ABSTRACT

The affective state of a player when playing a game has a significant effect on the players motivation and engagement. Recognizing players emotions during games can help game designers improve the user experience by providing sophisticated behaviors to the game characters and the system itself. This thesis presents work towards novel recognition of players emotions using posture skeleton data as input from non-intrusive interfaces. A database of samples of non-acted posture skeleton data was captured during active game playing using Microsoft Kinect sensor.

Four observers were asked to annotate the selected postures with an emotion label from a given emotion set. Based on Cohen's kappa, the agreement level of the observers was above or equal to good with overall agreement levels that outperform existing benchmarks. The data was used in a series of experiments for training the system in recognizing emotions. The results indicate that the compiled database of postures annotated with emotion labels performs considerably above chance level recognition of emotions and offers interesting research questions for improvements and future directions in the area.

Theocharis Zacharatos – University of Cyprus, 2012

AFFECTIVE RECOGNITION FOR GAMES USING BODY POSTURES

Theocharis Zacharatos

A Thesis

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Master of Science

at the

University of Cyprus

Recommended for Acceptance

by the Department of Computer Science

June, 2012

APPROVAL PAGE

Master of Science Thesis

AFFECTIVE RECOGNITION FOR GAMES USING BODY POSTURES

Presented by

Theocharis Zacharatos

Research Supervisor

Yiorgos Chrysanthou

Committee Member

Christos N.Schizas

Committee Member

Efstathios Stavrakis

University of Cyprus

June, 2012

ACKNOWLEDGEMENTS

I would like to thank my supervisor Dr. Yiorgos Chrysanthou for his significant help and guidance throughout my thesis.

A big thanks to Dr.Christos Gatzoulis for all his help through the duration of implementation and experimental part of my thesis.

In addition i would like to thank Dr. George Pallis and Dr. Chryssis Georgiou for their support.

Furthermore I would like to thank all the people that participated in the experimental part of this project , Panayiotis Charalambous, Nikolas Ladas, Dr. Efstathios Stavrakis, Panayiotis Hadjichristodoulou, Konstantinou Konstantinos, Demetriou Ioannis and Mitrousis Vasileios.

Last but not least, I would like to thank my wife Maria, for her support, love and tolerance.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Introduction

1.1 Project Rationale

The affective state of a player during a game has a significant effect on their motivation and engagement. Often, players lose their interest and stop playing a game, due to negative emotions such as frustration, anger, and so on. On the other hand, players who experience positive emotions during game playing are more likely to continue playing the specific game. Therefore, a system that recognizes players emotions during game playing can be a useful tool for designing games that can receive affective feedback from the player and respond to this through sophisticated artificial intelligence behavior to the game characters and the game itself.

1.2 Project Objectives

The main purpose of this work is to design and evaluate a system that can recognize emotions using posture data. The objectives of the project are to (a) investigate how to construct a database of postures labeled with emotions. (b) Record data from both single and multiplayer competitive games to capture the rich expressiveness of both game scenarios (c) Utilize state of the art non-intrusive interfaces such as Microsoft Kinect [32] to capture data and provide a system that can be used in todays games. This is particularly important as commercial systems are increasingly adopting such interfaces in contrast to traditional motion capture systems that are expensive, and have unrealistic space requirements to operate. Taking into consideration the limitation of movements and postures that an interface like Kinect can recognize due to its simple setup, it is interesting to see whether it performs similarly to other motion capture systems in terms of recognition accuracy.

1.3 Proposed Solution

The solution proposed contains the following steps:

1. Capture kinect's motion (video) and skeleton data from different participants (at least 5) playing kinect sports game.

2. Edit the video scenes in which emotion can be recognized by going frame by frame.

3. Annotate the captured data to 3 categories (Concentrating, Defeated or Frustrated and Triumph) with the help of two or more observers.

4. Run statistical test Cohen's Kappa to evaluate the agreement between the two observers

5. Given that the agreement is satisfactory, use one set of the data, to train the system with various classification algorithms and test the results against another set.

1.4 Thesis structure

The thesis is divided into sections. Chapter 2 contains the related theory of affective computing including emotion in games, and the explanation of how the skeleton data are captured. Then we describe the technology used in the experiments and finally the related work in the field separated by modality. In chapter 3 define the method that was used, participants selected, the training, the data collection emotion annotation and benchmark creation and finally how we make the recognition by using WEKA[31]. Next in chapter 4, we present the system installation for the live environment and system components and integration of the experiment. In chapter 5 we present the experiment implementation, the method and tips used before, during and after the experiment. Finally in chapter 6 we present the results and conclusion about the objective, project issues with proposed solutions and future work.

Chapter 2

Project Background

2.1 Related Theory

2.1.1 Affective Computing

The latest scientific findings indicate that emotions play an essential role in decision making, perception and learning, that is, they influence the very mechanisms of rational thinking. Not only too much, but too little emotion can impair decision-making. According to Rosalind Picard[34], if we want computers to be genuinely intelligent and to interact naturally with us, we must give computers the ability to recognize, understand, even to have and express emotions.

Affective computing is the study and development of systems and devices that can recognize, interpret, process, and simulate human affects.

Emotion is fundamental to human experience, influencing cognition, perception, and everyday tasks such as learning, communication, and even rational decision-making. However, technologists have largely ignored emotion and created an often-frustrating experience for people. Affective Computing research combines engineering and computer science with psychology, cognitive science, neuroscience, sociology, education, psychophysiology, value-centered design, ethics, and more.

The actual seat of emotion in human brain is not unambiguous. Paul MacLean has modeled the human brain as three regions: neocortex, limbic system(s), and reptilian brain [35]. The neocortex is traditionally the best studied, and contains the visual cortex and auditory cortex; it is where the majority of perceptional processing has been assumed to occur. The limbic system is the primary seat of emotion, attention, and memory.

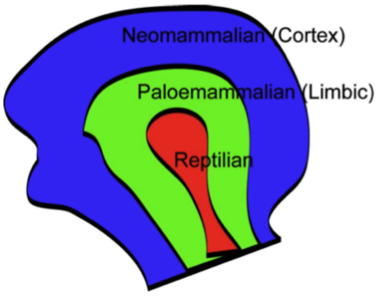


Figure 1: Paul MacLeans triune brain [35].

Emotions can hijack our rationality in situations of fear, panic, or love. They can strongly affect our decision-making even in normal situations, someone might be called very emotional person.

Evidence indicates that laws and rules do not operate without emotion in two highly cognitive tasks: decision making and perception. Detecting emotional information begins with passive sensors that capture data about the user's physical state or behavior without interpreting the input. The data gathered is analogous to the cues humans use to perceive emotions in others. For example, a video camera might capture facial expressions, body posture and gestures, while a microphone might capture speech. Other sensors detect emotional cues by directly measuring physiological data, such as skin temperature and galvanic resistance.

Recognizing emotional information requires the extraction of meaningful patterns from the gathered data. This is done using machine learning techniques that process different modalities speech recognition, natural language processing, or facial expression detection, and produce either labels (i.e. frustrated) or coordinates in a valence-arousal space.

2.1.2 Emotion in games and adaptive gameplay

Adaptive Gameplay[36] is an upcoming research area with the goal of making it possible for videogames to automatically adapt themselves to the preferences of the individual player. Adaptive Gameplay focuses on the possibilities to adapt, based on the personality of each player.

Over the years video gaming market has been growing explosively. The developers of a game create the game with a specific goal in mind. While for most games this goal is simply to present the players with a pleasurable and enjoyable experience, there are also games that are designed to give players an intellectual challenge or games that try to simulate realistic situations. The game offers the same gameplay experience for everyone and for that reason Adaptive Gameplay offers a tool to make videogames dynamic and able to adapt to the individual player.

The possibilities of adapting gameplay are wide. Almost every detail of a game can be changed based on a huge amount of information that can be gathered from the player. The

Adaptive Gameplay aims on changing the actual gameplay content, making it possible to serve the player with situations he prefers playing. Changing the affective side of a videogame can ensure a more complete, immersive, and engaging gameplay experience and can also be a very important tool for the future of Games.

Affective Gaming aims at augmenting the emotions experienced in the game and can be used in any situation imaginable where an interaction takes place. These are, however, changes that do not really influence the course or content of the game, but rather the overall feeling that it gives. Research is based around the emotions and the way they can be delivered through a virtual game, and focuses on the preferences of people with different personality traits.

2.1.2.1 Personality and Games

Whilst there is almost no research indicating that a player's personality will increase his entertainment, there is a link between personality and the environment the person is in[10]. Games take place in a virtual world. Players take on a virtual role and can do anything in this virtual world without having consequences on their actions. The personality of someone in a game, differs from their personality in the real world.

2.1.2.2 The future of Adaptive Gameplay

The more knowledge about players and respective personalities is essential to the improvement of Adaptive Gameplay [11]. Also the lack of a standardized measurement for entertainment should be addressed. The lack of such an instrument makes it very difficult to assess the functionality of adaptations.

- When the gameplay should be changed

  a) More information about the state of the player could help out.

  b) Gathering external, real-time information can be done in various ways.

- Revolution of large online communities

2.1.2.3 Game Adaptivity Impact on Affective Physical Interaction

Affective gaming has recently attracted significant attention within the affective computing community. Emotion recognition has been seen when monitoring children and adults playing. Some scientists over the years examined correlations between physiological signals, galvanic skin response, jay electromyography (EMG), respiration and cardiovascular measures and reported adult user experiences in computer games [36].

A sample game that used for experiments, is called Bug-Smasher[36]. The game is developed on a 6x6 square playground [12], comprising 36 tiles each incorporating processing power, communication, input (force pressure sensor) and output (light) as in Figure 2. During the game, different colored lights appear sequentially on the game surface and disappear again after a short period of time as a tiles light turns on and off respectively. The light position is picked randomly according to a predefined level of spatial diversity, measured by the entropy of the visited tiles and the goal is to smash as many lights as possible. Some characteristics of experiment are: normal weight participants, Heart rate, blood volume pulse and skin conductance recorded in real time.

Results of the experiment demonstrate the robustness and generality of the model in predicting reported entertainment preferences, even in adaptive physical interactive

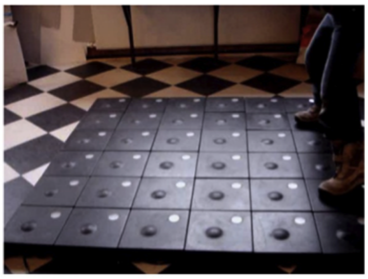


Figure 2: Snapshot at the 5th second of the game

games. Also it shows that adjustments of internal game controls (challenge, curiosity) have an impact on the performance of the adaptation mechanism and furthermore on expressed preferences of children.

2.1.3 Motion Capture Systems

Motion capture is a system that records movement of one or more objects or persons. It is used in military, entertainment, sports, and medical applications, and for validation of computer vision and robotics. In filmmaking, and games, it refers to recording actions of human actors, and using that information to animate digital character models in 2D or 3D computer animation. A performer wears markers near each joint to identify the motion by the positions or angles between the markers.

Optical motion capture systems utilize data captured from image sensors to triangulate the 3D position of a subject between one or more cameras calibrated to provide overlapping projections. Data acquisition is traditionally implemented using special markers attached to an actor; however, more recent systems are able to generate accurate data by tracking surface features identified dynamically for each particular subject. Tracking a large number of performers or expanding the capture area is accomplished by the addition of more cameras. These systems produce data with 3 degrees of freedom for each marker, and rotational information must be inferred from the relative orientation of three or more markers; for instance shoulder, elbow and wrist markers providing the angle of the elbow.

2.1.4 Body skeleton data

Skeleton tracking can be processed by depth image data, to establish the positions of various skeleton joints on a human form. For example, skeleton tracking determines where a user's head, hands, and center of mass are. Skeleton tracking provides X, Y, and Z values for each of these skeleton points as in Figure 3. Skeleton tracking systems analyze depth images employing complicated algorithms that use matrix transforms, machine learning, and other means to calculate skeleton points.

The skeleton data can be captured as a set of joints and the coordinate system is a full 3D system with values in meters. There are operations for converting between Skeleton Space and Depth Image space. The SDK supports up to two players (skeletons) being tracked at the same time. A player index is inserted into the lower 3 bits of the depth data so that you can tell which depth pixels belong to which player.

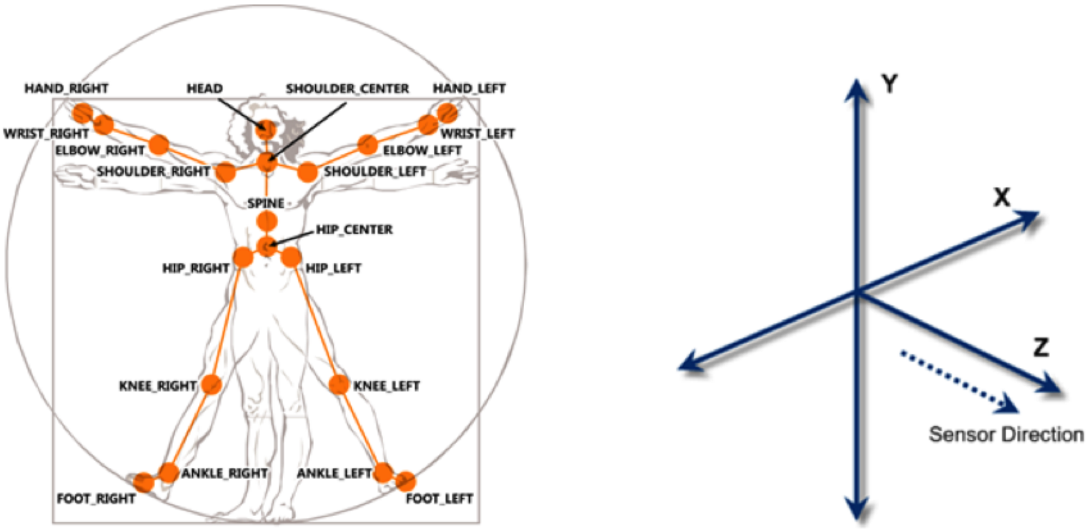


Figure 3: Joints of the skeleton data and coordinate system

2.2 Project Technology

2.2.1 Microsoft Kinect Hardware

Figure 4 displays the Kinect sensor[32]. It is a horizontal unit with a motorized pivot at the base. The device contains an RGB camera, a multi-array microphone and a depth sensor. Using these systems the Kinect is capable of full body 3D motion capture, face and voice recognition.

