Human Motion Analysis & Art and Music Generation

By

Shahqaan Qasim   *2008-NUST-BIT-54*

Department of Electrical Engineering
School of Electrical Engineering & Computer Science
National University of Sciences and Technology
Islamabad, Pakistan
2013

Human Motion Analysis & Art and Music Generation

By

Shahqaan Qasim   **2008-NUST-BIT-54**

**A Project report submitted in partial fulfillment
Of the requirement for the degree of
Bachelors in Information Technology**

Department of Electrical Engineering

School of Electrical Engineering & Computer Science
National University of Sciences & Technology
Islamabad, Pakistan
2013

CERTIFICATE

It is certified that the contents and form of thesis entitled "Human Motion Analysis & Art and Music Generation" submitted by Shahqaan Qasim (*2008-NUST-BIT-54*) have been found satisfactory for the requirement of the degree.

Advisor: _____

(Dr. Rehan Hafiz)

Co-Advisor: _____

(Mr. Maajid Maqbool)

DEDICATION

To Allah the Almighty

&

To my Parents and Faculty

ACKNOWLEDGEMENTS

Table of Contents

Human Motion Analysis and Art and Music Generation

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Since the release of Kinect, there have been many efforts from Microsoft, as well as, various other open source projects to track human motion. Many frameworks exist which give the developer an abstraction from tracking human movement and instead, provide them with parameters such as hand gestures. We aim to take it a step further. Using such frameworks, we analyzed hand and leg movements, combined with whole body's motion to get a score of user's emotions. Using the result of this analysis we developed a framework which other developers can use. While using this framework, they not only have hand and body movement data, but also the emotions of users which they can utilize in their applications. Applications can include creation of music and arts for the sake of advertisement, interactive installations and games which either use emotions as an input event or for studying emotional changes as a reaction to a certain event.

The second part of the project focuses on developing an application which demonstrates the full potential of this framework. We created a desktop application which will create real-time music and art based on the user performing intentionally or randomly. The music scale and saturation of the colors used is based upon the emotions of the performer.

Chapter 1

1 INTRODUCTION

Human beings are the most complex living organisms. Despite belonging to the mammals group, they are capable of evolving on such a fast scale and are able to redefine the aspects of their life through various means. Over the decades, human motions have shown such diversity that many scientists are trying to analyze and manipulate this knowledge to provide some benefit to the human race. In this regard, there are researches and attempts to read patterns in human motions and to use them to generate something useful such as art. This is a very promising avenue and can open doorways to more research and development.

DEFINING ART AND HUMAN MOTIONS

Art is defined as the expression or application of human creative skill and imagination. Art combined with human motions can be described and can be taken in various meanings such as people dancing or an artifact depicting human motion.

IMPORTANCE

The application of human motion analysis and utilization are vast. Some of the applications of such development can be seen as follows:

Table 1: Applications of Human Motion Analysis

Subject & environment Monitored	Parameters calculated	Output	Analysis/Application Domain
Person performing intentionally.	• Hands, arms, legs and body movement. • Mood and gender.	Animation and music generated and color calculated .	Fun sake, dance performance, advertisement, gaming, one minute competition...what else?
Person performing	• Hands, arms,	Pre-generated 3D	• Used as

Human Motion Analysis and Art and Music Generation

intentionally or randomly.	legs and body movement. • Mood and gender	scene, such as wind, rain, movement of fish and blooming of flowers controlled by input parameters.	installment in events, festivals, hotels' lobby and museum. • Uploading videos on social networking sites.
Person agrees to be monitored at workplace and/or home.	• Hands, arms, legs and body movement. • Mood and gender	An animation or picture generated.	Animation and picture can be uploaded to social networking websites. Or we can make a Facebook page where people upload such videos/photos and people rate each of them.
A group of people or even other objects moving in a specific area such as outside a building or at traffic signal.	Analyze movement of herds of objects?	An animation or picture.	Such a system can be installed outside a building which projects (maybe we can use projection mapping as well) movement of people inside/outside that biulding.

PROJECT GOAL

The goal of the project was to:

1. Develop a framework that gathers mood and motion data. It means that our module will:

   • Capture human motion

   • Carry out mood/emotional analysis on it

   • And present the results in a well-formed, consistent manner so that it's a breeze to use the module

2. Write a demo application that shows full potential of this framework

Chapter 2

2   Literature Review

Before starting to jump directly into developing this system, there was a need to see what had already been developed in the fields related to human motion analysis. In this regard, following papers were read and some conclusions were drawn depending on them:

Human Motion Analysis

2.1.1   Recognition and Re-synthesis of Human Motion with Personalized Variations

The purpose of this research paper was to analyze and recognize various human motions such as walking, jumping etc. It uses Hidden Markov Model recognize motion. This method was successful at recognizing different motions from a scene as well recognizing gender and mood of the actor with 100% accuracy.

In this research paper human motion data was gathered by infrared sensors placed at strategic locations on a human body but I chose to not write the details of the data gathering process as we are using Kinect and we will already have human motion data in the form of joints.

This paper also addressed the problem of transforming one type of motion into another. They used two different approaches to implement this and both were successful in transforming a male walk into a female walk.

2.1.2   3D Human Action Recognition and Style Transformation Using Resilient Back propagation Neural Networks

This paper was published by the same authors as above but it uses Resilient Back propagation Neural Networks instead of Hidden Markov Model to implement the same principles as above.

