

Media Engineering and Technology Faculty
German University in Cairo

Bachelor Thesis Title

Bachelor Thesis

Author:	Khaled Osman
Supervisors:	Prof. Georg Jung
	Sup 2
	Sup 3
Submission Date:	XX July, 20XX

Media Engineering and Technology Faculty
German University in Cairo

Bachelor Thesis Title

Bachelor Thesis

Author:	Khaled Osman
Supervisors:	Prof. Georg Jung
	Sup 2
	Sup 3
Submission Date:	XX July, 20XX

This is to certify that:

- (i) the thesis comprises only my original work toward the Bachelor Degree
- (ii) due acknowledgement has been made in the text to all other material used

Khaled Osman
XX July, 20XX

Acknowledgments

Text

Abstract

Today's technology is getting smarter producing amazing new opportunities and making our lives easier. One of the most amazing technologies emerging nowadays are touchless interface devices as they offer an easy way to allow the user to interact with an application with much less effort. Handicapped people have difficulties interacting with the electronic home devices they use daily. We could make their lives a little easier by offering several ways to interact with their home devices through a touchless interface sensing device like the Microsoft Kinect sensor. This project allows you to control home devices through a touchless interface saving power, time and effort. This is a thesis presented in the development of Microsoft Kinect sensor and home automation technology by using depth and skeletal tracking, voice recognition, gesture recognition and a touchless graphical interface.

Contents

Acknowledgments	V
1 Introduction	1
2 Background	3
2.1 Kinect sensor	3
2.1.1 history	4
2.1.2 Capabilities and constraints	4
2.2 Microsoft SDK	4
2.2.1 Requirements	5
2.2.2 OpenKinect history	5
2.2.3 Kinect vs OpenNI	6
2.2.4 What Kinect can't do	7
2.3 Microsoft XNA	7
2.3.1 Definition and Overview	7
2.3.2 Requirements	7
2.3.3 Documentation	7
2.3.4 Basics	8
2.4 Arduino	9
2.4.1 Requirements	9
2.4.2 IDE	9
2.4.3 How it works	9
2.4.4 IDE basics	9
2.5 Project Design	10
2.5.1 User Stories	10
2.5.2 Software Architecture	11
2.6 Project Features	11
2.6.1 Voice Recognition	11
2.6.2 Kinect Arduino Communication	13
2.6.3 Screen Manager	13
2.6.4 User Avatar	13
2.6.5 Music Player	13
2.7 Gesture Recognition	14
2.7.1 How it works	14

2.7.2	How to add a new gesture	17
2.7.3	Defects	17
3	Conclusion	19
4	Future Work	21
	Appendix	22
A	Lists	23
	List of Abbreviations	23
	List of Figures	24

Chapter 1

Introduction

Lately, touchless interface devices have appeared in the technology world making our lives easier. One of these newest technologies which is currently being used world-wide now is the Microsoft Kinect sensor. Microsoft Kinect sensor is a motion sensing input device by Microsoft for the Xbox 360 video game console and Windows PC, it was mainly used for gaming, however it is currently being used for developing several applications in different regions due to its several capabilities. It offers face detection, voice recognition and skeletal tracking which makes it better than other motion sensing devices. The idea is to extend the Kinect's potential uses from gaming to be used in different aspects and make more and better use of it's capabilities. Since touchless interface devices are mostly handicapped friendly and can be used to interact with different applications, they could be used in controlling electronic home devices we use everyday to make our lives easier and more elegant. Currently the average household pays over 100 dollars a month for electricity according to the U.S. Energy Information Administration. Half of those costs are to keep the motor and compressors operating in air conditioners and heating systems. If consumers had the ability to control their home's climate control systems remotely, they could save money without sacrificing convenience and comfort. This is a handicapped-friendly project that provides an easy way to control home devices using Microsoft Kinect sensor and Arduino board, (a simple microcontroller that can be used for taking inputs from a variety of switches and sensors, and controlling a variety of lights, motors, and other physical outputs), by saying voice commands and doing body gestures and interacting with a touchless graphical user interface.