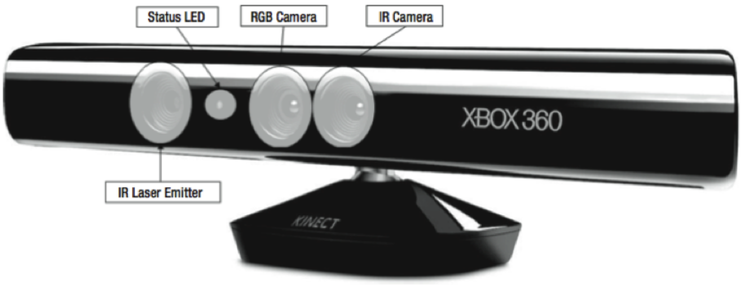


Figure 4: The Kinect Device

2.2.1.1 Kinect Camera

The Kinect camera is an RGB device with a resolution of 640x480 and a 32 bit color range. Working at a rate of 30 frames per second this camera, with the aid of software can recognize human gestures, facial expressions and 20 human joint movements per player. The response time of the camera is 33ms. A maximum of 6 players can be tracked, but only two players are monitored for skeleton capture.

The Kinect sensor is limited as to how far it can see and has a working range of between 1.2 and 3.5 meters. The horizontal field of view is 57° wide, which means at its maximum range it will be able to scan a scene 3.8 meters wide. The sensor has a vertical field of view of 43°. This field of view is enhanced by the vertical pivot system that allows the sensor to be tilted up or down by as much as 27° in either direction.

2.2.1.2 Kinect depth sensor

The Kinect consists of two parts: the IR laser emitter and the IR camera. The IR laser emitter creates a known noisy pattern of structured IR light at 830 nm.

The depth sensing works on a principle of structured light. Theres a known pseudo-random pattern of dots being pushed out from the camera. These dots are recorded by the IR camera and then compared to the known pattern. Any disturbance are known to be variations in the surface and can be detected as closer or further away. This approach creates three problems, all derived from a central requirement: light matters.

- The wavelength must be constant
- Ambient light can cause issues

- Distance is limited by the emitter strength

The wavelength consistency is mostly handled by the Kinect itself. Within the sensor, there is a small peltier heater/cooler that keeps the laser diode at a constant temperature. This ensures that the output wavelength remains as constant as possible (given variations in temperature and power).

Ambient light is the base of structured light sensors, and there are measures put in place to mitigate this issue. One is an IR-pass filter at 830 nm over the IR camera. This prevents stray IR in other ranges (like from TV remotes and the like) from blinding the sensor or providing spurious results. However, even with this in place, the Kinect does not work well in places exposed to a lot of light since sunlights wide band IR has enough power in the 830 nm range to blind the sensor.

The distance at which the Kinect functions is also limited by the power of the laser diode. The laser diode power is limited by what is eye safe. Without the inclusion of the scattering filter, the laser diode in the Kinect is not eye safe; its about 70 mW. This is why the scattering innovation by Prime Sense is so important: the extremely bright center dot is instead distributed amongst the 9 dots, allowing a higher powered laser diode to be used.

The IR camera operates at 30 Hz and pushes images out at 1200x960 pixels. These images are downsampled by the hardware, as the USB stack cant handle the transmission of that much data (combined with the RGB camera). The field of view in the system is 58° degrees horizontal, 45° degrees vertical, 70° degrees diagonal, and the operational range is between 0.8 meters and 3.5 meters. The resolution at 2 meters is 3 mm in X/Y and

1 cm in Z (depth). The camera itself is a MT9M001 by Micron, a monochrome camera with an active imaging array of 1280x1024 pixels, showing that the image is resized even before downsampling.

This system enables the Kinect to see in 3D in all kinds of ambient lighting conditions. The sensor range is automatically adjusted by software and the range can be tuned to take account of the room size and obstructions such as furniture.

2.2.1.3 Kinect Microphone

The Kinect microphone is capable of localizing acoustic sources and suppressing ambient noise. Physically the microphone consists of an array of 4 microphone capsules. Each of the four channels processes sound at in 16 bit audio at a rate of 16 KHz.

2.2.2 Kinect SDK

Kinect for Windows SDK supports applications built with C++, C#, or Visual Basic by using Microsoft Visual Studio 2010. The newly release Kinect for Windows SDK version 1 offers skeletal tracking, speech recognition, modified API, and the ability to support up to four Kinect for Windows sensors plugged into one computer. It enables raw sensor streams like depth sensor and there is also improved synchronization between color and depth, mapping depth to color, and a full frame API.

2.2.3 WEKA - Waikato Environment for Knowledge Analysis

Weka[31] is a software made by the University of Waikato, and contains collection of machine learning algorithms for data mining tasks. The algorithms can either be applied directly to a dataset or set using your own Java code. Weka contains a collection of visualization tools and algorithms for data analysis and predictive modeling, together with graphical user interfaces for easy access to this functionality. It also contains tools for data pre-processing, classification, regression, clustering, association rules, and visualization. It is also well-suited for developing new machine learning schemes.

WEKA was used for kinect skeleton and motion data classification. We give as an input an ARFF (Attribute Relationship File Format) text file that contains two sections. The header and the data section. In the header section is the list of attributes, each one associated with a unique name and type. The latter describes the kind of data contained in the variable and what values it can have. The classification attribute is the last in the list. In the data section there are skeleton data frames, that we want to classify line by line.

Chapter 3

Related studies of affect recognition

To date, many computer software solutions that can recognize a persons identity by using specific data such as video, voice, images, etc. have been proposed. But no sufficient explanation has been found in which the computer can accurately determine a persons feelings. Affective computing (AC) is the solution that will bring computer systems into a new era where they can recognize and respond to human feelings. The idea behind Affective Computing is that if a computer can adjust the software interface based on the users affective state, the quality of the interaction will enhance making the interface more enjoyable, usable and effective.

The last few years, the only people interested in human emotion were phychologists. In 1980 for the first time Turkle[14] suggested that computers might be used to study emotions but it took ten years for systematic research programs to emerge in this front. Picard[34] describes three types of Affective Computing Applications:

1. systems that detects the emotions of the user

2. systems that express what a human would perceive as an emotion and

3. systems that actually feel an emotion

There has been research in the field of emotion recognition using various modalities. The most widely used technique has been through facial expressions [30]. Nevertheless, body posture and movement recently started to interest researchers in the field. One approach uses cyclic arm movements to recognize fundamental emotions [24]. A number of experiments that use acted data as the training set for the recognition system have already been presented. Coulson [25] investigates recognition accuracy, confusions, and viewpoint difference while attributing emotions to postures. Kleinsmith et al. [28] use emotion recognition from low level posture data to report cross-cultural differences.

In another study [27], [29], automatic recognition models grounded on low-level posture descriptions were built and tested for their ability to generalize to new observers and postures using random repeated subsampling validation. The automatic recognition models achieve recognition percentages comparable to the human base rates.

Another approach deals with movement qualities such as amplitude, speed and fluidity of movement to infer emotions instead of low level posture data [26]. Similar to this, Savva and Bianchi-Berthouze use non-acted data captured during Nintendo Wii game playing to extract movement features to recognize emotion from animation rather than posture [33]. In this approach, human recognition of emotion is taken through observation of skeletal information only. In our approach, both video and skeleton data is given to the human observers during labelling, as during experimentation, a number of postures could not be predicted using skeleton data only, as humans are capable of using multiple modalities at the same time to express emotions.

Moreover, all the above approaches use body posture or animation data captured from traditional motion capture equipment, which is not as ubiquitous as Kinect. Kinect is not as accurate as traditional motion capture devices so this raises a question about whether it is practicable to capture training data for the recognition system using Kinect.

3.1 Affect Recognition

Recognizing human emotions is a very ambiquous and challenging process since they cannot be measured accurately, because the vary between individuals. We need to show how we can blend the complex scientific theories of human emotions with the practicality of engineering in order to develop affect sensitive interfaces.

3.1.1 Designing an Affect Recognition System

3.1.1.1 Databases

The most important issue to be addressed before designing an Affecting Computing application is the structure and the description of an affect. Both together provide critical information for the affect displays that emotion recognition systems are design to detect. In affect detection we will use the most common and old way that physiologists used to describe emotions which is in terms of discrete categories. An example of this description is the prototypical emotions categories such as happiness, sadness, fear, anger, disgust and surprised. One can argue (and will be right) that emotions are not limited to these six categories[37]. Further more it can also be argued that in many situations these categories appear together at the same instance. So what is the accurate size of labels to be used for emotions in an applications database? Are those six sufficient? What are the emotions we are interested in for the application we are going to develop? The size of the descriptions to be used, or labels as we are going to refer on them for now on, is the first and the most important characteristic of the systems database. If we choose a large size of labels we will increase the complexity of recognition, choosing a very small one on the other hand we might end up with a very simple and at the end not very useful system.

Another way of categorizing emotions is the use of dimensions. These dimensions include evaluation, power, activation, control etc. The evaluation dimension measures the persons feeling from positive to negative. The activation measures whether a person is more likely to take an action under an emotional state from active to passive. We mostly use this representation if we want to label a range of emotions.

The next question to be answered is what data we should use as a basic reference for the system to decide about an emotion. Data were collected consisting of videos and sound[1] from several different sources. Sound data are collected from interviews recordings, customer support recordings and meeting recordings. Those data are later analyzed by dictionaries, spectral characteristics, prosody, frequency of repeated words or phrases, speaker voice tension and pauses in speech. Dictionaries are used in order to search for words or phrases that suggest an emotion. Spectral characteristics describe the relation of a given audio weave with certain behaviors. Prosody, frequency of repeated word or phrases, speakers voice tension and pauses are the common everyday used criteria

[1] http://emotion-research.net/wiki/Databases

in deciding emotions from speech. Video data are collected from video recordings during certain activities (also known as video kiosk), video recording of interviews and videos that can be found on the web [2] . To analyze video data facial expressions, such as muscle movement in various parts of the face, are observed. But in many cases, video data suffer from poor lighting and bad angle of capture. In order to eliminate these phenomena, if they occur, we develop a 3D reconstruction of the video and later we proceed in analyzing it. The decision about what label should be given to each sound or video data is made by an expert in the field of psychology. This procedure may take a lot of type especially if we want to use a very large set of labels.

Spontaneous emotions do not always give us the clarity we need in order to decide about a certain emotion and at the end are not suited for the purposes of the system we design. A common approach in collecting data as mentioned above is the interviews recordings. In these interviews we deliberately present to the individual appropriate chosen material in order to record his reactions. An example of these is gathering a number of people to watch a comedy play in a theatre and during the play a camera records their reactions. However, increasing evidence suggests that deliberate behavior differs in visual appearance, audio profile, and timing from spontaneously occurring behavior.

Another important characteristic of a database is the number of entries. A large number of entries could lead to more accurate decisions but also in large computational or training time. Choose to use a very small number of entries in the database and you might end up with a wrong judgment, due to insufficient recordings. Choosing the correct size

[2] http://www.mmifacedb.com/

of the data entries lays many times in a trial and error procedure or the experience of the designer.

The last but not least of the issues of database characteristics is the accessibility of the database. Databases which contain these types of data can be either public or private. A very large effort is being made for a collection of databases containing both spontaneous emotions and deliberate emotions[3] .Public databases are more commonly to contain deliberate emotions. Databases which contain spontaneous emotions in many cases are dimmed of being published due to the lack of appropriate agreement of subjects. The reason is that spontaneous displays of emotions reveal personal and intimate experience. So privacy issues jeopardize their public accessibility.

3.1.1.2 Decision Techniques

Many techniques have been proposed over the years for decision making in Affect Recognition Systems. The most important technique used in a variety of different systems comes for the field of Artificial Intelligent. The first technique, is to train a neural network to work with a set of data which comes from the database described above in order to decide about the human emotions. A neural network can also be trained for classification of data. Mota and Picard[41] used this technique in their Tekscans Body Pressure Measurement System (BPMS).

The second technique that is used to evaluate a humans emotion via a system, is by using heuristics. In more precise terms, heuristics are strategies using readily accessible, though loosely applicable, information to control problem solving in human beings and

[3] http://corpus.amiproject.org/

machines. Scherers GENESE expert system was built on a knowledge base that mapped appraisals to different emotions as predicted by Scherers heuristics and theory[8]. The system was successful in achieving a high accuracy (77.9 %) at predicting the target emotions for 286 emotional episodes described by subjects as a series of responses to situational questions.

3.1.2 Multimodality

It is a common observation that emotions can be expressed by multiple physiological response systems. So trying to correctly decide for a persons emotions from by gathering information from a single modality may lead you to a false outcome. Think of a person being upset about a particular thing. That person will raise their voice, the motion of their hands will increase, their face probably will turn red etc. The hallmark of emotion theory is that during an emotion episode we have responses from multiple systems bound in space and time. Multimodality was widely advocated but rarely implemented until few years before. The problem was the unisensory detection that was needed which increase in multisensory environments.

To combine signals from different sensors we use the following 3 methods[23]:

- Data Fusion: Data Fusion is performed on the raw data for each signal and can only be applied when the signals have the same temporal resolution. This technique can be used for intergrading psychological signals coming from the same recording equipment. It could not be used to intergrade a video signal with a text subscript.