Chapter 2

2 Literature Review

Before starting to jump directly into developing this system, there was a need to see what had already been developed in the fields related to human motion analysis. In this regard, following papers were read and some conclusions were drawn depending on them:

██ Human Motion Analysis

2.1.1 Recognition and Re-synthesis of Human Motion with Personalized Variations

The purpose of this research paper was to analyze and recognize various human motions such as walking, jumping etc. It uses Hidden Markov Model recognize motion. This method was successful at recognizing different motions from a scene as well recognizing gender and mood of the actor with 100% accuracy.

In this research paper human motion data was gathered by infrared sensors placed at strategic locations on a human body but I chose to not write the details of the data gathering process as we are using Kinect and we will already have human motion data in the form of joints.

This paper also addressed the problem of transforming one type of motion into another. They used two different approaches to implement this and both were successful in transforming a male walk into a female walk.

2.1.2 3D Human Action Recognition and Style Transformation Using Resilient Back propagation Neural Networks

This paper was published by the same authors as above but it uses Resilient Back propagation Neural Networks instead of Hidden Markov Model to implement the same principles as above.

2.1.3 Conclusion

I have read, at a very abstract level both HMM and neural networks but both are fairly complex so a comparison at this time is not possible. I think we can decide on the algorithm to be used in the implementation phase after we know the exact form in which we have the data that is to be analyzed.

As far as re-synthesis is concerned, I don't think we need re-synthesis as mentioned in both of these research papers. We are to create totally different form of artifacts from our motion analysis but gathering mood and gender can come in very handy.

Color Detection and Optimization

2.2.1 Aesthetic Coloring for Complex Layout Using Genetic Algorithm

Various research papers focused on generating a final color palette which an artist uses to choose a color from. But this paper solved the problem of generating an optimized color scheme based on a certain input colors. This paper relies on Moon (G.D. Birkhoff. Aesthetic Measure. Harvard University Press, Cambridge, MA, USA, 1933) and Spencer (P. Moo n, D.E. Spencer. Aesthetic measure applied to color harmony. Journal of the Optical Society of America, vol. 34, Apr. 1944, pp. 234-242.) color harmony model which is based on psychological experiments. They argue that Genetic Algorithms the best method to solve this kind of problem.

In 1928, Birkhoff formalized the notion of beauty by the introduction of the aesthetic measure, defined as the ratio between order and complexity. Based on this measure Moon and Spencer proposed a quantitative model of color harmony, using color difference and area factor based on psychological factors.

Implementation is carried out using three phases. In the first phase image they read and evaluate the color image and initialize genetic algorithm parameters. The program reads the size of image, number of color and color pairs and area of each

2.1.3 Conclusion

I have read, at a very abstract level both HMM and neural networks but both are fairly complex so a comparison at this time is not possible. I think we can decide on the algorithm to be used in the implementation phase after we know the exact form in which we have the data that is to be analyzed.

As far as re-synthesis is concerned, I don't think we need re-synthesis as mentioned in both of these research papers. We are to create totally different form of artifacts from our motion analysis but gathering mood and gender can come in very handy.

Color Detection and Optimization

2.2.1 Aesthetic Coloring for Complex Layout Using Genetic Algorithm

Various research papers focused on generating a final color palette which an artist uses to choose a color from. But this paper solved the problem of generating an optimized color scheme based on a certain input colors. This paper relies on Moon (G.D. Birkhoff. Aesthetic Measure. Harvard University Press, Cambridge, MA, USA, 1933) and Spencer (P. Moo n, D.E. Spencer. Aesthetic measure applied to color harmony. Journal of the Optical Society of America, vol. 34, Apr. 1944, pp. 234-242.) color harmony model which is based on psychological experiments. They argue that Genetic Algorithms the best method to solve this kind of problem.

In 1928, Birkhoff formalized the notion of beauty by the introduction of the aesthetic measure, defined as the ratio between order and complexity. Based on this measure Moon and Spencer proposed a quantitative model of color harmony, using color difference and area factor based on psychological factors.

Implementation is carried out using three phases. In the first phase image they read and evaluate the color image and initialize genetic algorithm parameters. The program reads the size of image, number of color and color pairs and area of each

color (read in pixels). Genetic algorithm parameters for this phase include string size, number of generations, population size, mutation and crossover rate.

In the second phase evaluation of the aesthetic score for each possible solution takes place. This determines the possibility of survival and reproduction of each solution in the following generations.

Phase 3 is population generation. For each generation, three ages of population (parent, child and combined) are created. The best solutions in this combined population regardless of their origin are retained and passed to the following generation as a parent population.

In the experiment conducted, it took them 55 seconds to read an image and search for 6 unique optimized solutions.

2.2.2 Conclusion

There are also basic rules available to create color combinations. For example, if we hard code a certain color palette into out program and round-off each color read in the frame to one of those found in the color palette. This can be easily done in real time. While the solution mentioned in the research paper above is optimum, I think we are going to have a problem implementing that solution in real time.

Psychological Analysis of Human Motion

2.3.1 Affective States

Affect is described as feeling or emotion. Affective states refer to the different states of feelings or emotions.

2.3.2 Difference between Emotion and Mood

Emotion and mood both are types of affective states but emotion is focused whereas mood is unfocused or diffused.

2.3.3 Arousal, Valence and Stance

Arousal is defined as the level of energy a person possesses while displaying a certain emotion. Valence describes how positive or negative the stimuli is (which is causing a certain emotion in a person). Stance describes how approachable the stimuli are.

Together, these three terms form a model of quantitative analysis of emotion depicted by body language.

2.3.4 Gesture

It is a movement of body that contains information.

