Critical Appraisal	35
8.4.1	Planning	35
8.4.2	System Utilities Implementation.....	35
8.4.3	System Implementation	35
8.4.4	Result Generation.....	36
8.5	Further Work.....	37
9	References	38
10	Appendix	41
10.1	Project Log	41
10.1.1	Autumn Term.....	41
10.1.2	Winter Break.....	41
10.1.3	Spring Term	41
10.2	Dataset Review.....	42

10.3	Experiment Classes	45
10.4	Data Capture.....	46
10.5	Parameter Evaluation	48
10.5.1	Noise Reduction.....	48
10.5.2	Background Estimation.....	48
10.5.3	Optical Flow.....	50
10.5.4	DBSCAN	50
10.5.5	k-NN and Classification.....	51
10.6	Documents pertinent to participants in experiments	52
10.6.1	Experiment advert.....	52
10.6.2	BFI-10 Personality Measure big five	53
10.6.3	Demographics questionnaire.....	54
10.6.4	Engagement questionnaire	54
10.6.5	Participant Information Sheet	55
10.6.6	Participant Consent form	56
10.7	Research Governance and Ethics Committee Application Form	57
10.8	Research Governance Approval	60

Abstract

We document the process of developing a system which extracts a user's cognitive affective state through applying computer vision techniques to depth data. Identifying a user's current cognitive affective state is useful for programs and responsive systems sensitive to boredom and frustration. The problem is a 6-way classification task identifying the following states: Active engagement, Rapt engagement, Restless boredom, Lethargic boredom, Frustration and Neutral. We explore the development of taking variants of methods typically used on colour/luminosity based footage and apply them to depth footage. The depth footage is captured via a Kinect sensor and has a noise reduction step applied to it prior to analysis due to its unstable nature. We apply a variant of background estimation to the noise fixed depth footage to identify both a general idea of movement within the sequence and when these movements occur. Using background estimation as a pre-processing method, we then apply a dense optical flow method to extract a flow field from the data. This flow field is then clustered to distinguish between different movements. We use alpha-beta filters to predict their locations and thereby track the cluster's centroids at each step. We extract various features from these methods with which we train a k-NN classifier. The classifier is then run many times with each iteration using a random permutation of the dataset. Each iteration the dataset is submitted to 3-fold cross validation. The results show that the system is slightly better than random and exhibits promising correlations. The results and findings gathered by this research are made available to the community with source code on GitHub at <https://github.com/OscarTrott/Cognitive-Affective-state-detection-through-analysis-of-depth-ata-code>.

1 Introduction

Innately humans are excellent at determining their enjoyment of any activity they are currently undertaking. Currently, it is very tough for machines to recognise these feelings externally without the assistance of a human. That said, even humans are not infallible in this regard and can often misinterpret signals, making them believe that there has been a shift in engagement when one has not occurred.

Cognitive affective states alter how a subject reacts to experiences encountered. When comparing boredom and frustration, we can make use of Ryan and colleagues' work which found that boredom was the state which was most strongly associated with "poorer learning and problem behaviors" (Ryan et al., 2010). We can decompose these engagement and boredom states such that they have internal definitions which can be applied to different contexts. This is necessary because of the confusion which can arise due to the apparent crossover of these states. Therefore, each of the super-states has two sub-states, a restless one, and a restful one, e.g., restless boredom and lethargic boredom or restless engagement and rapt engagement. These states are demonstrated through the way that you can be cognitively engaged while motionless watching a film or you can be engaged whilst undertaking strenuous activities such as playing a game of football. Witchel and colleagues (2016) work indicate that one way to differentiate between the sub-states can be Non-Instrumental Movement Inhibition, NIMI for short. Non-Instrumental movements are movements undertaken by the subject which are not pertinent to the task at hand, for example, scratching an itch would be considered a non-instrumental movement, but comparatively moving the mouse using your arm would be an Instrumental movement as it is required by the media. This concept of NIMI is such that when a subject is cognitively engaged with an activity their non-instrumental movements are inhibited, and therefore the frequency and consistency are reduced.

Aviezer and Todorov indicated that although humans in self-reflection will identify the face as conveying the largest quantity of emotional information, it is the posture of the body which we humans use if the two are in conflict (Aviezer, H., & Todorov, A. 2015). This makes posture a valuable source of features for identifying cognitive affective states.

This project is motivated by the vision of Harry Witchel of the Brighton and Sussex Medical School of a computer system capable of automatically performing the kind of affective state analysis he put forward in his research. There are two significant reasons behind examining the relationship between nonverbal behaviour and a cognitive affective state. Firstly, the history of nonverbal communication is extensive and stretches back throughout human history; this background information is given in Section 3. Secondly, research in engineering and behavioural patterns is seeking to recognise cognitive affective states through analysis of a selection of nonverbal features expressed by humans in response to these states.

The work can be used as a foundation for making more effective HCI systems such as cognitively responsive learning tutors (Kapoor et al., 2007; D'Mello et al., 2007) or companion robots designed to both monitor cognitively-challenged patients (Gratch et al., 2014) and provide services to clients (Huth and de Ruiter, 2012). Alternatively, the system could be used in many areas where cognitively engaged programs would enhance the overall user experience, an idea fundamental to the principle of gamification and serious games.

Gamification or serious games are a genre of programs based on the concept that using ideas from game design in non-game contexts can increase productivity and efficiency. These programs have broad applicability from corporations using gamification software to increase productivity (Kumar, 2013) to enhancing e-learning environments to raise engagement in students (Muntean, 2011). Gamification is an ever-evolving paradigm that has been a buzzword in recent times, as it progresses it requires ever changing technology to assist in its use. One significant issue with gamification applications which are currently in use is that if the program is designed badly and is poorly tested, it may be implemented without proven ground truths. It would be useful to these systems for them to have feedback from users on the effectiveness of the gamification system allowing it to automatically adapt or pass the feedback on to the development team. This is where research into the automatic identification of cognitive affective states comes in. With its applications as described above, it is a useful medium through which feedback can be gathered from users allowing either the program or developers to change the functionality of the system and adapt it to be more intuitive and engaging to use.

2 Professional considerations

2.1 Code of Conduct

Ethics in Computer Science are set out in the Code of Conduct provided by BCS – The Chartered Institute for IT found at <http://www.bcs.org/category/6030>. These ethics ensure that research is undertaken in a professional manner such that the participants have full control over their participation and that no harm is allowed to come to them. The code of conduct defines the guide which BCS members should uphold at all times, over the course of this project the code is followed and maintained at all times. Sections of the Code of Conduct are particularly pertinent to different parts of the project. Section one describes how the members should act in the public's interest, sections 1.a and 1.b are of interest when carrying out experiments with human volunteers who must be anonymised, informed of their rights and allowed to opt out of the voluntary role at any point. Sections 1.c and 1.d are followed by not discriminating against any volunteers and advertising in multiple locations on the university campus.

Section 2.a, b and c are followed in the systems implementation by ensuring that the author is professionally competent to do so through background reading and thorough research on the problem area and possible solutions. 2.d, Legislation is followed through academic procedures such as submitting the ethical review before performing the experiments to ensure that the correct methodology is followed and that the work is ethical. 2.e, we will seek out reviews and criticism of our work such that any problems found can be addressed suitably. 2.f, all work and experimentation is carried out with utmost care to avoid damage to persons, their belongings or social and professional status. 2.g, the findings will be purely recorded based on the results extracted from the research conducted in this project and shall not be influenced by any person or inducement.

Section 3 is followed by accepting the University and the supervisors of the project as the recognised relevant authorities and acting in their wishes and ensuring that any issues or conflicts which arise are discussed and resolved before any subsection of part 3 of the Code of Conduct is broken.

Section 4 is followed throughout the projects span and following its deployment through taking responsibility for the work, informing BCS of any criminal offences or occurrences which a colleague or we commit and provide constructive criticisms and encouragement to others and their work.

2.2 Ethical issues

In undertaking this project, an ethical review was required to be submitted so data capture could be done using humans as participants. This review was submitted by Harry Witchel of Brighton & Sussex Medical School as an addendum to Tom Ranji's ethical review for his tests done to examine physical body reactions when exposed to predefined stimuli determined to evoke specific emotions in the user. The research approval was given by the Brighton and Sussex Medical School Research Governance and Ethics Committee (RGEC) under the full study titled "AV Engagement: Addition of Inertial Sensors" and the RGEC reference number "16/046/WIT", full approval document in Appendix 10, section 8. The data collection is handled by Harry Witchel and Tom Ranji during his experiments, the research was approved by the BSMS Research Governance and Ethics Committee, and the documentation involved in the ethical review is given in full in Appendix 10, section 6.

2.2.1 Data Security

The data stored shall not include any personal details other than the physical appearance of the user, but even then, it is only in a depth format¹. This format has been chosen, so the data is not affected by identifiable characteristics. Researchers should ensure data security by ensuring they, “record, process and store confidential information in a fashion designed to avoid inadvertent disclosure.” A range of security measures are in place and mentioned in the approved ethics proposal to RGEC given in Appendix 10, section 7.

2.2.2 Anonymity in film

Anonymity is maintained by providing each participant with a volunteer code such that any recordings made only use this code. The only location of personal details such as name and address are stored in separately stored consent files. Filming is mentioned in the PID and consent forms submitted and returned. In cases where the films are used to present representative examples to other researchers any identifying features, e.g., faces, tattoos, are blurred out via pixelation provided by mosaic tiling.

2.2.3 Informed consent

The researchers sought informed consent from all participants. Prospective participants were given the opportunity to read both the Participant Information Sheet and the Informed Consent form on arrival. The researchers also ensured “from the first contact that participants were aware of their right to withdraw at any time from the receipt of professional services or research participation.” This information is included in the Participant Information Sheet and the Informed Consent form Appendix 10, sections 6.5 and 6.6. Filming is included in the PIS and consent forms and participants are shown the cameras to be used. The consent forms used are given in Appendix 10, section 6.6.

2.2.4 Legally responsible adults

All participants used in the study were legally competent. We did not accept vulnerable groups or children from taking part in these experiments, and the experiment was only advertised within the University of Sussex community. The advert is given in Appendix 10, section 6.1.

¹ Data in depth format is a video file containing a sequence of images where each pixel holds the depth in millimetres to a point rather than the points colour information. There is no colour or luminance information stored.

3 Problem Background

Humanity has become fascinated by the nature of bodily indicators of emotion and gesture with an increase since the middle of the twentieth century, but interest goes back far further with some of the earliest recorded work being the teachings of Quintilian in 50AD (Katula, 2003). This fascination changed from intrigue into areas of research as knowledge of the body's capacity became more apparent, shown in Darwin's work on how emotions are expressed facially in humans and animals (Darwin, 1872). In the past century, psychologists have researched this further by identifying certain facial expressions which are unique to a given emotion. It became apparent that it was necessary to determine whether these facial expressions are unique to different cultures or whether there is some evolutionary bias to develop these combinations of muscle contractions in response to emotional affects. In the 1960's Paul Ekman researched facial expressions in Papua New Guinea to determine whether there was a cultural bias in facial expression. Ekman and Friesen did this by meeting members of preliterate tribes and collecting data from their reactions to stories designed to evoke emotions (Ekman and Friesen, 1971). They determined that facial expressions are ubiquitous in both definition and the type of emotion being expressed. As a result, it became apparent that it was possible to determine human emotion through external monitoring means.

As shown in Scherer's work (2005) it is difficult to have a concrete list of basic emotions so for this write-up we will say there are six basic facially expressed emotions with one other in contention. Happiness, Sadness, Anger, Sorrow, Surprise, Fear and, the contending emotion, Contempt (Ekman and Friesen, 1971). But for environments such as the classroom, it became apparent to D'mello and Graesser that another measure of engagement was required as mental states such as boredom, frustration and engagement were not accounted for in a high fidelity state using the seven base Ekman-emotions (D'Mello and Graesser, 2007).

When participating in an activity, it is highly likely that a subject will be in one of the above classroom cognitive affective states. Which state they are in depends on how the activity is progressing, their reaction to it and previous experiences with the events which occur. These affective states (confusion, frustration, boredom, engagement, etc.) can be used as a basis for identifying distinct states of cognition. Cognitive affective states can change how a subject reacts when exposed to an experience which does not adhere to the mere-exposure effect. The mere-exposure effect as described by Zajonc (1968) says that repeated exposure to a given stimulus increases the positivity of responses from a subject in each subsequent encounter, therefore changing the type of response given by them. For example, at Christmas dinners, the general reaction to a cracker joke is a resounding groan. These jokes provide an experience familiar but with repeatable results given the same preconditions. Cognitive affective states, however, can alter the subject's response when presented with the repeated experience, for instance, when confused it is unlikely that they will respond actively in a light-hearted manner.

Bianchi-Berthouze and Kleinsmith (2013) showed that current affective bodily recognition systems extract emotion information from dance sequences which can be effective but have difficulty being used in real world applications. There is still research which has been carried out in situations other than dance such as the work done by Mota and Picard (2003) recognising posture and associated affective states in children's interest levels during learning tasks on computers. The work carried out utilised two matrices of pressure sensors placed on the seat, from this the posture of the child sitting could be ascertained to a relative degree, the posture data was used in an HMM, a Hidden Markov Model, to procure an automatic classifier.

The cognitive affect state is judged either by another human trained for the role or by completing a test which is designed to reveal their cognitive affective state (D'Mello and Graesser, 2011). The external judgement of cognitive affective states has been done using a variety of methods. For example, using multiple judges to evaluate cognitive affective states at set intervals on a recording based on identifiable emotions displayed by the subject (D'Mello and Graesser, 2011) or using Non-Instrumental Movement Inhibition as a measure for identifying specific types of cognitive affective states as described below.

Witchel and colleagues (2016) described how observations of NIMI could be used to identify between both active and inactive types of both boredom and engagement cognitive affective states. Whereas a subject who is actively engaged their overall movements are large but their non-instrumental movements are small, a subject who is experiencing rapt engagement will have low overall movements and non-instrumental movements. Comparatively, for restless boredom, non-instrumental movements are high, and the overall movement allows for differentiation between active and lethargic boredom subclasses. By measuring how much NIMI a participant is experiencing, we can gauge whether they are bored or engaged, by conjoining the overall movements into this we get a solvable two-way classification problem.

4 Requirements Analysis

In early discussions with Harry Witchel it was agreed that although this research has few stringent requirements regarding input and output at any low level, there were still terms to be met in the development and implementation. These requirements were to allow for work to be understood and made use of in an interdisciplinary fashion without large quantities of confusion or requiring a background in computer science. Regardless of the investigative nature of the project, there were still general measures of success that can be used to identify the successfulness of the project upon completion.

1. Successfully able to record and playback files such that the data contained can be analysed and reviewed
2. Extraction and analysis of features from data files
3. Mapping recordings to a cognitive affective state using extracted features at an accuracy approaching 85-90%

4.1 Identify cognitive affective state and subclass

The system should identify both the cognitive affective state and the associated subclasses (active-engagement, rapt-engagement, restless-boredom, lethargic-boredom, frustration and neutral) of a participant from a given depth based footage with the correct setup. This classification should have an associated confidence value which describes the overall faith that the system has in its classification. After discussion with Harry Witchel, an average accuracy value which the system should aim for was decided on as 85-90% though this was unrealistically high in practice.

4.2 Use movement as the source of features

The features used in the machine learning methods were to be extracted from participant's posture and movement utilising depth data. Features could be the locations of body parts in the image, their speed (not velocity as the information should be independent of direction), the range of their movements over a period and some form of NIMI with an associated value.

4.3 Identify optimal feature set for classification task

Following the feature selection and extraction, an analysis would be performed identifying an optimal feature set which provides the highest average accuracy using each trialled machine learning method. These results were to be detailed and discussed in the results section with their associated accuracy, training and testing times based on the data sets used.

Concerning a list of initial features from which we can expand upon we can determine that a few elements will be extremely pertinent when encountering cognitive affective states and their effects on a person's body. As shown by multiple researchers there are various features which have been identified as conveying information relevant to judging affect such as:

- Angle of lean, orientation of the subject's upper body compared with the computer monitor (Sanghvi et al., 2011)
- Slouch factor, the curvature of the subject's back (Sanghvi et al., 2011)
- Quantity of motion, value of motion detected within a given scene, generally based on the changing silhouette of a person in the footage (Castellano et al., 2007)
- Contraction index, A measure of the angle between the upper and lower back (Castellano et al., 2007)

- NIMI (Non-Instrumental Movement inhibition), the amount of inhibited movement which is not instrumental to participating with the task at hand e.g. scratching (Witchel et al., 2016)

4.4 Programming language

Due to the research being conducted in an interdisciplinary environment with non-experts in computer science, the code written was in a language easy to learn so it can be understood by academics from various disciplines with adequate commenting and little research into the workings of the language of choice. As such a choice of languages were put forward: Python or Matlab. The language used should be enabled such that it can interface with the Kinect and run machine learning algorithms without becoming a bottleneck for research.

4.5 Kinect

The Sensor used was a Microsoft Kinect which was connected to a laptop machine which is kindly provided by the Sussex University IT staff. The files created when recording with Kinect studio were unmanageable regarding size, shown in particular when recording multiple streams of data (RGB, Infrared, depth) at 30Hz which caused several GBs of data to be produced in a minute. Thus, a lightweight recording script for both recording and reviewing data needs to be created. The script must record data at a reasonable size and allow for the Frame Rate of the recorded file to be reduced/increased as deemed necessary.

The development will make use of past works in Computer Vision whether it be through the use of Pattern recognition systems or another form of identification. Regardless of the computer vision technique used there must be a facet of temporal data included as the model states that movement is important concerning identifying cognitive affective states. The method used must apply to depth based images wherein edges of objects are often easier to locate and identify but also where colour is not made use of meaning that if a digital clock were included in the recording the time displayed would not be visible.