The main reason though that is not widely used is the sensitivity that has on noise when noise is produced by malfunction or misalignment of the different sensors.

- Feature Fusion: Feature fusion is performed on the set of features extracted from each signal.

- Decision Fusion: Decision Fusion is performed by merging the output of the classifier for each signal. Hence, the affective states would first be classified from each sensor and would then be integrated to obtain a global view across the various sensors. It is the most commonly used approach for multimodal HCI.

3.2 Facial Expression Recognition

The face plays an important role to emotion expression so most of the vision based affect recognition studies focus on facial expression analysis. In the current research on the machine analysis of facial expressions, we can distinguish two main streams. The first is the recognition of affect and the second the recognition of facial muscle action.

Most of the work done on the automatic facial expression recognition was based on deliberate or in many times exaggerated facial displays. However, several efforts have been recently reported on the automatic analysis of spontaneous facial expression data. Some of them study the automatic recognition of action units, rather than emotions, from spontaneous data displays. Studies have also been made in order to detect the differences between a spontaneous and an exaggerated data display. As an example of such studies is the system proposed by Valstar[5], which characterizes temporal dynamics of facial actions and employ parameters like speed, intensity, duration, and the co-occurrence of facial muscles activations to classify facial behavior present in a video as either deliberate or spontaneous. Other work includes the study of Zeng [37] for separating emotional state from non emotional state, separating posed from genuine smiles, and studies for the recognition of pain from facial behavior.

The best facial expression recognizers employ various pattern recognition approaches and are based on 2D spatiotemporal facial features. The most common facial features extracted are either geometric features such as the shapes of the facial components (eyes, mouth, etc.) and the location of facial salient points (corners of the eyes, mouth, etc.) or appearance features representing the facial texture, including wrinkles, bulges, and furrows. The best solution would be to use both geometric and appearance features. The best examples for this practice are those proposed by Tian[42] who used facial component shapes and the transient features like crow-feet wrinkles and nasal-labial furrows, and that of Zhang and Ji[43] who used 26 facial points around the eyes, eyebrows, and mouth, and the transient features proposed by Tian. Its worth to mentioning that most of the existing 2D Feature Based methods are suitable for the analysis of facial expressions under a small range of head motions hence the recognition is based on recordings in near frontal view. As last we mention the facial expression analysis based on 3D face models. One state of the art implementation on this field is the work of Huang and his colleagues[44]. They used features extracted by a 3D face tracker called the Piecewise Bezier Volume Deformation Tracker[45]. Given that the subject can be recorded in less controlled real-world settings, the progress of this methodology may yield view- independent facial expression recognition, which is important for spontaneous face recognition.

3.3 Audio Based Affect Recognition

The most common method of detecting emotions is through audio. This states that the system is recognizing a subset of basic emotions from speech signals. However recently new methods have found that propose interpretation of speech signals in terms of certain application-dependent affective states. Devillers and Vidrascu[4] are using lexical cues resulting in a better performance than using paralinguistic cues to detect relief, anger, fear, and sadness in human to human medical conversations. Also some studies introduce the idea of using a combination of acoustic and linguistic features to improve vocal affect recognition. [17],[46]

Although this can show an improvement on vocal recognition in many cases the automatic extraction of these related features can be a very difficult problem. Existing systems based on this idea cannot reliably recognize the verbal content of emotional speech. Furthermore extracting semantic discourse information is even more challenging. Most of these features are either extracted manually or by transcripts.

The state of the art in research at the moment for Audio Based Affect Recognition is including:

- Detect non-basic affective states, including coarse affective states such as negative and non- negative states, application-dependent affective states, and nonlinguistic vocalizations like laughter and cry.
- Integrate paralinguistic features and linguistic features such as lexical, dialogic, and discourse features.

- Use of contextual information to improve the affect recognition performance.
- Affective states across languages.
- Influence of ambiguity of human labeling on recognition performance and proposed measures of comparing human labelers and machine classifiers.
- Advanced techniques in feature extraction, classification, and natural language processing.

3.4 Affective Neuroscience Brain Imaging and EEG

Neuroscientists new techniques in brain imaging help map the neural circuitry that underlies emotional experience. An example of the contribution of neuroscience in understanding human emotions and detect them are the evidence provided to the discussions on dimensional models where valence and arousal might be supported by distinct neural pathways. One of the most popular methods that neuroscientists use is the Functional Magnetic Resonance Imaging [47]. Immordino-Yang and Damasio[48] used evidence from brain damaged subjects to prove that emotions are very important in decision making. Scientists observed that patients with a lesion in a particular section of the frontal lobe showed normal logical reasoning but yet they could not see the consequences of their actions and were unable to learn from their mistakes. By that observation scientist then could conclude that emotion related processes were required for learning, even in areas that had previously been attributed to cognition.

Unfortunately the cost, time resolution, and the complexity of setting up protocols that can be used in real world activities are still problematic issues that put the application development with use of these techniques to a hold. Nevertheless, signal processing[49] and classification algorithms[50] for EEG have been developed in the context of building Brain Computer Interfaces.

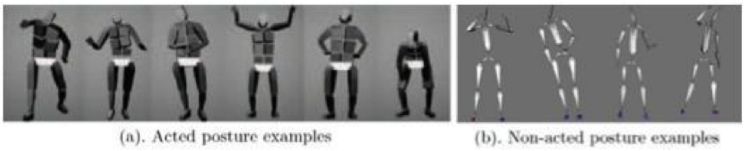
3.5 Posture and Motion Data Recognition

With the emergence of the affective computing field, various studies have been carried out to create systems that can recognize the affective states of their users by analyzing their body expressions in order to recognize, understand and model human emotion. Human understand emotions with language & sound of the voice, context and non-verbal communication, than includes facial expressions and body postures and movement.

3.5.1 Role of bodily information play in affect recognition

Bodily expressions have been recognized as more important for nonverbal communication and changes in a persons affective state are also reflected by changes in body posture. The type of information about the body that is necessary for recognizing the affective state displayed, according to a neuroscience study by Giese and Poggio[51], there are two separate pathways in the brain for recognizing biological information, one for form information (i.e. the description of the configuration of a stance) and one for the motion information.

Studies showed[52] that combination of face and gesture is 35% more accurate than facial expression alone. There are two type of Body Motion based affect recognition, using acted data which the entirety of the movement encodes a particular emotion and non acted data which are more natural data like knocking door, lifting hand, walking etc.


(a). Acted posture examples

(b). Non-acted posture examples

Figure 5: Acted and Non acted posture examples

3.5.2 An Architecture of a Posture Recognition System

An architecture of a system to recognize affect from non-acted subtle body expressions represented as a sequence of postures like Figure 5, of a person playing a video game by using only the information about the configurations of the postures without considering the temporal information[33]. A motion capture system is used to record the sequences of postures from the players. The system is implemented as a combination of an affective posture recognition module and a sequence classification rule to finalize the affective state of each sequence as you can see in Figure 6.

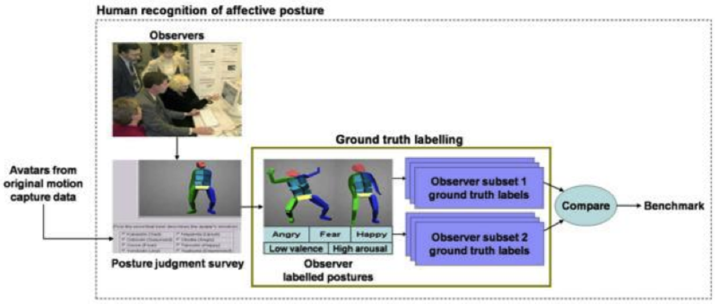

Human recognition of affective posture
Observers
Avatars from original motion capture data
Posture judgment survey
Ground truth labelling
Angry
Fear
Happy
Low valence
High arousal
Observer labelled postures
Observer subset 1 ground truth labels
Observer subset 2 ground truth labels
Compare
Benchmark

Figure 6: The Affective Posture Recognition System

3.5.3 Ground truth labeling

To test human performance in recognizing affect from posture, the ground truth must be built; labels need to be assigned to the affective postures. In the acted postures study, the actors labels and dimension ratings are not used because the actors may not portray what they intend to portray as they are not professional actors. In the non-acted postures study, the players are not used to label their own postures because self-reported feelings at the end of a task are notoriously unreliable, and it is not feasible to stop the players during the gaming session to ask them their current affective state. Furthermore, because the complete affective state is expressed through the combination of a variety of modalities in the non-acted scenario in particular, it is difficult for the players to be aware through which modality affect was expressed or if their bodies were expressing their true feelings. Thus, the approach used in this research is to build ground truth labels from outside observers judgments of the postures using posture judgment surveys.

3.6 Multimodal Affect Recognition

The past few years, scientists are focusing towards multimodal affect recognition for several reasons. Firstly, some psychological researches indicated that judging someones affective state, people mainly, rely on facial expressions and vocal intonations. Hence more than one method of recognition is insufficient. Another motivation for multimodal recognition is the fair engineering prospect of improved reliability. More specifically, current techniques for detection and tracking of facial expressions are sensitive to head pose, clutter, and variations in lighting conditions, while current techniques for speech processing are sensitive to auditory noise. Due to complementarity and redundancy of the data coming from the two channels, audiovisual human affect recognition is expected to perform more robustly than uni-modal methods. Thus, affect recognition should inherently be the issue of multimodal analysis.

An example of such a system is shown in Figure 7. This figure illustrates a prototype of such an affect-sensitive, multimodal computer-aided learning system.[53] In this learning environment, the user explores Lego gear games by interacting with a computer avatar. Multiple sensors are used to detect and track the users behavioral cues and his or her task. More specifically, a camera is used to record the users facial expressions, a set of cameras are used to track eye movements, another camera is used to monitor the progress of the task, and a microphone is used to record the speech signals employed subsequently to recognize the speech and analyze prosody. Multisensory information is then processed and visualized, including the users emotional state, engagement state, the utilized speech keywords, and the gear state. Based on this information, the avatar offers an appropriate tutoring strategy in this interactive learning environment.

To build a testing set for automatic recognition, posture sequences were automatically extracted from the remaining replay windows (i.e., those not used to build the training set). The automatic extraction yielded posture sequences ranging from two to 40 frames in length. 836 posture frames across 75 posture sequences were extracted. A set of observers recruited for labeling the set of sequences. As opposed to labeling each individual posture within a sequence as was done for the training set preparation, the observers viewed each sequence in the testing set as an animated clip of a simplistic humanoid avatar and were asked to assign a unique label to the sequence. This approach was taken to set the target system performance as the level of agreement obtained by observers who considered both form and temporal information.[29]

The within observers agreement was 66.67% with kappa reaching 0.162, indicating slight agreement. There were 14 posture sequences with the defeated ground truth label, 8 triumph, 39 neutral and 14 ties. In the case of ties, the ground truth was randomly selected between the two tied labels.

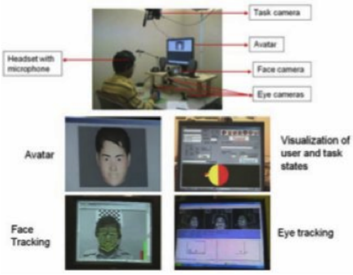


Figure 7: A Multi Modal Example

3.6.1 Audio - Visual Multimodal Affect Recognition

In this first application, we are focusing on the development of a computing algorithm that uses both audio and visual sensors to detect and track a users affective state to aid computer decision making[53]. Using a new Multi-stream Fused Hidden Markov Model (MFHMM), coupled audio and visual streams were analyzed to detect 4 cognitive states (interest, boredom, frustration and puzzlement) and 7 prototypical emotions (neural, happiness, sadness, anger, disgust, fear and surprise). The MFHMM allows the building of an optimal connection among multiple streams according to the maximum entropy principle and the maximum mutual information criterion.

This application was tested in a person- independent affect recognition problem by using leave-one- out cross validation scheme. For this test, all of the sequences of one subject were used as the test sequences, and the sequences of the remaining 19 subjects were used as training sequences. The test was repeated 20 times, each time leaving a different person out. The evaluation metric which was being used was the accuracy of each method. Five different cases were used to evaluate the ability of this algorithm. These cases were the following:

- The use of only face features
- The use of only audio features
- The use of an energy Hidden Markov Model
- The use of an independent Hidden Markov Model which assumes that face and pitch features are independent
- The use of the present application, of the Multistream Fused Hidden Markov Model which assumes statistical dependency between face and audio features and energy.

The results were very encouraging for this new method, MFHMM. As Figure 8 shows, the accuracies achieved from this new method under various audio SNR[4] conditions are the best between the five cases we have examined. As a result, it is clearly shown that

[4] Signal to Noise Ratio

these new multi modal methods have better results than using methods with only uni-modal information.

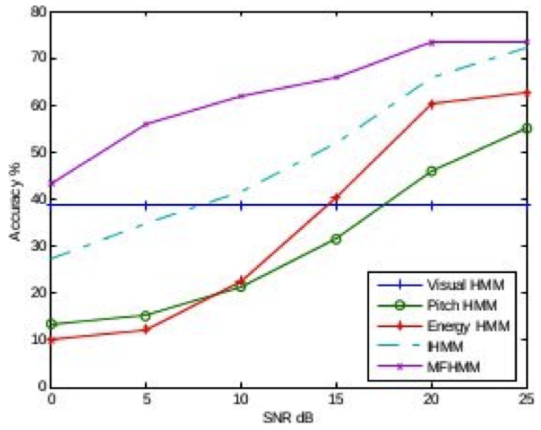


Figure 8: Accuracies of different methods under various audio SNR conditions

3.6.2 Using Face and Body Gestures for Affect Recognition

In this second application the Histogram of Oriented Gradients on the MHI-HOG [5] is combined with the Image-HOG [6] [54] to capture both motion and appearance information of. The MHI-HOG captures motion direction of an interest point as an expression evolves over the time. The Image-HOG captures the appearance information of the corresponding interesting point. Combination of MHI-HOG and Image-HOG can effectively represent both local motion and appearance information of face and body gesture for affect recognition. The temporal normalization method explicitly solves the time resolution issue in the video-based affect recognition.

