REAL TIME VIRTUAL FITTING ROOM WITH FAST RENDERING

Pubudu Tharaka Hewa Edirappulige

168270U

CS6997 - MSc Research Project

Department of Computer Science and Engineering

University of Moratuwa
Sri Lanka

March 2018

DECLARATION

I declare that this is my own work and this PG Diploma Project Report does not incorporate without acknowledgement any material previously submitted for a Degree or Diploma in any other University or institute of higher learning and to the best of my knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgement is made in the text.

Also, I hereby grant to University of Moratuwa the non-exclusive right to reproduce and distribute my thesis, in whole or in part in print, electronic or other medium. I retain the right to use this content in whole or part in future works.

Pubudu Tharaka Hewa Edirappulige Date

I certify that the declaration above by the candidate is true to the best of my knowledge and that this project report is acceptable for evaluation for the CS6997 PG Diploma Project.

Dr. Indika Perera Date

ACKNOWLEDGEMENTS

My sincere appreciation goes to my family for the continuous support and motivation given to make this thesis a success. I also express my heartfelt gratitude to Dr. Indika Perera, my supervisor, for the supervision and advice given throughout to make this research a success. Last but not least I also thank my friends who supported me in this whole effort.

ABSTRACT

With the rapid growth of the technology, virtual try-on applications have become quite popular in past few years because they allow users to see themselves wearing different garments without the effort of changing them physically. Due to this, which helps users to quickly decide whether a selected garment is suited or not in small amount of time and also allows retail shoppers to sell more in less time with a high customer satisfaction. The objective of the present research is to create a virtual fitting room with fast rendering that realistically reflects the appearance and behavior of the garment.

Presently, there are several commercial virtual try on systems, such as fitnect, Zugara…etc. can be seen in the market with different price ranges and variants, namely web cam based, Kinect based, 3D avatar based, photograph based …etc. can be listed. But yet there are lots of improvement are needed from the aspect of performance, graphics, real time processing, realistic user experience …etc. Therefore this research was inspired to full fill those requirements.

The research mainly focused on how to extract users' specific body parts from the video streams which are provided by the Microsoft kinetic, then body customization and creation, skin tone matching , superimposing of garments on to the user which is extracted from the video stream and speed up the rendering process to create more realistic virtual fitting room system.

Finally, Evaluation of the presented virtual system was done by seven independent evaluators. They were asked to go through the whole system and answer few questions. Each these questions carried a rating from 1-5 and categorized into two main sections, respectively interface evaluation and functionality evaluation. Both of the sections had good evaluation results by proving research has full filled its intended purpose with the overall success rate of 87.53%.

TABLE OF CONTENTS

Page

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

Abbreviation	Description
NI	Natural Interaction
PCL	Point Cloud Library
VFR	Virtual fitting room
AR	Augmented Reality
VR	Virtual reality
ASM	Active shape model
API	Application Programming Interface
SDK	Software Development Kit
RGB	Red, Green and Blue
NUI	Natural User Interface
IR	Infrared
TOF	Time-of-Flight
GPU	Graphical Processing unit
PCA	Principal Component Analysis
MVVM	Model –View-ViewModel
WPF	Windows Presentation Foundation
API	Application Programming Interface

Chapter 1

INTRODUCTION

1.1 Motivation

Traditional retail outlets are the main key to the fashion industry in living world. The main reason for this situation is the most customers have adapted to buy the things after physically trying them out. But this process consume considerable amount of time due to different activates such as standing in front of fitting rooms, taking clothes on and off…etc. On the other hand being trapping inside a busy lifestyle it is no possible to people to find a suitable time slot to go and shopping on traditional retail outlets. However, with the time the necessities and choices of them upon their outfits have become complicated despite of being busy.

Along with the globalization e-commerce became known among the people and eventually they tend to do comfortable online shopping over the internet. Online shopping tries to have some share of the market by allowing customers to buy goods using internet facilitates while saving a lot of time Even though being advantageous by saving time, it avoids the capability of users to try on their outfits before buy them. Thus online shopping became known to be residing in a doubtful state and as consequence it does not fit in with purchasing cloths with e-commerce web sites. Generally, internet shopping processed according to the customers clothing orders relative to different sizes and colors first, but returning the majority of them when they don't seem to be fit. Even today these high rates of returning have become one of the leading cost factors for online stores.

As always, people want to try on clothing that build up their figure and features. This is more complicated than think since people vary in their gender, age, body texture, body color and stature. Since fashion industry is frequently updating with the most fashionable and latest designs new items and styles become popular in every time. Most probably the concept of "Virtual Fitting Room" provides best solutions for aspects which have indicated above. The main objective hold in this thesis is to introduce the concept of" Virtual Fitting Room". According to this concept it can limit the most disadvantageous conditions such as long queues and long waiting time in front of the fitting rooms. Because it can reduce the time taken for changing cloths. In addition it is capable of increasing the selling rates on e-commerce web sites by reducing the high return rates of clothes since customer has the ability to try and confirm the size of cloths that suit properly. Easy access to cloths by using this system will satisfy the necessities of people with different gender, age and stature.

Usually, this technique can be grouped under augmented reality (AR), where a real-time view of the reality is extended and furthermore overlaid with additional information.

1.2 Aims and Objectives

The main objective of this research is to create a Virtual Fitting Room with fast rendering that is capable of showing the physical appearance and the behavior of garment more realistically. Virtual human bodies and clothing is useful in various kinds of applications such as 3D animated movies, games, and online fashion.

The following are the other objectives reside within this concept.

i. To investigate on how to extract models of human body and body shape, extract the human skeletal with coordination.
ii. To study some existing selected virtual try on systems (FITNECT, Zugara) and related scope with reference to literature review.
iii. To Study on Microsoft Kinect device and SDK.
iv. To develop a prototype system using a standard outfit.
v. Study Human Movement prediction algorithms.
vi. To make the program suits to the specific stature of the customer.
vii. To develop the system according to the gender by creating multiple outfits.

1.3 Scope

The scope of the Research includes, identify and develop a Real Time virtual fitting room (VFR) with fast rendering using Microsoft Kinetic sensor. When a person is standing in front of the VFR, that person is allowed to select desired garments from the given set of garments which are displayed on side of the display. Once a garment is selected, then that is virtually superimposed on to the user to check how it looks like. Due to this, it allows users to get a better idea about garments before buy it even without putting them on.

In the first step, Capturing and Extraction of the user from the video stream which is provided by the Kinect sensor. Extraction of user from the video stream allows to create an augmented reality environment .In the second step, track the user with the help of Kinect sensor and SDK. Finally, superimposing garments on to the extracted user by using skeletal data provided by the Kinect SDK and speed up the rendering process.

1.4 The Problem/opportunity

Existing researchers on virtual try on system are based on either non-customized avatar or customized and dynamically generated avatar using camera and sensors technologies. The user's avatar is then augmented with virtual garments. But yet, due to lack of rendering speed, the most of the VFR applications are presented with limited realism. Therefore this research is mainly focused on how to build a VFR with fast rendering and an optimized software architecture in order to create a more realistic VFR application.

1.5 Natural Interaction (NI) sensors and supporting software platforms.

Capability of ordering and controlling the digital world using hands, body and voice known as the Natural Interaction (NI). Depend on the technologies Natural User Interfaces (NUIs) may lapse considerably from the classic desktop paradigm, since they use body gesture identification to direct applications and complete the duties such as browsing, selecting and checking out. Presently there are several commercial products and technologies, that are using NUIs in their applications namely, Perceptive Pixel, Microsoft Pixel Sense [27] ,3D Immersive Touch [28], Microsoft Kinect [26], Asus Xtion …etc.

Figure 1.1. Virtual Fitting Room with a Natural User Interface [7]

Microsoft with Kinect and Asus with Xtion are known to be the major NI players in the hardware sensor consumer market and both of them are using PS1080 chip from Prime-Sense [2]. Exposing of augmented reality (AR) to a completely new level has launched by Real-time modeling of natural scenes with graphics hardware and commodity sensors. These multi-sensor bar products gather synchronized video, audio and depth data streams.

Microsoft Kinect was promoted to the market in November 2010 and it was present as two versions namely, Kinect for the Xbox game console and Kinect for Windows. Additionally, Kinect for Windows consists of some improvements such as “near mode” operation, API improvements, skeletal tracking control etc.: As well as Kinect corporates in a wide range of applications as a voice-activated console. Microsoft has also introduced the Kinect as an improvement to the Xbox Live experience. Most precisely, a webcam-like live chat can be done with multiple friends at once through Xbox Live Video Chat, since its use of Kinect's cameras and microphones.

As well as The Xtion [29] from Asus occurs in three versions: Xtion, Xtion Pro and Xtion Pro Live. Last two versions have wrapped with drivers. The Live version adds an SXGA RGB camera to capture live image/video. Adaptive depth detection technology, infrared sensors, color image sensing and audio stream are used by both of Kinect and Xtion PRO LIVE version. Color image sensing and audio stream useful in capturing user's real-time image, movement, and voice. The significant properties of the Kinect and Xtion sensors shown in Table 1.

Specification	**Microsoft Kinect**	**Asus Xtion Pro-Live**
Field of View	57° H, 43° V	58° H, 45° V
Sensors	RGB & Depth	RGB & Depth
Depth Range	1.2m – 3.5m	0.8m – 3.5m
Depth Steam	QVGA (320x240) 16-bit @ 30 fps	VGA (640x480) @ 30 fps QVGA (320x240) @ 60 fps
RGB Stream	VGA (640x480) 32-bit @ 30 fps	SXGA (1280*1024) @ 30 fps
Microphones	3 (left) + 1 (right)	2 mikes for stereo capture
Power Supply	12V DC + 5V USB connection	5V USB connection
OS Support	Win7 32/64	Win 32/64: XP , Win7 Linux Ubuntu 10.10: X86,32/64 bit
Dimensions / Weight	12" x 3" x 2.5" at 3.0 lb	7" x 2" x 1.5" at 0.5 lb

Table 1. MS Kinect and Asus Xtion (Pro-Live version) pertinent characteristics [22].

Both sensors consist of supporting software environments and they make easy to create custom gesture-based applications for builders.

- **Gesture detection**: The ability to capture motions of hands in real-time, hence converting the user's hands into controllers.
- **Full-body detection**: ability of capturing full body motions with support for multiple player identification.
- RGB and Depth image sensing capable of capturing of video scenes in real-time which can be used separately or in combination.
- Audio streaming provides support for voice recognition and voice control.

There are mainstream development environments for NI sensors namely, Kinect and Xtion .These development environments consist of OpenNI (Open Natural Interaction) Framework, the Microsoft Kinect SDK and the Point Cloud Library. OpenNI [1] [5] Corporation was an industry-led, non-profit corporation. It has formed to declare and stimulate the affinity and interoperability of NI devices, applications and middleware. The corporation includes collaborators, like Prime Sense [2], Willow Garage (personal robotics applications) [3], Side-Kick (motion control games) [4], Asus (provides hardware such as the Xtion for OpenNI applications) and AppSide (end-to-end content marketplace for motion-controlled entertainment devices).

Prime Sense manufactures the PS1080 SOC, the heart of NI hardware. The chip capable of integrating depth images, color images and audio streams which are get from device’s cameras and microphones. As well as able to perform in hardware all depth acquisition algorithms. Additionally, it supplies a USB 2.0 interface to exchange information to the host.

OpenNI is a cross-platform open-source which is known as a multi-language (C, C++, C#, Java). It can be freely downloaded and it specifies APIs (Application Programming Interface) for writing applications using Natural Interaction (NI). OpenNI issues (API) and it facilitate the communication with both low level devices and high-level middleware solutions. In addition to that API can be utilized to improve applications which are used natural interaction. OpenNI APIs includes with several interfaces in order to write NI applications. Forming of a standard API that capable of exchanging data with both vision and audio sensors, vision and audio perception middleware.

The OpenNI Framework enables providing the interface for middleware elements and for physical devices. Its architecture has constructed throughout modules, production nodes and production chains. In order to know how to create and upload OpenNI-compliant applications, information can be gathered by utilizing OpenNI Arena and it is like a convention to developers.

Chapter 2

LITERATURE REVIEW

Most of virtual fitting systems are currently using different kind of sensors, GPUs, various AR based algorithms...etc. in order to provide more interactive and realistic experience to their customers. For example, some researchers use multiple Kinect sensors instead of one in order to provide more realistic visualizations. But in general what virtual fitting room does is, first it extracts the user from the video stream provided by a Kinect sensor/Camera/ Xtion by using depth and user label data. Then, register the garment model with the skeletal tracking data and finally, detect the skin to adjust the order of layers. But other than this method there are several other methods which were founded by researches during the recent past years such as cloning yourself, Hybrid Reality Virtual Clothes Try-On System, pre-recorded garments database…etc. Each of these methods has its own pros and cons and yet more or less improvements need to be done.

The main objective of this research is to create a Realistic Virtual Fitting Room with fast rendering. Hence, in order to achieve this objective, have to conduct research on following areas and identify the most suitable techniques and approaches.

- Need to explore technological variants available for “Virtual Fitting Rooms”.
- Need to find a way to extract user from the video stream.
- Need to find ways to body customization and skin tone matching.
- Need to find a way to superimpose cloths on a user's body.
- Need to find ways to fast rendering with optimized architecture.

2.1 Technological variants available for Virtual Fitting rooms

In this section, six technological approaches that have identified from different VFR researches will show and described. Among these six approaches, first two approaches belong to the AR, because the user can see himself/herself on the given display and remaining are based on the VR due to the usage of avatars.