2.3.5 GEMEP

The Geneva Multimodal Emotion Portrayals (GEMEP[1]) is a collection of audio and video recordings featuring 10 actors portraying 18 affective states, with different verbal contents and different modes of expression. It was created in Geneva by Klaus Scherer and Tanja Bänziger, in the framework of a project funded by the Swiss National Science Foundation (FNRS 101411-100367) and with support of the European Network of Excellence "Humaine" (IST-2002-2.3.1.6 Multimodal Interfaces, Contract no. 507422). Rating studies and objective behavioral analyses are also currently funded by Project 2 and the Methods module of the Swiss Affective Science Center Grant (FNRS).

2.3.6 Importance of Body Language Analysis

There has been thorough research on motion analysis from facial recognition. But there has been little or no research on body language analysis. It is known that body expressions are as powerful as facial recognition when it comes to emotion analysis. In any social interactions, body language along with facial features communicates the mood/emotions of the person.

In chronic pain rehabilitation, specific movements and postural patterns inform about the emotional conflict experienced by the patients (called "guarding behavior") which affects their level of ability to relax. If doctors have a way to know that a person's emotional state is not letting him/her progress in his therapy, then they can formulate ways to treat such patients better.

Students lost motivation when high levels of affective states such as frustration, anxiety or fear are experienced. If such systems are developed which can read the body language of all students present in a class, it can point out when the teacher needs to change his/her tactics.

2.3.7 Survey

The whole point of this research is to answer two questions:

1. What bodily information is necessary for recognizing the emotional state of a person
2. Whether specific features of the body can be identified which contribute to specific emotions in a person

███ The Role of Spatial and Temporal Information in Biological Motion Perception.pdf

In the experiment conducted, 9 human walkers were fitted with point-lights on all major joints. There movements, both to the left and right, were recorded. From the movements, 100 static images were extracted based on the following four configurations:

1. Normal spatial and temporal
2. Scrambled spatial and normal temporal
3. Normal spatial and scrambled temporal
4. Scrambled spatial and temporal

The experiment was conducted by both, an algorithm and a human subject in two further stages. Stage 1 analyzed the spatial structure of the frame by matching with some templates of body shapes. Stage 2 analyzed the temporal arrangement. The task was to find the facing direction (form) of the point-light body and movement (motion) direction of the body.

The results show that form can be recognized when temporal data is scrambled and spatial data is intact. But movement cannot be analyzed when either of the data is scrambled.

Evidence for Distinct Contributions of Form and Motion Information to the Recognition of Emotions from Body Gestures.pdf

The research concluded that motion signals alone are sufficient for recognizing basic emotions.

Visual Perception of Expressiveness in Musician's Body Movements.pdf

Recognizing emotional intentions of a musician by their body movements. The results indicate that:

- Happiness, sadness and anger are well communicated but fear was not
- Anger is indicated by large, fairly fast and jerky movement
- Sadness by fluid and slow movements
- But expressions of the same motion varied greatly depending upon the instrument played

Automated Analysis of Body Movement in Emotionally Expressive Piano Performances.pdf

While playing piano, movement is related to both the musical score that is being played as well as emotional intention conveyed. In the experiment conducted the pianist was asked to play the same musical score with different emotional intentions. Two motion cues were studied using an automated system:

1. Quantity of motion of the upper body
2. Velocity of head movement

The paper states that a comprehensive account of emotional communication should consider the entire path from sender to receiver. On the sender side, emotions are expressed with appearance and behavior by means of cues which can be objectively measured. On the receiver side, these cues are processed based upon the perception of the receiver. Receiver's perception can be affected by many things such as culture and his own mood. So, although emotions perceived by the receiver are based on emotions expressed by the sender, but they are not necessarily equal.

This implies that a comprehensive account of emotion communication requires the conclusion of both expression and perception.

There are some distinctive patterns of movements and postural behavior associated with some of the emotions studied:

1. Lifting shoulders seemed to be typical for *joy* and *anger*
2. Moving shoulders forward is typical for *disgust, despair* and *fear*

Survey of various research papers concluded that head movement plays an important role in communication, as well as, perception of emotion.

For dance performances,

1. Overall duration of time
2. Contraction index
3. Quantity of motion
4. Motion fluency,

showed differences in four emotions: anger, fear, grief and joy. Another research indicated that quantity of motion and contraction index of upper body played a major role in discriminating between different emotions.

2.7.1 Conclusions:

No other emotion, except for sad had any impact on quantity of motion (but this is because of lack of movement space due to piano)

Another research indicates that quantitative analysis of body expressions was also possible. For example, it was concluded that arm was raised 17 degrees higher for angry movements than other emotions and expanded limbs and torso both signify content and joy.

Toward a Minimal Representation of Affective Gestures.pdf

- 12 emotions expressed by 10 actors
- Visual tracking of trajectories of head and hands were performed from frontal and lateral view
- Postural and dynamic expressive gesture features were identified and analyzed