[5] Motion History Image

[6] Image Histogram of Oriented Gradient

Figure 9 shows the general method to incorporate the temporal dynamics in expression recognition from both face and body gesture modalities. So, this method is separated into the following layers:

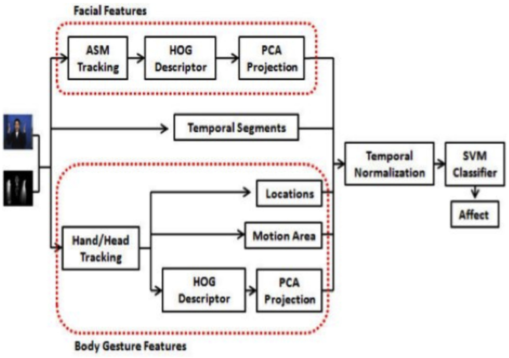


Figure 9: Flow chart of expression recognition from both face and body gesture modalities

3.6.2.1 Facial Feature Extraction and Representation Layer

There are three steps to extract facial features. The first step is to track the facial landmark points using the ASM[55] model as shown in Figure 10. The second step is to extract the Image-HOG and the third step to extract the MHI-HOG descriptors of the facial landmark points. As shown in Figure 11, the MHI image captures motion information of the facial landmark points, while the original image can provide the corresponding appearance information.

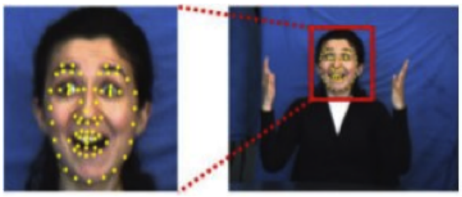


Figure 10: ASM facial landmark points

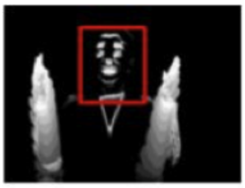


Figure 11: MHI-Image tracking

3.6.2.2 Body Gesture Feature Extraction and Representation Layer

For body gesture, the Image-HOG and MHI-HOG are being extracted for both hand regions. In addition, the positions and motion areas of both hand and head regions are employed to measure their trajectory and motion intensity. Before we extract the body gesture features, a simple skin color-based hand tracking is applied to detect hand regions as shown in Figure 12[56]. The center position of the head is extracted based on the ASM facial landmark points, as shown in Figure 13. Then the center points of the hand and head regions are employed, with reference to the neutral frames corresponding positions, to describe the location of the hands and the head respectively.

The hands and head positions are further normalized with the subjects height, which is measured from the center of the head to the bottom of each frame image. Motion areas of the hands and head regions are measured by counting the number of motion pixels from the MHI image within an NxN size window at each center, as shown in Figure 14.The MHI-HOG and the Image-HOG of both hand regions are extracted in the following steps. Firstly, uniform grid interest points within both hands skin regions are selected, which are also within the patch at each hands center. Secondly, the Image-HOG and the MHI-HOG descriptors for each selected skin interest point are extracted and finally a bag of words representations of the Image-HOG feature and the MHI-HOG feature respectively for each frame is being formed.

To evaluate this method, the accuracy gained from the combination of face and body gestures with the accuracy gained from face only and gesture only methods have been compared. The results shown in Figure 15 clearly show that Multi Modal methods of affect recognition have much better results than uni-modal methods.

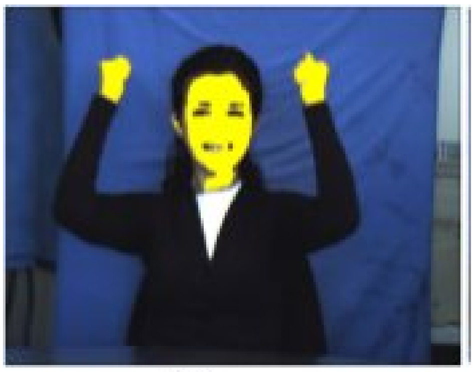


Figure 12: Hand tracking by skin color-based tracker

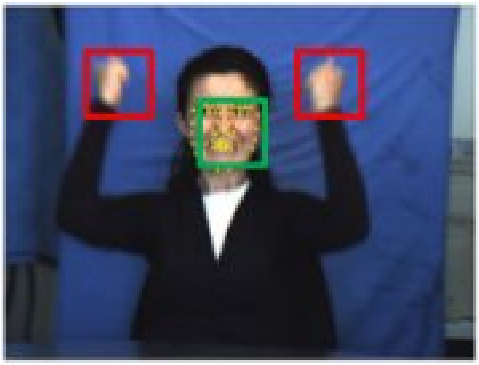


Figure 13: Position of hands using skin color tracking & position of head using ASM model

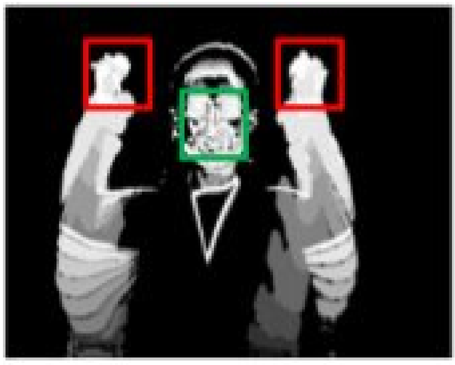


Figure 14: Extract motion areas of hand and head regions

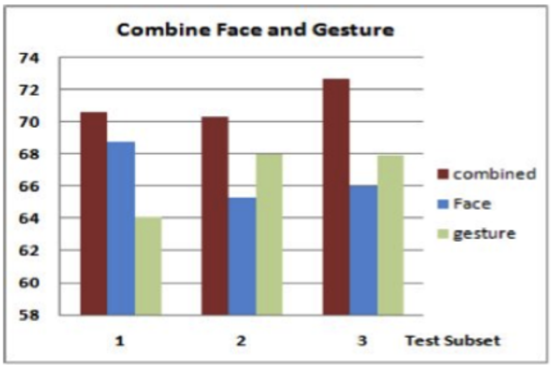


Figure 15: Comparison of affect recognition methods

Chapter 4

Method

4.1 Training & Testing Process

Before recording

- Decide and set the emotion set (confirm as Concentrating, Defeated/Frustrated, Triumphant)
- Decide a game (or more) to use in recording of non-acted data
- Develop the software to capture synchronized motion data, face and body video, and game video
- Test recording functionality and synchronization
- Confirm students participation in playing the game. Students should not be informed about the project purpose during recording

During recording

- Set up the system on good lighting conditions and with sufficient space for gameplay
- Provide students with a consent form and collect demographics info (Appendix E)
- Record more than 30 mins for each player both during gameplay and replay session
- Organise stored data in folders by student

After recording

- Develop a user interface for the observers (C# programming)
- Get observers determine motion clips with affective states both during gameplay and replay session.
- Get the observers to annotate postures with emotion labels. Method - online evaluation.
- Create a benchmark using observation agreement.
- Run cross-validation and compare results on all data types.

4.2 Participants

For the current experiment, seven different participants were instructed to play with Kinect sports game, while capturing their skeleton data and video. After the data collect, two observers were instructed to seperate the videos into image postures that according to their oppinion expressed some kind of emotion. Four different observers were used, in order to evaluate and annotate the postures into 3 different emotion categories.

4.3 Training

Participants were given instructions how to play the kinect sports game, and practiced a couple of times in order to familiarize themselves with it. They tried different sport games, like football, volleyball, running, javeling and bowling.

Figure 16: Training how to playing football in kinect

4.4 Data collection

The first step was to collect posture data. The Kinect SDK skeleton representation data was used to measure and store the posture of the players in each frame. Six male university students were recruited to take part in the data collection. The players were asked to play games with the Xbox integrated with a Kinect. A second Kinect was connected to a PC and was used separately to record the motion data for the players. The software used for the motion capture also had the capacity to capture and display both video and skeleton data in real-time. The PC screen was captured to provide replay capacity for the extraction of the apex poses.

In contrast to [27], skeleton and emotion postures were recorded during both actual game play sessions and replay sessions, as it may not be sufficient for an emotion recognition system to be trained with different data than the one it is aimed to recognize. After the data was recorded, the observers replayed the captured video to determine and select possible affective postures for four emotional states: Triumph, Concentrating, Frustrated and Defeated[27].

The observers were given the option to identify emotional states based on both video and skeleton data, watching captured video. Both observers agreed that the actual video made it easier for them to decide on the emotional state in many instances that the skeleton data was not sufficient enough. A total of 147 postures were extracted using this technique.

4.5 Emotion Annotation

For all the postures that were extracted, a screenshot of the corresponding frame of the software was taken for each posture. A digital questionnaire (Appendix F) was created and four observers were asked to annotate the screenshots (Figure 17) with an emotion label of one of the three above mentioned emotional states. After the questionnaire, all the observers expressed a concern on selecting a label between frustrated and defeated, as they all considered these emotional states very similar in many circumstances. For this reason, the emotion label set was reduced to three (triumph, concentrating, frustrated/defeated).

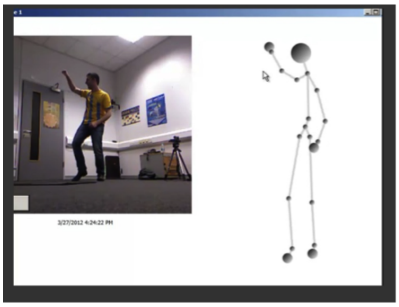


Figure 17: Screenshot example used by human observers

4.6 Cohen's Kappa

Kappa[38] provides a measure of the degree to which two judges, A and B, concur in their respective sortings of N items into k mutually exclusive categories. A 'judge' in this context can be an individual human being, a set of individuals who sort the N items collectively, or some non-human agency, such as a computer program or diagnostic test, that performs a sorting on the basis of specified criteria.

Comparison between the level of agreement between two sets of dichotomous scores or ratings (an alternative between two choices, e.g. accept or reject) assigned by two raters to certain qualitative variables can be easily accomplished with the help of simple percentages, i.e. taking the ratio of the number of ratings for which both the raters agree to the total number of ratings. But despite the simplicity involved in its calculation, percentages can be misleading and does not reflect the true picture since it does not take into account the scores that the raters assign due to chance.

Using percentages result in two raters appearing to be highly reliable and completely in agreement, even if they have assigned their scores completely randomly and they actually do not agree at all. Cohens Kappa overcomes this issue as it takes into account agreement occurring by chance.

How to compute Cohen's Kappa

The formula for Cohens Kappa is:

$$k = \frac{Pr(a) - Pr(e)}{1 - Pr(e)}$$

where Pr(a) is the relative observed agreement among raters, and Pr(e) is the hypothetical probability of chance agreement, using the observed data to calculate the probabilities of each observer randomly saying each category. If the raters are in complete agreement then k = 1. If there is no agreement among the raters other than what would be expected by chance (as defined by Pr(e)), k=0

4.7 Benchmark creation

Table 1 presents observer agreement for all possible pairs among the four observers. As can be seen, agreement strength is above or equal to good at all cases, demonstrating that the labelling quality is satisfactory for the data extracted for the recognition system. Observer 2 clearly performs better on all paired cases with the other observers, achieving not only very good agreement with observers 1 and 4, but also the highest score with observer 3. The average Kappa value for observer 2 on all three pairs is 0.812.

Table 1: Paired observer agreement strength

Observer pair	Kappa	SE of Kappa	95% confidence interval	agreement strength
1&2	0.806	0.043	0.721 - 0.890	Very Good
1&3	0.739	0.049	0.643 - 0.835	Good
1&4	0.749	0.047	0.656 - 0.842	Good
2&3	0.782	0.045	0.694 - 0.871	Good
2&4	0.849	0.038	0.775 - 0.923	Very Good
3&4	0.686	0.052	0.585 - 0.787	Good

Table 2 lists the actual number of agreements and percentage of agreement for all possible pairs that include observer 2. It also compares against chance level agreement. The score is significantly higher than the chance level agreement and above 86% on all pairs, making observer 2 a suitable candidate for the labeling of postures. Considering this and the Kappa value, the labeling from observer 2 was used for posture data annotation for the different emotion recognition tests.

Table 2: Agreement score for Observer 2

Observer IDs	Number of observer agreements	Percentage of agreement	Number of agreements expected by chance	Percentage of expected agreement
1&2	129	87.76	54.41	37.04
2&3	127	86.39	55.10	37.50
1&2	129	87.76	54.41	37.04

Finally, the overall agreement of all observers was far above chance level at 72.3%, agreeing on 107 out of the 148 postures. This is used as the benchmark for evaluating the performance of the collected skeleton data postures as input for the emotion recognition system. Figure 18 presents the distribution of labels across the 147 postures according to observer 2.

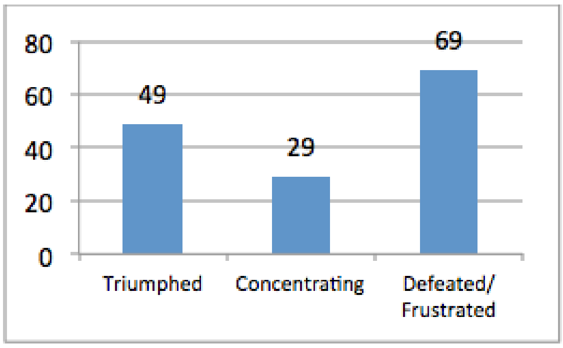


Figure 18: Distribution of emotion labels for the 147 postures, according to observer 2

4.8 Recognition Testing

After the observer performance is completed, the skeleton data for each labelled posture is extracted and annotated with the corresponding emotion label taken from the observer 2 annotation. The skeleton data comprises standard 3D rotational information for each of the joints of the body.