2.1.1 Variant 1 : Real Person / Kinect / Superimpose clothing

In this method, the user just stands in front of the VFR and select whatever clothes he/she is given using simple hand gestures. Once the clothes are selected, those will superimpose on to the user's body which has appeared on the screen according to his/her body shape. Moreover, it allows the user to check cloths with different movements in real-time. It uses depth cameras like Microsoft Kinect to keep tracking the user movements. This is the most generic and popular solution currently available. E.g. Fitnect [7]

2.1.2 Variant 2 : Real Person / Web Cam / Superimpose clothing

In this method, it uses a generic webcam instead of relying on depth cameras and different sensors. From the customer's point of view, this is a more realistic and easy approach because a customer doesn't have to use different sensors or cameras to use the system. But when it comes to seller's point of view, this might be really expensive due to some approaches such as the use of local clusters for VFR [30]. Another main problem with this method is the lack of precision and realisticity due to the usage of generic webcams. Commercial product Zugara [11] is using this approach.

2.1.3 Variant 3 : Avatar / Kinect / 3D Clothing

In this method, the user is previously scanned with a depth camera and mapped with a virtual avatar defined in the system. So that avatar can represent the real user movements with the 3D clothes. For example, the research [23] namely, "Real-time virtual fitting with body measurement and motion smoothing" that conducted by U. Gültepe and U. Güdükbay was relying on this approach.

2.1.4 Variant 4 : Avatar / Input Body Measures / 3D Clothing

In this method, it uses more simple approach to create a VFR. In this case, a user has to manually provide his or her body measures such as height, burst, weight and waist to the system to build an avatar which is satisfying the given measures. Currently available web VFR solution "trimirror" [12] uses this variant.

2.1.5 Variant 5 : Predefined Avatar / 3D Clothing

This is the simplest form of the VFR. In this case, it just uses a ready-made avatar to represent the listed 3D clothes [31]. Users are allowed to choose a suitable avatar, which fits better with real user's appearance. Currently available web solution "glamstorm" [13] is using this variant.

2.1.6 Variant 6 : Photo / Superimpose Clothing

All aforementioned VFR approaches are using either online solutions or in-store based. But in this variant, it uses mobile phones based approach. In this case, a user takes a photo of himself or herself and then superimpose the clothes on to the picture and do the necessary adjustments according to the body shape.

2.2 Virtual Fitting Room (VFR) Applications

VFR-enabled shopping is accessible from homes, stores and smart mobile devices. Outstanding VFR applications, victorious representatives of which will be examined recently, mechanically generate accurate body scan data to covenant the standard of fit. These "Body Shape" IDs can be handling to various purposes such as to make precise 3D avatar-type models of the customer's body on which to fit desired garments and as well as to incorporate 3D scan data with information such as age and gender of the customer.

VFRs give distinctive opportunities for retailers along various directions. Shopper guarantee of a "perfect fit" result in the enhancement of sales and reduce the return rates due to the capability of removing the barrier for online shoppers. Some of the most using commercial implementations in the VFR industry can be listed as follows.

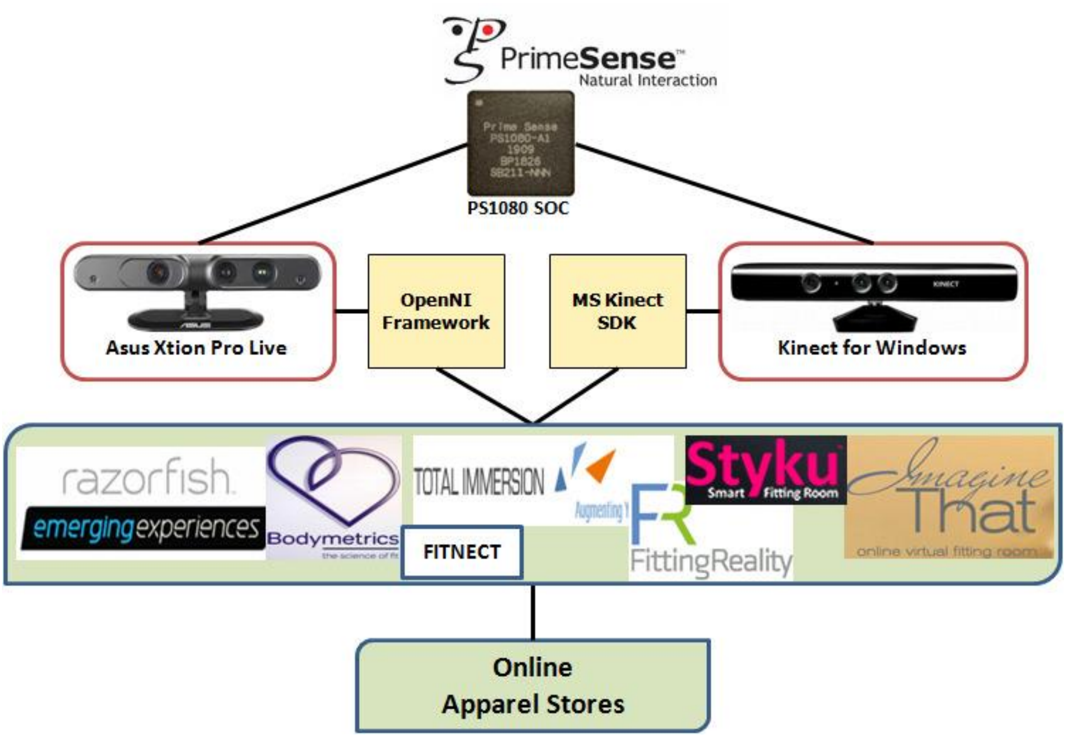

Figure 2.1. Various players in the Virtual Fitting Room Arena [22]

2.2.1 KinectShop from Razorfish Emerging Experiences

KinectShop is one of the most successful VFR application and also known to be an online AR shopping platform. Depends on the Xbox Kinect sensor, the platform converts customer's real movements into on-screen motions. It allows searching through a massive virtual shelf of accessories and allows to check it, in front of a big-screen TV which is capable of superimposing the garments with the user. As well as a customer can try to use the Kinect RGB camera to create a collection of different angles and multiple looks to share with friends via social networks.

2.2.2 Bodymetrics

Bodymetrics [7], a London-based technology company, has already invented 3D body scanning solutions for home customers and as well as for retail stores. The body scanner version which is using in retail stores uses eight PrimeSense 3D sensors to measure the customer's body precisely and allows to select the best fitting garment for the customer's body. The procedure is done subsequently by "Fit Stylists". The home customer version occupies the same sensor as used in retail stores but has based on NI sensors like Microsoft Kinect and ASUS Xtion.

These NI sensors provide facilities for scanning of body and then scanned data saved into an online profile. Additionally, when each piece of garment plotted on to the model’s body, a color map is covering the garment to represent the tight spots of the garment. So through this process customer is able to try on different sizes of garments and various types of clothing according to their body shapes.

2.2.3 Fitnect

Fitnect [7] is one the most successful AR based VFR application presently available in the commercial market. It utilizes the Kinect sensor and acts as a 3D device which is capable of capturing full-body. Widgets that have listed on the screen allows users to interact with the system by using Natural hand gesture Interaction and it allows customers to virtually try on garments according to their selection. Additionally, Fitnect utilizes 3D models of clothes and cloth physics to provide a stronger experience.

2.2.4 VIPodium from Fitting Reality

VIPodium is a shopping system and also known as an AR fitting room application. It is capable of analyzing full-body scanning image data to distinguish the physical contours of a user and make a ShapeID. Exploration of the desired garments and item selections can be done by using the 3D interface. The different versions of VIPodium available for home shoppers and for in-store usage. The previous versions of VIPodium occupied either Kinect for Xbox hooked to an Xbox 360 console or Kinect for Windows connected to a computer. In addition to above, it can also act as an apparel try-out place in order to increase the experiences of shopping.

2.2.5 “Imagine that” on-line Virtual Fitting Room

It provides [8] 3D models of clothing, a natural interaction user interface depends on hand motion recognition and social media combination and also facilitates the customer by giving them 360o view. Big-screen TV acts as a full-body mirror at in-store kiosk version. It creates models of the customer and programs the live video stream with the help of NI sensors RGB camera. Consequently, the user able to select different items to mix-and-match as well as to reduce the selection scope according to his or her interest. The online version state complete integration with occurring websites. Also utilizes investigative technologies and body scanning for virtual try-on, in the customer’s own home by giving them comfort. The combination of Social media also improves the shopping experience.

2.2.6 TryLive Eyewear from Total Immersion

Various trading AR platforms have released by the "Total Immersion" for mobile, web, and kiosk deployment following a usual TryLive logo: Virtual Fitting Room, Watches, Eyewear, Shoes and Jewelry. TryLive [9] Eyewear uses computerized face identification computations such as face shape classification. In order to do that it needs a camera-equipped computer, tablet or smartphone. Customers able to check the desired items, preview them cooperatively with different angles, exchange glasses and frames for a custom-built fit. Also, provide the ability to share the images with friends via social network media such as Facebook, Twitter etc.

2.2.7 Apparel Manufacturing Made for You – AM4U

Under this concept, the business structure based on “demand manufacturing” and follow a step moreover from other solutions by facilitating designing methods according to the customer preference and their body measurements by allowing for a garment to be cut, sewn, and shipped. Microsoft Kinect sensor is the major equipment used in the Styku’s smart fitting room. It is capable of scanning the shopper’s body and able to build an embodiment which is more suits to the dimensions of the shopper. With the help of Styku’s customization tool, users can create their own clothing in 3D while judging the standard of fit on their own custom embodiment. In addition by using Tukatech’s [10] TUKAcad and TUKA3D software, customers able to virtually check the desired clothes, determine the standard of the fit, recognize the sections that might not suits(loose or tight) and change the pattern or measurements of the garment based on customer preference.

The final output is then sent off to fabric printers namely "Critical Mass", where use the innovative waterless and chemical-free technology to print and dye the clothing. As consequence, with the help of traditional automated sewing machines, fabrics are then sewn into garments, pressed, and distributed to the customer.

2.3 Extract user from the video stream

In order to extract a user from the video stream provided by an input device (Microsoft Kinect, Web Camera), different researchers have used different approaches. For example, Research [14] that was conducted by F.Isikdogan and G.Kara used a method which, first extract and isolate the user from the background to create an augmented reality environment. To segment the foreground, they have used the depth images and user labels which are provided by the Kinect sensor as in Figure 2.2. Then detect skin color and bring it to the front layer to allow the user to fold arms or hold hands in front of the garment model. And they have threshold the image in YCbCr color space, using the values that were found to be effective for skin-color segmentation, as follows,

$$Y = 0.299R + 0.587G + 0.114B < 70$$
$$77 < Cb = 128 - 0.169R - 0.332G + 0.5B < 127$$
$$133 < Cr = 128 + 0.5R - 0.419G - 0.081B < 173.$$

YCbCr is a family of color spaces used as a part of the color image pipeline in video and digital photography systems. Y represents luma component and CB and CR represent the blue-difference and red-difference Chroma components [15] respectively. Other than that, in order to prevent any background pixel from being labelled as skin, they have applied the threshold only on the segmented foreground.

Figure 2.2 Depth image (left), color image (middle), extracted user image (right) [14].

Research [16] which was conducted by S. Saha, P. Ganguly, and S. Chaudhury, introduced another way of extracting user from the background by using a generic webcam instead of relying on depth cameras and different sensors. This is a more realistic and cheap approach because a user doesn't have to use different sensors or cameras to use the system. The functionality of the system has achieved in two steps.

In the first step a user selects an image of a cloth which he/she wants to try-out and in the second step system captures the user from the web cam and superimpose the cloth on to the user's body as shown in the Figure 2.3

This research can be divided into three main sections namely, data acquisition, pose estimation and apparel fitting. In data acquisition, first, the camera records the background without moving of the user. Once it is completed, the user is allowed to be in the field of view of the web camera. Finally use frame difference, background registration and subtraction techniques [32] to segment out the human body from the video stream. In the pose estimation, calculations are based on head, neck, elbows and end of the feet and hands. Finally, the selected cloth is superimposed on the user's body.

Figure 2.3 Cloth is superimposed on to a user [16]

2.4 Body customization and skin tone machine

When it comes to 3D avatar based variants, Body customization and skin tone machine have to play a big role in VFRs. Because those are mandatory factors that are required to preserve the realistic appearance of 3D based VFRs. Simply Body customization and skin tone matching can be divided into two main sections namely, Manual configuration and Automatic Configuration. In manual configuration method, a user has to manually provide his or her body measures such as height, burst, weight and waist to the system in order to build an avatar which is capable of satisfying the given measures. Research [17] which was conducted by Rong Li ,Kun Zou, Xiang Xu, Yueqiao Li and Zan Li, they have implemented a VFR, which can be used to easily evaluate wearing effects of clothes online, with the above mentioned manual capabilities.

In automatic configuration user doesn't have to bother about proving height, wrist or any other details. Instead body creation happens automatically according to a user who is stand in front of the VFR which based on data that are provided by the Microsoft Kinect. Researches M. Yuan, I. R. Khan, F. Farbiz, S. Yao, A. Niswar, and M.-H. Foo have presented a hybrid solution [18], which was implemented using both virtual avatar plus real user. In this method virtually clothed avatar body blended with the real user's face. Depending on the user's gender, the system will automatically show a male or female avatar model, correspondingly. After selection of an avatar model, at first, it gets automatically customized based on the size of the actual user and its skin color. The body skin color of the user is created by matching the face skin color. Because, the face is the only place that has less effect in different light conditions [33]. Furthermore, researchers have used a novel algorithm to accurately align the customized 3D avatar with user's image in real time. Finally merging of the real user's head with the virtually generated and customized headless avatar occurs. As a result, the user sees he/she look like an avatar on the display wearing virtual clothes.

2.4.1 Body customization

Avatar customization could be a more suitable and promising design decision in modelling and simulation applications rather than creating a model from the scratch with different body shapes and sizes. Because it is much economical in terms of time and effort. Research [18] has presented a method to modify a generic avatar based on twelve key human body measurements i.e. shoulder width, height, bust girth, waist girth, hip girth, thigh girth, ankle girth, waist height, crotch height, knee height, upper arm length and forearm length for their try-on system. Some of the measurements, like height, arm length, shoulder width and waist height can be directly received from a kinetic sensor and can be used on an avatar to properly configure it. They modified the avatar by scaling vertices locally and globally. This method is relatively faster and much promising to produce natural looking results compared to other solutions which were found in other researches.