2.8.1 Features tracked:

1. Overall amount of motion captured
2. The degree of contraction and expansion of body computed using its bounded regions
3. Motion fluency computed on the bases of the changes magnitude of the overall amount of motion over time
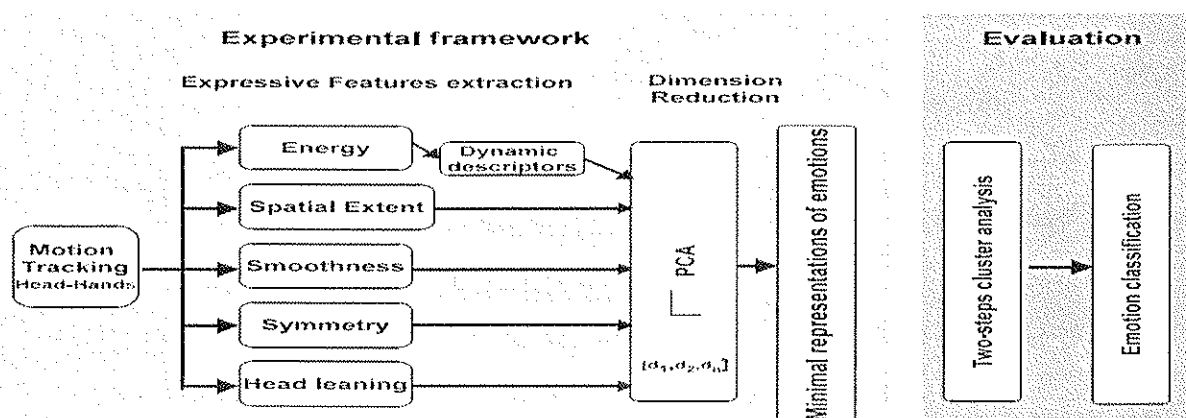
2.8.2 Framework used:

Figure 1: Framework Used for Gesture Representation

1. Module 1 computes low-level motion features i.e., the 3D positions and kinematics of head and hands

2. Module 2 computes a vector of higher-level expressive gesture features, including the following five sets of features:

    a. Energy (passive vs. animated)

    b. Spatial extent (expanded vs. contracted)

    c. Smoothness and continuity of movement (gradual vs. jerky)

    d. Forward-backward leaning of the head

    e. Spatial symmetry and asymmetry of the hands with respect to the horizontal and vertical axis

3. Module 3 reduces the dimensionality of the data, while highlighting the salient patterns in the data set

The paper also contains the details of how each of the features was computed as well as the Dimension Reduction Model.

Suggest that use of upper body only would be sufficient for classifying a large amount of effective behavior.

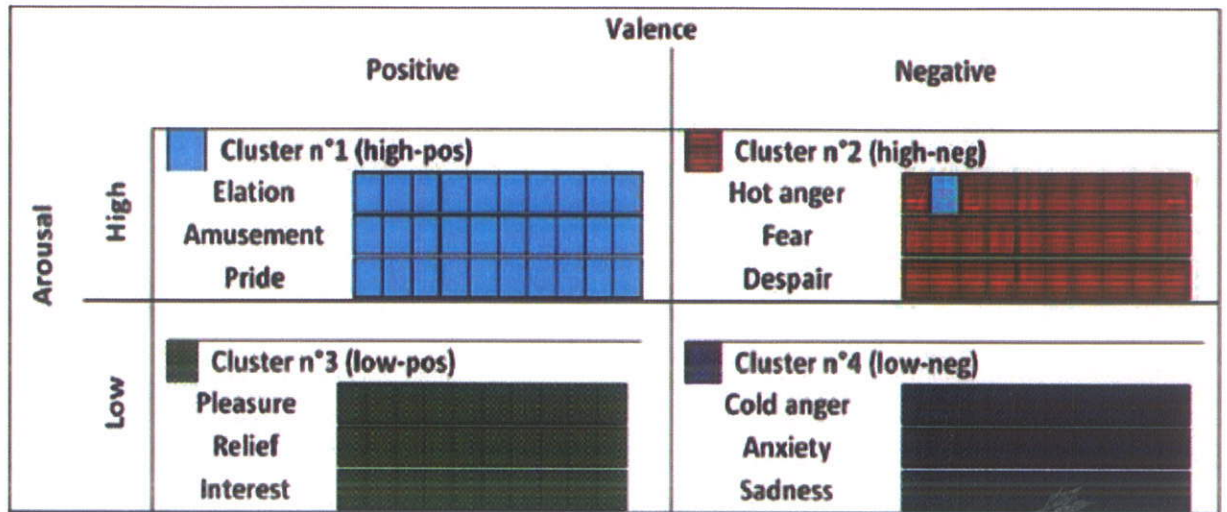*Human Motion Analysis and Art and Music Generation*

2.8.3 Results:

Figure 2: Cluster Membership of the 12 Emotion Classes with Each Cell Denoting a Portrayal

Attributing Emotion to Static Body Postures.pdf

Static body postures were shown to various observers and they categorized various emotions. The result is as follows:

Table 2: Showing Percentages of Successful Recognition of Static Postures by Observers

	Front	Side	Rear
Anger	90%	50%	36%
Disgust	6%	5%	43%
Fear	67%	67%	50%

Critical Features for the Perception of Emotion from Gait.pdf

Carried out a 3-step process to extract and validate the minimum set of spatial-temporal motor primitives that drive the perception of particular emotions in gait. They showed that perception of emotions is based on specific changes in joint angles

Human Motion Analysis and Art and Music Generation

amplitudes with respect to the pattern of neutral walk. They also studied the after-effects of emotions and concluded that there were after-affects in perception of sad and happy movements, indicating that the feature sets are complete.

2.10.1 Conclusion

Arms straight down, close to the side of the body indicates sadness (but head position also plays an important role, for example, bent forward for disgust and backward for pride).

While some emotion categories do share a core set of body expressions, they also exhibit a number of variations for other body parts. For example, happiness and elated joy are indicated by head bent back in some studies. In some studies arms are raised while they remain straight down in Critical Features for the Perception of Emotion from Gait.pdf (which may be due to the contextual factor of gait).