Automatic emotion recognition is tested based on the capacity to recognise new postures. The labelled data used in WEKA [31] to build and test a model using back-propagation algorithm with 10 fold cross-validation, similar to [27].

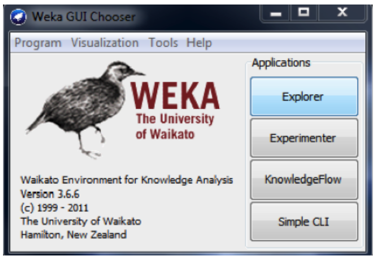
4.8.1 Building and testing an emotion recognition model in WEKA

4.8.1.1 Supervised-Learning Algorithms

WEKA provides a set of machine learning and data mining algorithms that can be used to build and test recognition models. Examples of algorithms that can be used in the current experiment are the back-propagation algorithm and classification using decision trees. Back propagation [40] is a supervised learning algorithm. This means that a set of examples of mappings from input to output must be provided as training data. The algorithm constructs an artificial neural network that can be used with unknown input data and make predictions, such as in our case providing emotion recognition. Decision trees are classification schemes created from training data (again mapping sets input-output) and comprise a tree and a set of rules. According to Hans & Kamber [39] a decision tree is a tree structure similar to a flowchart, in which an internal node denotes a test on one of the attributes, a branch represents the outcome to a test and leaf nodes represent classes or their distributions. This section describes how the back-propagation algorithm was used during the project to build and test an emotion recognition model.

4.8.1.2 Loading the data

WEKA allows researchers to input training data to the system using a standard file format that lists the parameters of the system and examples of mappings between input and output. WEKA can buse the explorer application from the GUI chooser to input, process and visualize data (Figure 19).

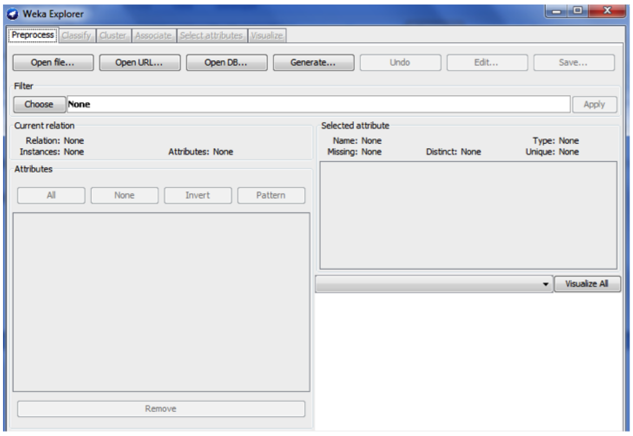


Waikato Environment for Knowledge Analysis
Version 3.6.6
(c) 1999 - 2011
The University of Waikato
Hamilton, New Zealand

Figure 19: WEKA's GUI chooser

Once the explorer is chosen and presented, initially only the preprocess tab is activated, as there is no data yet to classify, visualize and so on. The next step is to click on the open file button (top-left) and select the file that includes the data to be used in the experiment (Figure 20). This is done using a standard windows explorer for opening files in applications.

As soon as the data is loaded, the explorer prepares information about the set of the attributes used in the file. More specifically, it provides standard statistics such as mean value, min, max and standard deviation for each of the attributes, gives a visualisation option and lets the user filter out some of the data if desirable (Figure 21).

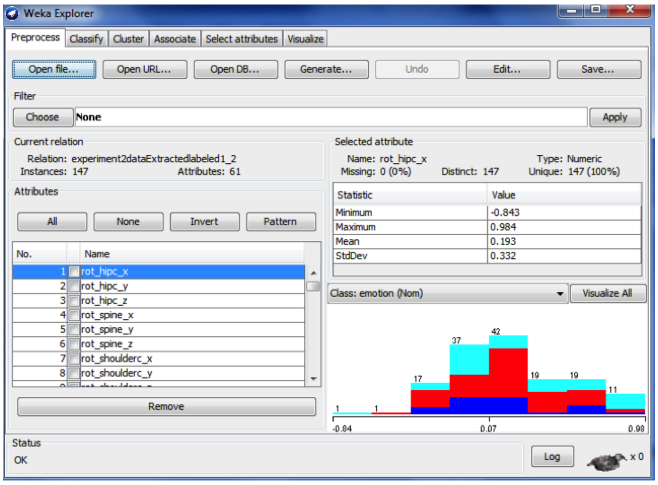


Filter: None

Relation: None
Instances: None
Attributes: None

Name: None
Missing: None
Distinct: None
Type: None
Unique: None

Figure 20: The explorer starting window



Filter: None

Relation: experiment2dataExtractedlabeled1_2
Instances: 147
Attributes: 61

Name: rot_hipc_x
Missing: 0 (0%)
Distinct: 147
Type: Numeric
Unique: 147 (100%)

Statistic	Value
Minimum	-0.843
Maximum	0.984
Mean	0.193
StdDev	0.332

No.		Name
1	□	rot_hipc_x
2	□	rot_hipc_y
3	□	rot_hipc_z
4	□	rot_spine_x
5	□	rot_spine_y
6	□	rot_spine_z
7	□	rot_shoulderc_x
8	□	rot_shoulderc_y

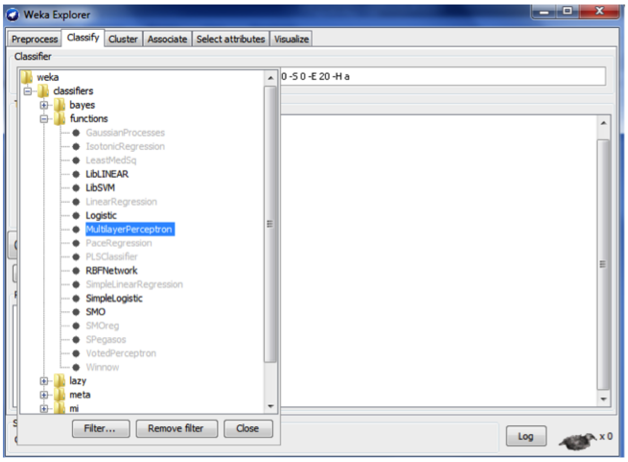
Class: emotion (Nom)

Status: OK

Figure 21: The WEKA explorer with the file loaded

4.8.1.3 Selecting the learning algorithm

After this step, we are now ready to select an algorithm to build a learning model using the provided data. This is achieved by clicking on the classify tab, and then in the new tab, click on the choose button (top-left) and select the algorithm of your choice. In this specific example, we used the MultilayerPerceptron, which is WEKAs implementation of the back-propagation algorithm (Figure 22).

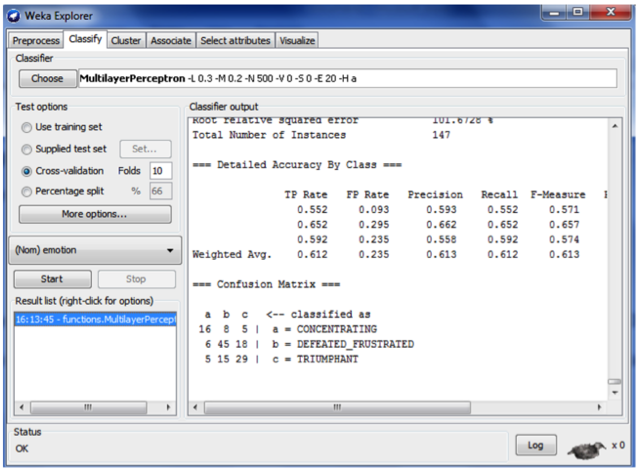


0 -S 0 -E 20 -H a

Figure 22: Selecting back-propagation for the model

4.8.1.4 Executing the test

As soon as the classifier has been selected, test options allow the researcher to select if he wishes to use separate test data, sample through all the data and separate it to test and training using cross-validation and other options. In Cross-validation the data is split into a number of subsets (folds) and the experiment is repeated a number of times equal to the number of sets. Each time a different subset is used as testing data, while the rest are used to train the model. At the end, the results are summarised. In the current experiment we performed cross validation using 10 folds. This means that the data was split to 10 sets and the model was built and tested 10 times. As soon as the testing options have been decided, the button start needs to be pressed and the algorithm starts building the model. Once built finishes, the testing follows. Results are then summarised in the classifier output textbox of the WEKA explorer (Figure 23).



MultilayerPerceptron -L 0.3 -M 0.2 -N 500 -V 0 -S 0 -E 20 -H a

Folds 10

% 66

(Nom) emotion

16:13:45 - functions.MultilayerPercept

```
Root relative squared error            101.6728 %
Total Number of Instances              147

=== Detailed Accuracy By Class ===

               TP Rate   FP Rate   Precision   Recall  F-Measure
                 0.552     0.093     0.593      0.552     0.571
                 0.652     0.295     0.662      0.652     0.657
                 0.592     0.235     0.558      0.592     0.574
Weighted Avg.    0.612     0.235     0.613      0.612     0.613

=== Confusion Matrix ===

  a  b  c   <-- classified as
 16  8  5 |  a = CONCENTRATING
  6 45 18 |  b = DEFEATED_FRUSTRATED
  5 15 29 |  c = TRIUMPHANT
```

Status: OK

Figure 23: WEKA results

The results can be then processed and evaluated. In the above description we explained how we executed the first experiment that used all the 147 sample postures and tested how it can generalise to new postures. The back-propagation algorithm split the 147 postures into 10 folds and used each one of them separately as testing data for one execution. In the second experiment in which we tested how to generalise to new players, the data was separated to two files. Each file contained postures of different players. The biggest file that contained 108 postures was used as the training set with a process similar to the one described above. The other file with 39 postures was used as a separate testing test to see if the model can generalise to new players. The difference using WEKA[31] is to simply select the test option supply testing set and browse to the file that is to be used as test data.

Chapter 5

Experiment Implementation

All of the experiments took place in the same location: Multimedia Lab, department of Computer Science, in University of Cyprus , which is pictured below. Situated in this lab were the Kinect and camera sensors used to collect information related to emotions.

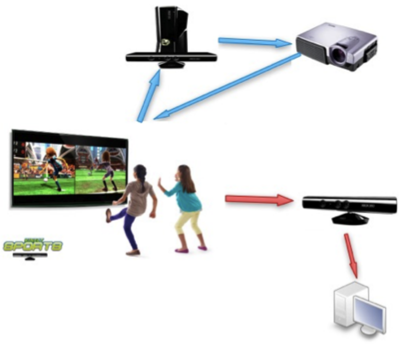


Figure 24: Experiment environment schematic

A schematic design of the experiment environment is showed in Figure 24 . Players are seeing a projector panel to play a kinect game called Kinect Sports, representing athletes. This is done using a normal Xbox Kinect that is connected with a projector in order to project the game view in the projector panel. In the same time a second kinect is capturing players skeleton data, and motion RGB data storing them in a WEKA [31] format on the pc.

In the table below we describe the equipment used to finalize the experiment.

Table 3: Equipment needed for the experiment

Equipment	Description
Motion Capture System	Kinect Device
Hi-Res Camera and Microphone	External JVC camera
Xbox and Kinect sensor	Xbox
Computer Game to be used	Kinect Sports
Subjects and actors to be used	Students
C# custom software	For capture kinect skeleton data
WEKA software	For classification

5.1 Data Files

The data file created for WEKA, contain the following:

- Attributes for skeleton joints containing their type for e.g attribute rot-footr-x numeric, which means the x coordinate in 3D space of the right foot
- Attributes for the emotions to be annotated e.g. attribute emotion Concentrating, Defeated-Frustrated, Triumph
- The skeleton data line by line, each line representing a whole skeleton with its respective joints in one frame.

You can find the a sample data file in Appendix A.

5.2 Software development for Data Collection

A custom software has been programmed in C# in order to capture the skeleton data and video data from the Kinect. Two seperate screens one displaying the skeleton of the moving player and the other displaying the real time video of the player. All the skeleton posture data for each frame is stored in real time in a local computer, using the formating that WEKA[31] needed.