4.6 Balancing quantities of data

As the project relies on recognising the participant's postural expressions and movement through depth based footage, this footage must be of a high enough fidelity regarding frame rate such that the greatest predictors of the cognitive affective states are not missed. The most likely location for this problem to arise is during identification of NIMI as many movements, such as itching, can consist of very rapid movements which, if the temporal fidelity is not high enough, may be lost in between frames. This requires that the footage must be able to record at a reasonably high rate without generating excessive amounts of data, while also showing miniscule movements of almost no relevance. Using the basic Kinect Studio system this is impossible as both the frame rate is non-adjustable and the files generated are completely unwieldy.

The Kinect Studio software outputs recordings of one minute with a file size of 1.5GB meaning that for a full 90 minutes' data capturing session the file, pre-compressed, is roughly 135GB which is un-storable at best. This is the reason for the lightweight script being produced. The result of implementing these scripts is a one-minute clip running at ten frames per second is reduced to 354MB, and a 90-minute recording is reduced to 31.8GB. When compressed, this volume is cut on average by a factor of four times, meaning that for a full 90-minute recording the post-compression size is now only around 8GB which is a highly significant size reduction.

4.7 Data can be analysed/classified post event

The system does not require on the fly classification of data as our analysis will take place after the experiments, similarly to Harry Witchel's experiments. As a result, the projects does not focus on online efficient processing likely branching into a multi-threaded system but rather an efficient implementation designed to run with some given input file/s and outputting labelled data files. This reduced heavy-weight real-time processing allows for higher data capture rates with less computing power.

4.8 Data capture efficiency in terms of computer power

Due to the multi-purpose nature of the location in which the experiments took place (including teaching and experiments), the recording system required setting up and taking down at the start and end of the experiments respectively. Therefore, the system was to be lightweight and capable of being run on a laptop which is limited regarding computing power while allowing for relatively high frame rates at recording, >10fps. If the system did not work within these constraints, there would have been interference with the data capture.

6 Implementation

This section details the methods used and things of note relevant to their implementations.

The system works in four discretised steps where each sub-program has a well-defined input and output and provides a concise, definable step in the program's execution. The components are split into four main groups: Storage Management, Pre-processing, Analysis and Classification.

These methods and implementations have been made to work with depth-based footage when they have been developed for use with colour/intensity data. Thus, we have adapted the methods to allow them to work as intended. Over the course of this work, we have abstracted the depth data for processing. This abstraction means that we treat each depth frame as a colour frame which happens to contain luminosity values in the range 0 – 8000. While this is obviously false, it allows us to apply the methods as normal with some alterations which we describe below.

We review the dataset on which we apply these methods in Appendix 10, section 2. Additionally, the method through which the data was gathered is documented in Appendix 10, section 4.

6.1 Data storage

We use the data storage model provided by the HDF5 group (<https://www.hdfgroup.org/>) to store and manage the two files required for processing².

One file stores the data being processed and the results of processing, the other file stores the results of analysis applied to all instances of the first file. Storing each of the stages between processing steps in the first file allows playback of data before and after processing has been applied to review effectiveness and quality. The data storage file contains two groups splitting the contents into two sets; one contains the analysis and one contains the data. Within the analysis group, there are two datasets, movement estimate, and optical flow; these contain the results of the analysis performed at background estimation and optical flow respectively. The data group contains a dataset for each step of the process. No set is required for optical flow as nothing performs processing on the graphical output from optical flow although we do save an mp4 video of the result.

The second file contains a dataset for each file analysed. The dataset's id is the name of the analysed file and contains the processed analysis of that file and an attribute determining that file's class. This file is the only required input for the classification function preventing cross-contamination of bugs from a single thread of execution.

² HDF5 files have a hierarchical format where the root is the base directory, groups act as sub-directories which can be created at will, datasets (similar to Numpy arrays) which are stored either in a nested set of groups or the directory, and attributes that can be attached to any of the elements as meta-data.

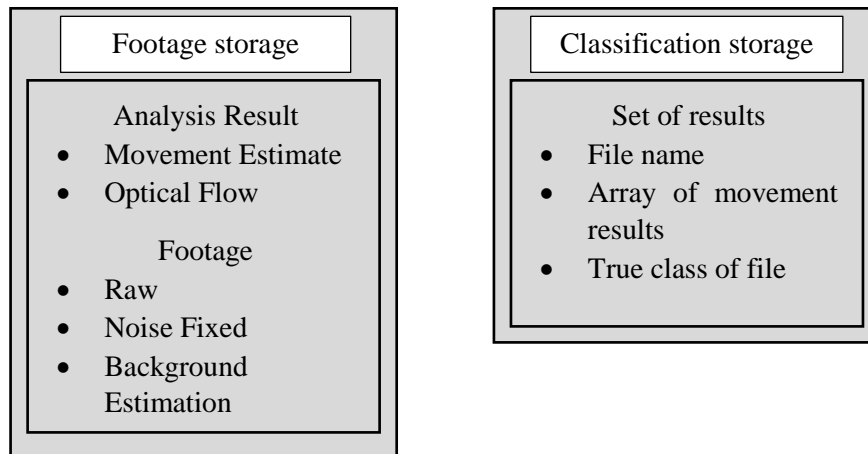


Figure 1, File storage diagram

6.2 Footage capture

The PyKinectv2 library provides functionality which our Python program utilises to access the Kinect data streams. The program continuously accepts frames as numpy arrays and records them into the corresponding HDF5 footage file's raw dataset until a cancellation event occurs. A cancellation event is either a null value returned by the PyKinectv2 interface or the user pressing ctrl+c which is caught by an exception handler. Appended with this data is meta-information such as the time the frame was recorded.

The HDF5 file is written using a file handler class. The file handler has some methods which allow the program to extract and write a single frame to and from a chosen dataset. A dataset is updated by incrementing a counter within the file handler each time a new frame is received indicating the location of the frame. The dataset is reshaped to accept the new data, and the program stores the data.

The method of data capture is detailed in Appendix 10 section 4.

6.3 Footage playback

Footage playback iterates over the dataset located within the selected HDF5 file and displays the data sequentially to the user, pausing execution as necessary to let the footage playback at the correct fps. Footage playback has standard media playback controls: Play, Pause, Fast Forward, Rewind, and Stop.

We display data through the Matplotlib PyPlot library. The Colour map used in our implementation is "jet". The colour maps available are defined at http://matplotlib.org/examples/color/colormaps_reference.html. When setting the colour map limits, there are several options available; we can set them manually for each instance, always use the minimum/maximum values within the data or utilise some measurement to display as much of the relevant data without outliers influencing the data presented. The way we have chosen is to show data between three standard deviations above and below its mean, subject to constraints. This method should, and in practice generally does, give a good range between which information is displayed. The constraints previously stated are that the values obtained by the summation of the standard deviations of the mean do not rise above the maximum value or below the minimum value observed in the data.

6.4 Footage trimming

At the start of each experiment, there is a minute of white noise displayed to the user. Harry Witchel indicates that once the stimulus has started, there is a short duration in which the user's cognitive affective state is in flux. We trim the footage as at the end of the recording there is a short period where the experiment has finished, but recording has not yet stopped.

For standardisation, the system always trims the footage to the same duration and a set period from the end of the footage. The experiments last 3 minutes, including one minute of white-noise, we trim the data such that it ends 5 seconds, or 100 frames before the recording was stopped. The total duration is 100 seconds or 2000 frames. We then save this back into the raw dataset of the processed HDF5 file.

6.5 Noise reduction

This method reduces edge-case noise induced by limitations in the Kinect system and crops the footage to the area containing the subject³. Cropping is useful for the system to reduce both unnecessary processing and data which does not include information pertinent to the task but which may induce noise within the data.

The method implemented here relies on most points having a value similar to those around it. This assumption can be justified by the fact that objects within the scenes viewed by the sensor are continuous regarding depth values. The method extracts a short sequence of frames at the start of the footage and iterates over each pixel. If a pixel has a non-zero value, then it is assumed reliable regarding depth. The point currently being processed is acted upon if it is zero in value. A search window is formed within the space and is searched to find a pixel containing a depth value, we then assign this value to the point. This method can be extended to account for pixel reliability concerning variance over time, however, issues are found when there is movement within the frame sequence being analysed. Cropping is performed using efficient NumPy array slicing.

This implementation works well in most cases. But, as the footage becomes more unreliable and the areas of zero values grow the larger the search window must become to provide the same level of coverage. This issue reduces the reliability of the method substantially and can create more artefacts than it removes in vast areas of uncertainty.

This method has a parameter which was subject to systematic investigation, the reasoning is given in Appendix 10, section 5, method 1.

³ The cropped area is determined by the user.

6.6 Background Estimation

Background estimation is a method which detects motion in footage by distinguishing anomalous data from the learnt background. Background estimation works by building a statistical model of a sequence of images or piece of footage over its duration. This model finds the mean and standard deviation of each pixel over its duration; these statistics can be for luminosity or RGB components depending on the implementation. Provided that the background is stationary, over a long enough period the average finds a resting point imitating the background. We can then extract information about any changes which occur in the footage using a form of background subtraction.

6.6.1 Applying Background Estimation to Depth

The standard mean calculation is used with a change to account for significant variations in the depth values while still representing the data accurately. We change the method of determining the standard deviation to account for noise estimation. Noise in depth data still exists and is similar to noise in RGB data although, in most cases, the size of relative changes can be many magnitudes larger in luminosity as the values can vary between the entire data range. Though in depth footage this can also occur in areas where the depth value varies between zero and their true depth value. Generally, the variations due to planar noise in depth will be relatively small. When viewing a static scene over a long enough period this would most likely average itself out. However, in our case we are dealing with a situation where the average length of our footage will be only a couple of minutes long and will contain a human participant who almost certainly will not rest in the same position for the entire duration. An example can be thought of as a participant who is leaning forward for the first half of the footage and leans back for the second half. In the pixel locations where this change occurs, the mean will rest at a value between the depth of the user and the background behind them. The standard deviation would be large to account for the change in depth. This would give us a useless statistical representation of the scene. So, as time proceeds and frames are processed, the statistical representation needs to tend towards the current value scene. We could implement this via a moving window, but this reflects the original issue just with a shorter interval. Thus, we have opted to perform a weighted, rolling average for both the mean and the standard deviation, where the standard deviation has an extra regularisation parameter which penalises it being large.

Our background estimation proceeds in two parts; the first generates initial estimates for the mean and variance at each pixel. We perform this over the first n frames within the footage. We produce the estimates using standard metrics for rolling mean and standard deviation (described below). We then continue updating our statistics with modified measures over the entire footage. We define any pixels outside one standard deviation of the mean as anomalous, indicating movement. The aggregate of abnormal pixels represents the movement in the scene. A threshold is used to determine where significant movements start and noise ends. When an object moves orthogonally to the depth sensor's view, the method often does not detect movement in areas other than those where significant change has occurred. We propose that this is due to the online variance at those locations learning the micro-movements performed by the subject and inaccuracies when dealing with non-planar surfaces (Yang, L. et al., 2015). It might appear that the best way to adapt when using depth is by calculating the mean as the last frame, reducing non-necessary movement. However, this creates two problems: first, this completely removes the desensitising effect of averaging out noise over time; second, when a foreground object occludes a background noise emitter any depth fluctuations which would usually indicate movement would be disregarded due to the aggregate of the mean and standard deviation.

We use the frame sequence output of background estimation as input for our optical flow method. The optimal output of the background estimation would be the background, without the participant, with the last stable component of the participant on top including any anomalous pixels. This would reduce the effect of the means at object edges assuming the mid-point between the two peaks of the bimodal distribution representing the depth values encountered where one peak is the static background, and the other is the participant/object. However, due to the short duration of the footage, differences between setups in experiment sets and the fact that the participants will always be occluding a section of the background we cannot use this.

This method is not background subtraction as the output is not the frame with the background subtracted. The output is the input frame's value where anomalous data is detected, and the remaining values are the current average. This provides a continuous footage stream upon which we can apply optical flow.

This method has some parameters which were subject to systematic investigation, the method used the values given in Appendix 10, section 5, method 2.

6.6.1.1 Online Mean

To calculate the mean of a list of values of length n when we only have access to x elements at time t we perform the following method. First, we calculate the mean of the initial list, in this case, $x = 1$ so the mean is the elements value. Then for each subsequent value the new mean is equal to:

$$\mu_{n+1} = \mu_n + \frac{x - \mu_n}{n + 1}$$

where μ is the mean, x is the new value and n is the current frame number.

The online mean implemented here has two forms. During the initial estimate of the background, it is calculated using the function above. During the remaining execution of the program, the online mean is the weighted version where each value is weighted as though it were the j th entry in the rolling mean. Where j is a constant value which can be thought of as the learning rate. Thus, the update rule used for all remaining mean updates is:

$$\mu_{n+1} = \mu_n + \frac{x - \mu_n}{j}$$

where μ_0 is the initial estimate generated. This has the effect that the learning rate determines how long a pixel is defined as an anomaly for. In practice this equation produces the following graph.

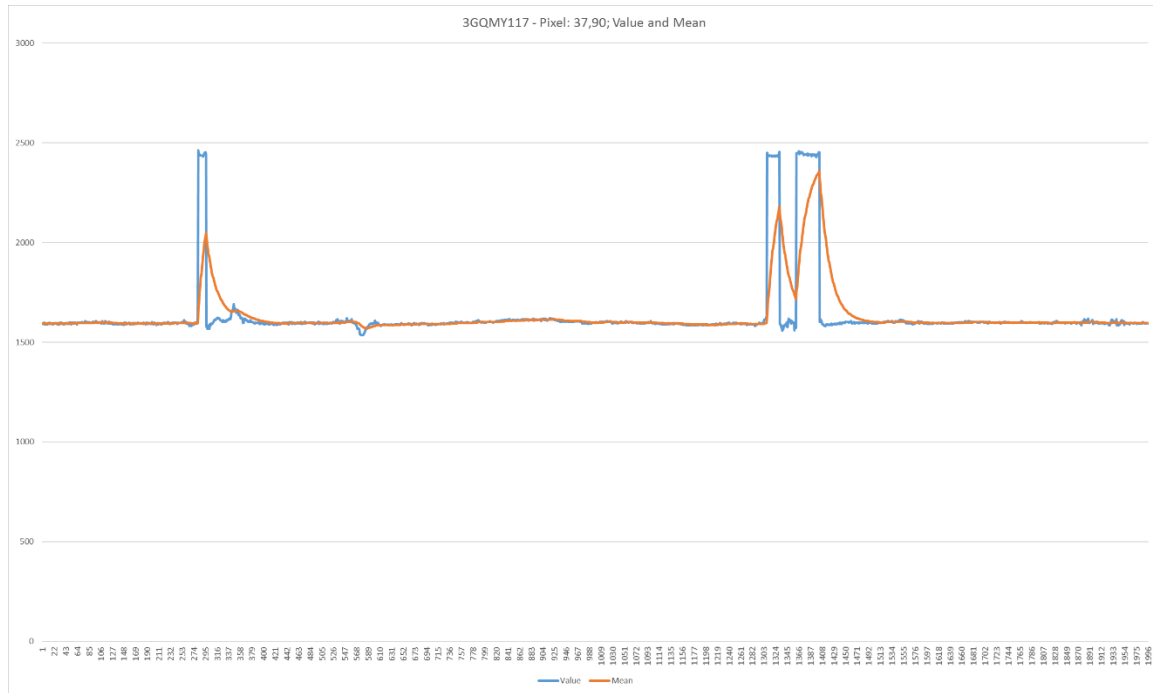


Figure 2, Pixel depth and mean estimate over time

6.6.1.2 Online Standard Deviation

Rolling standard deviation would be hard to do directly compared to variance due to the rooting. Our rolling variance is implemented using a modified version of Welford's algorithm (Welford, B. 1962). Rather than a set window size to estimate the mean and standard deviation over, we opt to implement a tuneable online weighted variance estimate. This change requires removing the denominator from the equation and inserting a learning rate as well as a regularisation parameter, allowing the system to react to new data and retain information from past data. The function implemented are:

$$var_{n+1} = var_n + a \cdot \tanh((x - \mu_n) \cdot (x - \mu_{n+1})/b) - c$$

$$\sigma_n = \sqrt{var_n}$$

where a is the learning rate parameter determining how fast new data is learned, b is a constant scaling parameter to allow the tanh function to perform effectively and c is a regularisation parameter set from past variances. Parameter b both reduces the variance into the activation range of the tanh function and acts as a learning rate parameter as it forces the function always to act as though the value calculated was the b^{th} element in the series. Parameter c is put forward to prevent the variance from continually increasing each time a new value is encountered. Removing c would eliminate the system's ability to adapt when sources of background noise are occluded by an object with reliable depth values. If the variance did not reduce then contours and edges within the new object would be ignored as noise when, if the system started with the object occluding the noise source, they would be identified as movement.