2.4.2 Skin – tone Matching

Another main requirement to achieve a more realistic appearance in a VFR is the skin-tone matching of an avatar according to a user who is standing in front of the virtual fitting room. Because in “Hybrid Virtual Clothes Try-On Systems” and “Full 3D avatar based try on systems” they use virtually generated avatar body to represent the real user.For example in research [18], researchers have used user’s face skin color to match skin tone of the dynamically generated avatar. In order to do that researcher have broken up the process into three main steps. In the first step, with the help of active shape model technique (ASM) facial features are located [19]. Then in the second step extract the cheek patches of a user to create the color of the skin by transferring RGB colors to YCbCr [15] color space and threshold the chrominance component. They have selected the cheek area, because it is the largest flat muscle on the face and least affected area with shadows (e.g. forehead, hair…etc.).Finally, have used a global color transfer method to apply the created color to the avatar body.

2.5 Superimpose cloths on a user's body

VFR applications allow users to see how the selected cloth may look like on his/her body without putting it on while standing in the in front of the system. Basically, depending on the cloth model that a VFR application is using, that can be classified into two categories. Therefore it can be in either 2D based cloth model or 3d based model. In the 2D model, it just displays a still image of the selected cloth on the user's body without any animation or rendering. Some existing commercial products such as Zugara, Swivel...etc. use this approach. In the 3D model, it provides more realistic cloth animations while a user is posing in front of a VFR. Some existing commercial systems such as fitnect use this approach.

Due to this differentiations researchers have to use different techniques to superimpose cloths on a real user or avatar model. The most simple way to superimpose clothes on a user is, first capture the user who is standing in front of a VFR using the Kinect sensor and get his/her skeleton data(Figure 2.4). Finally, overlay the selected cloth on to the skeleton by using necessary points of the skeleton. For example, to map a t-shirt, it is required to get shoulder and hip points, to map a handbag, it is required to get wrist left and right points.

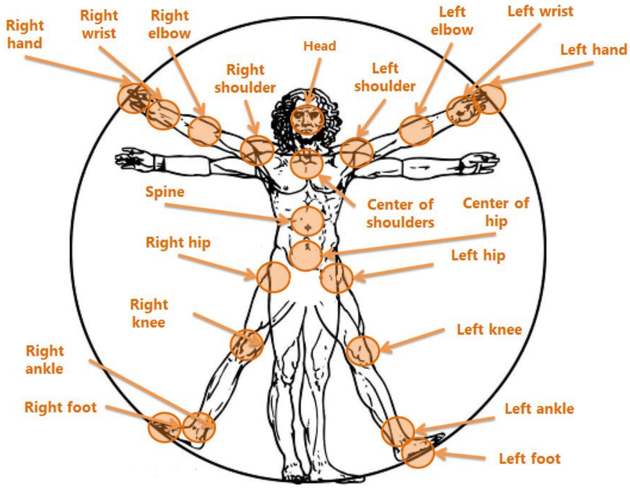

Figure 2.4 : Skelton point that are used for registering the cloth model [24]

Except for the above-mentioned method, there are several other techniques which can be used to achieve more realistic cloth superimposing in VFRs. For example, research conducted by S. Hauswiesner, M. Straka, and G. Reitmayr [20] have presented a way to use previously recorded garments in a real-time virtual fitting platform. In this method, simply it transfers recorded garments from one user to another by matching pre-recorded frames with the input from the Kinect sensor, using image-based visual hull rendering and online registration methods. Due to the usage of real garments, it improves the realistic appearance with high-quality rendering. On the other hand, it is very suitable for wide range of garments with complex appearances (e.g. wedding dresses), because it allows viewing garments from arbitrary viewing angles.

In this approach, it uses previously recorded garments to generate virtual clothes on to the user who is standing in front of the display and also allows moving freely and can see from arbitrary viewpoints. They have achieved this by creating a garment database that uses to store different perspective of a worn garment over time. The recording is doing with multi-camera devices used for augmentation and these recordings are not required to be from the same user. At the runtime, the best fitting frame is selected, for rendering by querying the database with silhouette images. The image-based visual hull algorithm is used to render users and clothes from arbitrary viewing angles. In order to fit the transferred clothes to a user and adapt the garment's shape to its new body, both non-rigid and rigid registrations are performed.

2.5.1 Virtual try on

Methods that indicated above only works on the best matching dataset which extracted from the previously recorded database and this database contains all possible configurations. The system contains an implemented method known as "pose similarity metric" in which first learn and then search a database of poses. The best match is selected and they operate in 2D. Hence, do not allow the user to view him- or her from arbitrary viewpoints.

3D scans of real garments are attained by color-coded cloth and real samples are used to measure the Cloth surface properties. MIRACloth is a one of the most significant clothes modeling application which is used to create clothing and also capable of superimposing them into human models. But, as similar as the virtual try on MIRACloth application also do not consists of mixed reality component. Mixed reality component is a kind of application that allows users to see realistic clothing on themselves immediately.

Kinect-based body scanning consists of a single sensor and requires multiple views, as well as modelled garment meshes, are generally used by the system. Usually, it performs a non-rigid registration to fit the desired garments with the user. A pose adaption process called "Volumetric Laplacian deformation" is used to Transfer the garment meshes from one human model to another human model in virtual try-on applications. Virtual try on based on three main steps namely, Motion capture, reconstruction and retexturing that are used to render dressed people while other virtual mirrors are restricted to specific tasks, like augmenting logos.

2.5.2 The augmentation process

The augmentation process consists of two phases namely, offline phase and online phase. Offline phase is responsible for recording garments and online phase is responsible for cloths augmentation. In the recording process, different poses of a clothed user have recorded after he or she entered the recording room. Then the features of the human body extracted and stored in a garment database with the Garment segmentation. An advantageous condition is, this phase can be controlled such as able to recapture the scene when it is incomplete or switch segmentation strategies. In the runtime phase (online phase), free movements of a user captured and features get extracted. As the next step best fitting pose of the selected garment and the captured user have rendered from the same viewpoint. This step is done by using image-based rendering. Both rigid and non-rigid registrations are used to compensate small pose mismatches. As a consequence, composite mirror image results from the two phases.

2.5.2.1 Offline: garment database construction

In the offline process, that involves both model-user and an end-user. The model-user capable of performing all the desired motions with dressed garments, while the end-user just accept or consume the outputs of the model user of as he or she wishes. Recordings are synchronized, and ten cameras are involved in the recording process. These cameras are able to capture the series of different poses performed by the model-user. Each garment database consists of information about every single piece of clothing. Thus, every single piece of clothing needs separate recording and database with respect to their size and the type of the garment.

Garment database construction mainly includes four essential steps namely, Preprocessing, Extracting features, reducing the dimensionality and extending the pose space. At the level of preprocessing, that takes all camera videos as inputs and segmented them separately. For instance, subtracting of previously recorded static background from another recorded image. In the process of extracting features, the most suitable frame must be spotted. The most suitable frame means the frame, which consists of model-user's pose that is more similar to the current user's pose. The most suitable frame can be detected, by comparing colors, silhouettes or higher level properties such as motion capture positions. In order to extract features, have to compare three different silhouettes sampling approaches. Usually, the first approach uses a radial Pattern. On the other hand, the second approach uses an axis-aligned sampling pattern while the third approach samples the silhouettes densely. In order to reduce the number of dimensions, have to use the method of principal component analysis (PCA), which is able to reduce the thousands of dimensions to around 50 dimensions for line sampling. In addition to all above steps, there is another possible step, which is useful in situations such as when model-user forgets to cover certain poses, or when the temporal resolution of the recording is not sufficient namely, extending the pose space. This task can be fulfilled, at the runtime by combining several poses into a single output.

2.5.2.2 Online (At runtime): clothes augmentation

Cloth augmentation is a process that Proceeds at the runtime. In this phrase, a user can enter to the virtual fitting room and try out different outfits which are stored in garment databases. In order to output the current user pose with the selected garments first, have to extract the poses of the current user which are corresponding to the creation of current configurations from the offline phase, and transformed them to PCA space by applying the pre-computed transformation matrix.

Figure 2.5: Creation of a new pose by combining the previously recorded poses of model user. (a) Illustrates the pose of current user. (b) and (c) best fitting poses for each image half. (d) Result in output by combining two halves together without non-rigid registration. (e) Final result image with non-rigid registration. The camera images which are closer to the pose of the current user most likely show the slightly different position than the current user. To solve this issue a rigid registration is performed.

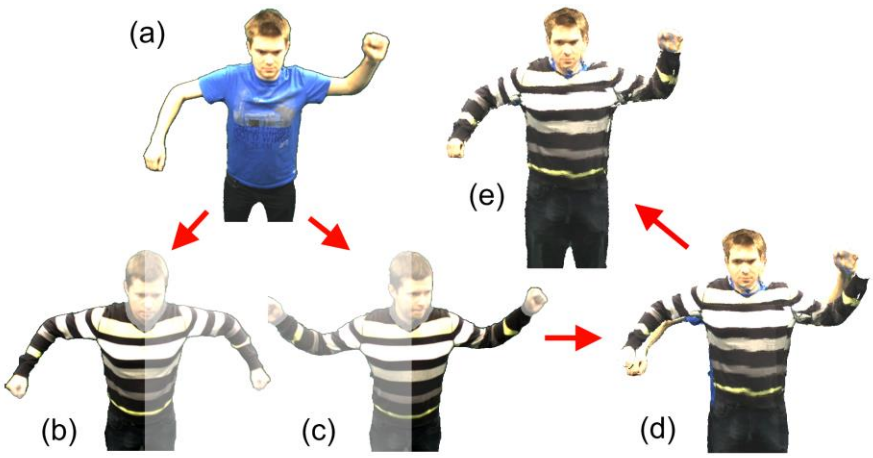

Figure 2.5 – Cloth superimposing process [20]

2.6 Fast rendering with optimized architecture

One common use of predictive tracking is to estimate the future orientation and position in order to minimize the perceived latency. That means, try to minimize the time between movement and when that movement is updated on the set up screen (On VFR Display).

2.6.1 Reasons for latency

Latency can come from different sources such as,

- Processing delays: In order to process data from different sensors [34], are often combined with some kind of sensor fusion algorithms. Therefore execution process of these algorithms can cause delays.eg calibration of RGB and depth data.

  The calibration methods can be divided into four main categories. In first category assumes that the three sensors (RGB camera, IR camera and IR projector) behave in the same way and that the pinhole camera concept is applicable to everyone. This method uses the RGB and photogrammetric beams to calibrate the internal and external parameters of the three sensors. Conventional distortion models for each sensor separately compensated for their distortion effects. Due to the difficulty of getting raw data from the IR camera, the author calculated such data from the disparity and the approximate baseline among the IR camera and the projector. The main limit of this category is the dependence between IR and camera and the related calibration parameters.

  In second category that combines disparity image produced by the depth sensor and the image produced by the RGB camera, and an empirical mathematical model to eliminate the distortions of the IR camera. The distortion model is based on the error behavior of the whole RGB-D sensor. Unlike the first category, this one is restricted to Kinect v1 and another limitation is that includes the lack of automation and rigor. The user has to select the points in the depth image manually, and the homography matrix is calculated using only four corresponding points.

  The third category is dealing with the refinement of the in-factory calibration procedure, where the manufacture parameters including the baseline between IR projector and camera as well as the standard depth were precisely calibrated along with the RGB camera. The author used the maximum likelihood estimation to calibrate color internal parameters and the manufacture parameters for a depth sensor without any distortion effect. The main limitation of this method is the distortion parameters for both IR cameras and projectors, which are not estimated or compensated.

Finally, in the fourth category that is mainly dealing with the depth calibration of the RGB-D sensor, the method adopts a mathematical error model derived from observation equation of depth measurements. The method used the output depth information from depth sensor and an adequate experiment setup to obtain the true depth of each pixel to model each pixel's depth error. The method is valid for both SL and ToF RGB-D sensor. Although the error model was precisely applying to the middle part of the depth image, the whole image's depth error model can be achieved and applied to the whole depth image. The error model has demonstrated a significant improvement while scanning small objects. As aforementioned based on these calibration algorithms, different calibration methods consumes different times to do the calibration. So based on the selected calibration algorithm difference processing delays may occur.

- Sensing delays: Sensors are failed to report updates (orientation or position changes) instantaneously. For example, the time gap between receiving of data of a user by kinetic sensors and sent it back to the host processor. This kind of things can be happen due to fast movements of user. For example in Kinect version, that uses IR projections to capture user movements has less efficiency compared to Kinect version 2. Because Kinect version 2 uses TOF method to captures user movements.

- Data smoothing & Filtering: Sometimes, in order to prevent jitter and noise [35] [37], hardware/software low pass algorithms are executed. For example, the joint filtering latency is, the time required for the filter output to detect the position of the actual joint when there is movement in a joint. In this kind of situation that the filter output is lagging behind the input when there are changes in the input. It is important to note that the latency introduced by the joint filter is not the CPU time required to execute the filtering routine. In general, the delay of the filter depends on the rate at which the input changes, and therefore it is not possible to assign a specific delay value to a given filter for all cases. Phase distortion is used in signal processing. A special class of filters called linear phase filters has the same delay for all input frequencies and a specific delay time can be assigned to the filter for all inputs. Reducing phase distortion is important in some signal processing applications, especially in audio processing. However, it is not necessarily important in NUI co-filtration, so having a linear phase filter is not a design criterion in NUI co-filtration.