More than one body expression pattern may be associated with the same emotion category. So, there are contextual factors which may affect the emotional state's expression. Components that make up an emotion expressed are not fixed and each emotion reaction is not unique. But, discriminative patterns may emerge when context is not considered.

Other Papers

- Multimodal Analysis of Expressive Gesture in Music and Dance Performances .pdf

- Gesture Based Affective Computing on Motion Capture Data.pdf

- Estimating the Efficiency of Recognizing Gender and Affect from Biological Motion.pdf

- Detecting Affect from Non-Stylised Body Motions.pdf

- Exploring Non-Verbal Communication in Human-Machine Interaction.pdf

- Automatic Recognition of Affective Body Movement in a Video Game Scenario.pdf

- Affective Computing.ppt

- Visual Informatics - Bridging Research and Practice.pdf

The above papers show automatic recognition of bodily expressions mono-modally.

I studied recognizing emotional states of people playing tennis on Wii. Individual idiosyncrasies were removed by normalizing each expression according to min and max value of the features for that participant.

Best results were obtained using angular velocity, angular frequency and amount of movement. Overall the system was able to correctly classify a high percentage of both the high and low-intensity negative expression and the happiness expressions, but considerably fewer of the concentration expressions.

In tennis, some played using their hands and wrists while some with shoulders and arms. In hand/wrists combination, negative emotions may be expressed by jerky and fast movements of the wrists while in latter, with jerky movements of the larger body.

- Probabilistic Combination of Multiple Modalities to Detect Interest.pdf

- Multimodal Emotion Recognition from Expressive Faces, Body Gestures and Speech.pdf

- Automatic Temporal Segment Detection and Affect Recognition From Face and Body Display.pdf

- Beyond Facial Expressions.pdf

The above mentioned papers show automatic recognition multi-modally.

- Emotion Recognition from Dance Image Sequences Using Contour Approximation.pdf

- Interactive System Design.pdf

- Extraction of Motion Characteristics Corresponding to Sensitivity Information Using Dance Mvoement.pdf

- Emotional Body Movements.pdf

Majority of above mentioned affective recognition systems have focused on dance performances.

Recognition of fear was most often misclassified as anger. Misclassification of grief and joy is interesting since joy movements are very fluid and grief movements are not.

These systems are relatively successful but dance movements are generally exaggerated while day-to-day body movements are generally subtle.

Chapter 3

3   Methodology

Figure 3 describes high level architecture of what the project was intended to accomplish:
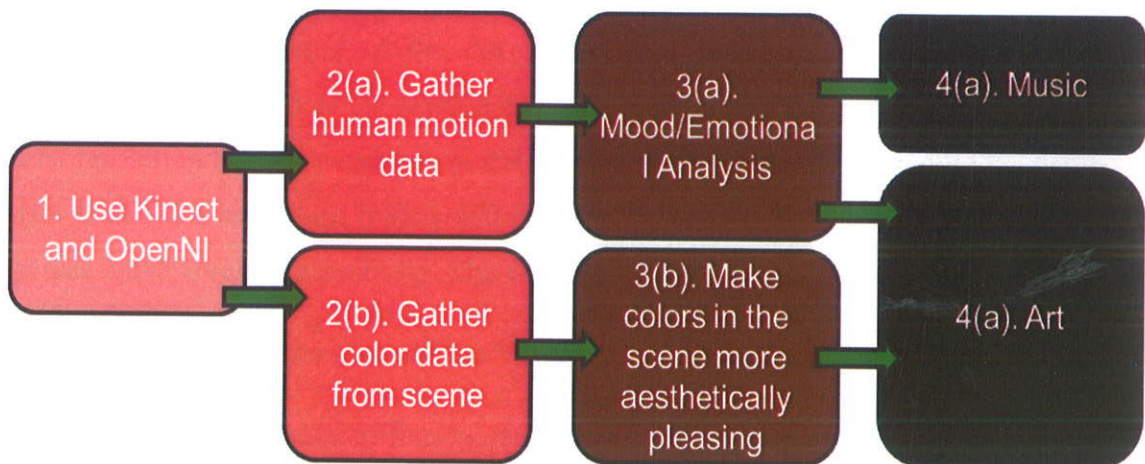
Figure 3: Showing the Basic Layout of the Project

■ Kinect

This portion of the document describes Kinect hardware and various software frameworks to be used with Kinect. Microsoft has a well-documented Kinect SDK (for windows only) but some third-party SDK's and drivers, as well as frameworks are also available. I propose that we use OpenNI coupled with NITE (both are explained below) instead of Microsoft's SDK as they are open source (you need to purchase a license to use Microsoft's Kinect SDK for commercial purposes) and can be easily ported to Mac and Linux.

3.1.1   Kinect Sensor Hardware

The Kinect sensor includes:

1. RGB camera

2. Depth sensor

3. Multi-array microphones

4. Tilt motor

5. Three-axis accelerometer

The Kinect's depth sensor consists of an infrared light source , a laser that projects a pattern of dots, that are read back by a monochrome CMOS IR sensor. The sensor detects reflected segments of the dot pattern and converts their intensities into distances. The resolution of the depth dimension (z-axis) is about one centimeter while spatial resolution (x- and y-axes) is in millimeters. Each frame generated by the depth sensor is at VGA resolution (640 x 480 pixels), containing 11-bit depth values which provides 2,048 levels of sensitivity. The output stream runs at a frame rate of 30 Hz.

The RGB video stream also utilizes VGA resolution and a 30 Hz frame rate.

The audio array consists of four microphones, with each channel processing 16-bit audio at a sampling rate of 16 KHz. The hardware includes ambient noise suppression.