In areas of high noise where values change rapidly and often throughout the footage sequence, the standard deviation should remain at a relatively high level where the constant influence of unexpected values outweighs the regularising c value. Still, the standard deviation should always increase more rapidly than it decreases. We include a regularisation parameter c as the following function which produces the graph below.

$$c_{n+1} = d \cdot (var_n - (x - \mu_n) * (x - \mu_{n+1}))$$

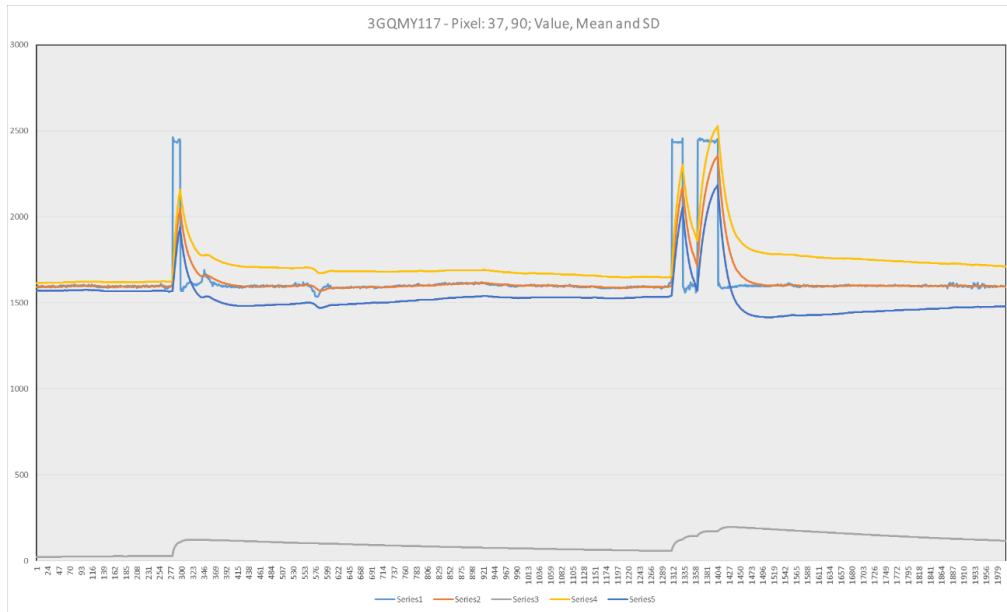


Figure 3, Pixel depth, mean, standard deviation and mean combined with standard deviation

In this equation d is the regularisation parameter, it determines how fast the variance tends to zero. Set too large the system reaches zero very quickly, too small and it has no effect on the variance, allowing it to grow indefinitely. The regularisation equation takes the instantaneous variance estimate and the cumulative variance estimate into account to justify the regularisation amount. Because attempting to regularise based purely on the previous rolling variance value does not account for the current variance within the footage. This has been found to be a useful modification and allows the system to recover from periods of noise to a stable state.

6.6.1.3 *Density-Based Point Validation*

As a pre-processing step for optical flow we only allow those anomalies which fulfil a density criteria wherein an anomalous value detected in an area free of other anomalous values is ignored.

6.6.1.4 *Validation*

The validation for background estimation is straightforward. We review the various experiments and test sequences generated by the project and compare the raw footage with the background estimate. An output of the background estimation is a total of the anomalous pixels in any given frame. These are listed in the Github repository online indicating movement size effects. They provide us with a measure of total movement within the scene. The resting value for these graphs should be 0 when there is no movement for an extended period. This is often not the case due to micro-movements and subtle changes in deformable objects throughout the scene; the resting value tends to zero. This can be seen in the data within the repository.

6.7 Farnebäck Optical Flow

Optical flow is a well-documented process of extracting apparent motion from the relative motion of the observer and the scene which is used in many applications (Prazdny, K. 1980). The output is a flow field. A flow field is a matrix, in this case 2-dimensional, with the same resolution as the input image where each element stores a tuple of magnitude and velocity. The two-versions covered by us here are those provided by OpenCV. Optical flow assumes that luminance is independent of the point's spatial location within the scene and relies on this fact when finding values for Δu , change in x coordinate, and Δv , change in y coordinate. Optical flow optimises for corner cases⁴ as these are the points which give the most reliable disparity between frames.

The following description of optical flow uses information in: Fleet & Weiss (2005), Farnebäck (2003), and OpenCV: Optical Flow. (n d).

Optical flow makes some assumptions:

- Each pixel is similar to its spatial neighbours in terms of luminance
- Most movement between frames is small
- Regardless of distance moved, the intensity value of a given point does not change

These assumptions let us form the following conceptual model of the problem to identify the distance travelled for each pixel. A given pixel (x, y) at time t will have a luminance value j at time t+1. If no movement has occurred then the point (x, y) will have luminance value j, if movement has happened then the pixel at (x, y) will have luminance value k and the pixel to which the original point has moved (x+ Δx , y+ Δy) will have the old luminance value j. We use Taylor series expansion to approximate the given input data, this is then compared over a window to find the most likely locations for it to have moved to. This vector is marked as the disparity between these points, giving us the Cartesian coordinates from the original pixel's location to its new location.

In formal terms, by taking the prime assumption that pixel intensity does not change over time we can say that the intensity of the same point in the space after time Δt will be the same, thus:

$$l(x, y, t) = l(x + u, y + v, t + dt)$$

Where: $u \equiv \Delta x$, $v \equiv \Delta y$ and $dt \equiv \Delta t$. The right-hand side can be approximated by an infinite Taylor series. The Taylor series output is equivalent to the left-hand side summed with several higher order terms derived from the differential of the added components u and v. Thus:

$$\begin{aligned} l(x, y, t) &= l(x + u, y + v, t + dt) \\ &= l(x, y) + \frac{\delta l}{\delta x} u + \frac{\delta l}{\delta y} v \end{aligned}$$

This lets us rewrite the equation such that we can form it as a problem solvable by simultaneous solutions. The way these solutions are found is determined by the implementation used.

⁴ A corner case is an area within the frame which contains two distinct lines with different orientations which either cross or meet.

We have also investigated an alternate form of optical flow known as Lucas Kanade optical flow. Lucas Kanade follows a slightly different methodology to the Farnebäck method whereby a set of points within the image are initially given which represent elements defined as being robust under transformation. In the OpenCV Lucas Kanade implementation of optical flow the method takes a 3x3 patch around the pixel in question providing a set of nine pixels which should all have the same motion. This variation of optical flow is not used in the final system for a few reasons:

1. It is limiting in the points which can be tracked. Lucas Kanade works best when the points given to track are those selected by the function provided by OpenCV as it finds areas which are robust in terms of tracking. However, the fact that the points are good to track does not mean that they will provide good information.
2. Tracked points do not change in between point updates. The base usage of Lucas Kanade would have the same points over the duration of the entire footage which introduces a whole ream of problems. To combat this, we would update the points tracked at discrete time periods or per some evaluation. The most effective points to be chosen can change over time however, meaning that new tracks can be created and old ones removed which exacerbates point 1.
3. Tracked points can drift regardless of local reliability. This is a big problem when extracting movement from data as this means that the method is detecting movement when there is none. This can be combatted slightly by reducing noise and limiting footage analysed to only areas where movement occurs, but the points generally still drift.

There is an extension of optical flow called Range Flow which has been investigated by Hagen Spies et. al (2002). In their work, the standard Range Flow estimation uses a bi-modal system which utilises the entire HSBD (Hue, Saturation, Brightness, Depth) image by constraining the flow vectors on the intensity image in conjunction with the depth. This is not applicable for our work due to the method's requirement of using HSBD frames which conflicts with our system requirements.

Optical flow deals well with small, consistent and well-defined movements. As movements grow larger when the movement source is ill-defined optical flow loses accuracy due to the widening feature space and the aperture problem. The aperture problem states that when an object moves if the subject space is small such that along the object's principal axis within the space the object's shape is invariant then it is impossible to tell the direction of movement. A good graphical representation can be seen at: https://en.wikipedia.org/wiki/Motion_perception#/media/File:Aperture_problem_animated.gif. This is one of the causes of uncertainty at object edges and can cause noisy flow fields and other anomalies. This is illustrated in the following image where the flow vectors at edges all follow general trends but the actual angle varies from vector to vector.

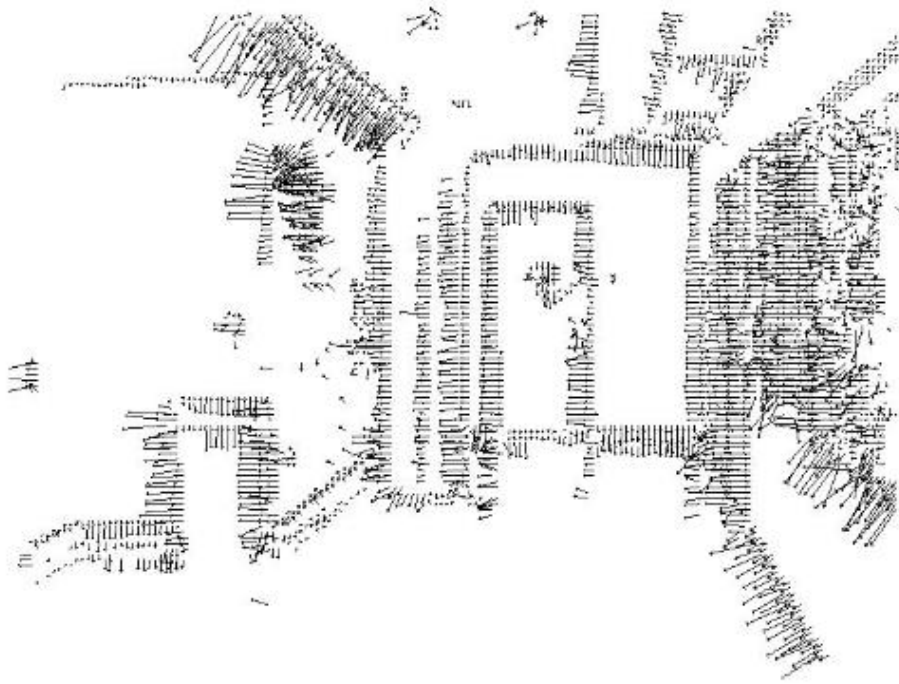


Figure 4, Flow field from corridor sequence (Barnum, P; Hu, B; Brown, C. 2003)

To mitigate the issue of not detecting great movements due to them moving outside of the searched space optical flow employs a form of pyramidal implementation. This is used in the OpenCV implementations, whereby the flow field is initially calculated on a subsampled form of the image, thereby allowing the method to find movements over greater distances within the same window. The output flow field of this iteration is then used as the starting flow field for the non-subsampled frame pair (Bouget, J.-Y., 2001). This pyramid can have an arbitrary number of levels with an arbitrary scaling provided that the scaling does not force super-sample, where the input of one level is the flow field of its subsampled frames.

6.7.1 Applying Farnebäck Optical Flow to Depth

Due to the fact that the method being used is provided by a library, much of the work performed concerning the application of Farnebäck optical flow upon depth data has been in the pre-processing, parameter selection and validation phases, ensuring that the optical flow method operates as it should.

When optical flow is applied to depth it looks for disparity between areas of constant luminance, however, the chances of the depth being consistent when moved is tiny as it would require the movements to be perpendicular to the sensors field of vision. This presents a number of challenges. For example, as depth does not change by much when the object moves perpendicular to the sensors view due to the depth variations within the objects mesh being relatively small when compared to the foreground vs background depth variations meaning only object edges will have movement detected. As with luminance footage, one of the hopes is that movements are small allowing high fidelity movement estimates which would let the constant luminance condition to hold.

OpenCV's Farneback optical flow function requires its input data to be a NumPy array with data type uint8. As our data has a range between 0 – 8000, we scale the data to be within the 0-255 range that is required by the functions. We scale based on the maximum and minimum values within the entire footage. This scaling has the effect of losing granularity in the data. However, without using an optical flow method built from scratch, we cannot prevent this. Dense optical flow is a slow process to run, and as such we only run it on frames where movement has been detected in background estimation. This scaling has the interesting effect of allowing use to perform a modulus upon the data, increasing the granularity and forming artificial textures on object contours. In practice, this allows a denser flow field which forms around the artificial contours, although this does not help our work much as it means that the movement detected around object contours is based on a multiple of the modulus.

Taking background estimation and noise reduction to be the pre-processing steps for optical flow we get the following. Optical flow does not react well to small, ill-defined changes in pixel groups. Thus, we apply the notion that informative and reliable movements will be those where multiple localised sources agree. As such we use a density based estimation, like that employed in DBSCAN (Ester, M. et al. 1996), on each anomaly finding the number of points within its neighbourhood which also are anomalies. We then only accept those anomalies which satisfy the density criteria, namely being in an area of relatively high anomalous pixel density.

The output of background estimation containing all non-zero values⁵ becomes apparent when we consider the following scenario. If we output a stream of zero valued frames, there will be no flow detected, which is correct. But, when an object moves the corresponding zero values change to be non-zero, optical flow then reacts by assuming that the object was previously at infinity and has moved to be at its apparent location. This results in the flow vectors not following the actual movement but diffusing away from the object's shape. Transitively, this influences our flow clustering method (described later) as the flow vectors point in all directions reducing the chances that they are grouped in the angle feature space.

This method has some parameters which were subject to systematic investigation, the method used the values given in Appendix 10, section 3, method 3. The validation for this section is combined with the following method as the two are intertwined.

⁵ Only true if all zero values are accounted for in the noise reduction stage of pre-processing.

6.8 Flow Clustering

At this stage the system extracts a flow field for each sequential frame pairings. From these flow fields we can extract quantitative information such as the direction and magnitude of movement at a point within the frame pairing. Any features extracted directly from the flow field suffer from problems similar to obtaining features from background estimation. Namely that the features would be over the whole frame meaning multiple movements could be occurring simultaneously or one complex movement and it would look the same. Therefore, to analyse movements in a more concise manner we cluster flow vectors which are similar in terms of both location, magnitude and direction. This lets us identify not only the magnitude and direction of movements but also the size of the object creating them⁶.

During the implementation of this method a number of parameters were subjected to systematic investigation, the method used the values given in Appendix 10, section 3, method 4.

6.8.1 DBSCAN (Depth Based Spatial Clustering of Applications with Noise)

DBSCAN is a clustering method based on the concept that elements which are densely packed belong to the same grouping and that outliers are those points or groups of points which reside in locations within low-density areas of the feature space (Esther, M. et al. 1996). This means that the only requirement for a point to be in any class is that it is in an area densely populated enough to be considered as a class rather than noise. DBSCAN is an unsupervised form of clustering; there is no pre-defined class to which each point belongs to. An understood limitation of the system is that instances within the footage can arise where the separability of two or more separate movements can be called into question. In other words, at times it can be difficult to distinguish between where one movement starts and another ends.

As we have not previously described the DBSCAN method we do so here. The standard implementation of DBSCAN is as follows.

1. Iterate over each instance within the problem space which are not currently defined as being part of a cluster and apply step 2.
2. Apply a radial function to each point within the window centred on the current point identifying all other instances within a predefined distance; the radial function determines the distance between two points. If the number of points found is above a set threshold value, then define the point and its neighbours as being part of a new cluster. Iterate over each of the points found by applying the radial function on each point in their window and carry out step 3 on each. If the number of locally found points is not above the threshold, then define the current point as noise and return.
3. Perform the radial function identifying local instances ignoring points belonging to a cluster. If the number of found points is not above the threshold, then define the current point a leaf of the cluster which found this point. If it's above the threshold, then perform step three to each locally found instances.

⁶ Assuming a continuous flow field

The data clustered is a matrix with shape $n \times m \times k$, where n and m are the pixel's location and k is the feature space for that pixel. The features contained within each pixel are the: x location, y location, angle and magnitude. The other inputs to the function are the epsilon value determining how similar two points must be to be clustered together, the minimum density for a point to be considered not noise and the minimum magnitude a vector must have to be clustered.

6.8.2 Alpha-Beta Filter

Many of the hypothesised “good predictor” features, such as jerk, rely on how movements change as time progresses. Thus, we track these clusters and record their changes as time progresses. There are many solutions available to solve this and a few issues when implementing them on the flow clusters.

Most methods rely on data which, in our case, are either unobtainable or difficult to estimate. This is reflected by the requirements in works such as Andriyenko and Schindler (2011) which relies on a sequence of instance points to estimate their paths post data capture; these points would need to be sampled from the data. A fundamental element of the method is based on the underlying objects shapes being mostly continuous through time. Our shapes, however, are not, as the flow field produced can change from frame to frame depending on estimates made. This gives rise to the need of an estimate to be obtained which is independent of cluster shape and dimensions. Held et. al (2015) discussed some methods but, again, they all require a shape which does not deform much over time. Thus, we only find the centroid for each movement cluster at the average location of the flows contained within that cluster. This idea is backed up by the way that human bodies are shaped, as each body-part which we detect movement for (feet, thighs, calves, hands, forearms, upper-arms and head) are single continuous shapes making the independent flow fields obtained convex giving non-overlapping centroids. This is not true in practice as, as revealed by our optical flow implementation, the clustered flows around the head often form a crescent shape with no filling which can be discontinuous along the plane of motion creating two separate crescent-shaped clusters which can combine as time progresses.

A method well-renowned for its simplicity and effectiveness is the alpha-beta filter (Kalata, P.R. 1984). The alpha-beta filter uses elementary observations to estimate following values by continuing the current trends within the data to the n th order. The standard implementation uses two base values which are assumed to be constant between time steps, and the first value is obtained from the integral of the second. The values are continued by incremented steps with a size determined by the second value. E.g., in predicting the change in position of a moving object we utilise the basic value of its position and estimate its velocity as the change in its position, we then can estimate the objects next object by projection. This can be extended into higher order approximations by gathering estimates for acceleration and jerk. We can apply this alpha-beta filter to our centroid flow clustering model to extend clusters temporally by estimating their next location and fitting clusters based on their likelihood of belonging to the same cluster. This is not a flawless solution and can suffer under very large movements and discontinuous flow clusters. It also has issues when dealing with clusters which split or join. For instance, a person leans forward and halfway through the movement they raise the arm closest to the sensor. What previously was one cluster containing the upper body moving forward, has now split into two separate movements, the reverse can also occur where two movements combine into one.