Chapter 2

Background

2.1 Kinect sensor

Kinect is a motion sensing input device by Microsoft for the Xbox 360 video game console and Windows PCs. It enables users to control and interact with the Xbox 360 without the need to touch a game controller through a natural user interface using gestures and spoken commands to provide full-body 3D motion capture, facial recognition and voice recognition capabilities. See `kinectsensor` `datastream`

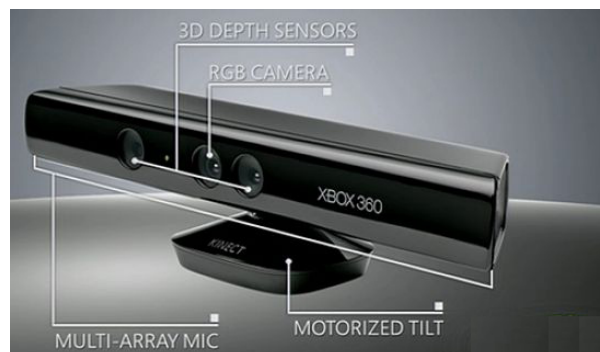


Figure 2.1: Kinect sensor

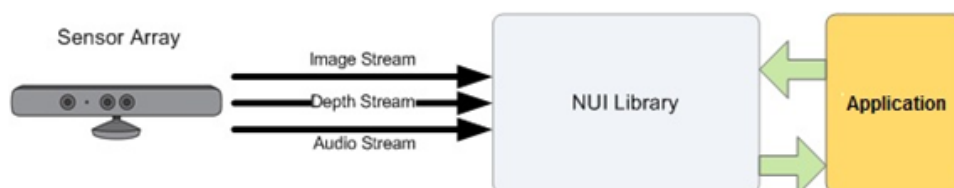


Figure 2.2: Data stream

2.1.1 history

Kinect itself was first announced on June 1, 2009 at E3 under the code name "Project Natal". It was officially launched on November 4th 2010. On June 17, 2011, Microsoft finally released the Kinect SDK beta to the public under a non-commercial license after demonstrating it for several weeks at events like MIX. As promised, it included the skeleton recognition algorithms that make an initial pose unnecessary as well as the AEC technology and acoustic models required to make Kinect speech recognition system work in a large room. Every developer now had access to the same tools Microsoft used internally for developing Kinect applications for the computer. In May 21, 2012, The newest Kinect for Windows 1.5 SDK was released. It adds new features, support for many new languages and debut in 19 more countries.

2.1.2 Capabilities and constraints

The Kinect sensor consists of an infrared projector, a LED light indicator, a color camera, an infrared camera, a microphone array and a tilt motor. The area required to play Kinect is roughly 6 meters squared, although the sensor can maintain tracking through an extended range of approximately (2.320 ft). The sensor has an angular field of view of 57 degrees horizontally and 43 degrees vertically, while the motorized pivot is capable of tilting the sensor up to 27 degrees either up or down. The horizontal field of the Kinect sensor at the minimum viewing distance of 0.8 m is therefore 87 cm, and the vertical field is 63 cm, resulting in a resolution of just over 1.3 mm per pixel. The microphone array features four microphone capsules and operates with each channel processing 16-bit audio at a sampling rate of 16 kHz. The Kinect sensor can detect up to 6 players at the same time, however it can only keep full track 2 skeletons at the same time while it only stores some basic information like position values for the other 4.

2.2 Microsoft SDK

The Kinect for Windows SDK is a set of libraries that allows us to program applications on a variety of Microsoft development platforms using the Kinect sensor as input. With it, we can program WPF applications, WinForms applications, XNA applications and even browser-based applications running on the Windows operating system, however we cannot create Xbox games with the Kinect for Windows SDK. The new Kinect for Windows 1.5 SDK includes 'Kinect Studio', a new application that allows developers to record, playback, and debug clips of users interacting with applications, support for new "seated" or "10-joint" skeletal system that will let apps track the head, neck, and arms of a Kinect user whether they're sitting down or standing and support for four new languages for speech recognition French, Spanish, Italian, and Japanese. Additionally it would add support for regional dialects of these languages along with English.