- Transmission delays: Transmission delay [34] [37] from a sensor to display (e.g. Data transfer over USB). In a VFR setup, transmission of video from a Kinect to computer can cause a considerable delay. Because video transmission needs high bandwidth connection and processing load of the system to transmit high quality images with low latency. So this kind of issues can be corrected through a improvement of hardware or by using more efficient video coding or decoding process or even using kind of prediction algorithms like, Kalman predictor, Dead reckoning algorithms.

- Frame Rate & Rendering delays: Time to render an image [36], which is ready to be sent to the display device. Obviously this can be experienced when you're playing a high end video game with low end computer. This scenario is also called lag. It is possible to remove lag by adjusting some parameters. Such as, use of video card to do the parallel processing instead only relaying on CPU.

2.6.2 Common prediction algorithms

- Dead reckoning: It is a very simple algorithm based on position and velocity details. For example, if the velocity and the current position is known, at a given time, then it is possible to predict the next position based on that details.

  E.g. Give position is 100 units and velocity is 10 units/sec. Then the predicted position after 10 mSec can be calculated as follows.

  100 + 10 x 0.01 = 100.1

- Kalman predictor [38]: Based on Kalman filter. Used to reduce sensor noise in systems.

2.6.3 Dead reckoning

As indicated above dead-reckoning [21] algorithm is a technique that can be used to reduce, the amount and size of the messages that are exchanging between the participants, and also useful for basic non-segmented objects in popular systems such as NPSNET. Dead-reckoning algorithm can be explained, by using dogfight game with n players as the example. In this game each player capable of controlling their own ship. If player A moves its own ship, it conveys the same message to other (n-1) players, including the new position. As the same, if all players are participating the game move, all about n*(n-1) messages are conveyed. So in order to decrease this over exchanging of messages the player A can sends the position and velocity of its own ship to other players. By the use of velocity and current position data, other participants are able to guess the new position of the player A. This estimation procedure is known as the dead-reckoning technique.

As well as, each participant able to stock alternative copy of its own model and it is known as" ghost model". The player A should share the original position and velocity with the other players if the difference between the original position and this extra copy is larger than a presume maximum amount. This message sending is only occurring if there is a huge difference between the real position and the estimated position. Through the message sending other participants able to accurate their copy of participant A's object.

The potential of the dead-reckoning algorithm has based on the accurate predictions about the upcoming frames. In order to improve the dead-reckoning algorithm, body model and the properties of the simulation must be taken into consideration.

2.6.4 Virtual human representation

An accurate controlling of the body pose requires a similar segmented structure related to the human skeleton [21]. Animation of virtual human figures and the Real-time representation have recently become as most outstanding areas involve in computer graphics. Format related to the body configuration has to be linked, with the skeleton, and clothing might be covered around the body configuration. HUMANOID is one of the segmented human body model that can be used.it consists of 74 degrees of freedom without the hands, with additional 30 degrees of freedom for each hand.

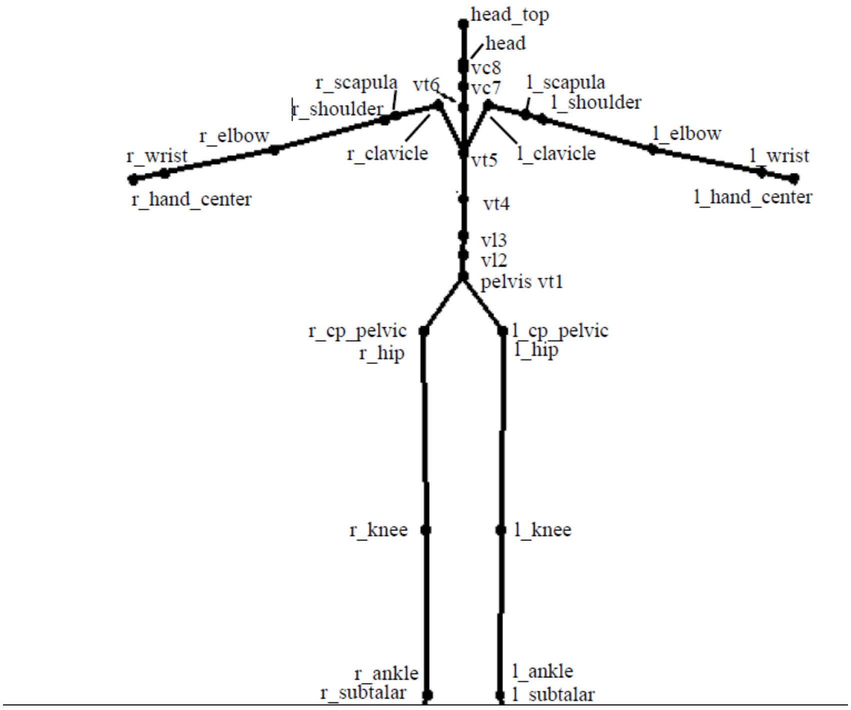

Figure 2.6: Virtual Human Figure Representation [21]

Generally, the same algorithm which is valid for the body joints also can be applied to the hand joints. A 3D segmented pecking order of joints has used to present the skeleton with accurate maximum and minimum limits. The skeleton has enclosed with geometrical, topographical, and inertial properties of various body limbs. Global scaling, frontal scaling, high and low lateral scaling, and the spine origin ratio between the lower and upper body are the escalating parameters, that can be used to form various body instances. As well as the structure of the body consists with a stable topographical arrangement of joints.

When considering the skeleton meatballs in the second layer act as muscle and skin. Hence most of the body can be represented by the blobs and the major satisfaction depends on enabling us to coat the complete human body with only a small amount of blobs. Based on this, the body can be divided into 17 parts, namely head, neck, upper torso, lower torso, hip, left and right upper arm, lower arm, hand, upper leg, lower leg, and foot. The head, hands and feets have constituted by triangle meshes, but not with the blobs because of the complication of these parts. The cross-sectional tables have used for distortion for the other parts. These cross-sectional tables have created by dividing each and everybody components into a number of cross-sections and by calculating the outlying crossing points with the blobs. And additionally, these tables only formed once for each body. The outlying points constitute the skin outline. The skin outline has connected to the body during the runtime and every step interposes around the connection based on the joint angles. New body triangle mesh is created by distorted element using interposed skin contour. Hence, body data includes in a single frame is capable of representing as the rotational joint angle values.

2.6.5 Denoising of kinect depth data

For past years, computer science domains such as motion capture and 3D reconstruction used two main types of devices namely, marker-based [44] and marker-less. By considering the marker-based device, since it is not suitable for the daily use their usage has limited to only certain applications. However, in marker-less devices since webcams, DSLRs and video recorders can be customized and combined together to fit almost any kind of application. Hence it may be considered as more convenient technology contrast to the marker-based devices. Even though marker-less devices are profitable, it bears several limitations such as problems in achieving a good calibration to obtain a reliable depth map in stereo-camera systems and have to bear a high computational cost.

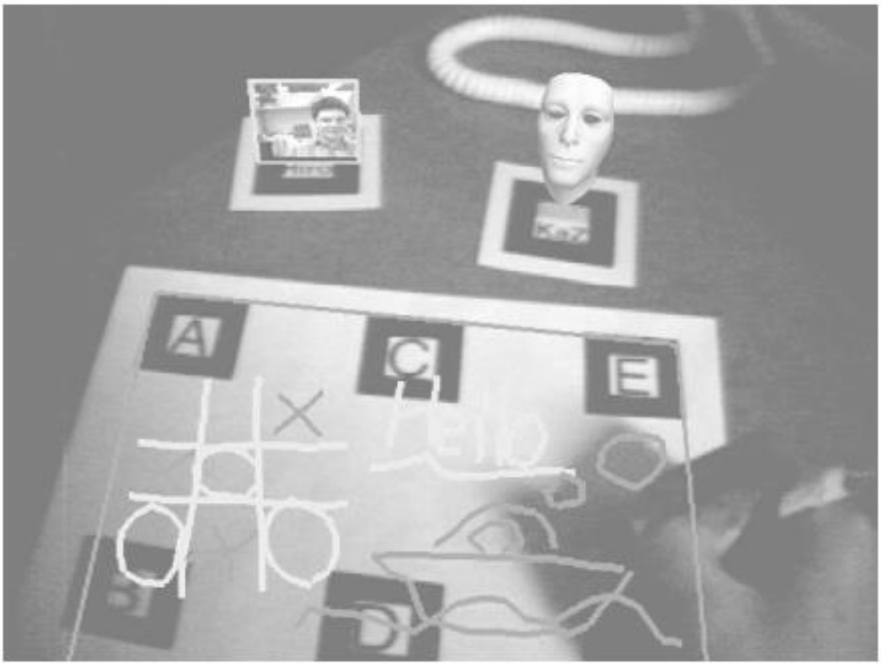

Figure 2.7 Maker Based Augmented Reality [44]

The release of the Microsoft Kinect has attracted the attention of researchers in variety of computer science domains in terms of hardware. The Kinect is an economically convenient device, easy to set up and handle unlike laser scanners and ToF(Time of Flight) cameras. As well as it can be used as a replacement for stereo-cameras and multi camera systems, which provides a depth map of the scene. In recent years, Microsoft Kinect has used in different applications namely, robotics, human motion analysis and human-computer interaction due to its outstanding performances.

In building of human models, most researches are struggled to build them with conventional devices since it requires more accurate motion analysis. Hence the Kinect can be profitably used for the human motion analysis since it seems to be one of the most promising. Even though the Kinect holds the considerable amount of advantages it has its own limitations and disadvantages. Since most of the studies have been based on how to extract the full capacity of the Kinect in different kinds of applications, but very few studies have done on its limitations such as in the precision of depth measurements. Furthermore, problems such as calibration of multiple Kinects, with the associated interference effect, combining of multiple Kinects are difficult tasks, which is still far from being solved. According to the studies carried out by comparing Kinect-based systems with marker-based multiple-camera systems, results indicate that the Kinect is promising, but still not as good as the other systems. Especially, depth instability, resulted by the nature of structured light-based sensors, avoids the utilization of Kinect depth data in applications that require a high level of accuracy. Research "Temporal denoising of Kinect depth data" [45] is based on presenting an algorithm to solve the instability of depth measurements by developing the stability and reliability of the Kinect depth map with low cost.

Basically, The Kinect consists of three main parts: an IR projector, an IR camera and an RGB camera. When computing the depth map of a scene, the Kinect functions creates some problems due to the lighting conditions and interferences of the structured light and the objects in the scene. To overcome the limited coverage area of 3.5 meters, multiple Kinects can be combined if there is no any interference between the calibration and the projected patterns. Hence Interference is a serious challenge in multi-Kinect set-up since the intersection of the IR projector fields of two or more Kinects results misplace of essential data due to the less capability of the device to recognize the projected IR pattern correctly.

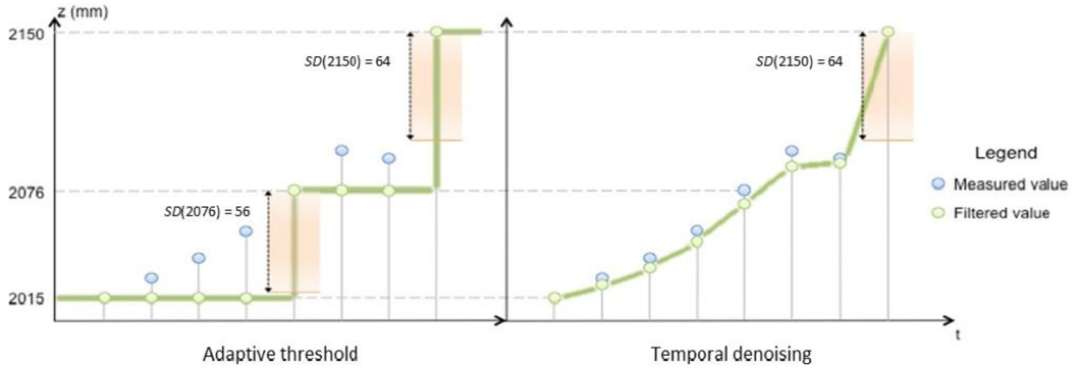

Figure 2.8 Adaptive threshold filtering vs. temporal denoising filtering [45]

The problems which Kinect suffer from can be noticed as a flickering and vibration of the depth values. The modified the Smoothed Pointing algorithm (DEVELOPED TEMPORAL DENOISING ALGORITHM), which can be identified as a velocity-based precision-enhancing technique for distal pointing. Smoothed Pointing algorithm is suitable to filter out Kinect depth vibrations since it works by dynamically adjusting the control-to-display (C-D) gain according to the average velocity of the user movements. The first part of the denoising algorithm is based on a sinusoid function and is in charge of reducing the vibration and smoothing the slow movements. Unlike adaptive threshold suppression algorithm, this algorithm works by verifying if the difference between the previously recorded value and the current measured value is greater than a depth adaptive threshold; it changes the depth value to the current measured value, but if the difference is lower than the depth adaptive threshold it keeps the depth value as the previously recorded value. According to the preliminary experimental results, adaptive threshold suppression algorithm completely discards the slow movements, while proposed temporal denoising algorithm is not filtering out the slow movements and does not cause the discontinuities in the depth measurements. Hence temporal denoising algorithm is suitable for use in applications that need to capture small details in movements, like in the field of human motion analysis. As well as the results also represent the real-time applicability of the whole filtering process.

2.7 Image based 3D sensing

3D models of the human body can be captured by using imaging sensors and is considered as one of the most significant tasks of the Virtual-try-on system. In many application areas, image-based sensing techniques have widely used with the improvement of inexpensive and best-quality digital cameras in recent decades. Moreover, single and multi-camera setups and active cameras are the some of the techniques that exist to fulfil the task of image-based sensing. This section deals with the existing camera technologies, their possible uses, main fundamentals regarding each of technology and represents a camera technology which is able to capture a metric 3D representation of a human body in real-time.