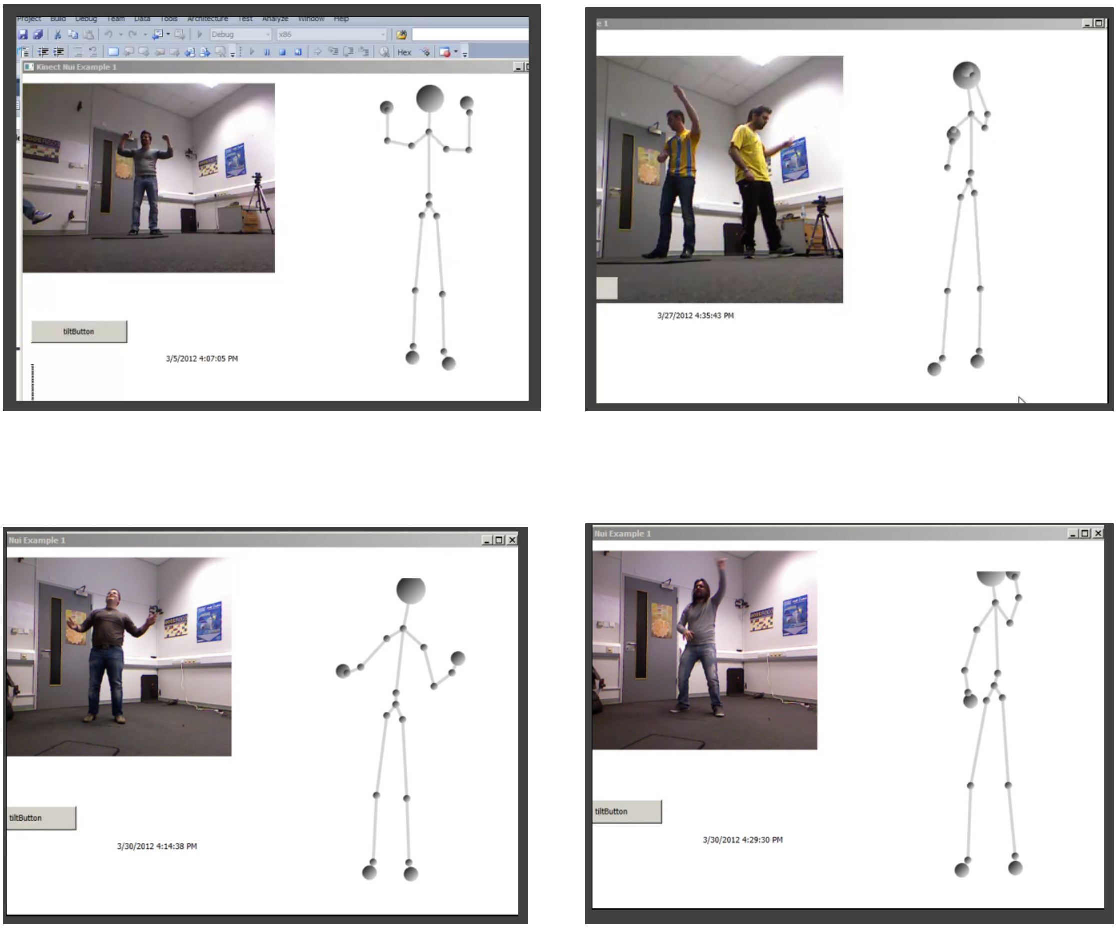
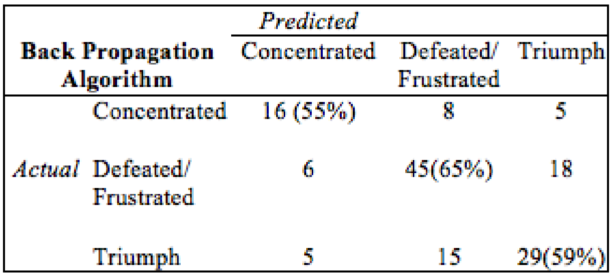


Figure 25: software to capture skeleton and video data from kinect

Chapter 6

Results and Discussion

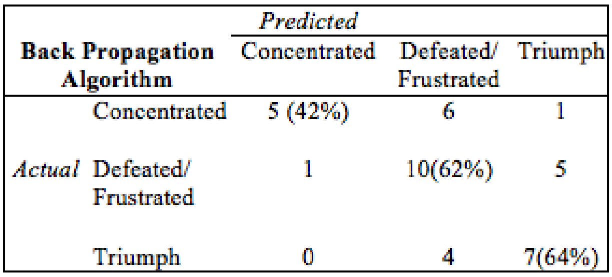
The model recognised successfully 90 out of the 147 postures, resulting in overall recognition rate of 61.22%, which is approximately double than chance level recognition. Figure 26 presents the confusion matrix for the conducted test. The recognition rates are balanced across all three emotional labels, with defeated/frustrated slightly higher than the other two labels. It can be seen that as the number of postures increases, the recognition rate improves. It is possible that a more balanced database of sample can result into more balanced and improved recognition rates.



		Predicted		
Back Propagation Algorithm		Concentrated	Defeated/ Frustrated	Triumph
	Concentrated	16 (55%)	8	5
Actual	Defeated/ Frustrated	6	45(65%)	18
	Triumph	5	15	29(59%)

Figure 26: Recognition rates for new postures using the back-propagation algorithm

Another automatic emotion recognition test evaluated the capacity to recognise emotions on new players. The data was separate to two sets (a) the first set comprises 108 postures that was captured by four players (b) the second test includes 39 postures captured by two separate players. The first set was used to train the back-propagation algorithm, while the second one was used as test data. Figure 27 presents the confusion matrix for the conducted test. The correctly classified postures were 22 out of the 39, resulting in 56.4% overall recognition rate. Recognition rate is almost the same for triumph and defeated/frustrated labels, but is significantly smaller for concentrating (42%). However, this is expected due to the small amount of training data for the new concentrating label, which was reduced to 17, as 12 of the 29 are now part of the test data. In general, it appears that the database again performs above chance level and can generalise to new players, although the training data sample is small.



		Predicted		
Back Propagation Algorithm		Concentrated	Defeated/ Frustrated	Triumph
Actual	Concentrated	5 (42%)	6	1
	Defeated/ Frustrated	1	10(62%)	5
	Triumph	0	4	7(64%)

Figure 27: Recognition rates for new players using the back-propagation algorithm

6.1 Conclusion

The current work presents initial results towards recognising emotions based on posture data captured using Microsoft Kinect. To our knowledge, there is no previous research that uses data captured from Kinect to recognise emotions. The results are similar to those in existing literature that use traditional motion capture equipment. This suggests that Kinect provides sufficiently valid data to construct such a model, which can be used in todays games. Improved recognition of concentrating labels compared to [27] can be attributed to smaller emotion set (three instead of four) or to better observer annotation due to the provision of camera input apart from skeleton posture data.

Compared to the benchmark rate taken from the observers in the current study, the overall recognition was 11.08% lower. This again, is probably due to the fact that observers had the advantage of camera input, while the algorithm uses only skeleton data. It would be interesting to test whether observers agreement is influenced by this, as in [27] agreement is lower than in our experiment. This could test and indicate the superiority that a multimodal system may have against single modalities such as posture or animation data.The recognition system yielded satisfactory results at the current build. However, we believe that it can be further optimised.

6.2 Summary of Achieved Objectives

As we have described in chapter 1.2 we have investigated how to construct a database of postures labeled with emotions. We did this for non-acted data. We have recorded data from single player games to multiplayer games in order to capture the different expressiveness in each category. We have used non-intrusive interfaces, Microsoft Kinect, in order to show that we can use emotion in todays games with todays devices used. The performans of Microsoft Kinect device similarly to other motion capture systems is quite good and can be used with some extra modifications on the software.

6.3 Project Issues(with proposed Solutions-Actions)

- The data captured were not equal balanced in 3 experiments for correct testing. For future experiments the same amount of data will be captured in all the experiments to keep consistency.

- Participating players often moved outside the field of view of the kinect causing difficulty to the kinect sensor to collect skeleton data (appeared noisy). For future experiments we can define boundaries for the game, to the players as long as to the software too. Another way is to increase the line of sight of the kinect, by modify the field of view of the kinect's software.

- We have recorded players alone and multiplayer, but the recording was made only for the one skeleton in multiplayer. For future experiments both of the skeletons will be recorded , to capture more data.

- The observation software did not have an automatic extraction of annotate postures in order to be used in WEKA. We can implement in future automatic extraction of annotate postures from the observers software.

- we have used back-propagation classification algorithm, in future we can use , test and compare more algorithms.

6.4 Future Work

Currently the system does not make use of mirror postures. Adding such functionality could offer slight improvements to the recognition rate. Moreover, the training data size can be adjusted to improve performance. Further to this, the current data was collected using specific sports games. It would be interesting to see if the constructed database can recognise emotions captured during game playing of different game genres. We are also looking to capture acted posture data and compare the benchmark and recognition of acted, non-acted and hybrid databases. Further to this, the captured and labelled posture data needs to be further studied. It is possible to detect correlation of rotational data for specific joints and emotion expression. Such patterns can also be present in energy patterns. For this, we plan to investigate animation data instead of single posture. Finally, we plan to test more machine learning algorithms for their ability to build emotion recognition models with the given database.

Bibliography

[1] Ioannou S, Raouzaiou A, Tzouvaras V, Mailis T, Karpouzis K, Kollias S (2005) Emotion recognition through facial expression analysis based on a neurofuzzy method. Neural Netw 18:423-435

[2] Caridakis G, Malatesta L, Kessous L, Amir M, Raouzaiou A, Karpouzis K (2006) Modeling naturalistic affective states via facial and coval expression recognition. In: International conference on multimodal interfaces, pp 146-154

[3] Kapoor A, Burleson W, Picard RW (2007) Automatic prediction of frustration. Int J Hum-Comput Stud 65(8):724736

[4] Devillers L, Vasilescu I (2006) Real-life emotions detection with lexical and paralinguistic cues on human-human call center di- alogs. In: International conference on spoken language processing

[5] Valstar MF, Gunes H, Pantic M (2007) How to distinguish posed from spontaneous smiles using geometric features. In: ACM inter- national conference on multimodal interfaces (ICMI07), Nagoya, Japan, pp 3845

[6] Littlewort GC, Bartlett MS, Lee K (2007) Faces of pain: auto- mated measurement of spontaneous facial expressions of genuine and posed pain. In: International conference of multimodal inter- faces, pp 1521

[7] Pantic M, Patras I (2006) Dynamics of facial expression: recog- nition of facial actions and their temporal segments from face profile image sequences. IEEE Trans Syst Man Cybern, Part B 36(2):433449

[8] K.R. Scherer, Studying the Emotion-Antecedent Appraisal Process: An Expert System Approach, Cognition and Emotion, vol. 7, pp. 325-355, 1993.

[9] Ekman, P., Friesen, W.: Nonverbal behaviour in pschotherapy research. Research in Pschotherapy 3, 179216 (1968)

[10] Sonny Schreurs: The Road to Adaptive Gameplay: Early Insights into the Link between Personality and Experienced Game Entertainment, (2011)

[11] Fang, X., & Zhao, F. Personality and enjoyment of computer game play. Computers in Industry, 61, 342-349. (2010)

[12] H. H. Lund, T. Klitbo, and C. Jessen. Playware technology for physically activating play. Artifical Life and Robotics Journal, 9(4):165174, 2005.

[13] G. N. Yannakakis and J. Hallam. Entertainment Modeling through Physiology in Physical Play. International Journal of Human-Computer Studies, 66:741755, October 2008.

[14] S. Turkle, The Second Self: Computers and the Human Spirit. Simon & Schuster, 1984.

[15] Robison, J., McQuiggan, S., Lester, J.: Evaluating the consequences of affective feedback in intelligent tutoring systems, pp. 16 (2009)

[16] DMello, S., Taylor, R.S., Graesser, A.: Monitoring Affective Trajectories during Complex Learning. In: Proceedings of the 29th Annual Meeting of the Cognitive Science Society, Austin, TX, pp. 203208 (2007)

[17] Litman, D., Forbes, K.: Recognizing Emotions from Student Speech in Tutoring Dialogues. In: Proceedings of the ASRU 2003 (2003)

[18] DMello, S., Jackson, T., Craig, S., Morgan, B., Chipman, P., White, H., Person, N., Kort, B., el Kaliouby, R., Picard, R.W., Graesser, A.: AutoTutor Detects and Responds to Learners Affective and Cognitive States. In: Workshop on Emotional and Cognitive Issues at the International Conference of Intelligent Tutoring Systems (2008)

[19] Ekman, P., Friesen, W.V.: Facial Action Coding System: a technique for the measurement of facial movement. Consulting Psychologists Press, Palo Alto (1978)

[20] Kaliouby, R., Robinson, P.: Real-time Inference of Complex Mental States from Facial Expressions and Head Gestures. In: Real-Time Vision for Human-Computer Interaction, pp. 181200. Springer, Heidelberg (2005)

[21] Sobol-Shikler, T., Robinson, P.: Classification of complex information: inference of co- occurring affective states from their expressions in speech. IEEE Transactions on Pattern Analysis and Machine Intelligence 32, 12841297 (2010)

[22] Pfister, T., Robinson, P.: Speech emotion classification and public speaking skill assessment. In: Salah, A.A., Gevers, T., Sebe, N., Vinciarelli, A. (eds.) HBU 2010. LNCS, vol. 6219, pp. 151162. Springer, Heidelberg (2010)

[23] Sharma, R., Pavlovic, V.I., Huang, T.S.: Toward a multi- modal human computer interface. In: Beun, R.-J. (ed.) Multimodal Cooperative Communication, pp. 89112. Springer, Heidelberg (2001)

[24] Bernhardt, D., Robinson, P.: Detecting Affect from non-stylised body motions. In Proceedings of ACII07, 2007

[25] Coulson, M.: Attributing Emotion to Static Body Postures: Recognition accuracy, confusions, and viewpoint difference. Journal of Nonverbal Behaviour, 28, 117-139, 2004.

[26] Castellano, G., Villalba, S., Camurri, A.: Recognising human emotions from body movement and gesture dynamics. Affective Computing and Intelligence Interaction, LNCS 4738, 71-82, 2007.

[27] KLEINSMITH, A., BIANCHI-BERTHOUSE, N.: Modelling non-acted affective posture in a video game scenario, Proc. Intl Conf. Kansei Eng. Emotion Res., pp. 19641974, 2010.

[28] KLEINSMITH, A., DE SILVA, R., BIANCHI-BERTHOUSE, N.: Cross-Cultural Differences in recognising affect from body posture, Interacting with Computers, 18, 6, 1371-1380, 2006.

[29] KLEINSMITH, A., BIANCHI-BERTHOUSE, N., STEED, N.: Automatic Recognition of Non-acted postures, IEEE Transactions of Systems, Man and Cybernetics, Part B, 2011.

[30] MANDLER, G.: History of Psychology. Chapter 8: Emotion, Vol. 1, Wiley, 2002.

[31] HALL, M., FRANK, E., HOLMES, G., PFAHRINGER, B., REUTEMANN, P., WITTEN, I.: The WEKA Data Mining Software: An Update, SIGKDD Explorations, Volume 11, Issue 1, 2009

[32] Kinect controller for Xbox360, Microsoft, http://www.xbox.com/en-US/kinect

[33] SAVVA, N., BIANCHI-BERTHOUSE, N.: Automatic Recognition of affective body movement in a video game scenario. International Conference on Intelligent Technologies for interactive entertainment. (Vol LNICST 78 pp.149-158). Springer, 2012.