6.8.3 Validation

Farnëback optical flow has a limitation which reduces the effectiveness of flow clustering and can have issues when calculating the size of movements which have been made. Consider the following two sequences which have been computer generated. They are free of artefacts and corruptions. They are generated such that in a perfect world they could be recreated. The sequences depict a sequence of 500 x 500 depth frames containing a rectangular object moving left to right at a constant velocity. The depth value of the background is 1000, and the depth value of the object is 2500. Technically this makes the objects further away than the background but this is just relative, and the function which will approximate it would just flip if they were the other way around.

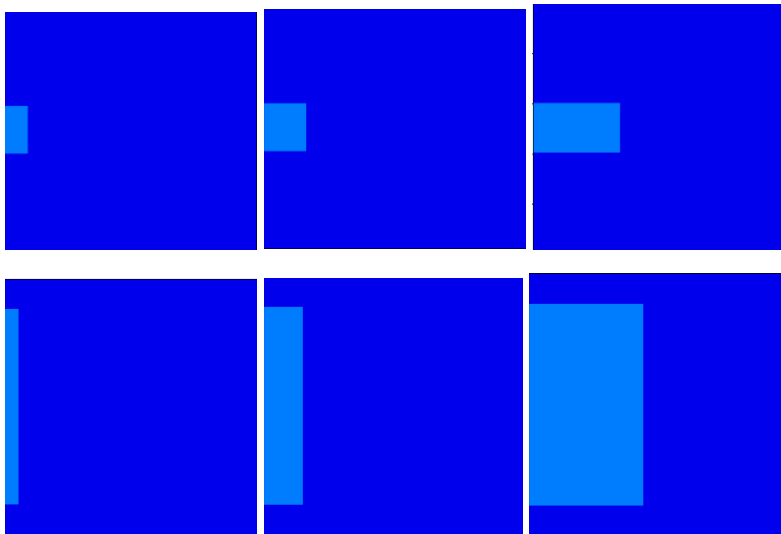


Figure 5, Depth test sequence, cube moving left to right

If we now apply Farnëback optical flow with our flow clustering, originally I would assume that the entire wave of motion would be represented by flow vectors with those which are undefined in each component axis as with the aperture problem having flow vectors along the definite movement axis, i.e., perpendicular to the edge. However, what we get is the following.

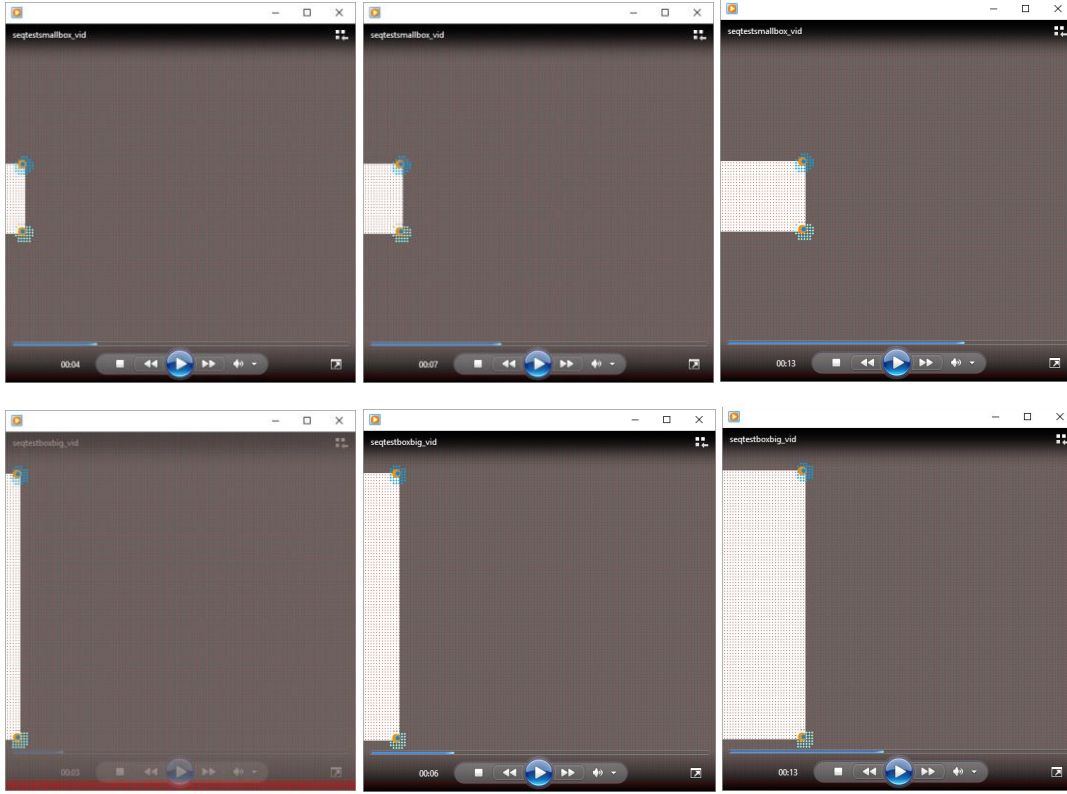


Figure 6, Depth test sequence, cube moving left to right, with clustered flow vectors

This means that the Farnëback implementation only assigns flows where it is confident of the direction of motion over all axes. This creates issues when attempting to form models around the concept that the flow field gives the definitive interpretation of motion within the scene when not all the motion is accounted for.

This is a recurring issue as it produces discontinuous flow fields which are ill-handled by clustering of any sort. This is not limited to our clustering solution as there is nothing to cluster between corner cases. This effect is not found in much of real world data as along object's principal axis the width is generally variant. The issue is exacerbated in our situation, however, as edges are the locations at which this problem is found, and edges are also often the source of zero values in the Kinect system. There is little we can do to fix this issue, and as such we apply our method in the hope that there is enough variance in the objects within the scene to effectively recover a flow field from the data.

An example of the clustering issue is shown below. The frame is from participant y117 and depicts a discontinuous flow field surrounding the participant's leg.

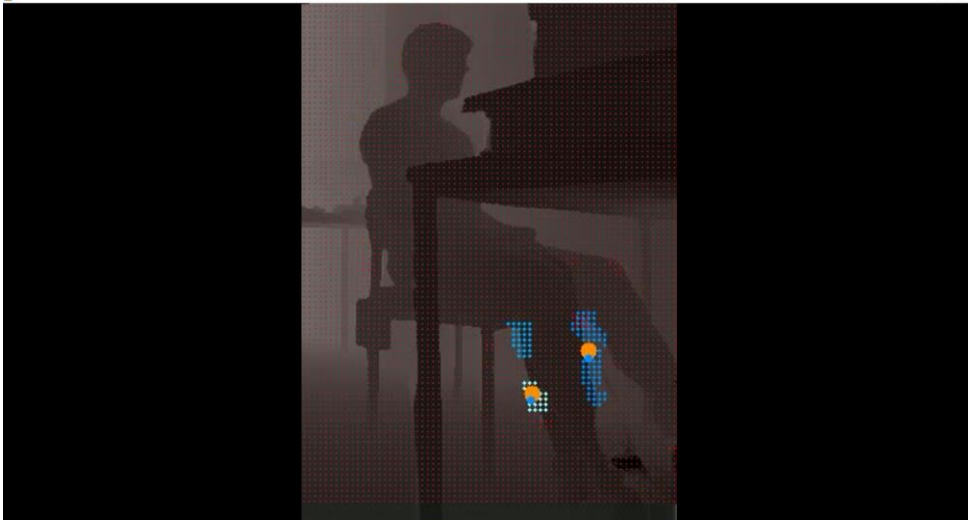


Figure 7, Experiment from set Y117, shades of blue indicate distinct clusters of flow vectors

Applying clustering upon the flow field, we show that we can identify between distinct movements where the flow-field between them is non-zero. For example, given two boxes where one of their corners are approaching one another as time proceeds, we can distinguish the movement of one from the movement of the other including when their vector fields overlap as is the case below.

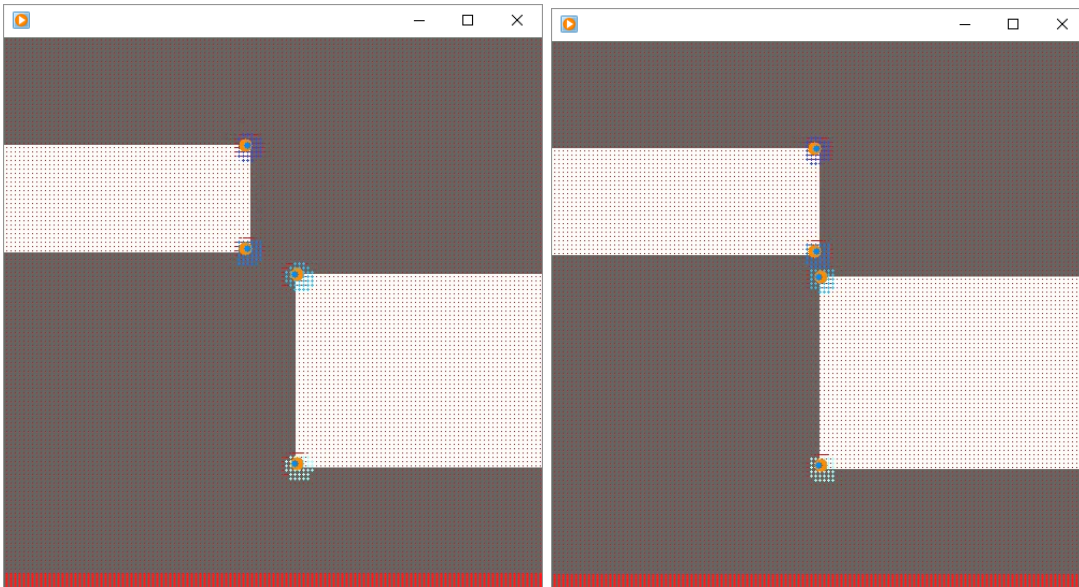


Figure 8, Overlapping flow field clusters independently identified

6.9 Classification

To obtain predictions from the features selected and generated we apply the reasoning upon which the project is based, namely, that people who experience similar cognitive affects make similar movements. This means that if we extract a good measure of what movements are pertinent to cognitive affect, the values should be grouped according to what class they belong to.

This section has a few parameters which were subject to evaluation, the implementation used the values given in Appendix 10, section 3, method 5.

6.9.1 k-NN (k-Nearest Neighbours)

k-NN provides a simple and effective method of identifying the class of a given feature set. It is quick to train and, depending on the distance function, fast to run. It goes through each of the testing set and builds a list of the k closest training sets as defined by the distance measure. The predicted class is then chosen as the majority class within the list. When two are tied, there are various ways of choosing the class. Without extra information, this will always be random. Thus, the class chosen in this case is based on the proportion of the class in the training set, biasing more the more frequent classes in the population.

6.9.1.1 Distance Measure

The typical distance measure for k-NN is Euclidean distance. We cannot use this in at least one of our cases as there are instances where the feature length is variable. This is shown in the information extracted from individual movements where there can be any number of different movements within an experiment.

This results in a difficult decision. We can apply standard statistical methods upon these data to get an aggregate of their qualities. However, this loses the data independence acquired for us by clustering the optical flow, bringing us to back a similar situation as background estimation. If we apply prior knowledge gathered from discussion with Harry Witchel, we can identify important aspects related to the detection of cognitive affects from NIMI. From NIMI we can see that footage with similar movements should give similar cognitive affect results. Therefore, aggregating the movements loses this measurement of similarity between independent movements. Thus, we utilise the information extracted from each movement such as the jerk and movement complexity. We then classify by using the currently implemented movement measurement:

1. Iterate over each movement estimate in the testing instance.
2. For each movement estimate in the testing instance we loop through each movement estimate in the testing instance comparing the difference between the measurements in each instance, keeping track of the closest estimate.
3. Sum the total difference between all estimate pairs
4. Return distance as the sum of differences

This provides us with a tunable distance measurement which can be altered with various parameters and functions to react differently depending on the input instances. For example, the starting value for summing the differences between estimates normally starts at zero. However, it can be changed such that the value it starts at depends on the difference between some other parameter contained within each of the instances. This then can be set using the standard Euclidean distance method when multiple different features are involved. The exact set of features and the justification behind them is given in both the discussion and results sections.

6.10 Libraries and Software

During the execution of this project multiple libraries and tools have been made use of for the implementation of various elements. These are listed here with their usage, version and source to provide information critical to replicability and maintenance of the system.

Library/System	Version	Usage	Source
<i>Python</i>	3.6.0	Language used to write all code in	https://www.python.org/
<i>Microsoft Visual Studio</i>	14.0.25431.01	Used as IDE for code writing	https://www.visualstudio.com/
<i>Python tools for Visual Studio</i>	2.2.50113.00	Allows Visual Studio to act as an IDE for Python	https://microsoft.github.io/PTVS/
<i>NumPy</i>	1.12.0	Manipulating and holding data being processed and analysed	http://www.numpy.org/
<i>MatPlotLib</i>	2.0.0	Displaying footage and results to user	http://matplotlib.org/
<i>H5PY</i>	2.6.0	File storage API	http://www.h5py.org/
<i>CV2</i>	3.2.0	Optical Flow function implementations used	http://docs.opencv.org/3.0-beta/doc/py_tutorials/py_gui/py_image_display/py_image_display.html
<i>Glob</i>	*	Used to find all files in directory with a given file extension	https://docs.python.org/2/library/glob.html
<i>Threading</i>	*	Provides parallelism in processing a directory of files, although implementation is not reliable so is not used	https://www.python.org/
<i>Thread Collections</i>	*	As above	https://www.python.org/
	*	Deque functionality used in DBSCAN when adding elements for Breadth First Search	https://www.python.org/
<i>DateTime</i>	*	Used to insert time data into frames during recording	https://www.python.org/
<i>OS</i>	*	Used in old implementations to identify file's existences and in new implementation when switching file system	https://www.python.org/
<i>Math</i>	*	Used in multiple locations and for various functions when various operations and checks are required on scalar values	https://www.python.org/
<i>Sys</i>	*	Provides access to arguments passed into program by batch file execution	https://www.python.org/
<i>Time</i>	*	Debugging purposes, pauses thread execution for inspection of variables printed to the console	https://www.python.org/
<i>PyKinect2</i>	0.1.0	Wrapper providing access to retrieve frames from the Kinect.	https://github.com/Kinect/PyKinect2
<i>TKinter</i>	*	Python's standard GUI library	https://www.python.org/

(* - Version is determined by the Python distribution version)

Figure 9, Libraries and software used in implementation

7 Results

Several results have been generated from the analyses of the data. This section defines them, justifies them and identifies their usefulness. We also quantify their accuracy in prediction using our k-NN classifier.

7.1 Background Estimation

Background estimation provides a generalised view of the total movement within a scene. It should be noted that the amount of movement detected is generally twice of that which occurred. This is due to these differences occurring whenever an object changes its position. When an object moves, the depth values of its new location change as well as the depth values of its original location.

Some of the hypothesized good features we can extract from background estimation that we test are:

- Frequency of resting periods
- Average size of movements (average magnitude of pixel displacement)
- Complexity of total movement (Average jerk)

Using these features as predictors, we produce the following results regarding their respective accuracy values. The predictions are averaged over 200 dataset permutations, ignoring those determined as corrupted, and the average accuracy is listed below⁷.

	Rest frequency	Average rest duration	Rest duration standard deviation	Inverse of average total jerk
True accuracy	16.4%	16.9%	26.3%	28.1%
Modified accuracy	11.3%	12.2%	19.1%	18.4%
Class 1 accuracy	26.3%	22.4%	43.4%	38.9%
Class 2 accuracy	30.8%	29.1%	39.9%	51.4%
Class 3 accuracy	7.37%	15.9%	22.7%	16.6%
Class 4 accuracy	1.41%	5.4%	8.09%	3.00%
Class 5 accuracy	0.00%	0.30%	0.00%	0.00%
Class 6 accuracy	1.10%	0.00%	0.20%	0.70%

Figure 10, Background estimation accuracy values

⁷ Orange colour indicates highest accuracy model for that measure

Per these results, we can see that we can predict class one and two marginally better than random and we can almost never predict class 5 suggesting that per these measures it is distributed too sparsely along another class's range. We can see that the standard deviation of the rest periods gives us greater predictive power over less frequent classes when compared to total jerk, but total jerk works more effectively on instances of class two. The rest frequency has an interesting quality of its predictive power being below random chance; there are various reasons this can occur and a likely one is statistical anomalies. Interestingly the rest frequency is well distributed over all classes excluding five (Restless Boredom), though this is likely to come from inadequate quantities of data.

7.2 Clustered Optical Flow

In theory, the movement clusters contain all the information of a movement over its duration. From the movement cluster, we extract information about the movement including; how it varies over time, its magnitude and its average jerk. As has been stated, the clusters are quite unreliable in terms of time, space and contents, this is mostly due to the discontinuous nature of our optical flow. We use the following measures extracted from the tracked clusters to generate our accuracy estimates:

- Movement Complexity (Average jerk)
- Movement Duration

	Movement complexity	Movement duration
True accuracy	22.4%	21.3%
Modified accuracy	14.6%	15.0%
Class 1 accuracy	35.4%	46.1%
Class 2 accuracy	50.0%	27.3%
Class 3 accuracy	0.93%	12.7%
Class 4 accuracy	0.36%	4.10%
Class 5 accuracy	0.08%	0.00%
Class 6 accuracy	0.00%	0.00%

Figure 11, Optical flow accuracy values

7.3 Results Review

We can see that we do not get high accuracies from our system. This is due to the features that we used do not have much in the way of differing distributions for the values of each class. For example, the rest frequency feature from background estimation has essentially the same distribution for each of the separate classes.