2.7.1 Monocular cameras

A monocular camera system mainly consists of two parts namely, single image sensor and an optical lens. Without the information about the scale, 3D data cannot be captured by a single static camera. Even though it is not able to capture the whole 3D information, it is still capable of taking simple 3D body measurements using a single camera. Without the need of clear 3D information, Anna Hilsmann and Peter Eisert [54] have used a single camera to augment the camera image with new information without the help of explicit 3D data. Their explanations mainly suit for the real-time visualization of garments in a virtual try-on environment and their attempt is to describe it through dynamic texture overlay method from a monocular image.

2.7.2 Multi-view systems

In a Multi-View System, more than one camera has used and these cameras are able to detect or capture the same object from different angles. If there are two cameras are involves in the in the system, it is known as stereo setup and for more than two cameras the system is called as a multi-view setup. By knowing the camera geometries and orientations between the cameras, it is possible to easily derive scale and depth information from the scene using more than one viewpoint over single camera system in contrast to the Monocular Camera system.

In order to fulfil the task accurately, cameras which are used to derive depth data should capture the same scene at the same time. As well as the object which is capturing by the camera system must not move or deform until each camera has captured an image. In general, with the help of either synchronized cameras or fixed objects, this requirement can be easily achieved. A typical example for multi-view camera setup is shown in Figure 2.9.

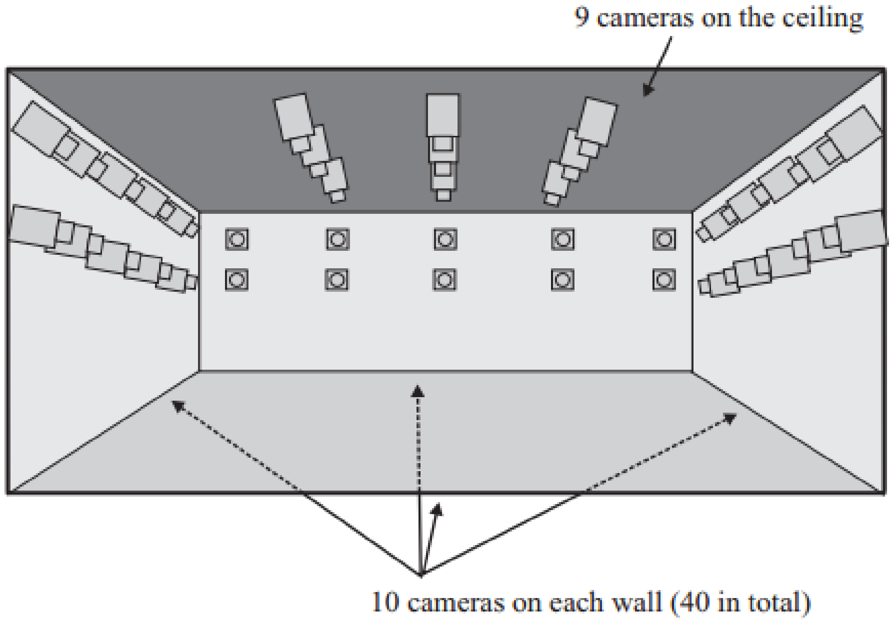

Figure 2.9 Multi-view camera setup [46]

By processing silhouette images, 3D data of multi-view images can be easily and quickly obtained. A silhouette is known as the background segmented binary image of an object. The 3D visual hull of a particular object can be calculated by using a set of multi-view silhouette images. Closest convex approximation of an object is known as the Visual hull and it can be obtained with an approach namely, volume intersection. Not only above indicated approach but also there are several approaches available to compute the human pose or human shape from silhouette images [47].

By utilizing epipolar geometry [48] it is possible to derive the depth information from a multi-view setup (Figure 2.10). As well as there are several algorithms exist to calculate dense depth maps from multi-view images. Surface estimation of the human body can be improved using such depth maps. But, when it comes to the real-time applications, computing of dense depth maps for multi-view images is still too slow even if, parallel computing platforms are used. Point clouds (sparse representation of a 3D object) consider a collection of points and they describe the surface of the object using these points. Basically, such points are acquired by triangulating the 3D position of key-points from two or more image inputs [53]. Through a set of subsequent steps a sparse point cloud can be densified around key points.

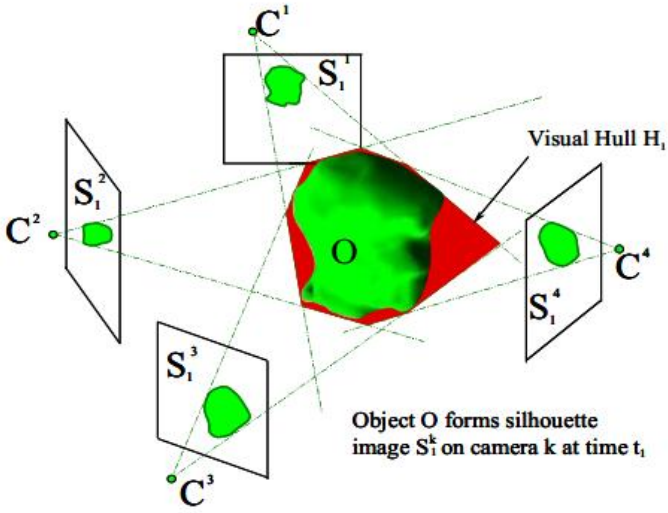

Figure 2.10 an example of classical Shape-From-Silhouette with a head-shaped object and four cameras at time t_1. [47]

2.7.3 Active depth cameras

Active depth cameras are just not an image capturing cameras; they consist of a projector which projects visible or invisible light into a scene. An active depth camera can capture the pattern and depth information of a scene by analyzing projected light that has directed towards the scene. The output is either a depth map of objects in the scene or dense 3D point cloud. However, a point cloud does not consist of points at the backside of the object and only contains points visible for the camera. Even today there are several active depth camera technologies exist. The following paragraph indicates a brief description of each technology.

Structured light is one of the method which can obtain depth images by projecting multiple patterns in rapid succession. Every projected pattern is captured by using a synchronized camera (Figure 2.11). A code map can be produced by combining these patterns and it enables stereo matching algorithms to generate a depth map without using existing image features. A laser scanner able to capture the scene while sweeping one or more laser lines across an object. At every sweep step, deformation or alteration of the laser lines can be used to obtain the depth information of an object. Even though laser scanner results in measurements with more accuracy it is considerably slower with respect to the other active scanning methods. Orderly projecting patterns or laser lines has a main disadvantage which is known as that the motion of the object while scanning can destruct scan completely or may results artifacts

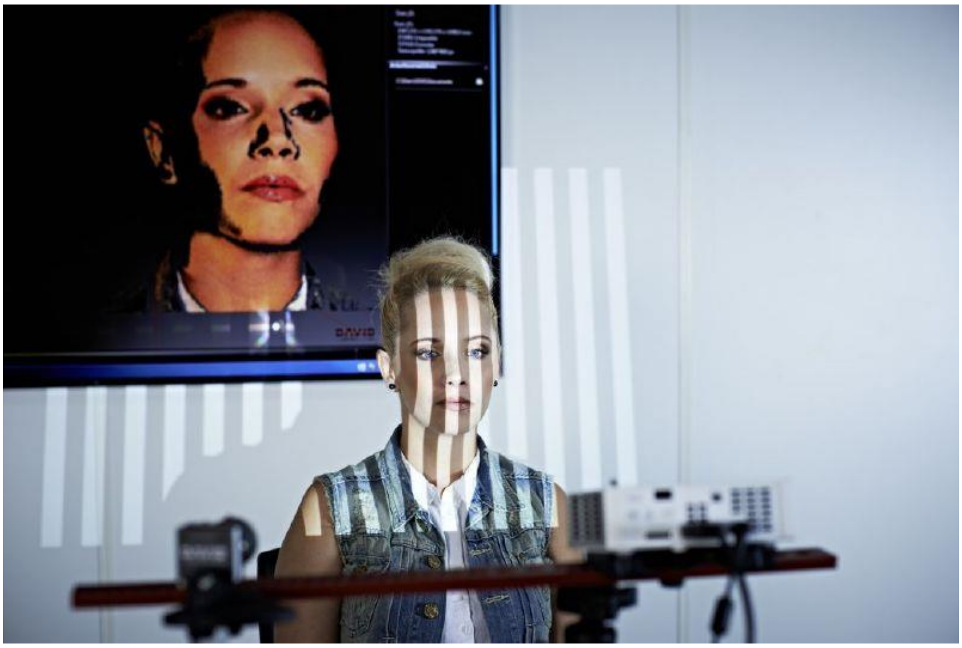

Figure 2.11 Laser light Scanning [49]

Microsoft Kinect is a modern camera which uses a single static infrared projection with a coded dot pattern. These patterns inherit a major advantage which is, in a decoding stage it helps to uniquely identify a neighboring set of dots and this mechanism helps to minimize errors and allows for more efficient depth computations. By using a single pattern, depth data can be extracted from a single camera frame and one of the major advantage of this is the motion of a particular object does not influence the quality of the scan. However, since a single pattern code requires multiple projector pixels, depth image resolution become reduced. Hence, depth image obtained has very reduced resolution and it equals to the half of the resolution of the camera [50].

Figure 2.12 Depth Image captured using Kinect Depth camera.

The third type of camera doesn't base on any pattern based methods instead it uses time measures. In this method, the time taken by the light to travel from its source through the object to the sensor or detector (Figure 2.13). For instance, cameras such as Basler, Senz3D and the MESA SwissRanger 4000 can be indicated. These cameras operate according to the principle called time-of-flight (TOF) principle which is capable of deriving depth information from the time-related phase shift of a modulated light source [51] [52]. According to this principle, it is possible to calculate the depth values of each and every camera pixel. As well as objects can be captured while they are in the motion since all depth values are calculated at the same time using the same modulated light source.

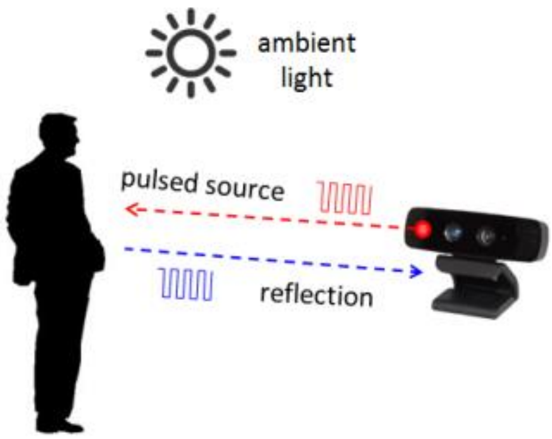

Figure 2.13 Time-of-flight camera operation [55].

In addition to above indicated advantages, these cameras also carry disadvantages with them. Followings are the some of the disadvantages in brief:

- Production of noisy images.
- Highly expensive.
- Reduced resolution of images.

CONSIDERATIONS	STEREO VISION	STRUCTURED-LIGHT	TIME-OF-FLIGHT (TOF)
Software Complexity	High	Medium	Low
Material Cost	Low	High	Medium
Compactness	Low	High	Low
Response Time	Medium	Slow	Fast
Depth Accuracy	Low	High	Medium
Low-Light Performance	Weak	Good	Good
Bright-Light Performance	Good	Weak	Good
Power Consumption	Low	Medium	Scalable
Range	Limited	Scalable	Scalable
APPLICATIONS			
Game		X	X
3D Movies	X		
3D Scanning		X	X
User Interface Control			X
Augmented Reality	X		X

Table 2.1 Comparison of 3D Image Sensing Technologies [55].

Chapter 3

METHODOLOGY

3.1 Methodology

This section will dedicate to explaining design components of VFR. This virtual system processed on inputs which are known to be as user motions and translate them into motion data by the aid of Kinect sensor. All the elements included within the virtual system such as user interface components and dress models along with the motion data that has produced are integrated to create a virtual scene. Consequently, the integrated virtual scene is then displayed on a screen converted by a graphics card. This section shows the system preparation and the software design.

3.2 High level flow of the system

When a user is standing in front of the VFR as in Figure 3.1, then he/she will be able to select desired clothes which are displayed side of the screen using his/her hand gestures. The selected garment is then virtually merged with the user recorded by the camera. Then he or she able to perform different postures, select different cloth items with sizes, take pictures with superimposed items …etc. and can check whether it is fine or not. It is very helpful to get better understand about "does it suit" and "does it fit" and reduce much of the guesswork involved in shopping.

3.3 Prototype system

Microsoft kinect sensor

1. Recording of the depth data
2. Capturing the RGB video stream
3. Capturing Skeleton data stream

Display / Screen

1. outputting the recorded video stream (mirrored)
2. The output is superimposed with the selected garment

Computer

1. Executing algorithms for skeleton tracking
2. Controlling the movement of cloth colliders
3. Combining of video stream and skeleton data (same viewpoint)
4. Computation for cloth physic simulations.
5. Controlling of the UI elements on screen

The Kinect camera captures the user motion and tracking will start when the user is in front of the VFR. Once The Kinect captured the whole figure or body of the human, then Kinect can start the skeletal tracking of the user. During this time customer can follow the menu icons and cloth categories on the screen.

The output is the mirrored image, which has superimposed with the desired clothing. In order to capture a non-interpolated skeleton, the full body tracking distance need be indicated. In the conditions such as when the person is closer than the desired minimum distance, then skeleton points have to be interpreted.

Most importantly, to get a stable skeleton tracking, a simple plane or a wall can be used as the background behind the user. The setup of a VFR mainly consists of 3 components namely, Kinect device, a display as well as a computer. The Kinect device is responsible for interacting with the user, who is in front of it, and the display simply shows the superimposed image with selected clothing.