2.2.1 Requirements

Unlike other Kinect libraries, the Kinect for Windows SDK, as its name suggests, only runs on Windows operating systems. Specifically, it runs on x86 and x64 versions of Windows 7. It has been shown to also work on early versions of Windows 8. Because Kinect was designed for Xbox hardware, it requires roughly similar hardware on a PC to run effectively. Hardware Requirement:

- Computer with a dual-core, 2.66-GHz or faster processor
- Windows 7 compatible graphics card that supports Microsoft DirectX 9.0c
- 2 GB of RAM (4 GB or RAM recommended)
- Kinect for Xbox 360 sensor
- Kinect USB power adapter Software Requirements
- Microsoft Visual Studio 2010 Express or other Visual Studio 2010 edition:
- Microsoft .NET Framework 4
- The Kinect for Windows SDK (x86 or x64):
- For C++ SkeletalViewer samples:
- DirectX Software Development Kit, June 2010 or later version:
- DirectX End-User Runtime Web Installer:

To take full advantage of the audio capabilities of Kinect, you will also need additional Microsoft speech recognition software: the Speech Platform API, the Speech Platform SDK, and the Kinect for Windows Runtime Language Pack. Fortunately, the install for the SDK automatically installs these additional components for you.

2.2.2 OpenKinect history

Almost as soon as the Kinect was released, people began trying to figure out how Kinect worked. Then came the hacks, then the alternate applications. On November 6, a hacker known as AlexP was able to control Kinects motors and read its accelerometer data. On Monday, November 8, AlexP posted video showing that he could pull both RGB and depth data streams from the Kinect sensor and display them. Hector Martin, a computer science major in Bilbao, Spain, had just purchased Kinect and managed to write the driver and application to display RGB and depth video. Martin became a contributor to the OpenKinect group and a new library, libfreenect, became the basis of the communitys hacking efforts. Throughout November, people started to post videos

on the Internet showing what they could do with Kinect. Kinect-based artistic displays, augmented reality experiences, and robotics experiments started showing up on YouTube. Sites like KinectHacks.net sprang up to track all the things people were building with Kinect. By November 20, someone had posted a video of a light saber simulator using Kinect another movie aspiration checked off. Microsoft, meanwhile, was not idle. The company watched with excitement as hundreds of Kinect hacks made their way to the web. On December 10, PrimeSense announced the release of its own open source drivers for Kinect along with libraries for working with the data. This provided improvements to the skeleton tracking algorithms over what was then possible with libfreenect and projects that required integration of RGB and depth data began migrating over to the OpenNI technology stack that PrimeSense had made available. Without the key Microsoft Research technologies, however, skeleton tracking with OpenNI still required the awkward T-pose to initialize skeleton recognition.

2.2.3 Kinect vs OpenNI

The Official SDK is developed by Microsoft which also develops the hardware and therefore should know internal information about the device that the open source society must reverse engineer. Also having economy behind it gives many advantages. On the other hand, "The OpenKinect open source community consists of over 2000 members contributing their time and code to the Project. OpenKinect was released long before the official SDK as the kinect device was hacked on the first or second day of its release.

- Programming languages supported: Official SDK: C++, C, or Visual Basic by using Microsoft Visual Studio 2010. OpenKinect: Python, C, C++, C, Java, Lisp and more.. Obviously not requiring Visual Studio.
- Operating systems support: Official SDK: only installs on Windows 7. OpenKinect: runs on Linux, OS X and Windows
- License: The Official SDK is in its current beta state only for testing. The SDK has been developed specifically to encourage wide exploration and experimentation by academic, research and enthusiast communities. commercial applications are not permitted. Note however that this will probably change in future releases of the SDK. OpenKinect appears to be open for commercial usage, but online sources state that it may not be that simple. I would take a good look at the terms before releasing any commercial apps with it.
- Documentation and support: Official SDK: well documented and provides a support forum OpenKinect: appears to have a mailing list, twitter and irc. but no official forum. Documentation on website is not as rich as Microsoft SDK. The official SDK provides drivers that installed and came with examples and code for easy getting started.