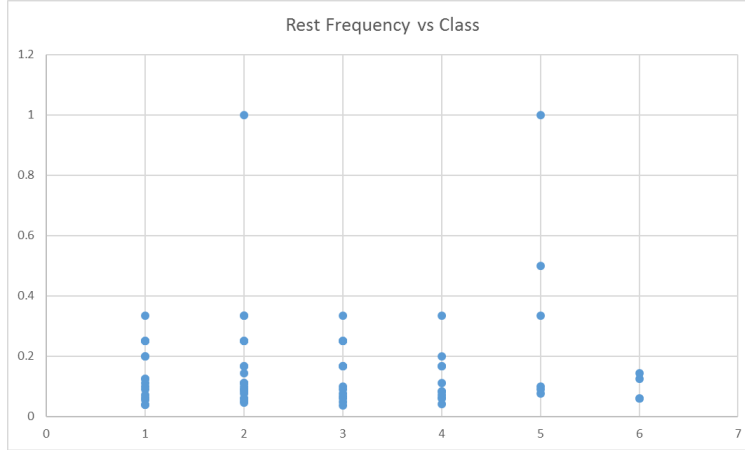


Figure 12, Rest frequency vs class

Interestingly we have accuracy values which are below random chance. These indicate a form of negative accuracy which indicates that, while there is a connection between the feature and its class it is not in the way we are currently utilising it. This is likely to be from the uneven distribution of classes within our dataset combined with the nature of our classification task being 6-way.

Evidently clustering optical flow, us almost random accuracy values in both measures we utilise. This is likely to be because our clusters are unreliable and discontinuous. Resulting in clusters joining and splitting as time proceeds, creating feature values which do not represent the movement as a whole but rather only a component of it. These components can also be created multiple times as there can be multiple clusters formed around various locations on a moving object which all provide the same information. Finally, an element of the different resolutions comes into play here. As we recorded the experiments from various positions, the percentage of the sensor's resolution representing the participant's image varies which affects the optical flow and therefore the clustering and classification. Thus, we do not feel that these values can be trusted and as such we revert to the features produced by the background estimation.

Background estimation can generally be trusted to a relatively high margin in its results. Although, as previously discussed, it is not effective in all cases, where there are frequent movements which outweigh the regularisation factor the system can become desensitised to movement and treat it as if it were noise. As evident by the accuracy values, background estimation's predictors are still not of high quality. This could be due to the dataset or various other reasons, such as us not taking smaller movements into account or movements concerning the rotation of body parts keeping them almost the same regarding depth and outline but still presenting movement.

In conclusion, the best predicting model we regarding true accuracy is the inverse of average total jerk and the rest duration standard deviation concerning our modified accuracy measure. Though, these are still not good predictors.

8 Conclusion

This project sought to develop a system which automatically labels the cognitive affective state experienced by a participant in a piece of depth footage. As there are no available methods to do so, we have explored developing variants of methods typically used on colour/luminosity data for use with depth data. These methods have been validated. We identified features upon which a simple k-NN classifier was developed to classify. Although we were unable to predict ratings to the expected level of accuracy, there were encouraging signs as shown in the results section. In the below, we provide a comprehensive review of both successful and not so successful outcomes of this project.

We have seen that we can identify and quantify movement from depth and can define a statistical model which can adjust and fit itself to the anomalous values from the hardware. We have shown that noisy and unreliable data can be reduced into a form capable of being processed without much interference and that this data can be stored and retrieved efficiently and quickly, provided that the data is relatively clean already. We have shown that the Farneback implementation of optical flow can be applied onto depth based footage with reasonable success and minimal pre-processing. We have then shown that we can cluster the resultant dense flow field to identify independent movements and track them over time with mixed results.

8.1 Project Objectives

In this section, we evaluate our work two ways. First we evaluate in accordance with our original objectives. Second, with the benefit of hindsight, we evaluate the work in relation to objectives that would have been more appropriate for this project.

8.2 Original Objectives

8.2.1 Identify cognitive affective state and subclass

As stated originally an accuracy value in the range of 85-90% was always an impossible goal due to the limitations such as quantity and quality of classes in the dataset. The accuracy values we extract when applying our k-NN classifier is disappointing, however, we have also shown that there are definite correlations between the features we have used and class membership.

8.2.2 Use movement as the source of features

We have demonstrated that we can identify, quantify and analyse individual and groups of movements and then use these extracted values as input for the classification task.

This is not to say that this works flawlessly as there are various issues with the solution. Generally, one of the biggest issues is the discontinuous nature of the Farneback optical flow which forces 0 magnitude flow vectors when the surface is non-variant along one of the principle axes.

8.2.3 Identify optimal feature set for classification task

This objective has obviously not been met. Due to various limitations, both with the solution and the quantity of data there is no way we can back up claims legitimately saying that we have found optimal or even “good” features. However, we can say that there are definite correlations between the results and the classes.

8.2.4 Programming language

The system was completely written in Python and all libraries are easily available on the Python Package Index website via the PIP function.

8.2.5 Kinect

In collecting data, the Kinect and our recording system were used in all instances aside from those generated by the computer for system validation.

8.2.6 Balancing quantities of data

By trimming the start and ends of each experiment, cropping out unnecessary areas of the frames and using a file storage system capable of holding and, if required, chunking the data which further reduces its size.

8.2.7 Data can be analysed/classified post event

The data have been analysed after collection and no analyses have been performed on the fly. This is not to say that these analyses cannot be done in a shorter response time, however.

8.2.8 Data capture efficiency in terms of computer power

In our experience of the system the number of dropped frames has been very low and it has almost always run as intended allowing easy extraction and analysis of the data. The solution itself is lightweight and requires little processing on our end. It is difficult to quantify how lightweight the solution is as it will run differently on various systems with code run times changing from hardware versions and various optimisations which can be performed.

8.3 Extended Objectives

As stated above the original objectives were not full nor expressive enough of the required system.

8.3.1 Quantify movement

We can assess movement differences on a frame by frame basis where the movement is determined by the cardinality of the pixel difference. The total movement is estimated via background estimation although a simple method of frame subtraction which ignores the background may be more effective as background estimation learns new data relatively slowly. Movements can be further quantified via optical flow clustering which generates values for individual movements retrieved from the clustered flow field.

8.3.2 Evaluate classical luminance-based methods on depth data

We have successfully applied a small sample of methods to depth data through introducing small changes. We have shown that depth can have pros and cons over luminance in the various situations.

8.3.3 Evaluate advantages and disadvantages of various systems performing the same functionality

We have mostly not done this aside from comparing Lucas Kanade and Farneback implementations of Optical Flow.

8.4 Critical Appraisal

This project has not been executed without fault nor error. This section provides an executive summary of the errors and issues encountered during the body of this work, focusing on each of the following four sections: Planning, System Utilities Implementation, System Implementation and Result Generation. This is an informal review of the work and has much conjecture and direct thoughts with the benefit of hindsight.

Overall my feelings about the project are mixed, I'm annoyed that I could not fulfil the original aim of predicting cognitive affect accurately and that there are issues with most stages of the system. But, I have learnt many techniques and had the chance to work on a research project attempting to automate a classification system in a novel way.

8.4.1 Planning

During the plan, I was quite naïve regarding image processing in terms of the methods and theories behind them. I made a plan assuming that tools would be readily available and that much of the work I would be undertaking, on the computer vision side, would be covered extensively. This, of course, was not entirely the case as each method was well documented by itself, but combinations of the systems were lacking, and there was little dealing with extracting movement information directly from depth sensors. Most movement analyses use intensity and depth as an afterthought or as an information source once the movements were already identified. This meant I gave myself unreasonable expectations on the amount and reliability of the work which could be undertaken.

8.4.2 System Utilities Implementation

This is where my unfamiliarity with various systems became an issue. I started by writing a data recording system which recorded the data as byte files, this is obviously bad practice, and once I found HDF5, I immediately swapped over the recording solution. Before I found HDF5, however, the system could not handle writing both colour and depth frames for some reason which was one of the reasons we worked purely with depth. When displaying data, I made use of, and never stopped, Matplotlib's display plots. While this is not necessarily bad in and of itself, the implementation was a bit messy, and while it works, I'm sure that there is a better and cleaner way of doing it. The file storage solution using HDF5 was not without its flaws of course as HDF5 is a very fast and extensible library for data storage but does not provide functionality for the removal of datasets. This means that every so often the current data needs to be transferred into a new dataset to prevent the constant increase in file size.

8.4.3 System Implementation

As the work proceeded and I became more acquainted with various footage processing techniques I kept on going back over various elements and realising that I had done something wrong or was not effective. For example, the background estimation initially followed the normal rules for online statistics generation, but this caused multiple issues such as the desensitisation to repeated movements over the same spatial location. This is what prompted the regularisation parameter, which I am also sure is not optimal but works well enough. Even now I'm not sure if background estimation as a feature extractor is better than a simple noise resistant form of frame differencing.

Our noise reduction solution is a simplistic and naïve method and relies on the footage being mostly clean. It also does not take into account the fact that, as most corruption is at object edges, edges are generally sloped.

In retrospect, it may have been more useful and practical to have implemented the range flow estimator as defined by Spies et al. (2002) upon our pre-processing methods. Though this may suffer from similar issues to the Farnëback implementation of optical flow, namely the aperture problem. We could have focused on implementing a model based system identifying connections between the different flows and as a requirement of non-variant areas is linearity we can directly connect the two flow fields and thereby infer the vectors in-between. I think an RGB-D system could have been much more effective and may have made elements of the system easier and more reliable.

I think that the DBSCAN clustering was a good proposal but could have been much more effective if the flow fields were proper. The centroids tracking method leaves much to be desired. It may have been best to process the entire footage and then produce a number of spatial and temporal points which can then be connected in post-processing using a method such as that proposed by Andriyenko and Schindler (2011).

8.4.4 Result Generation

The generation of results was one of the most frustrating parts of the project. It is evident that there are correlations between the statistics we have extracted and the class of the instance those features are drawn from. However, we have not been able to get a reliable classifier or set of features. This is mostly my fault due to inexperience and poor choices, such as choosing an online method of cluster tracking rather than a post-processing form. However, optical flow and the dataset are also to blame due to its limited in number and the distribution of classes throughout it.

With more, cleaner data and an unrestricted amount of time, I'm convinced that this problem is solvable, but I think I'd need to understand the physiological side of things more.

8.5 Further Work

There are a large number of improvements and extensions which can be made on this work, below, we elaborate on the most relevant and/or immediate ones.

- Generate much more data in a cleaner location with a more equal distribution of classes.
- Extend optical flow into range flow utilising a non-variance linear estimator⁸ which would produce a cleaner, non-discontinuous low field with vectors which provide information about each of the three dimensions.
- Utilise RGB-D footage rather than straight depth data as in Spies et al. (2002) work.
- Make use of Kinect Trilateration to improve accuracy of all points in the scene and perhaps form a 3d mesh upon which an $n \times m \times d$ flow field matrix can be obtained from assuming objects are hollow and that a cell which contains an objects surface has a Boolean value of True and all cells not containing an edge have a value of False.
- Define good features prior to implementation such that the system can be optimised for them, e.g., if micro-movements are important then no thresholding should be performed.
- Utilise a more robust tracking method which runs once all clusters have been performed, for example, Andriyenko and Schindler (2011).
- The area of the footage cropped should be implemented via a GUI whereby the user selects a bounding box around the subject. No doubt this could also be implemented to perform automatically.

⁸ A non-variance linear estimator is described as the process of inferring the flow vectors along an edge which falls victim to the aperture problem from the vectors at each end of the edge. Provided that the entire length is connected, each flow vector's value is set according to its relative position along the edge to the two calculated flow fields.

9 References

- Andriyenko, A., & Schindler, K. (2011). Multi-target tracking by continuous energy minimization. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (pp. 1265–1272). <https://doi.org/10.1109/CVPR.2011.5995311>
- Aviezer, H., & Todorov, A. (2015). Body Cues, Not Facial Expressions, Discriminate Between Intense Positive and Negative Emotions. *Science*, 338(November 2012), 1225–1229. <https://doi.org/10.1126/science.1224313>
- Barnum, P., Hu, B., Brown, C., (2003). Exploring the Practical Limits of Optical Flow (Report No. 806). Rochester, New York. The Univeristy of Rochester.
- Castellano, G., Villalba, S. D., & Camurri, A. (2007). Recognising Human Emotions from Body Movement and Gesture Dynamics. In *Affective Computing and Intelligent Interaction* (pp. 71–82). https://doi.org/10.1007/978-3-540-74889-2_7
- Darwin, C. (1872). The expression of the emotions in man and animals. *The American Journal of the Medical Sciences*, 232(4), 477. <http://doi.org/10.1097/00000441-195610000-00024>
- D’Mello, S., Chipman, P., & Graesser, A. (2007). Posture as a predictor of learner’s affective engagement. *Proceedings of the 29th Annual Meeting of the Cognitive Science Society*, 1, 905–910.
- D’Mello, S., & Graesser, A. (2007). Mind and Body: Dialogue and Posture for Affect Detection in Learning Environments. *Artificial Intelligence in Education*, 158, 161–168.
- D’Mello, S., & Graesser, A. (2011). The half-life of cognitive-affective states during complex learning. *Cognition & Emotion*, 25(7), 1299–1308. <https://doi.org/10.1080/02699931.2011.613668>
- Ekman, P., & Friesen, W. V. (1971). Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology*, 17(2), 124–129. <http://doi.org/10.1037/h0030377>
- Ester, M., Kriegel, H. P., Sander, J., & Xu, X. (1996). A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. In *Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining* (pp. 226–231). <https://doi.org/10.1.1.71.1980>
- Farnebäck, G. (2003). Two-Frame Motion Estimation Based on Polynomial Expansion. *Lecture Notes in Computer Science*, 2749(1), 363–370. https://doi.org/10.1007/3-540-45103-X_50

- Fleet, D., & Weiss, Y. (2005). Optical Flow Estimation. *Mathematical Models for Computer Vision: The Handbook*, 239–257. doi:10.1109/TIP.2009.2032341
- Gratch, J., Lucas, G. M., King, A. A., & Morency, L. P. (2014, May). It's only a computer: the impact of human-agent interaction in clinical interviews. In *Proceedings of the 2014 international conference on Autonomous agents and multi-agent systems* (pp. 85-92). International Foundation for Autonomous Agents and Multiagent Systems.
- Held, D., Levinson, J., Thrun, S., & Savarese, S. (2015). Robust real-time tracking combining 3D shape, color, and motion. *The International Journal of Robotics Research*, 35(1–3), 1–28. <https://doi.org/10.1177/0278364915593399>
- Huth, K., and de Ruiter, J. (2012). “Ordering a beer without your hands,” in *Proceedings of the 5th Conference of the International Society for Gesture Studies (ISGS 5)* (Sweden: Lund).
- Kalata, P. R. (1984). The Tracking Index: A Generalized Parameter for $\alpha - \beta$ and $\alpha - \beta - \gamma$ Target Trackers. *IEEE Transactions on Aerospace and Electronic Systems*, AES-20(2), 174–182. <https://doi.org/10.1109/TAES.1984.310438>
- Kapoor, A., Burleson, W., & Picard, R. W. (2007). Automatic prediction of frustration. *International Journal of Human Computer Studies*, 65(8), 724–736. <https://doi.org/10.1016/j.ijhcs.2007.02.003>
- Katula, R. a. (2003). Quintilian on the Art of Emotional Appeal. *Rhetoric Review*, 22(1), 5–15. http://doi.org/10.1207/S15327981RR2201_1
- Kleinsmith, A., & Bianchi-Berthouze, N. (2013). Affective body expression perception and recognition: A survey. *IEEE Transactions on Affective Computing*, 4(1), 15–33. <https://doi.org/10.1109/T-AFFC.2012.16>
- Kumar, J. (2013). Gamification at work: Designing engaging business software. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (Vol. 8013 LNCS, pp. 528–537). <https://doi.org/10.1007/978-3-642-39241-2-58>
- Mota, S., & Picard, R. W. (2003). Automated Posture Analysis for Detecting Learner's Interest Level. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops* (Vol. 5). pp. 49-49 <https://doi.org/10.1109/CVPRW.2003.10047>
- Muntean, C. C. I. (2011). Raising engagement in e-learning through gamification. *The 6th International Conference on Virtual Learning ICVL 2011*, (1), 323–329. Retrieved from http://icvl.eu/2011/disc/icvl/documente/pdf/met/ICVL_ModelsAndMethodologies_paper42.pdf

- OpenCV: Optical Flow. (n d). Retrieved April 2, 2017 from World Wide Web:
http://docs.opencv.org/trunk/d7/d8b/tutorial_py_lucas_kanade.html
- Prazdny, K. (1980). Egomotion and relative depth map from optical flow. *Biological Cybernetics*, 36(2) (pp. 87-102). <https://doi.org/10.1007/BF00361077>
- Ryan, S. J., Mello, S. K. D., Mercedes, M., Rodrigo, T., & Graesser, A. C. (2010). Better to be frustrated than bored : The incidence , persistence , and impact of learners ' cognitive – affective states during interactions with three different computer-based learning environments. *Journal of Human Computer Studies*, 68(4), 223–241. <https://doi.org/10.1016/j.jhcs.2009.12.003>
- Sanghvi, J., Castellano, G., Leite, I., Pereira, A., McOwan, P. W., & Paiva, A. (2011). Automatic analysis of affective postures and body motion to detect engagement with a game companion. In *Proceedings of the 6th International Conference on Human-robot Interaction* (pp. 305–312). <https://doi.org/10.1145/1957656.1957781>
- Scherer, K. R. (2005). What are emotion? And how can they be measured? *Social Science Information Sur Les Sciences Sociales*, 44(4), 695–729. <https://doi.org/10.1177/0539018405058216>
- Spies, H., Jähne, B., & Barron, J. L. (2002). Range Flow Estimation. *Computer Vision and Image Understanding*, 85, 209–231. <https://doi.org/10.1006/cviu.2002.0970>
- Welford, B. (1962). Note on a Method for Calculating Corrected Sums of Squares and Products. *Technometrics*, 4(3), 419–420. <https://doi.org/10.2307/1266577>
- Witchel, H. J., Santos, C. P., Ackah, J. K., Westling, C. E. I., & Chockalingam, N. (2016). Non-Instrumental Movement Inhibition (NIMI) Differentially Suppresses Head and Thigh Movements during Screenic Engagement: Dependence on Interaction. *Frontiers in Psychology*, 7(February), 157. <https://doi.org/10.3389/fpsyg.2016.00157>
- Yang, L., Zhang, L., Dong, H., Alelaiwi, A., & Saddik, A. (2015). Evaluating and improving the depth accuracy of Kinect for Windows v2. *IEEE Sensors Journal*, 15(8), 4275–4285. <https://doi.org/10.1109/JSEN.2015.2416651>
- Zajonc, R. B. (1968). The attitudinal effects of mere exposure. *Journal of Personality & Social Psychology*, 9, 1–27. <https://doi.org/10.1037/h0025848>

10 Appendix

10.1 Project Log

10.1.1 Autumn Term

Week	Tasks
1	Met supervisors and discussed project and background, introduced to the office space and various concepts
2	Discussed project proposal and set weekly meeting time
3	Discussed project proposal and applicable machine learning techniques. Started implementation of simple recording solution
4	Reviewed project proposal and referencing styles. Evaluated and altered recording solution. Discussed report formatting, contents and styling conventions
5	Started Interim report and researching various papers and subject matters. Begun detailing system requirements
6	First draft of interim report completed. Created visualisation system to review data collected
7	Second draft of interim report
8	Submitted interim report and improved recording system
9	Researched methods of noise reduction and their sources
10	Started system implementation, naïve form of noise reduction, does not work well. Learning tools required
11	Improved noise reduction system and finishing data collection.
12	Finished data collection and reviewed data.