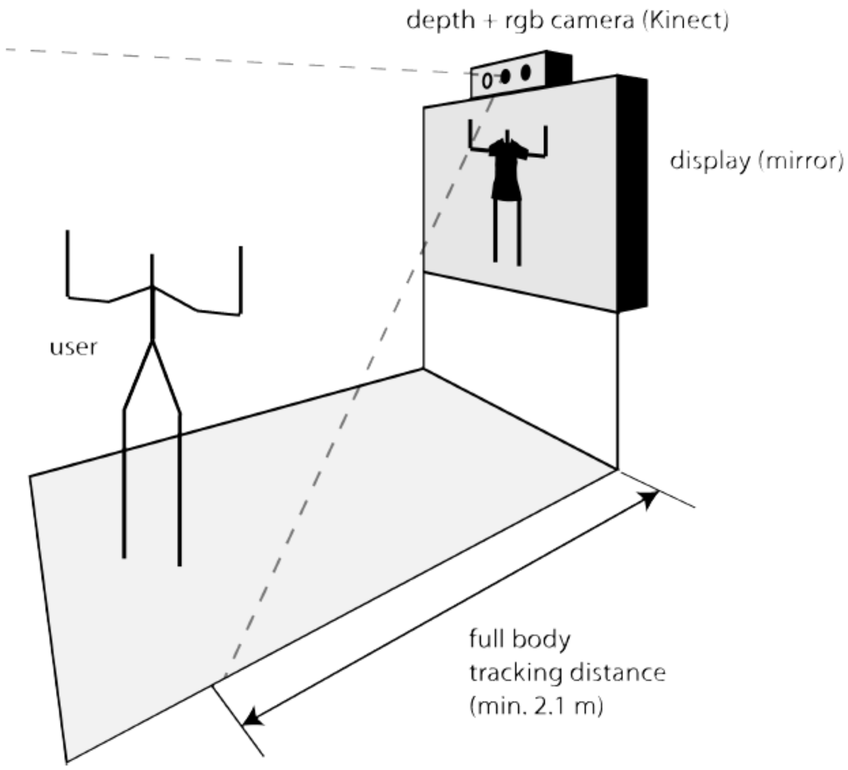

Figure 3.1: An intended basic setup of the fitting room.

3.4 Main flow of the system

By using the data captured by Kinect sensor, first, a depth model of human will be created according to the dimensions of the user. The user's moves will be reflected by the skeleton model since human model always follow the motion of the skeleton model captured by the Kinect.

As the next step, pre-determined cloth-model will be represented, and through this cloth-model, can build various types of clothes for dressing on. Cloth fitting is done through the interaction between the cloth model and the human model.

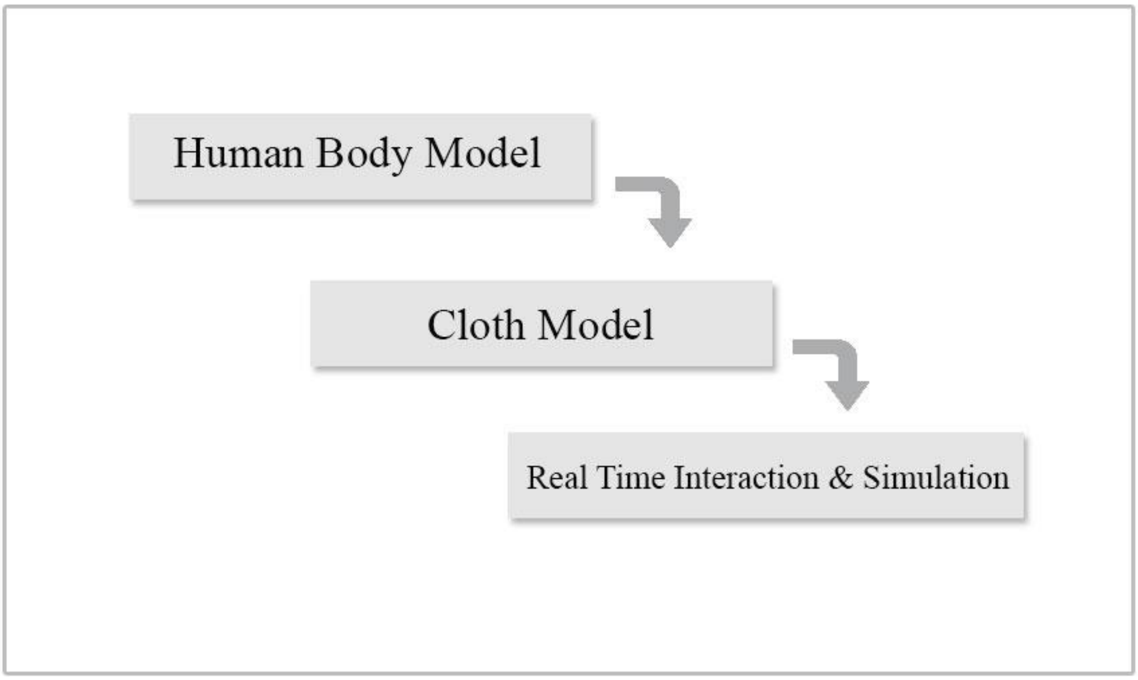

Figure 3.2: Approach for fitting room

3.5 Design patterns and concepts

MVVM [39] [Figure 3.3] design pattern is a widely used pattern in WPF applications. It is a derivation from the MVC pattern. It separates an application into three main modules called, Model, View, and the ViewModel. Simply in this pattern View communicates with components in ViewModel by a mechanism called Binding. View produces some methods in ViewModel using the Prism library interface ICommand. All components in model communicate with the ViewModel using the mechanism of events. However, there is no direct access to its resources. But ViewModel can communicate with a model in the same way but has access to its resources. Another main advantage of using MVVM is, it provides separation of components (loosely coupled), especially between the view and model. So it is more flexible to do application development and modification

Figure 3.3: MVVM pattern

Chapter 4

SYSTEM ARCHITECTURE

4.1 System architecture

As per the descriptive diagram in Figure 4.2, that shows the architecture or system components of the VFR. Generally, the basic model consists of a client application, Kinect SDK and Kinect device. In addition to aforementioned basic components Prism and Nu get libraries also can be found within the basic model. For the implementation purpose, a modern structure of the MVC design pattern has selected. Which is called MVVM (Model-View-View Model).The main goals are to provide a clear separation between presentation and domain logic layer, and gain efficiency. For example, if you want to do a change to a component in a view you can do it easily without affecting to its model and vice-versa.

The Kinect Xbox 360 sensor consists of two cameras, IR sensor, motor and an array of microphones. Data retrieving from the Kinect can be done easily through the Kinect SDK, which has provided by Microsoft Corporation. For example, sending data to the connected device, calibration of the camera, data de-noising...etc. However, to develop a VFR application, we only need the cameras and IR sensor. The RGB camera is using to collect RGB video stream and, IR camera and IR sensor are using to collect depth video stream. By using the NUI API can track the whole skeleton of the user with joints. The coordinates of each joint are responsible for doing the main manipulation of the video and depth streams.

Figure 4.1: Microsoft Kinect

A graphical user interface has provided to the user to in order to interact with the VFR using simple hand gestures. The application is controlled by user behaviors, recognized and supported by internal system components.

The component, Prism library plays the mediating role between the views of the application and its logical layer. As well as, it is a part of the middleware between these two components. The objectives of the project are, Implementation of the client application, the detailed architecture and components are represented in below.

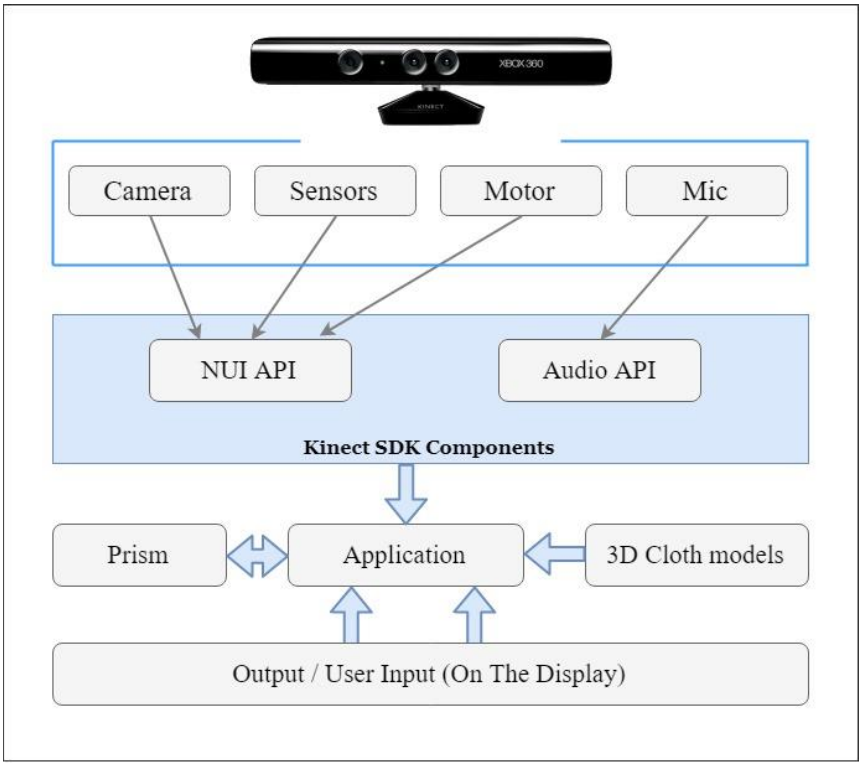

Figure 4.2: High-level Architecture Diagram

4.1.1 The functional block diagram of virtual fitting room application.

The functional block diagram in Figure 4.3 represents an overview of the virtual try-on system. Principally, through the Kinect SDK can access 3 different streams called, Skeleton data stream, Depth data stream and Color data stream. Depth stream is used to extract the skeleton of the user by using frame by frame. NUI API provides information about the location of the user who is standing in front of the Kinect sensor, with detailed position and orientation information after capturing the user. The information has provided by the NUI API is then transferred into application code as a set of 20 points. Most probably skeleton positions. Since skeleton represents a user's current position and pose, skeleton data can be used to get measurements of the different dimension of users' body parts. Skeleton data are filtered using the aforementioned image retrieval method calling a frame retrieval method and passing a buffer while VFR application can then use an event model by hooking an event to an event handler in order to capture the frame when a new frame of skeleton data is ready.

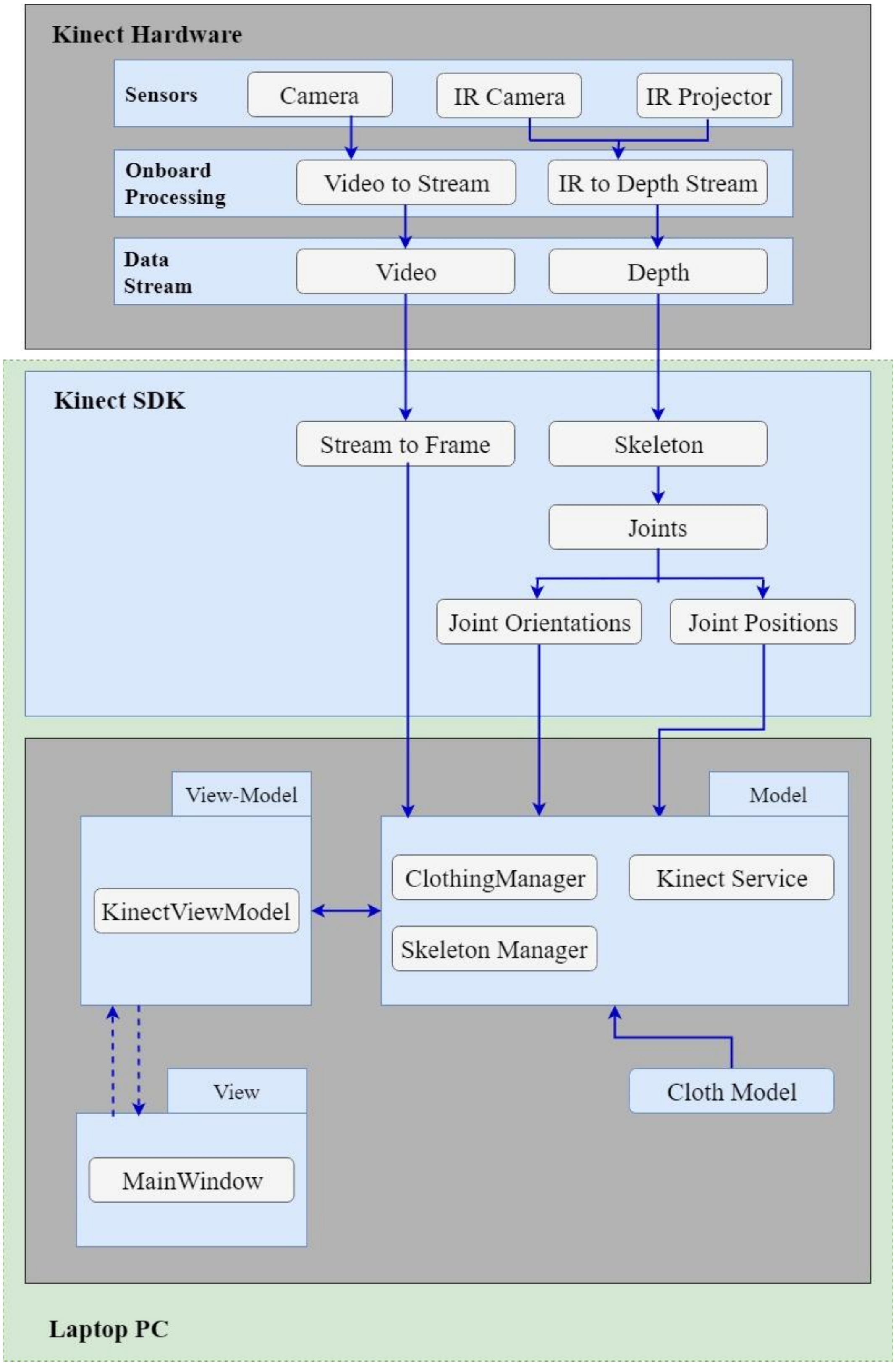

Figure 4.3: Detailed Architecture Diagram

4.1.2 Solution implementation

The main application window is the main component of the view module and It consists of all the user controls, the RGB video stream from the Kinect camera and all the listed garment items. Except above details, it displays auxiliary elements such as the skeleton data of the user and the relevant event notifications. Moreover, auxiliary elements very are useful during implementation.