[34] Picard, Rosalin. Affective Computing. United States. The MIT Press. 1998

[35] MacLean, P. The triune brain, emotion, and scientific bias. In Schmitt, F. (editor): The Neurosciences: Second Study Program, pages 336-349. New York, United States. Rockefeller University Press. 1970

[36] G. N. Yannakakis, Game Adaptivity Impact on Affective Physical Interaction, in Proceedings of the Int. Conf. on Affective Computing and Intelligent Interaction (ACII09), Amsterdam, The Netherlands, September 2009

[37] Zhihong Zeng, A Survey of Affect Recognition Methods: Audio, Visual, and Spontaneous Expressions, IEEE Transaction on pattern analysis and machine intelligence , VOL. 31, NO. 1, JANUARY 2009

[38] Carletta, Jean. Assessing agreement on classification tasks: The kappa statistic. Computational Linguistics, 22(2), pp. 249254, 1996

[39] Han, J., & Kamber, M. Data mining: Concepts and techniques. San Francisco: Morgan Kaufmann.(2001)

[40] Russell, Stuart J.; Norvig, Peter, Artificial Intelligence: A Modern Approach (2nd ed.), Upper Saddle River, New Jersey: Prentice Hall (2003)

[41] S. Mota and R. Picard, Automated Posture Analysis for Detecting Learners Interest Level, Proc. Computer Vision and Pattern Recognition Workshop, vol. 5, p. 49, 2003.

[42] Y.L. Tian, T. Kanade, and J.F. Cohn, Facial Expression Analysis, Handbook of Face Recognition, S.Z. Li and A.K. Jain, eds., pp. 247-276, Springer, 2005.

[43] Y. Zhang and Q. Ji, Active and Dynamic Information Fusion for Facial Expression Understanding from Image Sequences, IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 27, no. 5, pp. 699-714, 2005.

[44] L. Cohen, N. Sebe, A. Garg, L. Chen, and T. Huang, Facial Expression Recognition from Video Sequences: Temporal and Static Modeling, Computer Vision and Image Understanding, vol. 91, nos. 1-2, pp. 160-187, 2003.

[45] H. Tao and T.S. Huang, Explanation-Based Facial Motion Tracking Using a Piecewise Bezier Volume Deformation Mode, Proc. IEEE Intl Conf. Computer Vision and Pattern Recognition (CVPR 99), vol. 1, pp. 611-617, 1999.

[46] C.M. Lee and S.S. Narayanan, Toward Detecting Emotions in Spoken Dialogs, IEEE Trans. Speech and Audio Processing, vol. 13, no. 2, pp. 293-303, 2005.

[47] J. Ledoux, The Emotional Brain: The Mysterious Underpinnings of Emotional Life. Simon & Schuster, 1998.

[48] M.H. Immordino-Yang and A. Damasio, We Feel, Therefore We Learn: The Relevance of Affective and Social Neuroscience to Education, Mind, Brain, and Education, vol. 1, pp. 3-10, 2007.

[49] A. Bashashati, M. Fatourechi, R.K. Ward, and G.E. Birch, A Survey of Signal Processing Algorithms in Brain-Computer Interfaces Based on Electrical Brain Signals, J. Neural Eng., vol. 4, pp. R32-R57, June 2007.

[50] F. Lotte, M. Congedo, A. Le cuyer, F. Lamarche, and B. Arnaldi, A Review of Classification Algorithms for EEG-Based Brain- Computer Interfaces, J. Neural Eng., vol. 4, pp. R1-R13, 2007.

[51] Giese, M.A., Poggio, T.: Neural mechanisms for the recognition of biological movements. Neuroscience 4, 179191 (2003)

[52] Kapoor, A., Picard, R.W., Ivanov, Y.: Probabilistic combination of multiple modalities to detect interest. In: 17th Int. Conf. Pattern Recog., pp. 969972 (2004)

[53] Zhihong Zeng, Jilin Tu, Brian Pianfetti, Thomas S. Huang, Audio-visual Affective Expression Recognition through Multi-stream Fused HMM, IEEE Transactions on Multimedia 2008

[54] N. Dalal, B. Triggs, Histogram of Oriented Gradients for Human Detection, CVPR 2005.

[55] T. Cootes, C. Taylor, D. Cooper and J. Graham, Active Shape Models Their Training and Application, Computer Vision and Image Understanding, 1995.

[56] J. Kovac, P. Peer, F. Solina, Human Skin Colour Clustering for Face Detection, EUROCON Computer as a Tool, 2003.

Appendix A

Weka Sample Data file

```
    attribute rot_hipc_x numeric
attribute rot_hipc_y numeric
attribute rot_hipc_z numeric
attribute rot_spine_x numeric
attribute rot_spine_y numeric
attribute rot_spine_z numeric
attribute rot_shoulderc_x numeric
attribute rot_shoulderc_y numeric
attribute rot_shoulderc_z numeric
attribute rot_head_x numeric
attribute rot_head_y numeric
attribute rot_head_z numeric
attribute rot_shoulderl_x numeric
attribute rot_shoulderl_y numeric
attribute rot_shoulderl_z numeric
attribute rot_elbowl_x numeric
attribute rot_elbowl_y numeric
attribute rot_elbowl_z numeric
attribute rot_wristl_x numeric
attribute rot_wristl_y numeric
attribute rot_wristl_z numeric
attribute rot_handl_x numeric
attribute rot_handl_y numeric
attribute rot_handl_z numeric
attribute rot_shoulderr_x numeric
attribute rot_shoulderr_y numeric
attribute rot_shoulderr_z numeric
attribute rot_elbowr_x numeric
attribute rot_elbowr_y numeric
attribute rot_elbowr_z numeric
attribute rot_wristr_x numeric
attribute rot_wristr_y numeric
attribute rot_wristr_z numeric
attribute rot_handr_x numeric
attribute rot_handr_y numeric
attribute rot_handr_z numeric
attribute rot_hipl_x numeric
attribute rot_hipl_y numeric
attribute rot_hipl_z numeric
attribute rot_kneel_x numeric
attribute rot_kneel_y numeric
attribute rot_kneel_z numeric
attribute rot_anklel_x numeric
attribute rot_anklel_y numeric
attribute rot_anklel_z numeric
attribute rot_footl_x numeric
attribute rot_footl_y numeric
attribute rot_footl_z numeric
attribute rot_hipr_x numeric
attribute rot_hipr_y numeric
attribute rot_hipr_z numeric
attribute rot_kneer_x numeric
attribute rot_kneer_y numeric
attribute rot_kneer_z numeric
attribute rot_ankler_x numeric
attribute rot_ankler_y numeric
attribute rot_ankler_z numeric
attribute rot_footr_x numeric
attribute rot_footr_y numeric
attribute rot_footr_z numeric

    attribute emotion CONCETRATING,DEFEATED_FRUSTRATED,TRIUMPHANT

    data

    @data 0.254314,0.05976069,3.58193,0.2502755,0.1161106, ......,CONCENTRATING
0.1116866,0.2802979,3.644001,0.1239048,0.346866,3.697731,...........,DEFEATED_FRUSTRATED
..........................
..........................
0.0888944,0.2455028,3.554908,0.09765651,0.3049689,3.6339,.................,CONCENTRATING
0.1386756,0.4298375,2.981369,0.1636763,0.4967238,3.04132,.................,TRIUMPHANT
```

Appendix B

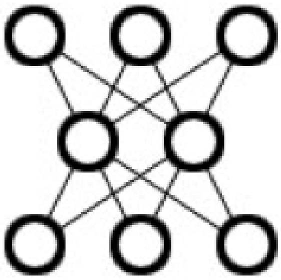
Back propagation algorithm

In many real world situations, we are faced with incomplete or noisy data, and it is important to be able to make reasonable predictions about what is missing from the information available. This can be an especially difficult task when there isn't a good theory available to help reconstruct the missing data. It is in such situations that Back propagation networks may provide some answers.

A Back Propagation network consists of at least three layers of units: an input layer, at least one intermediate hidden layer, and an output layer . Connection weights in a BackPropagation network are oneway. Typically, units are connected in a feed-forward fashion with input units fully connected to units in the hidden layer and hidden units fully connected to units in the output layer. When a Back Propagation network is cycled, an input pattern is propagated forward to the output units through the intervening input-to-hidden and hidden-to-output weights.We can interpret the output of a Back Propagation network as a classification decision.

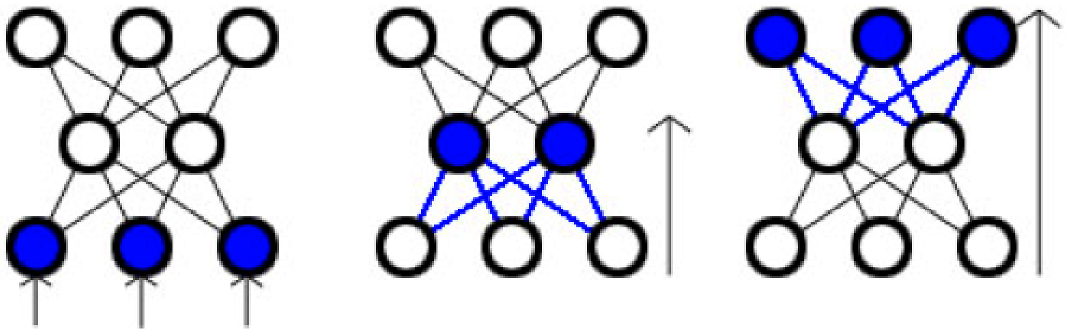
With Back Propagation networks, learning occurs during a training phase in which each input pattern in a training set is applied to the input units and then propagated forward. The pattern of activation arriving at the output layer is then compared with the correct (associated) output pattern to calculate an error signal. The error signal for each such target output pattern is then backpropagated from the outputs to the inputs in order to appropriately adjust the weights in each layer of the network. After a Back Propagation network has learned the correct classification for a set of inputs, it can be tested on a second set of inputs to see how well it classifies untrained patterns. Thus, an important consideration in applying Back Propagation learning is how well the network generalizes.

The backpropagation algorithm is a multi-layer network using a weight adjustment based on the sigmoid function, like the delta rule. The backpropagation method is example of supervised learning, where the target of the function is known.The following is an example of the backpropagation algorithm working on a small Artificial Neural Network.

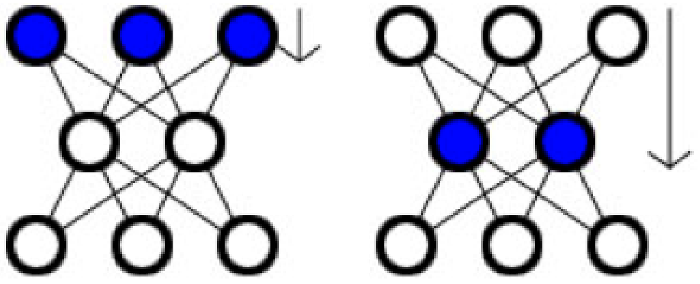


The Network has a single hidden layer of size two and input and output nodes of size 3.

Initialize the weighted links. Typically the weights are initialized to a small random number.Then, for each training example in the testing set:



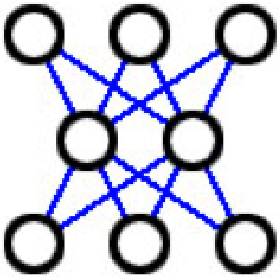
Input the training data to the input nodes, then calculate Ok, which is the output of node k. This is done for each node in the hidden layer(s) and output layer.



Then calculate k for the each output node, where tk is the target of the node and calculate k for each hidden node:

$$\delta k \leftarrow Ok(1 - Ok)(tk - Ok)$$

$$\delta k \leftarrow Ok(1 - Ok)\sum Wh, k \cdot \delta k$$



Finally adjust the weights of all the links, where xi is the activation and η is the learning rate:

Wi,j $\leftarrow$ Wij + $\eta\delta$Xi

Most likely the neural network will need to be trained at many iterations of the training set to find an acceptable approximation of the function it is being trained on.

Appendix C

Paired Agreement among all observers - Kappa

A: Triumph
B: Concentrating
C: Defeated-Frustrated
Quantify agreement with kappa

Observer 1 Observer 2

	A	B	C	Total
A	44	0	5	49
B	2	24	3	29
C	3	5	61	69
Total	49	29	69	147

Number of observed agreements: 129 (87.76% of the observations)
Number of agreements expected by chance: 54.4 (37.04% of the observations)
Kappa= 0.806
SE of kappa = 0.043
95% confidence interval: From 0.721 to 0.890
The strength of agreement is considered to be ’very good’.

Observer 1 Observer 3

	A	B	C	Total
A	42	0	7	49
B	3	22	4	29
C	7	3	59	69
Total	52	25	70	147

Number of observed agreements: 123 (83.67% of the observations)
Number of agreements expected by chance: 55.1 (37.50% of the observations)
Kappa= 0.739
SE of kappa = 0.049
95% confidence interval: From 0.643 to 0.835
The strength of agreement is considered to be 'good'.

Observer 1- Observer 4

	A	B	C	Total
A	41	3	5	49
B	5	21	5	31
C	2	3	60	65
Total	48	27	70	145

Number of observed agreements: 122 (84.14% of the observations)
Number of agreements expected by chance: 53.4 (36.81% of the observations)
Kappa= 0.749
SE of kappa = 0.047
95% confidence interval: From 0.656 to 0.842
The strength of agreement is considered to be 'good'.

Observer 2 Observer 3

	A	B	C	Total
A	44	0	5	49
B	3	22	4	29
C	5	3	61	69
Total	52	25	70	147

Number of observed agreements: 127 (86.39% of the observations)
Number of agreements expected by chance: 55.1 (37.50% of the observations)
Kappa= 0.782
SE of kappa = 0.045
95The strength of agreement is considered to be 'good'.