2.2.4 What Kinect can't do

Kinect's voice recognition sometimes doesn't work as expected especially when you're surrounded by ambient noise from other guests. The Kinect sensor can't detect gestures that does not require position changing of joints, differentiate between similar gestures or tell whether you're facing the kinect or not. Also in order to let the Kinect see you do something, you have to move in very pronounced motions because it seems only large movements can be seen by the stereoscopic cameras on the sensor. This works great for exercise games where you have to run in place or otherwise be full body active, but when all youre doing is making a pushing motion, the Kinect cant see this and doesnt do anything. Additionally, the Kinect doesnt like when theres a bunch of movement taking place all at once. its just too much input and it may simply not detect the motion.

2.3 Microsoft XNA

2.3.1 Definition and Overview

Microsoft XNA is a set of tools with a managed runtime environment provided by Microsoft that facilitates video game development and management. XNA is based on the .NET Framework for Xbox 360 development and .NET Framework on Windows. It includes an extensive set of class libraries, specific to game development, to promote maximum code reuse across target platforms.

2.3.2 Requirements

Since XNA games are written for the runtime, they can run on any platform that supports the XNA Framework with minimal or no modification. Games that run on the framework can technically be written in any .NET-compliant language, but only C# in XNA Game Studio Express IDE and all versions of Visual Studio 2008 and 2010 (as of XNA 4.0) are officially supported.

2.3.3 Documentation

- `Microsoft.Xna.Framework`: Provides commonly needed game classes such as timers and game loops.
- `Microsoft.Xna.Framework.Audio`: Contains low-level application programming interface (API) methods that can load and manipulate XACT-created project and content files to play audio.
- `Microsoft.Xna.Framework.Content`: Contains the run-time components of the Content Pipeline.

- `Microsoft.Xna.Framework.Design`: Provides a unified way of converting types of values to other types.
- `Microsoft.Xna.Framework.GamerServices`: Contains classes that implement various services related to gamers. These services communicate directly with the gamer, the gamer's data, or otherwise reflect choices the gamer makes. Gamer services include input device and profile data APIs.
- `Microsoft.Xna.Framework.Graphics`: Contains low-level application programming interface (API) methods that take advantage of hardware acceleration capabilities to display 3D objects.
- `Microsoft.Xna.Framework.Graphics.PackedVector`: Represents data types with components that are not multiples of 8 bits.
- `Microsoft.Xna.Framework.Input`: Contains classes to receive input from keyboard, mouse, and Xbox 360 Controller devices.
- `Microsoft.Xna.Framework.Input.Touch`: Contains classes that enable access to touch-based input on devices that support it.
- `Microsoft.Xna.Framework.Media`: Contains classes to enumerate, play, and view songs, albums, playlists, and pictures.
- `Microsoft.Xna.Framework.Net`: Contains classes that implement support for Xbox LIVE, multiplayer, and networking for XNA Framework games.
- `Microsoft.Xna.Framework.Storage`: Contains classes that allow reading and writing of files.

2.3.4 Basics

Any class that extends `Microsoft.XNA.Game` should override 4 main methods and call them in order which are

- `Initialize()`: Allows the game to perform any initialization it needs to before starting to run. This is where it can query for any required services and load any non-graphic related content.
- `LoadContent()`: `LoadContent` will be called once per game and is the place to load all of your content eg: textures, animations, fonts or any external files.
- `UnLoadContent()`: works as a destructor
- `Update(GameTime)`: Allows the game to run logic such as updating the world, checking for collisions, gathering input, and playing audio. it works as a `while(true)` loop according to certain flags. `gameTime` Provides a snapshot of timing values.
- `Draw(GameTime)`: This is called when the game should draw itself.

2.4 Arduino

Arduino is a simple microcontroller designed to make the process of using electronics in multidisciplinary projects more accessible. It can sense and control more of the physical world than your desktop computer and can be used to develop interactive objects by taking inputs from a variety of switches or sensors, and controlling a variety of lights, motors, and other physical outputs. The hardware consists of a simple open source hardware board designed around an 8-bit Atmel AVR microcontroller, though a new model has been designed around a 32-bit Atmel ARM. The software consists of a standard programming language compiler and a boot loader that executes on the microcontroller. Arduino projects can be stand-alone, or they can be communicate with software running on your computer.