Microsoft suggests that you allow about 6 feet of empty space between you and the sensor otherwise you can confuse the sensor.

3.1.2 Kinect Development Software

There are four main Kinect development libraries:

1. OpenKinect's libfreenect

2. CLNUI

3. OpenNI

4. Microsoft's Kinect for Windows

3.1.3 OpenKinect's libfreenect

Libfreenect is derived from a reverse-engineered Kinect driver and works across all OS platforms. OpenKinect Analysis library communicates with the OpenKinect

API and analyzes the raw information into more useful abstractions. They aim to to implement the following features but **most of them have not been implemented yet:**

- hand tracking
- skeleton tracking
- other depth processing
- 3D audio isolation *coning?*
- audio cancellation (*perhaps a driver feature?*)
- Point cloud generation
- Inertial movement tracking with the built in accelerometer or an attached WiiMote
- 3d reconstruction
- GPU acceleration for any of the above

CLNUI

CLNUI is aimed for windows only but **allows multiple kinects** to work together.

OpenNI (Open Natural Interaction)

OpenNI is a software framework and an API and provides support for:

1. Voice and voice command recognition
2. Hand gestures
3. Body Motion Tracking

OpenNI is a multi-language, cross-platform framework that defines API's for writing applications utilizing Natural Interaction. The main advantage of this framework is that you write software independent of the hardware. So, for example, we can aim to write a Human Motion Analysis program which can analyze human motion using kinect if it is available or the same program would use a camera if Kinect is not available.

Figure 4 shows the three-layered view of the OpenNI concept:

Figure 4: Three-Layered View of the OpenNI
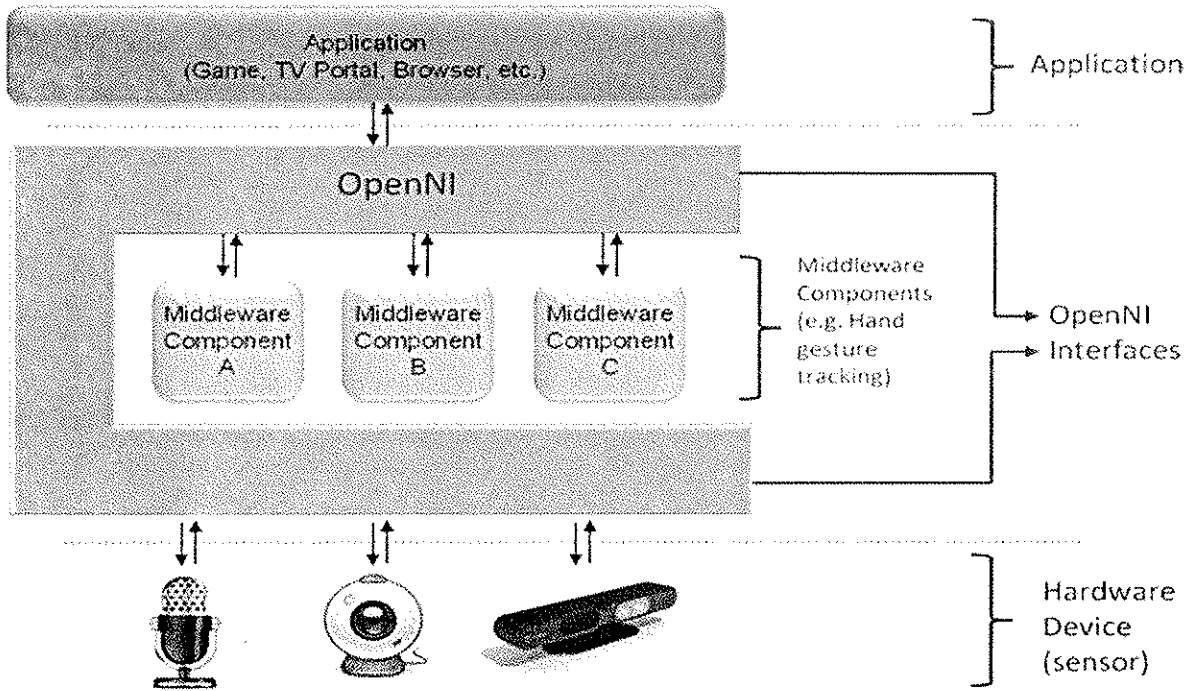
Sensor modules that are currently supported are:

1. 3D sensor
2. RGB camera
3. IR camera
4. A microphone or an array of microphones

Middleware components that are supported are:

1. Full body analysis middleware: a software component that processes sensory data and generates body related information (typically data structure that describes joints, orientation, center of mass, and so on)

2. Hand point analysis middeware: a software component that processes sensory data and generates the location of a hand point

3. Gesture detection middleware: a software component that identifies predefined gestures (for example, a waving hand) and alerts the application

4. Scene analyzer middleware: a software component that analyzes the image of the scene in order to produce such information as:

    a. The seperation between the foreground of the scene and the background

    b. The coordination of the floor plane

    c. The individual identification of figures in the scene (and output its current location and orientation of joints of this figure)