Observer 2 Observer 4

Number of observed agreements: 133 (90.48% of the observations)
Number of agreements expected by chance: 54.0 (36.76% of the observations)
Kappa= 0.849

	A	B	C	Total
A	44	4	1	49
B	3	24	2	29
C	1	3	65	69
Total	48	31	68	147

SE of kappa = 0.038
95% confidence interval: From 0.775 to 0.923
The strength of agreement is considered to be ’very good’.

Observer 3 Observer 4

	A	B	C	Total
A	41	4	7	52
B	3	19	3	25
C	4	8	58	70
Total	48	31	68	147

Number of observed agreements: 118 (80.27% of the observations)
Number of agreements expected by chance: 54.6 (37.17% of the observations)
Kappa= 0.686
SE of kappa = 0.052
95% confidence interval: From 0.585 to 0.787
The strength of agreement is considered to be ’good’.

Appendix D

Generalising for new postures: WEKA Back Propagation Built Model and Results

```
    === Run information ===
    Scheme:weka.classifiers.functions.MultilayerPerceptron -L 0.3 -M 0.2 -N 500 -V 0
-S 0 -E 20 -H a
Relation: experiment2dataExtractedlabeled1-2
Instances: 147
Attributes: 61
rot-hipc-x
rot-hipc-y
rot-hipc-z
rot-spine-x
rot-spine-y
rot-spine-z
rot-shoulderc-x
rot-shoulderc-y
rot-shoulderc-z
rot-head-x
rot-head-y
rot-head-z
rot-shoulderl-x
rot-shoulderl-y
rot-shoulderl-z
rot-elbowl-x
rot-elbowl-y
rot-elbowl-z
rot-wristl-x
rot-wristl-y
rot-wristl-z
rot-handl-x
rot-handl-y
```

```
rot-handl-z
rot-shoulderr-x
rot-shoulderr-y
rot-shoulderr-z
rot-elbowr-x
rot-elbowr-y
rot-elbowr-z
rot-wristr-x
rot-wristr-y
rot-wristr-z
rot-handr-x
rot-handr-y
rot-handr-z
rot-hipl-x
rot-hipl-y
rot-hipl-z
rot-kneel-x
rot-kneel-y
rot-kneel-z
rot-anklel-x
rot-anklel-y
rot-anklel-z
rot-footl-x
rot-footl-y
rot-footl-z
rot-hipr-x
rot-hipr-y
rot-hipr-z
rot-kneer-x
rot-kneer-y
rot-kneer-z
rot-ankler-x
rot-ankler-y
rot-ankler-z
rot-footr-x
rot-footr-y
rot-footr-z
emotion
Test mode:10-fold cross-validation

    === Classifier model (full training set) ===

    Sigmoid Node 0
Inputs Weights
```

```
Threshold -2.0301581832788984
Node 3 -1.999636175877659
Node 4 -4.180808763118338
Node 5 -2.343652261247904
Node 6 -0.7383641978652318
Node 7 -0.9373216238668436
Node 8 -2.866576741713843
Node 9 -2.289289743173148
Node 10 2.2191492337867498
Node 11 -1.7143395098795269
Node 12 -1.256272882048777
Node 13 -1.9923460406403113
Node 14 5.9760276958741505
Node 15 0.20962452277750768
Node 16 -1.618454327017899
Node 17 4.6666903934809305
Node 18 2.3593821620797586
Node 19 -3.501634629466943
Node 20 1.03471446431564
Node 21 7.901061627496437
Node 22 -1.9338812079030867
Node 23 -0.5090498769525801
Node 24 -2.509435635211622
Node 25 -3.7071150831008115
Node 26 6.244360286947088
Node 27 -1.323578455965381
Node 28 -1.449737882052147
Node 29 -1.9883997516401744
Node 30 -3.391602388482881
Node 31 4.394462286682217
Node 32 -1.0967742160852532
Node 33 -1.7766185460514905

   Sigmoid Node 1
Inputs Weights
Threshold -1.0819305987717072
Node 3 0.5973619222930937
Node 4 -5.248845698270944
Node 5 1.9080557194997498
Node 6 0.2375157903935067
...............................................
...............................................
Node 33 -4.120244240573109
```

```
    Sigmoid Node 2
Inputs Weights
Threshold 0.7146986774732528
Node 3 0.6879588822135472
Node 4 6.390612474431644
Node 5 -0.014357903342767032
Node 6 -0.04483549854622705
..............................................
..............................................
Node 33 3.0454169030892007

    Sigmoid Node 3
..............................................
..............................................
Sigmoid Node 33

    Class CONCENTRATING
Input
Node 0
Class DEFEATED-FRUSTRATED
Input
Node 1
Class TRIUMPHANT
Input
Node 2

    Time taken to build model: 3.79 seconds

    === Stratified cross-validation ===
=== Summary ===

    Correctly Classified Instances 90 61.2245 %
Incorrectly Classified Instances 57 38.7755 %-
Kappa statistic 0.3833
Mean absolute error 0.2701
Root mean squared error 0.4659
Relative absolute error 64.2597%
Root relative squared error 101.6728 %
Total Number of Instances 147
```

Table 4: Detailed Accuracy By Class

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.552	0.093	0.593	0.552	0.571	0.854	CONCENTRATING
	0.652	0.295	0.662	0.652	0.657	0.741	DEFEATED-FRUSTRATED
	0.592	0.235	0.558	0.592	0.574	0.741	TRIUMPHANT
Weighted Avg.	0.612	0.235	0.613	0.612	0.613	0.763	

Table 5: Confusion Matrix

a	b	c	classified as
16	8	5	a = CONCENTRATING
6	45	18	b = DEFEATED-FRUSTRATED
5	15	29	c = TRIUMPHANT

Appendix E

Informed Consent Form

The Department of Computer Science of University of Cyprus supports the practice of protection of human participants in research. The following will provide you with information about the experiment that will help you in deciding whether or not you wish to participate. If you agree to participate, please be aware that you are free to withdraw at any point throughout the duration of the experiment.

In this study we will ask you to play a Microsoft Kinect Game called Kinect Sports. If you have any back problem issues or any other injury, please inform the experimenter and the study will end now. All information you provide will remain confidential and will not be associated with your name. If for any reason during this study you do not feel comfortable, you may leave the laboratory and receive credit for the time you participated and your information will be discarded. Your participation in this study will require approximately 30-40 minutes. When this study is complete you will be provided with the results of the experiment if you request them, and you will be free to ask any questions. If you have any further questions concerning this study please feel free to contact us: Haris Zacharatos at University of Cyprus. Please indicate with your signature on the space below that you understand your rights and agree to participate in the experiment.

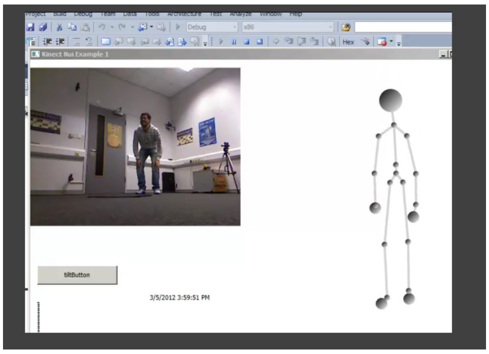
Your participation is solicited, yet strictly voluntary. All information will be kept confidential and your name will not be associated with any research findings.

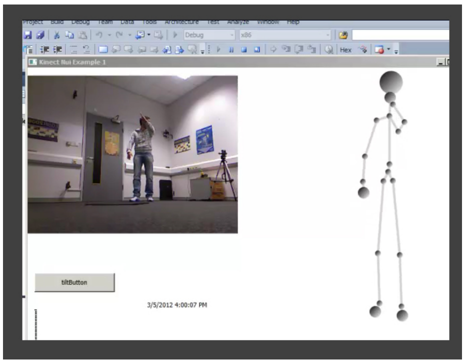
Signature of Participant

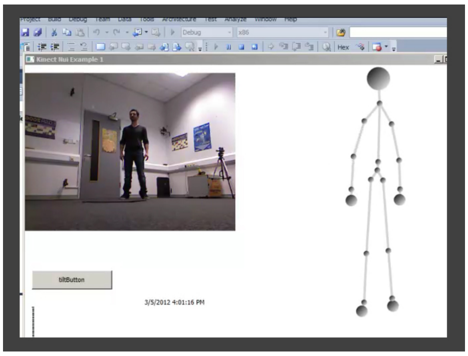
Appendix F

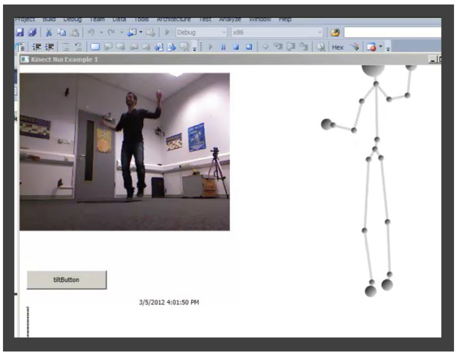
Affective posture observation quiz - sample

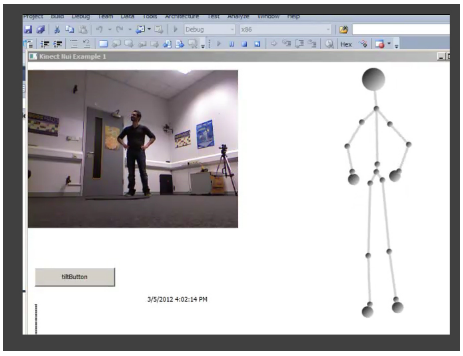
Affective Posture Experiments



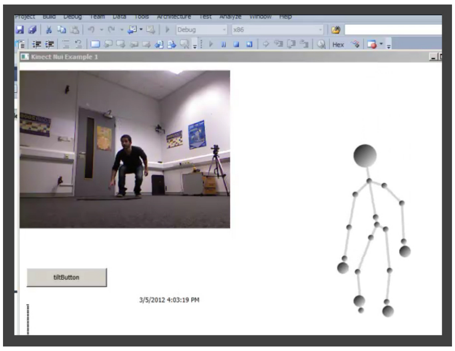




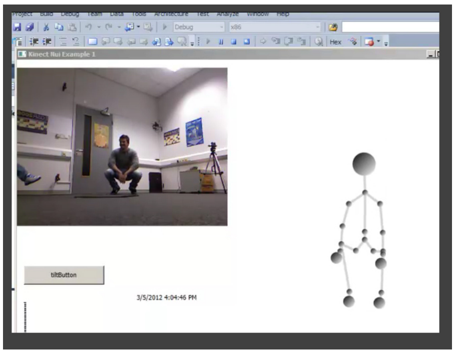




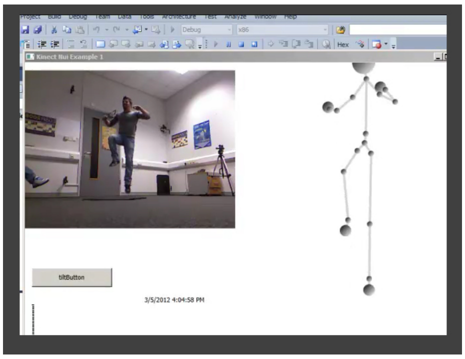
3/5/2012 4:02:14 PM



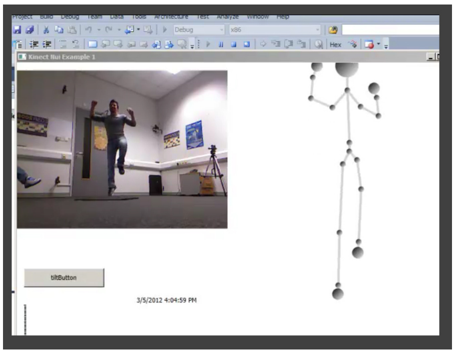
3/5/2012 4:03:19 PM



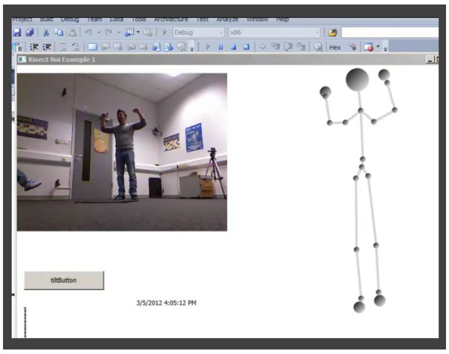
3/5/2012 4:04:46 PM



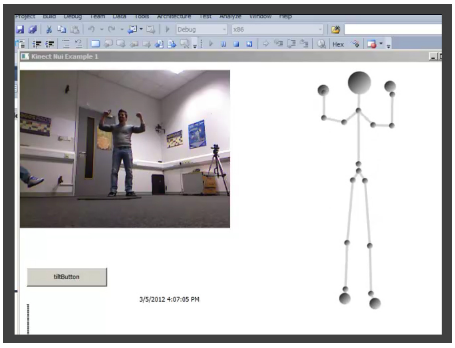
3/5/2012 4:04:58 PM



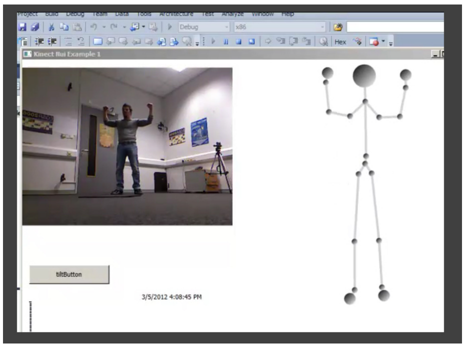
3/5/2012 4:04:59 PM



3/5/2012 4:05:12 PM



3/5/2012 4:07:05 PM



3/5/2012 4:08:45 PM