Figure 13, Project log for autumn term

10.1.2 Winter Break

Finished implementation of noise reduction and investigated forms of frame differencing and methods of tracking objects in space. Almost impossible to differentiate between markers and noise.

10.1.3 Spring Term

Week	Tasks
1	Research background estimation and implement simplistic version. Relative success, a number of bugs and unexpected occurrences
2	Finished background estimation and performed it upon dataset. Researched optical flow and its implementations in OpenCV
3	Considered Farneback implementation of optical flow and compared it to Lucas Kanade version. Lucas Kanade is promising but has too many large shortcomings in our system. Identified measures extractable from Flow Field
4	Implemented Farneback optical flow on background estimated dataset
5	Bug fixed Farneback optical flow and considered clustering techniques. Started generating some results from background estimation, look promising.
6	Decided upon DBSCAN and implemented on top of flow field. Needed to track clusters through temporal space. Used naïve implementation where track using alpha-beta filter applied to cluster centroids. Reviewed results, look less promising now, too much overlap between class features
7	Went back over background estimation and implemented regularisation parameter. Validated optical flow and performed it upon recorded and computer generated test sequences
8	Evaluated method parameters and implemented k-NN classifier. Started generation of results proper. Bugs found in classifier and after discussion adjusted k-folds split of data. Started final report writeup, began with writing implementation, used as a form of validation/bug checker
9	Extract results and reviewed them. Continue report writeup.
10	Finish report. Review and edit report. Generated final results. Sent draft to Supervisor

Figure 14, Project log for spring term

10.2 Dataset Review

The data gathered and the classes they contain are described here along with any corruptions found. We refer to the entire set of experiments used as the dataset; a single experiment-participant combination shall be referred to as an experiment and all experiments performed using a single user shall be referred to as the experiment set.

The dataset used in this work was gathered by Dr. Harry Witchel and Tom Ranji. Within the dataset, there are 86 experiments in sets of 9 to 11 where each set uses a single participant, and all controllable parameters are the same throughout the set. There are six distinct classes of data within the dataset: Active Engagement, Rapt Engagement, Restless Boredom, Lethargic Boredom, Neutral, and Frustration. These categories are disjointed enough to justify being separated into distinct sets. The entire dataset list and class membership is given in Appendix 10, section 3.

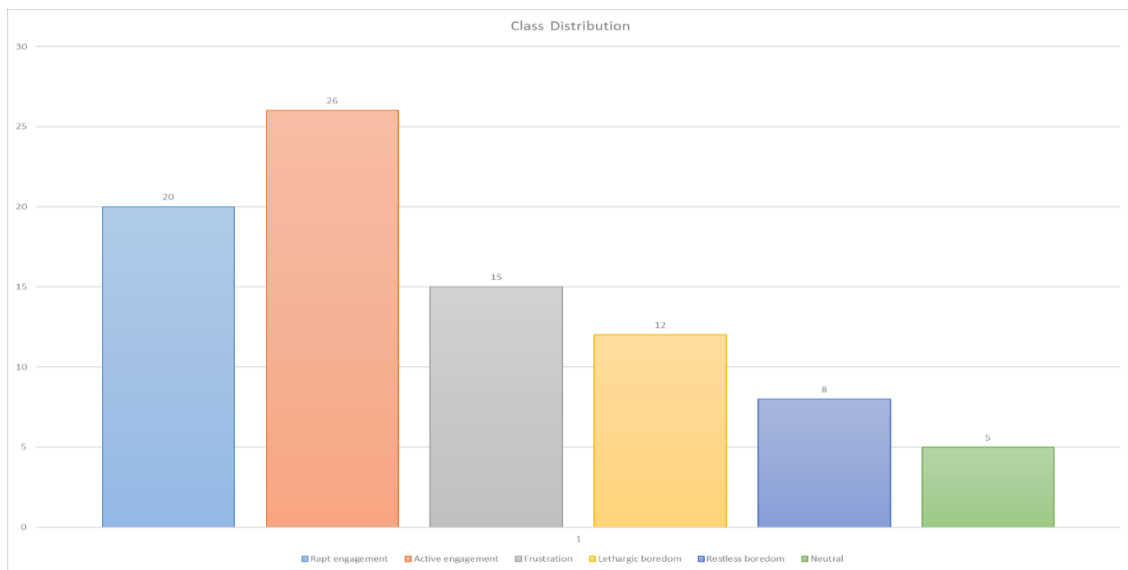


Figure 15, Class distribution in dataset

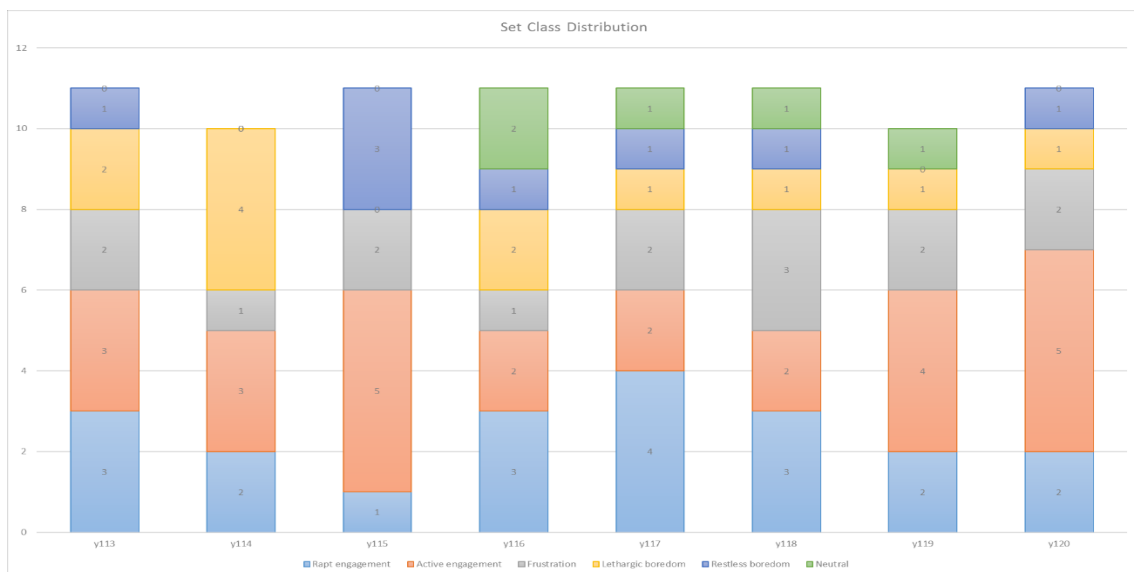


Figure 16, Class distributions by experiment set

We can see that the distributions of data are skewed, favouring engagement over the other four classes by having them comprise almost half of the data.

These data are not all recorded using the same setup, and as such we are required to take this into account during this review. The data were recorded in three separate configurations: Side on, 45° and front on. The method used also affects the distance of the sensor from the participant. The reason for this was to give various situations in which we can evaluate our methods and determine the best style for future work. However, it does reduce the uniformity of data and almost certainly influences the final classification prediction and effectiveness. It all forces a certain amount of abstraction away from the apparent movements as the response will change between methods. For instance, the silhouette of someone raising their hand to wave will be different from the side-on perspective when compared to the front-on view. The splits in the data between the recording methods are:

Side-on	45°	Front-on
Y113	Y117	Y119
Y114	Y118	Y120
Y115		
Y116		

Figure 17, Experiment recording locations

Not all recordings went flawlessly. Those files which are corrupted are given below.

Experiment	Issue	Class
nq1t10y113	Recording ended unexpectedly	5 (Restless Boredom)
zut9y113	Personnel on edge of sensor field of view	2 (Active Engagement)
inf8y115	Movement of personnel other than participant	3 (Frustration)
ipsk11y115	Chair in front of sensor	5 (Restless Boredom)
zum_y115	Chair placed in front of sensor	2 (Active Engagement)
zuty117	Movement of personnel other than participant	2 (Active Engagement)

Figure 18, Corrupt recordings

Removing the above corrupted files we now have the following distribution of data.



Figure 19, Class distribution in experiment sets without corrupt instances

We now generate the expected results from base predictions made both randomly and predicted every instance as being from a single class. This prediction gives us a starting point from which we can assess our accuracy. As specified previously we have two methods of evaluating accuracies due to the non-uniform weighting of data within the classes: Normal and modified. We use the frequency of the class in the dataset as its probability of being generated when generating values from the class distribution. The following values are produced from over 200 random permutations of the entire dataset and demonstrate well why we introduce our modified accuracy value which normalises the results regardless of distribution.

	Normal accuracy	Modified accuracy
Random	16.7%	16.8%
Random from class distribution	20.5%	16.7%
All class 1	25.7%	16.7%
All class 2	27.1%	16.7%
All class 3	19.9%	16.6%
All class 4	11.4%	16.7%
All class 5	8.6%	16.6%
All class 6	7.2%	16.7%

Figure 20, Basic accuracy values

10.3 Experiment Classes

Class key					
1	Rapt Engagement				
2	Active Engagement				
3	Frustration				
4	Lethargic Boredom				
5	Restless Boredom				
6	Neutral				

Experiment	Class	zum_y115	2	nq1t4y118	1
3gqm6y113	1	zut6y115	2	ok1_y118	1
bsr3y113	2	3gqt11y116	1	zum9y118	2
crh2y113	4	bsr7y116	4	zut4y118	2
dgj7y113	1	crh2y116	4	3gqt6y119	2
gq2t11y113	3	dgj3y116	3	bsr11y119	3
inf4y113	4	gq2m6y116	6	crh2y119	2
ipsk8y113	3	inf10y116	1	dgj9y119	1
nq1t10y113	5	ipsk9y116	6	gq2m5y119	2
zum5y113	1	nq1t8y116	1	inf8y119	3
zut9y113	2	ok1y116	5	ipsk4y119	6
3gqmy114	2	zum4y116	2	nq1t3y119	4
bsry114	3	zut3y116	2	zum10y119	2
crhy114	4	3gqmy117	1	zut7y119	1
dgjy114	1	bsr9y117	3	3gqt4y120	2
gq2t11y113	2	crhy117	4	bsr9y120	3
infy114	4	dgj5y117	3	crh2y120	4
ipsky114	4	gq2t10y117	1	dgj10y120	1
nq1ty114	4	inf11y117	5	gq2m7y120	2
oky114	1	ipsk8y117	6	inf6y120	1
zummy114	2	nq1t6y117	1	ipsk3y120	5
zuty114	2	ok_y117	1	nq1t5y120	3
3gqm7y115	2	zum7y117	2	ok1y120	2
bsr5y115	3	zuty117	2	zum11y120	2
crh_y115	5	3gqt8y118	3	zut8y120	2
dgj_y115	2	bsr11y118	3		
gq2t10y115	2	crh2y118	4		
inf8ay115	3	dgj7y118	3		
ipsk11y115	5	gq2m3y118	1		
nq1t9y115	5	inf10y118	5		
ok_y115	1	ipsk6y118	6		

Figure 21, Experiment class distribution in dataset

10.4 Data Capture

Data capture was done in combination with Tom Ranji's experiments on recording user responses to stimuli which cause objective engagement types. The data capture was performed using a depth sensor; this is to reduce the variability that is produced by luminance changes (compared to traditional light cameras) because the depth sensing camera works in the infrared spectrum with a Time of Flight depth estimator.

Participants are seated in a standard chair with arms facing a display situated on a desk in front of them. The user interacts with the machine using either a standard keyboard and mouse or trackball depending on the stimulus. The subject has a series of reflective tracking markers placed on them at set locations each of which represents the physical location of the body part they are on. The depth sensor is placed depending on the experiment facing the user; it is connected to a laptop computer running recording software. The subjects are then subjected to a series of predetermined stimuli designed to make them exhibit a given response. Each subject's session lasts approximately 90 minutes with each tests duration varying based on its requirements.

The depth sensor used is a commercially available Microsoft Kinect Sensor for Xbox One; the sensor has a depth based resolution of 512 x 424 with 12 bits per pixel representing the depth to that point in mm from the focal plane. The sensor required a windows machine with the Kinect SDK V.2 installed, early testing revealed that recording using the Kinect Studio application provided with the SDK was not appropriate for the task at hand due to the file sizes produced by recorded data. A 90-minute recording that should only contain depth data would come out at 130GB, this was unusable and as such a lightweight recording platform was written in Python by the author.

The recording software is a simple looping program which: checks if there is a new frame from the Kinect, records the time of that frame since midnight in milliseconds, stores time and frame in a predefined HDF5 file, waits a time appropriate for the frame rate to be kept periodic, repeats until either no more frames arrive or the user ends the program. Alongside this program, another Python script was written which allowed for a visualisation of the recorded data to be performed proving the data could be retrieved. With this new Python program, a 90-minute recording has been brought down to 3GB without using chunk compression.

When connected, the data stream produced by the Kinect sensor can be unstable nearing the edges of objects and the borders of the footage (Fig.2) where zeroes are recorded. As such we implement a noise reduction step which is applied to raw footage before analysis. This issue with pixel reliability is well documented by Yang et al. (2015). Their account gives a full and detailed account into the pixel interference and its distribution throughout the frame space.



Figure 22, Example of Kinect depth data, Contents is room and subject is author

We trim and crop the footage sequence before analysis to reduce the chance of interference from movements unrelated to the experiment and reduce the amount of data required to be analysed and stored. Stimuli duration is two minutes with an added minute of white noise at the start and several seconds at the end prior to the footage being stopped. Thus, we trim the footage to 1 minute 40 seconds' duration; the starting time is based on the end of the footage which is much more reliable than the start. We don't trim to the start of the stimuli exposure as Harry Witchel indicates that there is a period in which the Cognitive Affective State is in fluctuation after switching experiences. This gives us the following decomposition of the recordings.

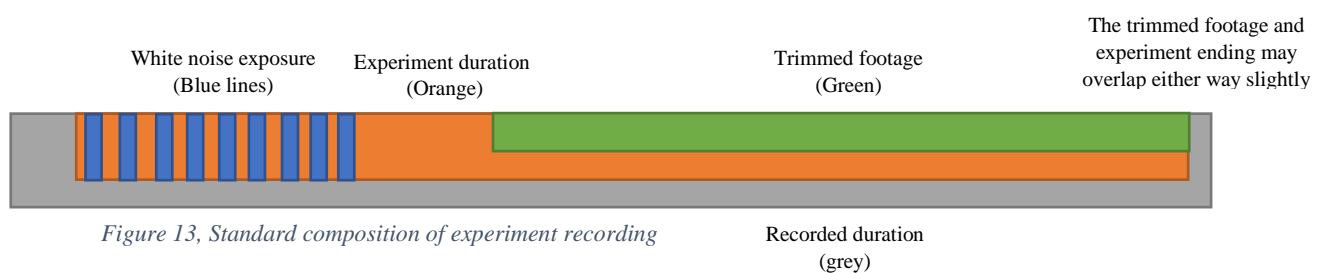


Figure 13, Standard composition of experiment recording

10.5 Parameter Evaluation

In this section, we evaluate and state the parameters used within the programs execution. There are many parameters in each of the processing steps which affect their function and thus the overall effectiveness of the system. It is understood that parameters should be evaluated in relation to how they affect classification performance and its generalisation to new data.

10.5.1 Noise Reduction

The window size value depends on the footage quality being analysed. All experiments using the same participant have the same general noise level as the sensor and scene do not change between experiments excluding the participant who is in roughly the same position each time. Thus, we set a single value for the window size over all experiments with a single participant. The value at which the parameter is set does not affect the other components of the footage evaluation and is simply there to remove most zero-values.