2.4.1 Requirements

The Arduino software runs on Windows, Macintosh OSX, and Linux operating systems. Most microcontroller systems are limited to Windows.

2.4.2 IDE

Simple, clear programming environment - The Arduino programming environment is easy-to-use for beginners, yet flexible enough for advanced users to take advantage of as well. For teachers, it is conveniently based on the Processing programming environment, so students learning to program in that environment will be familiar with the look and feel of Arduino.

2.4.3 How it works

The Arduino programming language is an implementation of Wiring, a similar physical computing platform, which is based on the Processing multimedia programming environment. The Arduino is a microcontroller board based on the ATmega328 . It has 14 digital input/output pins of which 6 can be used as PWM outputs, 6 analog inputs, a 16 MHz ceramic resonator, a USB connection, a power jack, an ICSP header, and a reset button. The ATmega328 has a 32 KB memory. It also has 2 KB of SRAM and 1 KB of EEPROM. See fig:arduinoboard

2.4.4 IDE basics

The Arduino integrated development environment (IDE) is a cross-platform application written in Java, and is derived from the IDE for the Processing programming language

and the Wiring projects. It is designed to introduce programming to artists and other newcomers unfamiliar with software development. It includes a code editor with features such as syntax highlighting, brace matching, and automatic indentation, and is also capable of compiling and uploading programs to the board with a single click. There is typically no need to edit makefiles or run programs on a command-line interface. A program or code written for Arduino is called a sketch. Sketches are saved with the file extension .ino. It has features for cutting/pasting and for searching/replacing text. The message area gives feedback while saving and exporting and also displays errors. The console displays text output by the Arduino environment including complete error messages and other information. The bottom righthand corner of the window displays the current board and serial port. The toolbar buttons allow you to verify and upload programs, create, open, and save sketches, and open the serial monitor. Arduino programs are written in C or C++. The Arduino IDE comes with a software library called "Wiring" from the original Wiring project, which makes many common input/output operations much easier. Users only need define two functions to make a runnable cyclic executive program:

- `setup()`: a function run once at the start of a program that can initialize settings
- `loop()`: a function called repeatedly until the board powers off

to execute your code you only need two buttons: Verify which Checks your code for errors. Upload which Compiles your code and uploads it to the Arduino I/O board.

2.5 Project Design

This section explains an overview of the project design and architecture.

2.5.1 User Stories

The user stories and the program features are as follows:

- As a user, I can turn the lights on/off by entering/leaving the room.
- As a system, I should be able to recognize certain gestures and control the device accordingly.
- As a user, I can interact with a touchless interface.
- As a user, I can give certain voice commands.
- As a system, I should have a circuit connection between the arduino board and the device I want to control.

- As a system, I should be able to control the device and switch it on/off when needed.
- As a system, I should create an interface to allow the communication between the kinect sensor and arduino board.
- As a user, I should see pop-up screens to show feedback from the interface at certain circumstances.
- As a user, I should see an avatar which represents my distance to the kinect.
- As a user, I should see an introductory fading screen to the application.
- As a user, I can play music.
- As a user, I should see a main screen where I can edit settings and enable/disable features like face/depth/voice/skeletal tracking.
- As a user, I should see a screen for each controlled device where I can edit it's settings manually.
- As a system I should provide a way for the user to enter a password to be able to manage the application.

2.5.2 Software Architecture

The software architecture of the project is shown in the figures below See fig:softwarearchitecture

2.6 Project Features

2.6.1 Voice Recognition

The kinect's voice recognition system is implemented with the help of Microsoft.Speech library and accessing the audiostream of the kinect sensor. The system takes as an input an array of commands which is the grammar of words it is expected to detect and compares each of them to the heard word, then eventually does some calculations that shows how much do they match according to a ratio ranging from 0 to 1 which is the confidence. However it is not fully efficient since the confidence differs according to different words and accents. The better the accent the less the error ratio is, so you may want to lower the confidence for easier tolerance and recognition, but make sure it is not too low to detect words that should not be detected aswell.. It also may fail depending on the surrounding noise. Adding that speech recognition is only supported in the following languages English,French,German,Italian,Japanese and Spanish.