Figure 5 describes an example scenario in which multiple modules are registered to work with an OpenNi installation:
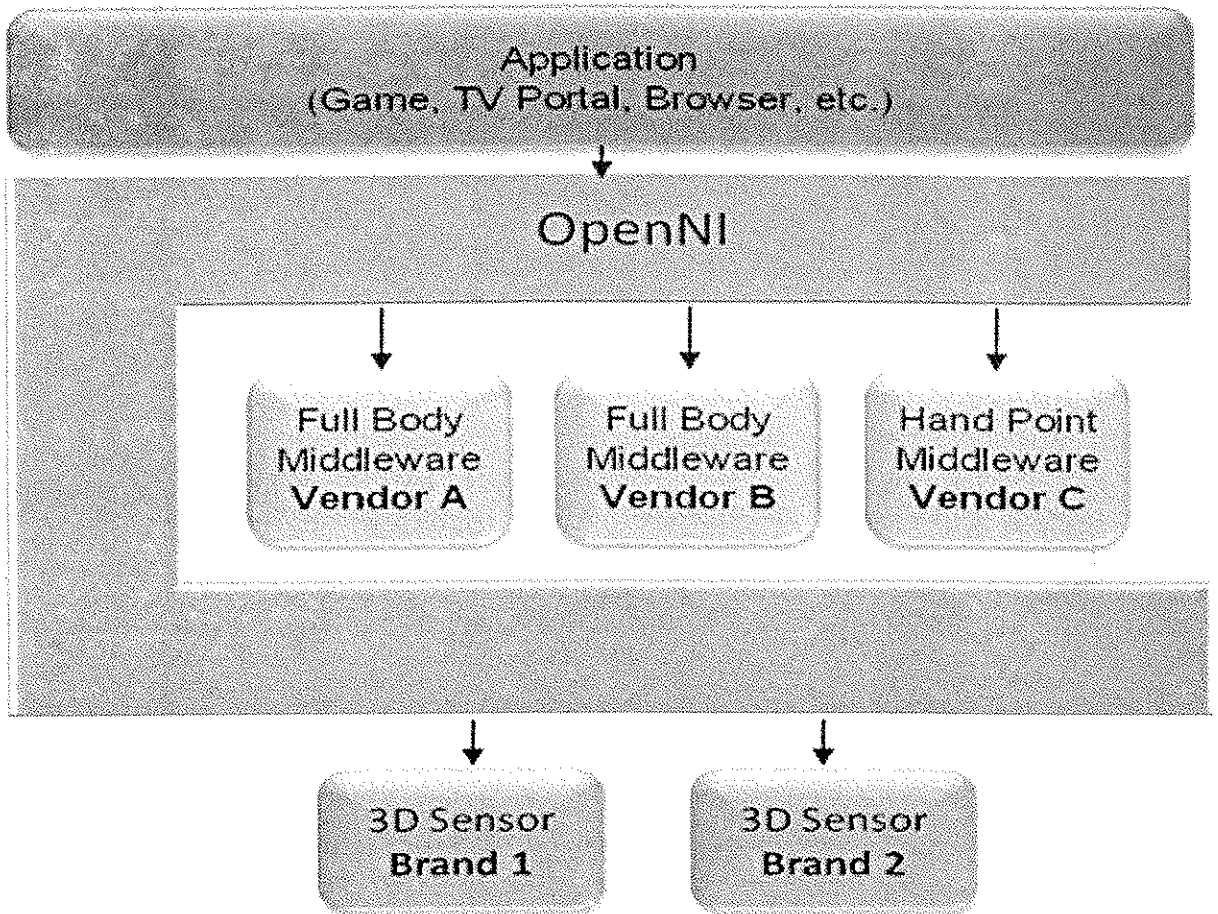
Figure 5: Example Scenario with Multiple Modules Registered to Work with an OpenNI Installation

OpenNI's Kinect driver is called SensorKinect. **One drawback of using OpenNI will be that tilt motor, status light and accelerometer will not be available.**

NITE

An important reason for using OpenNI is its support for middleware. The NITE library understands the different hand movements as gesture types based on how hand points change over time. NITE gestures include:

1. Pushing
2. Swiping

Human Motion Analysis and Art and Music Generation

3. Holding steady
4. Waving
5. Hand circling

"Primesense's NITE middleware allows computers or digital devices to perceive the world in 3D. NITE comprehends your movements and interactions within its view, translates them into application inputs, and responds to them without any wearable input."

Including computer vision algorithms, NITE identifies users and tracks their movements, and provides the framework API for implementing Natural-Interaction UI controls based on gestures. **NITE can detect when you want to control things with only hand gestures, or when to get your whole body involved.**

Hand control: allows you to control digital devices with your bare hands and as long as you're in control, NITE ignores what others are doing.

Full body control: lets you have a total immersive, full body video game experience. NITE middleware supports multiple users, and is designed for all types of action.

▉Comparison between Microsoft's SDK and OpenNI

Microsoft's Kinect SDK covers much the same ground as OpenNI. The low-level API gives access to the depth sensor, image sensor, and microphone array, while higher-level features include skeletal tracking, audio processing, and integration with the Windows speech recognition API.

The main area where SDK wins over OpenNI is audio. Other pluses for Microsoft's Kinect SDK are its extensive documentation and ease of installation on Windows 7. The main drawback for Microsoft's SDK is that it only works for Windows 7, not even Windows XP. The SDK is free but limited to non-commercial purposes.

███Development Configurations

3.6.1 OpenNI

3.6.1.1 Installing OpenNI and Kinect Driver on Windows 7 (x86)

1. Go to http://openni.org/Downloads/OpenNIModules.aspx

    a. Select OpenNI Packages

    b. Select Stable

    c. Select PrimeSense Package Stable Build for Windows x86 Development Edition

2. While installing, select OpenNI and NITE middleware. DO NOT check PrimeSense hardware as that driver is not for Microsoft Kinect

3. Download Kinect driver from https://github.com/avin2/SensorKinect (make sure that neither Microsoft's nor any other driver for Kinect is installed on your computer) and install it.