Generally, the window size used is ~ 8 . This breaks down slightly in some of the experiment sets, such as our wall rotation test sequence where there are large areas of zero values with locations of intermittent reliability. This is shown below.

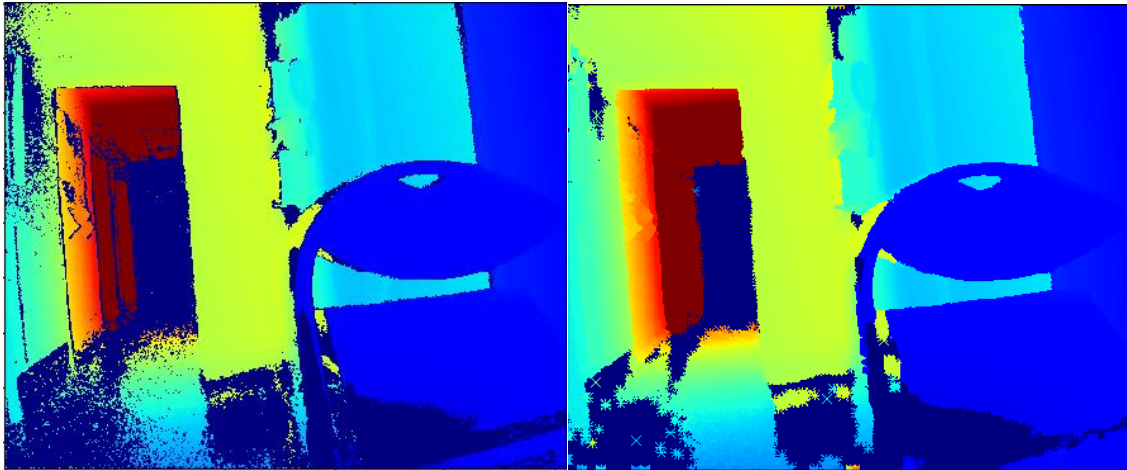


Figure 24, Breakdown of noise reduction when large areas of zero values occur

10.5.2 Background Estimation

Background estimation has a few highly tuneable parameters providing us with functionality to alter many parts of the effectiveness and how the system reacts to various states and changes within the data. These parameters are as follows:

1. Initial estimate analysis duration
2. Standard deviations from the mean for anomalous data
3. Initial estimate mean learning rate
4. Initial estimate standard deviation learning rate
5. Mean learning rate
6. Standard deviation learning rate
7. Regularisation amount

The parameters used in the initial mean and standard deviation estimates are simple to give. We set the duration to be short, 50 frame (2.5 seconds), to reduce the chance that there is movement contained within the sequence and long enough such that most data will be well represented by the statistical model. The learning rates are then set accordingly to:

- Mean: 0.04 (1/25)
- SD: 0.05 (1/20)

The mean learning rate is set slightly higher than the standard deviation learning rate as its values are initialised to have the values of the first frame. This value means the system requires 25 frames with a different value to the current average to shift the mean by half. The standard deviation is set at 0.05 as we are not dividing by the number of frames already analysed, by setting to this value we get the effect of each new value influencing the total as though it were the 20th frame being analysed. This is effective as we initialise the standard deviation values all to be 0. After the initial estimates, we use the method described in the implementation section. Each of the following variables alter a part of the learning rate. We set them at:

- Mean: 0.05 (1/25)
- SD: 0.02 (1/50)
- Regularisation: 0.001 (1/100)

Changing the mean learning rate affects how fast anomalous depth values are learnt. The learning curve is similar to a Brachistochrone where the convergence duration takes the same amount of time regardless of the distance.

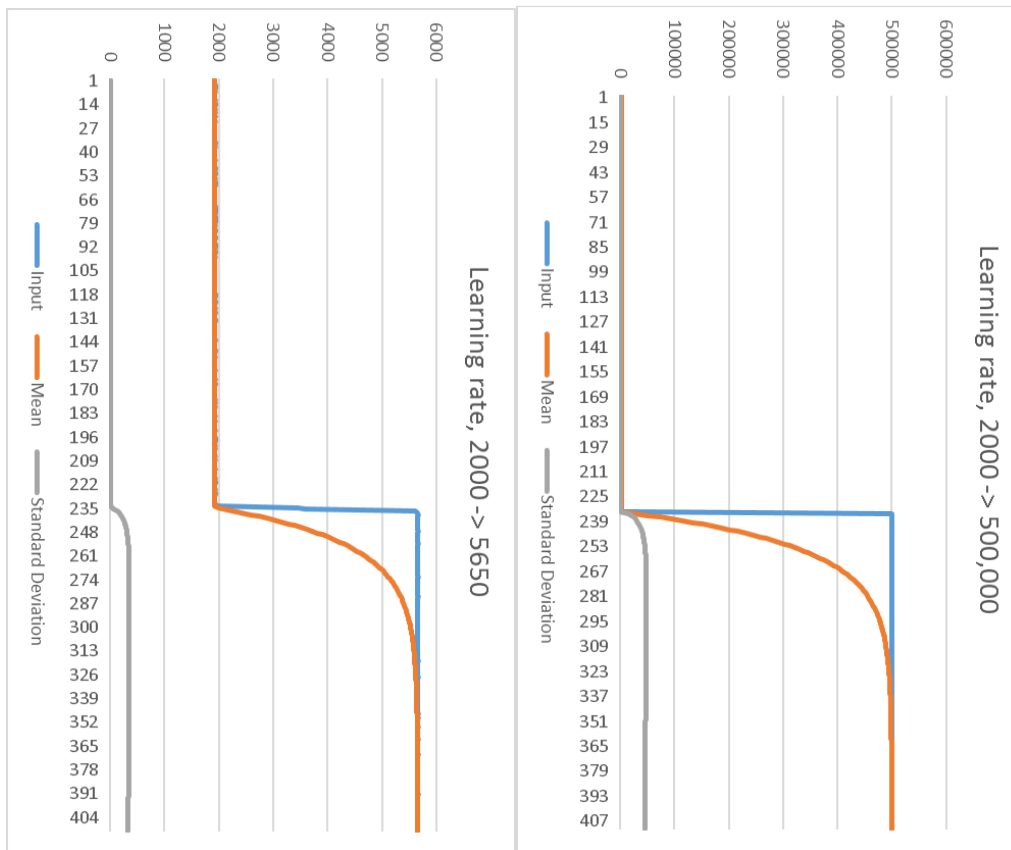


Figure 25, Constant time to learn any changes

Changing the standard deviation learning rate affects both how fast the system learns noise and how quickly the system becomes desensitised to repeated movements. This can have dramatic effects on the application as having it set too high means the system will not recognise movements in areas where many movements have been made.

10.5.3 Optical Flow

The parameters which require selection and evaluation are:

- Pyramid Scale (Resolution reduction between each level)
- Number of Pyramid Levels
- Window Size (Increases chance of detecting large movements, blurs motion field helping slightly with discontinuous nature)
- Iterations per level (Each iteration uses the previous iterations estimate as priori)
- Polynomial Expansion size (Size of pixel neighbourhood used when estimating with Taylor series)
- Polynomial sigma (standard deviation of Gaussian used to smooth derivatives for use in polynomial expansion)

The resolution of the footage is already quite low so we use a small pyramid with parameters as follows.

- Pyramid scale 0.5
- Levels 2

The window size is a balance between reducing the discontinuous nature of the flow field and ensuring that it does not become too large such that details are lost.

- Window size 15

The number of iterations does not seem to have any impact on effectiveness.

- Iterations 1

Polynomial Expansion is like window size but polynomial expansion affects the underlying polynomial estimation of the neighbouring feature space.

- Expansion 20
- Sigma 1.8

10.5.4 DBSCAN

Density Based Spatial Clustering for Applications with Noise has a few parameters which require being defined.

When determining epsilon, we need to consider the various feature components which will be evaluated against it. These features must also be normalised such that they contribute the same amount to the task, unless it is decided that a feature is more important than its peers at which point it will be weighted more highly requiring a consensus within the others to outweigh it. The parameters are:

- Minimum cluster size (in pixels)
- Epsilon
- Neighbourhood window size
- Minimum flow magnitude
- Minimum flow field density

Determining the minimum cluster size is a form of thresholding and is somewhat linked to the window size and polynomial expansion in optical flow. It is a means of attempting to reduce the number of unreliable clusters by ignoring those which are unlikely to provide reliable and consistent data.

- Minimum size 250

Epsilon is simple to define. When determining similarity, the true way it should be done is by taking their relative spatial locations, magnitude and angle, and giving a difference value based on these. In practice this is not very effective and in some cases the most effective method is to simply cluster those vectors whose magnitude is above some threshold. Our similarity measure uses only the difference in angle between the two vectors to identify similar clusters.

- Epsilon $\pi/2$

With this value, we get the effect that we can cluster motion which is similar within the high-density space. If we required that all vectors had the same direction, then an expanding oval would not be clustered but with our measure we get the following clustering.

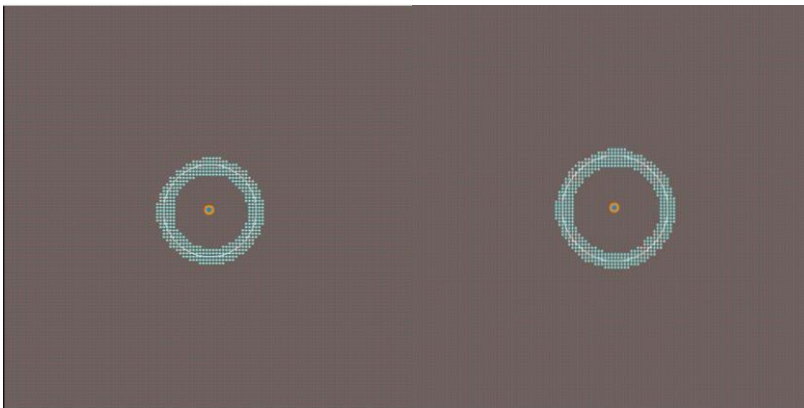


Figure 26, Clustering on circle with increasing radius

Window size gives a way to overcome the discontinuous nature of our flow fields. By extending the window we can overcome areas of zero magnitude and find continued flow. Although increasing the window size increases run time substantially.

- Window size 4

A simple thresholding to prevent searching of vectors in areas with no movement.

- Minimum magnitude 0.4

Minimum density, number of local vectors which must be found for the method to identify them as relevant. Outlier points like this should not exist but the method provides the functionality so we make use of it for those boundary cases.

- Minimum density 4

10.5.5 k-NN and Classification

k-NN is a simple algorithm and its parameters are similarly humble. We also include the parameters for classification within this section as they are closely related. The parameters required are as follows:

- Cross-validation folds
- K-neighbours
- Distance function

We use 3-fold cross validation when evaluating the accuracy. Due to the distribution of classes and how effectively instances can be classified depending on the spread we perform the classification many times, ~200, randomising the order of elements each time and giving the average accuracy value produced.

K's value is chosen by evaluating the generalisation at each value, therefore, it can change depending on the measure being used.

The distance function is a very changeable method which generally relies on finding the Euclidean distance between the training testing instance pair over the normalised feature set. Although, we implement a novel distance measure for identifying similarities between sets of distinct movements within each experiment.

10.6 Documents pertinent to participants in experiments

10.6.1 Experiment advert

Title: Audiovisual Engagement: Addition of Inertial Sensors

Subject: Psychology Experimental Volunteers: Emotional Responses to Movies & Games

We are looking for participants who are over the age of 18 and in good health to take part in an exciting psychology experiment. You will be paid £15 for attending one 90 minute session. Location: University of Sussex Campus (we can pay modest rail transport fares) Time: Thursdays, some time between 9:30 AM and 5 PM, to be arranged to fit in with your schedule.

The experiment will involve rating your emotional experiences to different stimuli. The stimuli will include playing computer games and watching non-threatening videos whilst having sensors attached to your wrists and ankles and being filmed. Please wear trousers/shorts for the session (no dresses or skirts), and your ears must be visible during the experiment (we will ask you to tie back long hair). We prefer people to wear solid (unpatterned) clothing, which is easier for the computer to recognise.

If you are interested, please contact us with your number and the dates you are available.

Our contact details:

Email: psychology-experiments@bsms.ac.uk

Phone: 01273 697 917

Advert jK9P3W - 2016-10-01

10.6.2 BFI-10 Personality Measure big five

BFI-10 Personality Measure

Instruction:

How well do the following statements describe your personality?

I see myself as someone who ...	Disagree strongly	Disagree a little	Neither agree nor disagree	Agree a little	Agree strongly
...is reserved	(1)	(2)	(3)	(4)	(5)
...is generally trusting	(1)	(2)	(3)	(4)	(5)
...tends to be lazy	(1)	(2)	(3)	(4)	(5)
...is relaxed, handles stress well	(1)	(2)	(3)	(4)	(5)
...has few artistic interests	(1)	(2)	(3)	(4)	(5)
...is outgoing, sociable	(1)	(2)	(3)	(4)	(5)
...tends to find fault with others	(1)	(2)	(3)	(4)	(5)
...does a thorough job	(1)	(2)	(3)	(4)	(5)
...gets nervous easily	(1)	(2)	(3)	(4)	(5)
...has an active imagination	(1)	(2)	(3)	(4)	(5)
...is considerate and kind to almost everyone	(1)	(2)	(3)	(4)	(5)

Big five Inventory 10

For the researcher:

Full description of BFI-10 is available at

[Barron et al., B., & John, O. P. \(2007\). Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German. Journal of Research in Personality, 41\(1\), 203-212.](#)

Volunteer Number _____ Date _____

10.6.3 Demographics questionnaire

Demographic Questionnaire

Name: _____

Subject #: _____

Date: _____

Demographic Information

Date of Birth (DD / MM / YYYY) _____ / _____ / _____ Age _____ Gender _____

Occupation _____

Height (in cm if possible) _____

Left-handed or right-handed (which hand do you use with a computer mouse)? _____

Highest Level of education completed or currently registered in? (circle one: grade school, some GCSE/ O Levels, A Levels, trade school, some university, BS/BA, some graduate school, Masters Degree, Doctorate, Professional Degree, other) If other, explain: _____

What area of Study? _____

Are you a native English speaker?

Yes No

If no, at what age did you begin formal education in English? _____

On average, how many hour(s) of TV do you watch per week? _____

On average, how many hour(s) of video games do you play per week? _____
















On average, how many hour(s) do you spend surfing the web per week? _____

10.6.4 Engagement questionnaire

Participant number _____ Date _____ Film _____ P43
Film number _____

While you were watching/experiencing the previous stimulus, what did you feel?

For each of the three rows of images below, fill in the circle that represents how you felt during the preceding stimulus.

				
INDEPENDENT	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	DEPENDENT
				
SAD	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	CHEERFUL
				
QUIET	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	ACTIVE

For each of the following questions, circle on the line below it how much you agree with the statement. You may instead write a number between 0 and 100 just after the question.

DS4. I felt motivated _____

Not at all 0 10 20 30 40 50 60 70 80 90 100 Extremely

TH3. I felt interest _____

Not at all 0 10 20 30 40 50 60 70 80 90 100 Extremely

WA6. I felt it was challenging for me _____

Not at all 0 10 20 30 40 50 60 70 80 90 100 Extremely

AD1. I felt anxiety _____

Not at all 0 10 20 30 40 50 60 70 80 90 100 Extremely

Participant number _____ Date _____ Film _____ P43
Film number _____

QC9. I felt restlessness _____

Not at all 0 10 20 30 40 50 60 70 80 90 100 Extremely

Z9W. I cared about it _____

Not at all 0 10 20 30 40 50 60 70 80 90 100 Extremely

G5V. I wanted to see more _____

Not at all 0 10 20 30 40 50 60 70 80 90 100 Extremely

RF3. I felt lethargy _____

Not at all 0 10 20 30 40 50 60 70 80 90 100 Extremely

G6K. I felt boredom _____

Not at all 0 10 20 30 40 50 60 70 80 90 100 Extremely

X2K. I was engaged by the experience _____

Not at all 0 10 20 30 40 50 60 70 80 90 100 Extremely

HE9. I felt apathy or detachment _____

Not at all 0 10 20 30 40 50 60 70 80 90 100 Extremely

PW8. I felt frustration _____

Not at all 0 10 20 30 40 50 60 70 80 90 100 Extremely

KJ7. I wanted it to end earlier _____

Not at all 0 10 20 30 40 50 60 70 80 90 100 Extremely

BV8. I felt curiosity _____

Not at all 0 10 20 30 40 50 60 70 80 90 100 Extremely

JK9. Circle one of the following that is the closest description of what you felt:

boredom, interest, confusion, frustration, curiosity, anxiety, neutral, another emotion quite strongly _____

10.6.5 Participant Information Sheet



Dear Participant

Audiovisual Engagement: Addition of Inertial Sensors

You are being invited to take part in a research study. Before you decide it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully. Talk to others about the study if you wish.

- Part 1 tells you the purpose of this study and what will happen to you if you take part.
- Part 2 gives you more detailed information about the conduct of the study.

Ask us if there is anything that is not clear or if you would like more information. Take time to decide whether or not you wish to take part.

1. What is the purpose of the study?

The purpose of this study is to measure how people react to different stimuli such as movies, videogames and picture montages. The measurements will be made using a variety of equipment including wearable inertial sensors, cameras and subjective questionnaires that measure how you reacted to each stimulus.

2. Do I have to take part?

No. It is up to you to decide whether or not to take part. If you do, you will be given this information sheet to keep and be asked to sign a consent form. You are still free to withdraw at any time and without giving a reason.