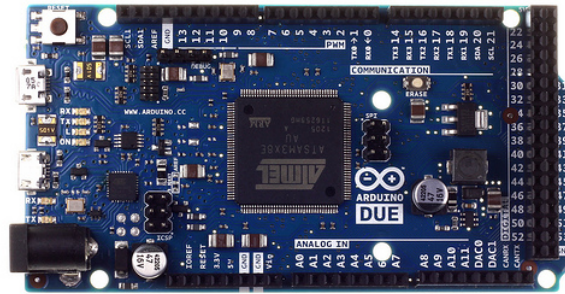


Figure 2.3: Arduino Board

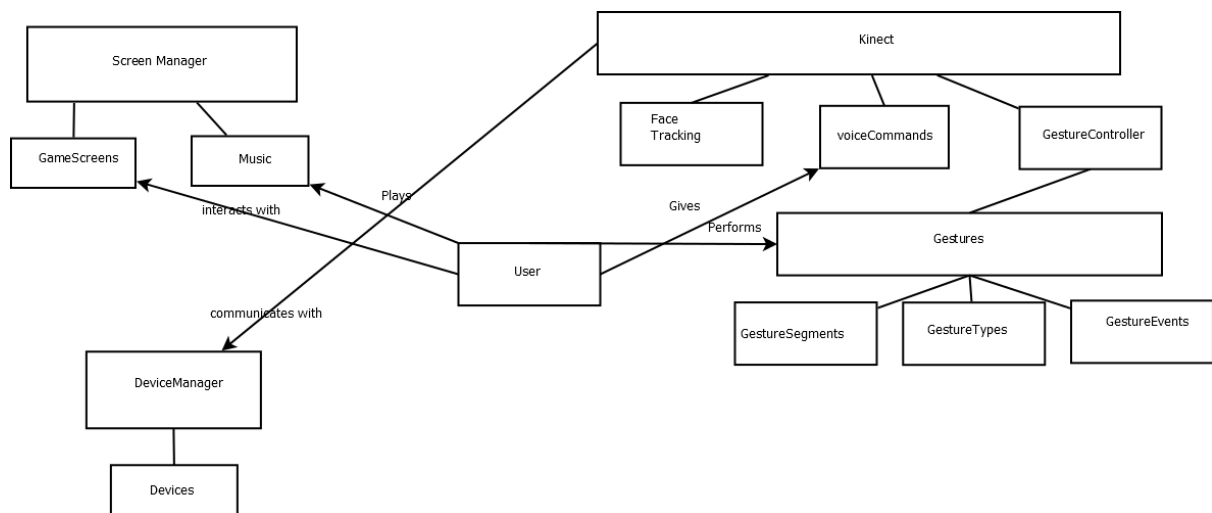


Figure 2.4: Software Architecture

2.6.2 Kinect Arduino Communication

The connection between the kinect and the arduino is done through serial ports with the help of the System.IO.Ports library. See fig:kinectarduinoconnection Serial ports, also called communication (COM) ports, are bi-directional. Bi-directional communication allows each device to receive data as well as transmit it. Serial devices use different pins to receive and transmit data using the same pins would limit communication to half-duplex, meaning that information could only travel in one direction at a time. Using different pins allows for full-duplex communication, in which information can travel in both directions at once. The connection manager system manages the connection details such as baudrate, portname, start/stop bits. It manages the connection between the arduino and the kinect as it opens the port, writes and receives data through it and finally close it. when a certain gesture or voice command is detected, the application opens the port, sends certain string to the arduino, the arduino receives it and does the expected command accordingly for example a "1" turns the internal LED attached to pin 13 on while a "0" turns it off. However the problem with this approach is that you have to specify the port name the arduino board is connected to, so the user should have the ability to choose from a list of available ports.

2.6.3 Screen Manager

The XNA framework allows communicating to the application through a touchless graphical user interface and showing feedback from the application. The ScreenManagers handles how to add a screen on top of another, When to show it, hide it, pause/unpause the screen and what screens comes next, it basically manages the flow of the screens. The class GameScreen is the super class which extends osoft.XNA.Game interface class to override the draw,update,initialize and loadcontent methods to add all the common features that will be added to all or most of the screens such as the useravatar, voice recognition and the music player.