The main class of the VFR application can be found in View-model and it is known as the "KinectView-Model". It is responsible for mediating between the main window and classes which are containing application data. Mediation among the other components are managed by the helper classes that are involving in view-model. The model consists of classes with data and the entire group of managers which are considered to be managing those data and pass them to the View-Model.

- **SkeletonManager** - It is a helper class. It is responsible for the service's backbone which Functioning to debug the application.
- **KinectService** – Check availability of the Kinect device and handles streams read from the Kinect device.
- **ClothingManager** – Capable of managing clothing collections such as cloth initialization, fitting selected cloth models, scaling, rotation handing and responds to user gestures.

4.2 Class structure

Represent the class hierarchy of the Virtual try on system. An overview of the class structure including their names, their casts and integrated connection with each other has considered in this subtopic.

4.2.1 View layer

- **MainWindow** - It is the main class that implements the View module. It keeps and frequently updates all the visual elements. In addition displays the image of Kinect, cursor, buttons and other visual elements visible on the screen.
- **HandCursorEventArgs** - The class which contains arguments passed during the Event associated with the cursor. As well as it has the ability to transfer the coordinates of the cursor.
- **KinectButton** - Which represents the button based events from the

Kinect. Such as, implementing the events responsible for the actions hovers on the button, defining the time until the button is considered pressed, moving the cursor on to the button, leaving the button for the cursor and click the button can be listed.

- **KinectInput** - Known as cache register and removes events related to the movement of the cursor. Capable of supporting events such as input controls the cursor in the area, the cursor movement in the area of control, leaving the cursor controls and gesture clicks.

- **ButtonsManager** - Responsible for the cursor. Follow the Singleton design pattern. Global access is provided to the created object. This created object consists of a method which separates events related to the cursor.

- **PropertiesObserver** - Properties which cannot be handled by the steps of Binding is updated using a contributive class. Also responsible for the other properties such as the height or width of the control.

4.2.2 View-Model Layer

- **View-ModelBase -** Known as the base class for the view-model module. Which implements the INotifyPropertyChanged interface, making the presentation layer (view) informed that the model has changed.

- **HandTracking -** Represents the cursor, it has the information about the positions of the left and right hand. The class is responsible for updating the position of hands on the screen and also implemented with the method that allows tracking of parts of the skeleton.

- **KinectViewModel -** Initializes an object of class KinectService and facilitates the way to access to a collection of clothing and clothing categories (hats, bags, tie…etc.). As well as it creates all objects with the categories of clothing buttons that will appear on the screen.

- **KinectViewModelLoader** - Responsible for initialization of KinectViewModel property.

- **ButtonViewModelBase -** provides access to the image displayed on the button.

- **ClothingButtonViewModel -** Class is representing the buttons of clothes in this category. Responsible for providing the path to the wearing type the clothes, the size of which is the proportion of the width of an image to a specific width of the pitch. Works on a method that keeps the connection or interaction between the view and the view-model. The method can be implemented after a single click and it updates a list of all the desired pieces of clothing and displays the model with the clothing on the screen.

- **ClothingCategoryButtonViewModel-** Included with various categories of clothing buttons and also consist of a list of clothes models in the currently selected user categories. Implements a method that enables interaction between the view and the view-model. After the click event, the method updates the on-screen buttons with models of clothes, for those that are associated with the currently selected by the user category.

- **TopMenuButtonViewModel-** Consists of buttons located in the top menu of the application. As well as provides access to information about the functionality that will ensure a button and calculate all possible functions. Involves with a specific method capable of updating the list displayed on-screen buttons and carrying out specific functions, such as manually resize the last selected garments by the user, distinguishes between category for men and women, superimposing user photo in the whole outfit, as well as enable or disable sounds and exit the current program.

- **TopMenuManager –** This class is representing the manager of the top menu buttons. Follow the Singleton design pattern like the "buttonsManager"**,** which means for the entire project there is only one instance of the class. The class gives global access to the created object.

4.2.3 Model Layer

- **IKinectService**- The interface consists of methods that are implemented by the Kinect Service.

- **KinectService -** This is an interface class which works on the service. Initializes an instance of the sensor in the program, also supports the events of Kinect. An instance of the class is responsible for collecting, updating and processing of data collected from Kinect. Capable of finding the frame closest to the camera and mapping the frame to the accurate points of the space.

- **SkeletonManager -** Known to be class of the manager of parts of the skeleton. Provides access to, interconnect all parts of the body. Implements methods that form parts of the body such as arms, legs, shoulders.

- **ClothingItemBase -** This is known to be as an abstract class representing a single piece of clothing. Held in the model of the clothes. Provide access to the height and width of the model and its position at the top and left side of the screen. Also, this class contains the enumeration of all the available types of clothes. It has data about the category of cloth and the relative size corresponding to it, such as the ratio of the width of an image to a specific width of the pitch. Consists of a method for updating the position of the clothes on the screen is expressed an abstract method which capable of tracking of parts of the skeleton.

- **HatItem -** This class consists of an implemented method which tracks the position of the head and shoulders. The class inherits “Clothing ItemBase” and represents the object hat.

- **SkirtItem -** It represents objects skirts and, tracks the position of the head, feet and spine.

- **GlassesItem -** The class inherits ClothingItemBase, which represents objects glasses. As same as the “**HatItem**” this class also consists of an implemented method which tracks the position of the head and shoulders.

- **ClothingManager -** This class is known to be the manager of a fashion collection. Follow singleton design pattern.Global access has provided to the created object. Generally, this class includes two collections of clothes. The first clothing stores the selected clothing category while second consists of some clothes, which are ready to be superimposed to the skeleton.

- **ModelBase -** Capable of representing the clothing in a 3D manner. Provides an implemented method of updating the position of the models on the screen.

4.3 Selection of technologies

4.3.1 Writable Bitmap vs. BitmapSource

In order to create a real-time video stream in the VFR "Writable Bitmap" has been used over "BitmapSource" due to considerable efficiency gaining. Writable Bitmap has gained more efficiency, due to its reusability of frames, whereas "BitmapSource" always creates new 30fps per every second.

4.3.2 Depth data smoothing in real time

One of the considerable drawback in the Kinect Version 1 is the poor quality of rendered Depth Frame images. But when it comes to the Kinect version 2 this is not an issue, due to its new hardware and software improvements. Therefore compared to Kinect version 1 it has much more sharper edges and less noise in its depth frames. But in Kinect version 1 there is a lot of noise in Depth Frames, with missing bits, which are indicated by value 0 and a pretty serious flickering issue.

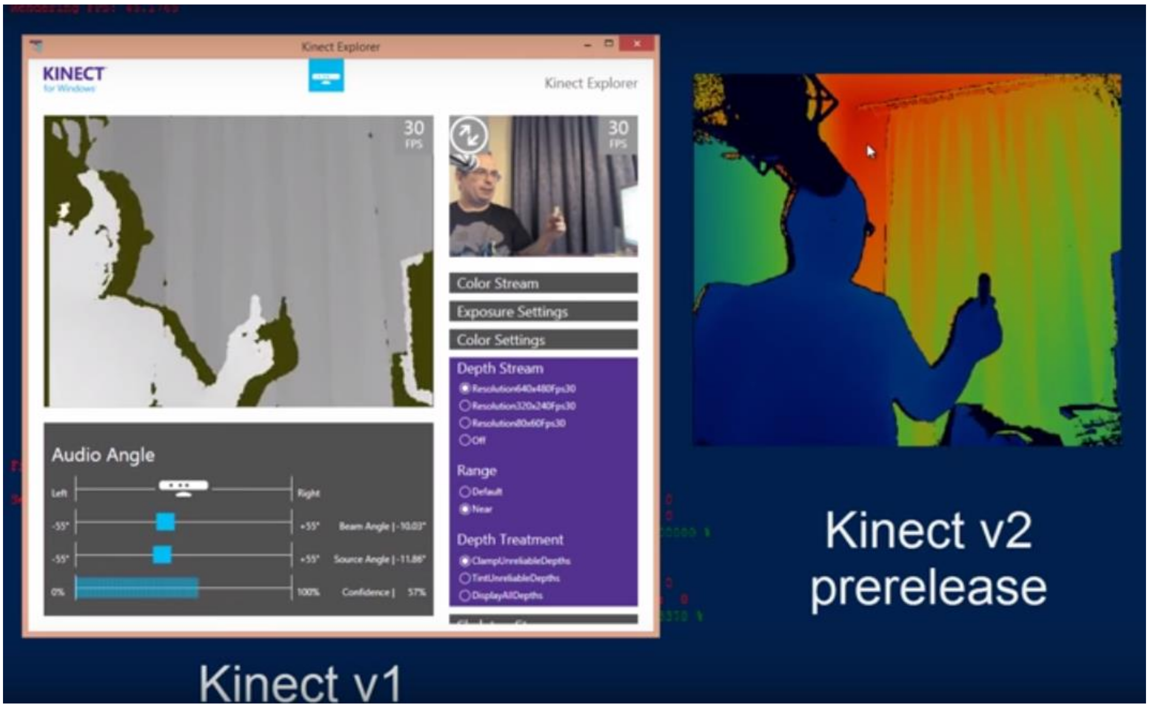

Figure 4.4: Kinect v1 VS Kinect v2

In order to prevent these issues in Kinect version 1, a combination of following 2 techniques has used, called Zero pixel filtering and weighted moving average. The first step of Depth Data Smoothing is called the Zero pixel filtering process, which transforms the depth data from the Kinect into something that is with less unwanted Zero Bits. In This method, first, creates an array of depth values from each new frame that has sent by the Kinect sensor. Once an array of depth values is ready, then it is ready to be filtered. In order to do that, have to scan through the whole array to looking for Zero values. Zero means that the Kinect couldn't process properly e.g. reflections from objects. So need to remove as many of these as realistically without affecting the performance.

When there is a Zero value in the array, then it is an eligible pixel for filtering. The filter basically has two bands around the selected candidate pixel. Each of which band consisted of neighboring pixels of the selected pixel. These bands are used to search for zero pixel values and identifies how many Zero pixels were found in each band. Then those will compare with the threshold value of each band and determine whether it should be filtered or not. If the threshold for either band has broken, then the mode of all the non-Zero depth values of the band will be applied to the candidate, otherwise left it as it is (Figure 4.5).

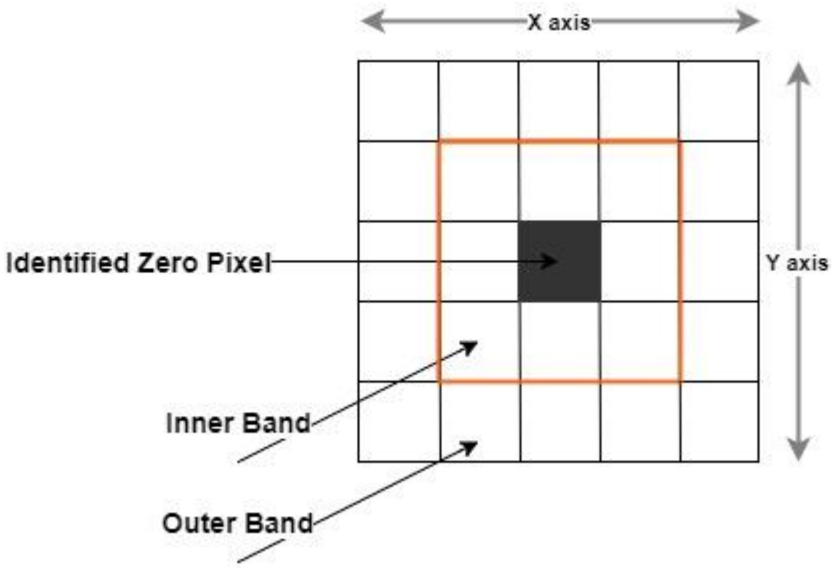

Figure 4.5: Selected Pixel with its neighboring pixels

4.3.3 Model positioning and rotation

The skeletal tracker able to returns the 3D coordinates of 20 body joints as x y pixel coordinates and meters for the depth. Due to the frame based recognition approach, flickers and vibrations on the joints can occur. This issue can solve up to some extent by adjusting the smoothing parameters of the skeleton engine of the Kinect SDK. The rotation angles of the model usually can be defined as the angles of the main axes which reside within the body and arms.

```
D = (Math.Atan(((double)jointLeftPosition.Depth - jointRightPosition.Depth)
/ ((double)jointRightPosition.X - jointLeftPosition.X)) * 180.0 / Math.PI)
```

In here the main body axis denotes the line between the shoulder center and the hip center, as well as axes for arms, are defined as the lines between the relative shoulders and elbows and a tan2 is a function equal to the inverse tangent but only defined in the interval [0; pi). Center of rotation is set to the middle of the corresponding line of the axis and it is defined as an anchor point for each model part.

4.3.4 Model scaling

The most basic approach for model scaling in VFR is, use of Skelton joints as a scaling factor. For example use shoulder center and hip center to calculate the upper body height. However, calculations only with the joints are not sufficient enough to accurately scale the model as user moves. So In order to perform more accurate model (cloths, glasses, hat...etc.) scaling in VFR, had to use an approach, that uses both depth and joints data as scaling factors.

Depth based scaling = M / Z

Where M denotes the model that has vector of default width and height when user is in front of the sensor and Z denotes the distance between the spine and the sensor.

Joints based calculations used in VFR.