3. What will happen to me if I take part?

The participants in this experiment will watch a series of short movies or play video games. Each stimulus will last about 3 minutes. These stimuli may be excerpts of Hollywood movies or television programmes, video games, sets of photographs, or movies constructed by the experimental team. These movies are meant to elicit emotions (as per watching typical movies and television), but the movies should NOT cause you any psychological stress or distress, and to further guarantee that no such issues should arise, you shall be free to withdraw from the study at any time.

At the beginning of the experiment, we will attach inertial sensors to your wrists and ankles in order to measure subtle movements.

We will also record your movements using a video camera. As with all electrical measurements made on people (e.g. in a hospital), these electrical measurements are absolutely safe, because they involve a safety feature called

Audiovisual Engagement: Addition of Inertial Sensors
2016-09-08
Version 4.01

optical isolation, which guarantees the safety of the participants by using light (rather than electricity) as the signal being conducted.

The measurements that the scientific team are performing will involve filming the participants as they watch the movie. The films of the participants will be purely for characterisation of the participant's responses, and will NEVER be shown to individuals other than the experimenters analyzing the responses (see below).

4. What do I have to do?

During the experiment, you will watch various stimuli or play video games on a computer in front of you, and afterward rate each stimulus in terms of how it made you feel while viewing it. None of the stimuli are offensive or upsetting. At the end of each stimulus, a short questionnaire will be presented. This will ask questions about the emotions you felt during the film; you may choose not to answer these questions.

5. What are the side effects of any treatment received when taking part?

There are no anticipated side effects or risks.

6. Will my taking part in the study be kept confidential?

Yes. All the information about your participation in this study will be kept confidential. The details are included in Part 2.

7. Contact Details:

Harry Witchel - psychology-experiments@bsms.ac.uk

Part 2

1. Will my taking part in this study be kept confidential?

All paper information that associates your name with you participant number will be stored under lock and key. Computers with this information will require a password to access, and this password will only be available to the research team.

Information will be anonymized when the participant begins.

2. What will happen to the results of the research study?

The results of the research study will be written and up and published in a scientific journal. The information from individual volunteers will be anonymized, and your name or anything that might indicate who you are or your participation will not be included.

3. Who is organising and funding the research?

Audiovisual Engagement: Addition of Inertial Sensors
2016-09-08
Version 4.01

The research is funded by Brighton and Sussex Medical School's IRP programme.

Thank you for taking the time to read this information sheet.

Audiovisual Engagement: Addition of Inertial Sensors
2016-09-08
Version 4.01

10.6.6 Participant Consent form

Participant ID Number _____

Participant Consent Form

As stated in the participant information sheet, the main aspects of standard ethical safeguards used in this study are as follows:

- ☐ The main procedures are as announced at the beginning.
- ☐ Your participation is totally voluntary; you do not have to take part.
- ☐ You may withdraw from the research at any time and for any reason.
- ☐ With questionnaires you have the option to refuse to answer any question you do not want to answer.
- ☐ Your data will be treated with full confidentiality and, if published, it will not be identifiable as yours.
- ☐ We also offer you the opportunity to find out more about the study and its results, after your participation is complete.

Your Consent

I have read and understood the text above, about the nature of the experiment, and the safeguards in place for me. I consent to participate in this experiment:

Name _____ Phone _____ Email _____

Signature _____

10.7 Research Governance and Ethics Committee Application Form

<p>Carina Westling, c.e.i.westling@sussex.ac.uk Carina Westling is a mature PhD student at University of Sussex in the School of Media, Film and Music. Her PhD is on immersive digital art, web experiences, and immersive performances, and she is interested in metrics of engagement</p>			
Does this project require NRES approval?		Yes	No X
Proposed start date: 08/10/2016		Proposed completion date: ongoing (or 30/10/2018)	
Please indicate in the box below whether the request is for:			
Sponsorship (for studies applying to NRES)		Sponsorship & Ethical Approval (for studies by BSMS/University staff/students utilising BSMS premises, staff or students)	
Governance (for projects proposed by external staff previously reviewed by an Ethics Committee)		Governance and Ethical Approval (for studies proposed by BSMS staff/students taking place overseas)	
<p>Please provide a summary of the project written in language accessible for a lay audience (approximately 350 words)</p> <p>This study hopes to measure subjective responses to a variety of stimuli that are meant to controllably increase or decrease engagement. We will be utilising very engaging stimuli, very boring stimuli, and a range of medium engaging stimuli that should predictably elicit incremental increases in engagement. We will also be comparing the measured engagement between stimuli that are interactive, passive with narrative, and passive without narrative. The participants will be asked to watch a variety of video montages and/or play a variety of computer games. Overall, the study is non-invasive. All stimuli are from a computer and will be experienced like movies or video games. None of the stimuli are worrying, offensive, or potentially upsetting – the stimuli are almost invariably of a positive subject matter. After each 3 minute stimulus, the participant is asked a series of questions (lasting ~3 minutes) concerning their emotional response to the stimulus. During each stimulus, participants will be video taped (to measure postural micromovements) and be recorded with a depth camera technology (Kinect), and will be wearing inertial sensors on the wrists and ankles. We hope to compare our questionnaire results to a participant's body movement. The total number of stimuli will be limited such that participants spend less than 1.5 hours. Most of the measurements would be subjective VAS measurements. An entry questionnaire about liking video games and recording their age, gender will also be included. They will also be given an initial questionnaire self-assessing their personality (the 11-question BFI-10, lasting one minute [Ramnstedt, B., & John, O. P. (2007). Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German. Journal of Research in Personality, 41(1), 203-212.]).</p>			

-2-
BSMS RGECC Application Form



Research Governance and Ethics Committee (RGECC) Application Form

Section A – to be completed for ALL projects

Title of Project: Audiovisual Engagement: Addition of Inertial Sensors			
Is the project a: (please highlight or tick box)	PHD/MSc/PhD/Phil study	BSc/BAMSc/MSc study	
	UG student project	Module 404 Individual Research Project (IRP)	X
	Staff Research	Other	
For student research projects please specify below the name of the course the student is enrolled on:			
IRP Module 404			
Name of Principal Investigator / Supervisor: Harry Witchel			
School/Division: BSMS			
Contact Details – Email: h.witchel@bsms.ac.uk Telephone: 01273 873 549			
Names of all Researchers/Students:			
Collaborators: Dr. Luc Berthouze, Reader in Engineering, University of Sussex, L.Berthouze@sussex.ac.uk Prof Nachi Chockalingam, Professor of Clinical Biomechanics, University of Staffordshire, n.chockalingam@staffordshire.ac.uk			
Although students change from year to year, this project represents ongoing research. The RGECC committee will be informed by letter of new staff when they enter the project			
For 2016-17 IRP: Thomas Ranji, BSMS Year 4 Student in IRP			
Contact Details – Email: t.ranji1@uni.bsms.ac.uk			
Oscar Trott - ojt21@sussex.ac.uk			
For 2016-17: Oscar is a year 3 Bachelor's student in Computer Science at Univ Sussex on course for a first. He is doing a year 3 final project and is interested in nonverbal behaviour and in the Kinect technology. He will not be meeting volunteers, but will be involved in seeing data and analysis.			

-1-
BSMS RGECC Application Form

Section A continued

Risk Assessment (Please tick or highlight the appropriate boxes)		
Will the study involve:		
Causing participants physical damage, harm or more than minimal pain	Yes	No
Manual handling of participants, vigorous physical exercise, or physical activity from which there is a likelihood of accidents occurring?	Yes	No
Physiological interventions or procedures outside of standard practice - These might include the administration of drugs or other substances; taking bodily samples or human tissue (e.g. blood, saliva, biopsy or urine) from participants; use of probes or other equipment to measure or monitor bodily performance	Yes	No
Psychological interventions or procedures outside of standard practice - These might include techniques such as hypnotherapy, psychometric testing	Yes	No
Exposure of participants to hazardous or toxic materials, such as radioactive materials	Yes	No
Inducing psychological stress, anxiety or humiliation	Yes	No
Questioning of participants regarding sensitive topics, such as beliefs, painful reflections or traumas, experience of violence or abuse, illness, sexual behaviour, illegal or political behaviour, or their gender or ethnic status	Yes	No
Children under 16	Yes	No
Incapacitated adults and/or people with learning disabilities or mental health problems	Yes	No
Groups where permission of a gatekeeper is normally required for access to its members, for example ethnic groups?	Yes	No
Access to records of personal or confidential information?	Yes	No
Storage and analysis of tissue samples	Yes	No
Have you considered the possibility that your investigations might uncover unexpected and possibly clinically relevant findings?	Yes	No
Any other risk not identified above	Yes	No
If you have answered 'Yes' to any of the above questions please describe the safeguards and monitoring procedure.		
The probes (wearable inertial sensors) to be used will be low-voltage battery powered sensors that are certified as safe for low voltage by European regulations. All equipment will be battery powered and optically isolated from the main electrical sources (including the computers); this means that the maximum electrical shock would be from AA batteries. All probes will be on the skin, away from orifices.		
The psychological testing will be of a minimal nature using questionnaires (or questionnaires) in the public domain, e.g. the Visual Analogue Scale (VAS) and BFI-10. All data will be anonymised by volunteer number, and data will be stored in locked rooms on password-protected		

computers.

We have considered the possibility that we may detect something clinically relevant, and this is not possible based on movement recordings or simple personality devices, because these do not have established indices to define pathology (unlike the ECG).

Section B – Project Proposal & Protocol – to be completed for ALL projects
Please complete the template below ensuring you cover all the points listed fully. Your protocol should also be included in this template.

What is the purpose of this study? Please clearly state the aims of the study or hypothesis to be tested.
This study hopes to measure subjective and postural responses to a variety of computer-based stimuli that are meant to controllably increase or decrease engagement. We also hope to compare the measured engagement between stimuli that are interactive, passive with narrative, and passive without narrative. The hypothesis is that engagement diminishes non-instrumental movements.
What is the methodology
All of our stimuli will be presented on a PC. The study includes very engaging stimuli, very boring stimuli, and a range of medium engaging stimuli that should predictably elicit incremental increases in engagement. Before any stimuli are presented, participants will fill in a BFI-10 personality questionnaire. After each stimulus is presented, participants fill in a questionnaire gauging their interest/engagement with the stimulus. The stimuli are made up of a variety of (positive and controversial) movie fragments (freely available on Youtube), video montages and a variety of computer games that are uncontroversial and suitable for all ages (Angry Birds, Zuma), and some quizzes of reading comprehension and general knowledge (eg geography). While viewing each stimulus, participants will be video taped and wired up to various skin conductors and an electrocardiogram. All equipment is PAT tested equipment from Dr. Witchel's laboratory.
What sort of participants will be involved? (i.e. how many, gender, ages)
Participants will be healthy volunteers over the age of 18. Around 60 participants will be involved. It is anticipated that we will have approximately equal numbers of males and females. Participants will be recruited from the University community.
If vulnerable groups (i.e. children, incapacitated adults) will be involved please give full details and outline steps that will be taken to protect them.
N/A
What are the inclusion/exclusion criteria?

and wearable inertial sensors.	When: Experiments will take place during normal working hours in BSMS. Monday to Friday.
	Where: The experiments will take place in the BSMS teaching building, in tutorial rooms (e.g. room 1.12) that are not in use by the medical school during the duration of the experiment. For example, on Thursdays, which are IRP days, neither year 1 nor year 2 students are usually scheduled to use any of the tutorial rooms; the curriculum office staff, who oversee the booking of these rooms, will be asked for permission as to when the rooms are available.
	What facilities will be needed and who will provide them?
	The room, tables and chairs will be provided by BSMS. Computers, stimuli, video cameras and consumables will be supplied by funds attached to the IRP or by the primary investigator; this equipment is portable (laptops, etc), and is locked away in the primary investigator's cupboard between experiments. Money for these small expenses may also be supplied by small research grants.
	How will the results be analysed and by whom?
	The results will be analysed by Harry Witchel and his team using computers; his team will consist of his IRP students, scientific collaborators, student volunteers, and other members of the university community working on this project. The Films will be analysed by motion tracking systems. Statistics will be analysed by programmes including Matlab, Stata, Excel and Graphpad Prism.
	What are the expected benefits of the research to participants or researchers?
	This study will help us prove that certain body movements correspond to engagement and boredom. In addition, we will determine if there are consistent ways of making stimuli sets that elicit predictably different levels of engagement/boredom.
	What means of dissemination will be used?
	Journal articles, invited talks. We have regularly present at European Conference on Cognitive Ergonomics, as well as at various departments in computer science.
	What arrangements will be made for giving the participants access to the results?
	While participants will be invited to ask for results when the experiments are completed (which may take as long as 3 years), participants will not have access to any results but their own until a final publication is made (at which time anonymized summary data will be available). For those participants asking for more information about what the study is about, at the end of their participation, the primary investigator (or the scientist running the experiments) will explain the specific hypotheses of the work and how the stimuli and measurements that the participant has just experienced fit into those hypotheses.
	What results/end points are to be measured/noted?
	The data from the questionnaire and an objective analysis of body movement/emotional response from the video recording.
	How will this project be funded? List all sources of funds e.g. grants, commercial sponsorship, school's funds etc.
	The funding requirements for this project are very modest: payments to the volunteers (< £200 per year) and consumables (<£200 per year). The project receives part of this funding from BSMS as an IRP. Some funding comes from the primary investigator

Exclusion: current hospitalisation, postural inability to sit still in a chair, under 18, vulnerable adults Must be fluent in spoken English	
Please state your rationale for your participant choice	Healthy participants will be recruited from the university community as stated below. It is hoped that university community members will be able to watch and appreciate simple computer video games and popular web based short videos. The University community is an ideal source of healthy volunteers, and much of the scientific literature on psychology of healthy volunteers is based on university communities.
How will participants be identified and recruited?	Participants will be recruited from the University community. Three approaches to recruitment will be pursued: 1) the Department of Psychology at the University of Sussex has a participant programme for their undergraduate students (SONA). 2) advertising by flyer, email list or by university publication; a sample advertisement is attached to this application. 3) friends of the student running the project may be asked to participate — if so, they will be provided in advance with complete written information and documentation allowing them to make a proper decision as to informed consent.
What measures will be taken to ensure confidentiality, privacy and data protection? <i>Data should be secure against unauthorised access and comply with data protection legislation. Where possible the data should be anonymised, where this is not possible confidentiality should be maintained.</i>	Information will be anonymized when the participant begins. All paper information that associates the participant with his/her subject number will be stored under lock and key. Computers with this information will require a password to access, and this password will only be available to the research team.
What is your procedure for obtaining informed consent? If it is not possible to obtain informed consent, full reasons must be given.	We will provide the participant with a clear participant information sheet before he/she starts the experiment (see attached). The information sheet will inform the participant that they are allowed to stop the experiment at anytime if he/she is uncomfortable with the study. The information sheet will also state how he/she will be compensated upon completion of the study.
What are the risks to participants or researchers, and how will these be managed?	N/A
Will participants be reimbursed for expenses or given any inducements? <i>If so, please give details.</i>	The participants will be paid £15 for their 1hour and 30 minutes of contact time.
How, where and when will the data be collected? <i>Please include a copy of any questionnaire that will be used or sample questions used in structured or semi-structured interviews.</i>	How: The data will be collected via two questionnaires: one filled out at the beginning of the experiment (after informed consent) and the other after each stimulus is presented. The questions (or questionnaires) will be from the public domain, e.g. the Visual Analogue Scale (VAS) or International Personality Item Pool. Body position and body acceleration will be measured via video tracking, depth camera tracking (Kinect).

10.8 Research Governance Approval

BSMS Research Governance & Ethics Committee (RGEC)
Chair: Professor Kevin Davies
Deputy Chair: Professor Bobbie Farsides
Secretary: Miss Caroline Brooks
Tel: 01273 872855 c.e.brooks@bsms.ac.uk
Applications and general enquiries: rgec@bsms.ac.uk



Brighton and Sussex Medical School
Medical Teaching Building
University of Sussex
Falmer
Brighton
BN1 9PX

6th October 2016

Dr Harry Witchel
Brighton and Sussex Medical School
Department of Neuroscience

Dear Dr Witchel

Full Study Title: AV Engagement: Addition of Inertial Sensors
RGEC Ref No. : 16/046/WIT

I am writing to inform you that the Brighton and Sussex Medical School Research Governance and Ethics Committee (RGEC) has now assessed your new application and granted **Research Governance Approval** to proceed with the above named project.

This letter acknowledges that you have the necessary internal regulatory approvals. *Please note this project is regarded as a separate project from project 11/165/WIT, for which amendments approved in 2013 and 2014 remain in place.*

Conditions of Approval

The approval covers the period stated in the Research Governance & Ethics Committee (RGEC) application and will be extended in line with any amendments agreed by the RGEC. Research must commence within 12 months of the issue date of this letter. Any delay beyond this may require a new review of the project resources.

Amendments

Project amendment details dated after the issue of this approval letter should be submitted to RGEC for review and formal approval. Please submit your application for an amendment to the Committee (via rgec@bsms.ac.uk) using the 'Request for an Amendment Form'.

Monitoring

The Medical School has a duty to ensure that all research is conducted in accordance with the University's Research Governance Code of Practice. In order to ensure compliance the department undertakes random audits. If your project is selected for audit you will be given 4 weeks notice to prepare all documentation for inspection.

It is your responsibility to inform me in the event of early termination of the project or if you fail to complete the work.

I wish you luck with your project.

Yours sincerely

A handwritten signature in blue ink, appearing to read 'Kevin Davies'.

Professor Kevin Davies
Chair of the BSMS Research Governance and Ethics Committee