4. To run the samples included with NITE, copy all .xml files from "[PrimeSense root directory]/NITE/Data" to "[PrimeSense root directory]/SensorKinect/Data"

3.6.1.2 Configuring OpenNI Development in Visual Studio 2010

1. Create a new or open an existing Visual Studio 2010 project

2. Open project properties

3. Go to **C/C++ -> General -> Additional Include Directories** and add "[OpenNI root directory]/Include"

4. Go to **Linker -> General -> Additional Library Directories** and add "[OpenNI root directory]/Lib"

5. Go to **Linker -> Input -> Additional Dependencies** and add **OpenNI.lib**

6. Your code should include **XnOpenNI.h** if you are using C interface, or **XnCppWrapper.h** if you are using C++ interface

7. Optionally, you can use the namespace "xn" or you can reference objects using scope operator (For example, "xn::Context context")

3.6.1.3 Configuring OpenNI Development in Visual Studio 2012

Installation and configuration in Visual Studio 2012 is exactly the same as Visual Studio 2010. But OpenNI doesn't let you use their library in compiler version greater than VS 2010. But it can be overridden using the following steps:

1. Within the OpenNI libraries directory, locate the file XnPlatform.h

2. At the top of the file you will find the code "if defined(_WIN32). Beneath this you will find another condition which checks the compiler version

3. Comment out that piece of code and you will be able to compile the project

3.6.1.4 Configuring OpenNI Development in Java IDE's

1. Create a new project in Eclipse or Netbeans

2. Add "[OpenNI root directory]/Bin/org.OpenNI.jar" and "[NITE root directory]/Bin/com.primesense.NITE.jar" to "additional libraries

Technologies Used:

1. Microsoft Windows 7 (x86)

2. PrimeSense's SensorKinect driver for Kinect

3. NITE middleware for OpenNI

4. OpenNI

6. Your code should include **XnOpenNI.h** if you are using C interface, or **XnCppWrapper.h** if you are using C++ interface

7. Optionally, you can use the namespace "xn" or you can reference objects using scope operator (For example, "xn::Context context")

3.6.1.3 Configuring OpenNI Development in Visual Studio 2012

Installation and configuration in Visual Studio 2012 is exactly the same as Visual Studio 2010. But OpenNI doesn't let you use their library in compiler version greater than VS 2010. But it can be overridden using the following steps:

1. Within the OpenNI libraries directory, locate the file XnPlatform.h

2. At the top of the file you will find the code "if defined(_WIN32). Beneath this you will find another condition which checks the compiler version

3. Comment out that piece of code and you will be able to compile the project

3.6.1.4 Configuring OpenNI Development in Java IDE's

1. Create a new project in Eclipse or Netbeans

2. Add "[OpenNI root directory]/Bin/org.OpenNI.jar" and "[NITE root directory]/Bin/com.primesense.NITE.jar" to "additional libraries

Technologies Used:

1. Microsoft Windows 7 (x86)

2. PrimeSense's SensorKinect driver for Kinect

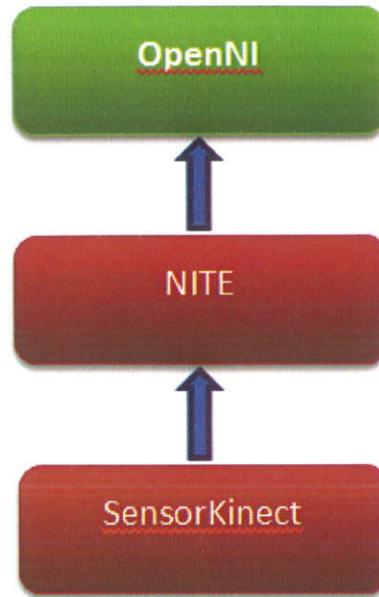3. NITE middleware for OpenNI

4. OpenNI

3.7.1 Architecture

Figure 6: Project Architecture

3.7.2 OpenFrameworks

1. Download the latest release of OpenFrameworks from http://www.openframeworks.cc/download/

2. There you go! Run any of the example from the openFrameworks' directory

Chapter 4

4 Results

Two major emotions, happiness and anger were taken into consideration. These two emotions were distinguished by only using the upper body data. Music was generated using motion of hands of the user as input. Emotional data was used to normalize the music generated to map it onto a musical scale so that it sounded aesthetically pleasing.

Chapter 5

5 Discussion

Where body language is a crucial part of emotional detection, either by another human or a computer device, it is not complete. Current state of body language largely depends on the context. For example, people playing tennis would depict different emotions differently through body language than people playing piano. This is because in each of the situation, body language will have different constraints and the subject can only show movement in certain directions.

Secondly, real-time generation of music according to the emotion, and which also sounds that way to an untrained ear also lacks accuracy.

Chapter 6

6 Conclusion

We conclude that body language alone is not sufficient for accurately determining the emotion of a person. But, coupled with facial expression analysis and vocal analysis, these three complete the method in which emotions are perceived by human beings and hence, it does have the potential of increasing the computer-aided emotion detection.

Chapter 6

6   Conclusion

We conclude that body language alone is not sufficient for accurately determining the emotion of a person. But, coupled with facial expression analysis and vocal analysis, these three complete the method in which emotions are perceived by human beings and hence, it does have the potential of increasing the computer-aided emotion detection.

Chapter 7

7 References

[1] http://www.affective-sciences.org/gemep

[2] http://www.openframeworks.cc/download/

[3] https://github.com/avin2/SensorKinect