2.6.4 User Avatar

The avatar on top right of the screen represents the user's distance from the kinect device and checks if the user is standing in correct position according to the Kinect field of view (fov) and constraints as stated earlier. The avatar was implemented according to how accurate the Kinect SDK can keep track of the skeleton in this range. The grey color means there's no user detected (out of range/fov), red means partially detected(too close), green means the user is in correct position, white means too far(starting to lose track of skeleton).

2.6.5 Music Player

The music player is implemented with the help of using Microsoft.Xna.Framework.Media library. You can play, pause, stop skip to next or previous songs by saying the correct

voice commands. To add your own music, just copy and paste them to the folder named "music" in the project content folder.

2.7 Gesture Recognition

The Microsoft SDK provides the player's skeleton as an array of joints of a vector4 value each, which contains each joint's x,y,z and w values. The x,y,z position values in 3d space and the quality level (Range between 0-1) of the position that indicates the center of mass for that skeleton. This w value is only 1 for passive fully tracked players. (See fig:skeletonjoints)

There's different ways you could build a gesture recognition system; you can record your gestures as frames and set them as a standard to compare the newly performed gestures to, however image processing would be expensive and may make the real time application lag depending on the application itself and its capabilities. Furthermore it would be very difficult to set the standards and the error ratio when comparing between these gestures in terms of the almost never matching joints' positions and gesture performing speed, adding that there are no gesture recording applications compatible with the Microsoft SDK yet. Gesture recognition in this project is done by comparing joint positions relatively to each others, draw imaginary vectors between the joints and compare them to the expected directions. You may also need to measure the angle between the moving body joints and add the expected range for each gesture to make your system more efficient. Each gesture is divided into smaller gesture parts/segments if needed and is checked for acceptance according to the method stated above in order through each frame. When all gesture segments are detected in order the gesture is recognized.

2.7.1 How it works

A gesture controller contains all gestures defined and updates them with the newly updated skeleton data and constantly checks if any is detected through a certain amount of frames. A gesture is defined by an array of gesture segments and the gesture type, a segment is a division of a gesture such that a gesture can be one segment or more. Basically a gesture is a sequence of segments performed in certain order. Each segment gesture check can return three values; fail for when it is not detected, pause for when the part of the segment is done, meaning the gesture is being performed at the moment allowing the system to wait for the next few frames and succeed for when the segment is detected. see fig:gesturearchitecture1 and fig:gesturearchitecture2

The check is done by comparing the skeleton joints x,y and z positions relatively to each others and/or by measuring the angle between joints.

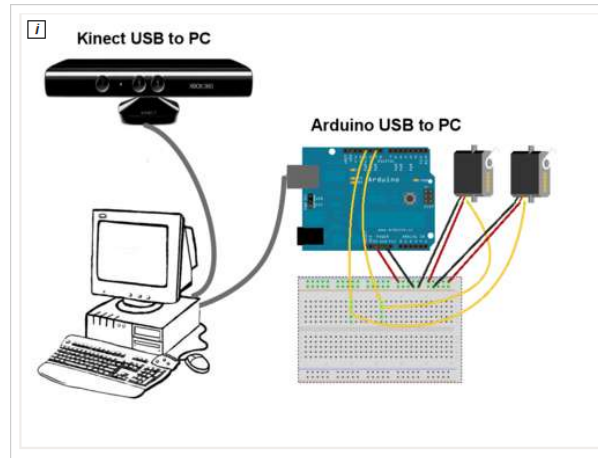


Figure 2.5: Kinect Arduino Connection

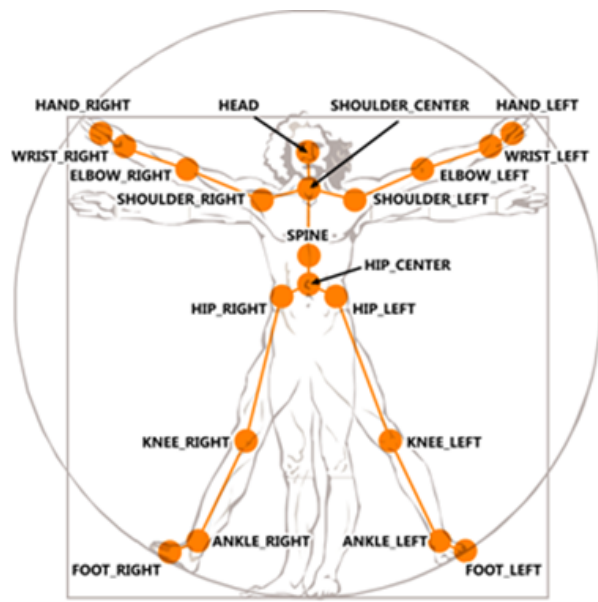


Figure 2.6: Skeleton Joints

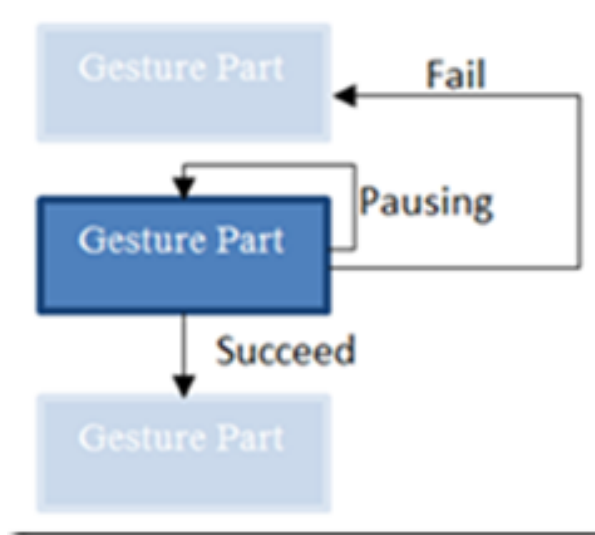


Figure 2.7: Gesture Architecture1

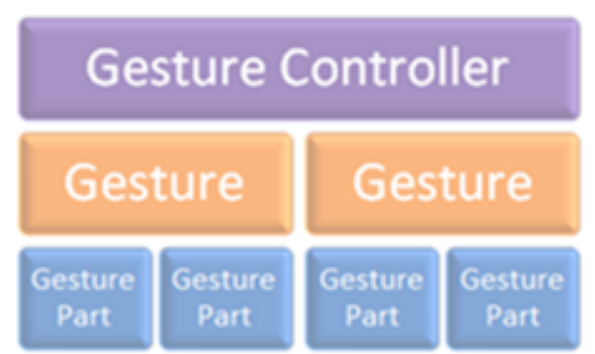


Figure 2.8: Gesture Architecture2

2.7.2 How to add a new gesture

To add a new gesture simply create a class for each gesture segment implementing the `IRelativeGestureSegment` interface and override the `checkgesture` result and write the algorithm for your gesture, then add these segments to a gesture made of an array of these segments, then use `GestureController.Add` to add your new gesture to the recognizer.

2.7.3 Defects

However this system may have some disadvantages if the gestures are somehow similar or overlapping, if a gesture sequence of segments is a prefix or a suffix of another then both gestures will be detected. Also some gesture may not be able to be detected due to the fact that the Microsoft SDK detects the user as a skeleton as an array of joints, so for an example the system will not be able to differentiate between a fist bump gesture and a clap gesture. however its really rare the user may want to implement two similar gestures so this will not be a problem in this project.

Chapter 3

Conclusion

Conclusion

Chapter 4

Future Work

Text

Appendix

Appendix A

Lists

Beginning kinect programming with microsoft kinect sdk by Jarrett Webb and James Ashley. Arduino cookbook by Michael Margolis.
<http://msdn.microsoft.co> <http://arduino.cc/en/Main>

List of Figures

2.1	Kinect sensor	3
2.2	Data stream	3
2.3	Arduino Board	12
2.4	Software Architecture	12
2.5	Kinect Arduino Connection	15
2.6	Skeleton Joints	15
2.7	Gesture Architecture1	16
2.8	Gesture Architecture2	16

Bibliography