- Calculate length of an arm - Length of an arm is calculated using Euclidean distance algorithm with the help of corresponding shoulder and the elbow.
  - Arm length = $\text{sqrt}\ (((X_{\text{shoulder}} - X_{\text{elbow}})^2 + (Y_{\text{shoulder}} - Y_{\text{elbow}})^2))$
- Calculation of width and height of the upper body - The width and height of the upper body are also calculated using the distance between the shoulders and the distance between the shoulder center and the hip center

correspondingly as follow

- Body width = $X_{\text{left shoulder}} - X_{\text{right Shoulder}}$
- Body Height = $X_{\text{shoulder center}} - X_{\text{hip center}}$

4.4 Appearance

The application consists of the main window with all the cloth models and other required menus [Figure 4.4.1]. From the menu on the left side of the window, you can select different categories of clothing such as hats, shirts, ties, eye glasses...etc. At the top of the screen, you have presented set of icons, which provide access to basic functions such as resizing clothes, taking pictures of dresses, removing of the whole set of clothes and the exit from the entire program. All the menus are accessible through hand gestures and pressing events are handled by holding the cursor over it for a certain time, accompanied by the sound of confirming click..

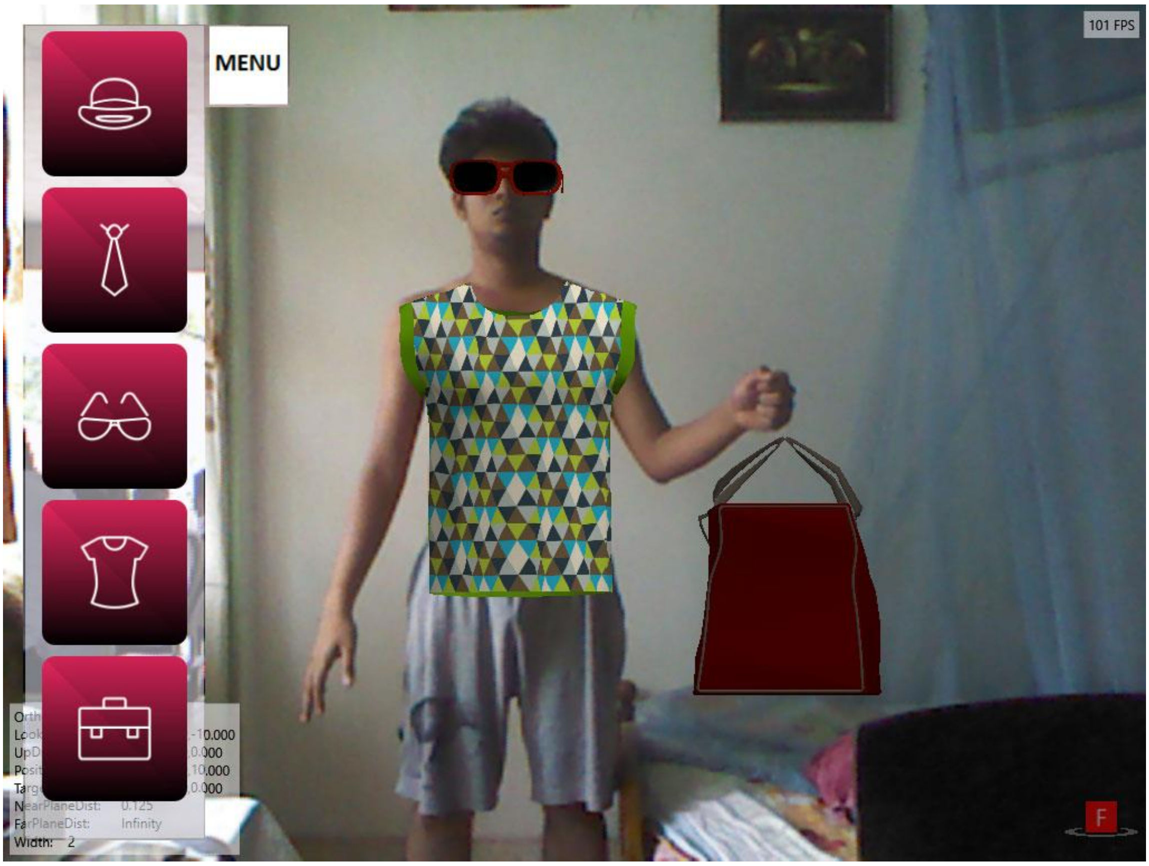

Figure 4.6: VFS interface

Chapter 5

EVALUATION

5.1 Evaluation

Evaluation is the only way to prove the willingness, compatibility and the affinity of a solution for the desired problem. Thus, the evaluation does a significant part by involving in a research project. In This research, Evaluation is broken down into two sections namely, the interface evaluation and the functionality evaluation. In interface evaluation section it will be focused on the user interface of the VFR, while functionality evaluation will focus on performance, features, user-friendliness...etc.

In order to conduct a successful system evaluation, seven evaluators with different age gaps (3 females and 4 males) were invited and asked them to try out the fully functional system for a considerable time.

5.2 Hardware specification

- Evaluation of the VFR was conducted on a personal computer with following specification.
- Processor: Intel(R) Core(TM) i5-2450M CPU @ 2.50GHz
- Random Access Memory: 8.00 GB
- Operating System: Windows 10, 64-bit Operating System
- Video Graphic Card : nvidia GeForce 525m.

5.3 Evaluation process steps and result

5.3.1 Before system evaluation questionnaire (Appendix A)

In this section, a questionnaire is used to evaluate the shopping habits of each individual who are participating in the evaluation process. The survey results are listed as follows.

1) 2 out of 7, go shopping as necessary and the rest go shopping less than once per month.
2) Every one shopping on weekends or public holidays.
3) 3 out of 7, spend half a day for shopping activities, another 2 of them spend as necessary and the rest spending 1-3 hours for their shopping activities.

4) 5 out of 7, try less than 5 clothes on a shopping trip and the rest try as necessary.

5) Out of 7, only 3 people know about virtual try-on systems and the rest never heard about it. No one has ever tried out a virtual try-on system.

6) According to participants, the most time-consuming activities are the queue to cashier, queue to fitting rooms, and searching for cloths.

5.3.2 Interface evaluation and results

Interface evaluation has conducted in order to provide the best user experience for the virtual try-on system. In this process of evaluation, selected people (previous seven evaluators) who are capable of evaluating the system interface will be invited and allow them to try out the VFR. Each individual, who is participating in evaluation process should answer to specific questions (Appendix B) after utilizing the VFR. The questions were about, screen management and some other general questions. General questions were based on the number of components used and the number of garments tried. The evaluation results are shown as follows.

1) 2 out of 7 were not happy with the screen design due to lack of color combinations and used images for buttons.

2) 6 out of 7 were clear about the system interface and think it is very easy to use and one person had some difficulties with navigation with hand gestures.

3) 2 out of 7 pointed out that the interface needs to be improved if it intended to introduce as a commercial product

4) 3 out of 7 were asked, is it possible to press menu buttons by clicking on it. Because it will easy to interact with the system rather than using hand gestures.

5.3.3 Functionality evaluation and results

The major objective of the Functionality evaluation is to estimate the usefulness and the performances of the Virtual try on system. In order to do the Functionality evaluation, both technical and non-technical parties (previous seven evaluators) were invited to try out the complete functional system for at least 5 minutes and to complete the following task list to evaluate the system.

1) Try to move the hand cursor pointer using hand gestures.
2) Go through the system menu options and identify the available features in the system.
3) Try out different garments such as shirts, eye glasses, bags…etc.
4) Try to resize, move up, move down and clear cloths features.
5) Take photos with virtual dresses.

Once all the participants completed their system walkthrough, they were asked to fill another questionnaire (Appendix C) to evaluate the functionality of the system in different aspects such as satisfaction, usability, performance…etc.

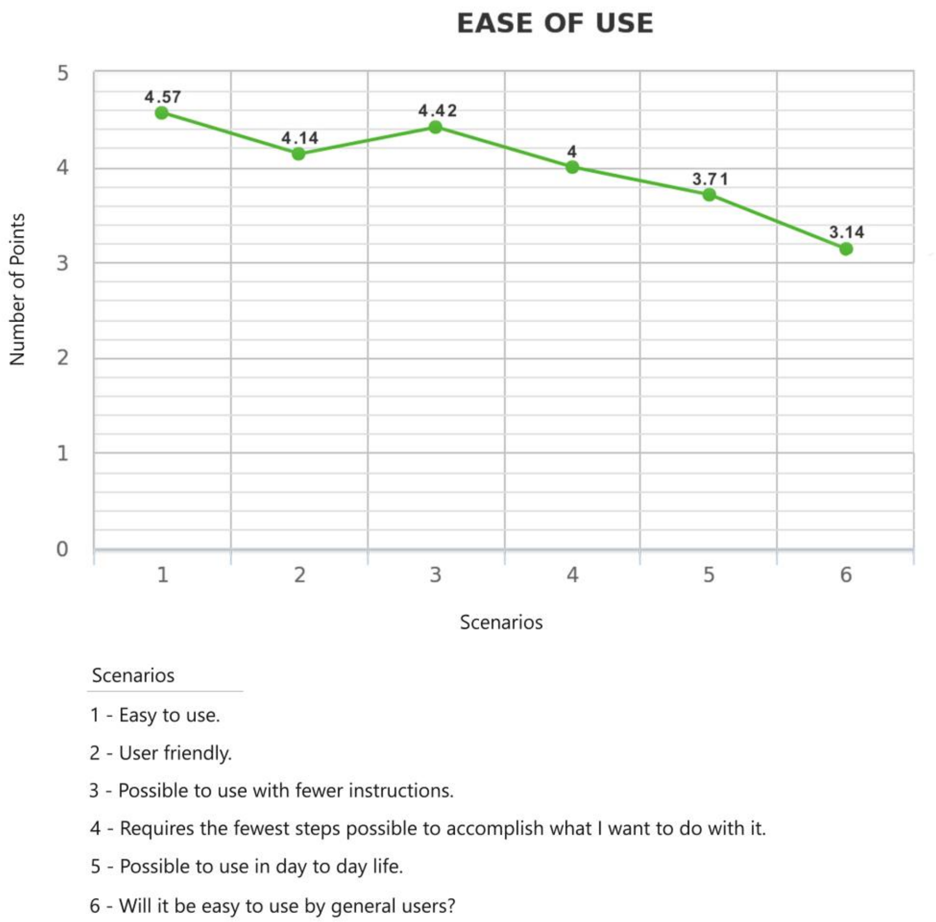

Figure 5.1 Ease of use

In Figure 5.1, except scenario 5 and 6, all other 4 scenarios show considerably good averages within the boundary of point 4 and 5.which indicates the system is relatively easy to use according to the seven evaluators. According to the question 5, it asks is it possible to use the Virtual fitting room in the day to day life. For this question 2 out of 7, have a neutral opinion and one has a disagreed opinion. Because that person thinks this system needs further improvements.eg. Suggested, use touch try method instead of using hand gestures. According to the question 6, it asks about will it be easy to use by the general users. For this question 4 out of 7, has 2 neutral opinions and 2 disagreed opinions. This because they think people who don't know how to interact with Digital devices may experience various difficulties in their usual day to day shopping activities.

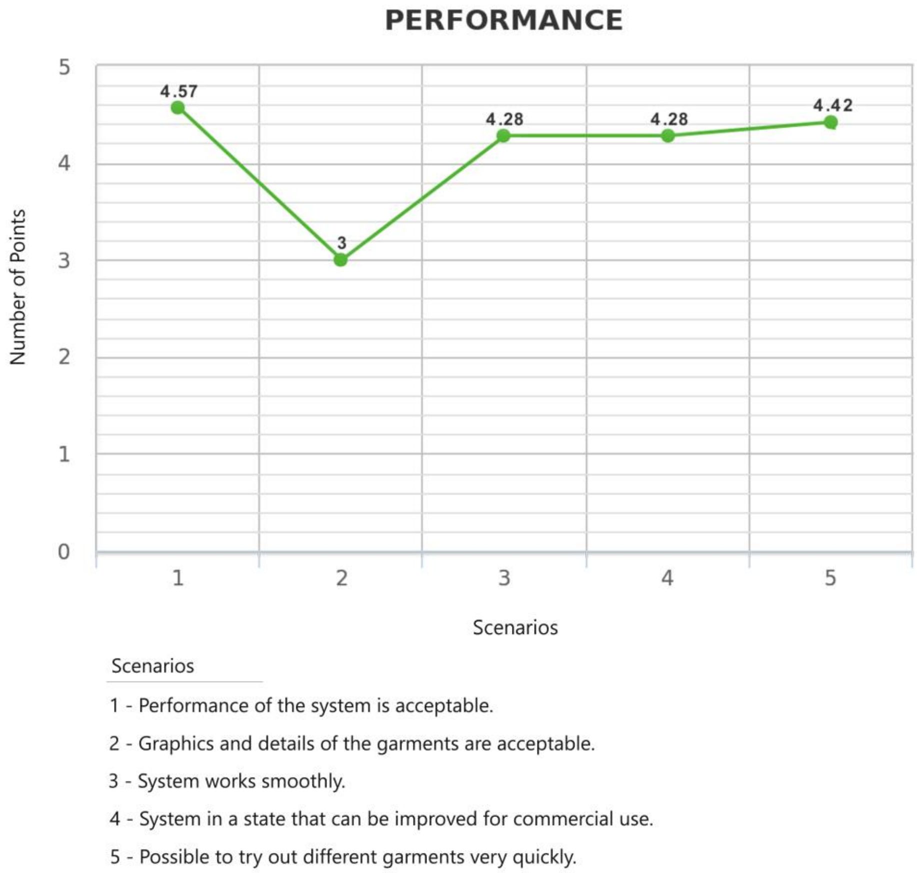

Figure 5.2 Performance

In Figure 5.2, except scenario 2, all 4 other scenarios shows considerably good averages within the boundary of point 4 and 5. That indicates the system is relatively better in the aspect of performance. According to the question 2, it shows lower point 3 due to the system's primitive 3D graphics and lack of details in the clothes, and this will consider in future work of the system.

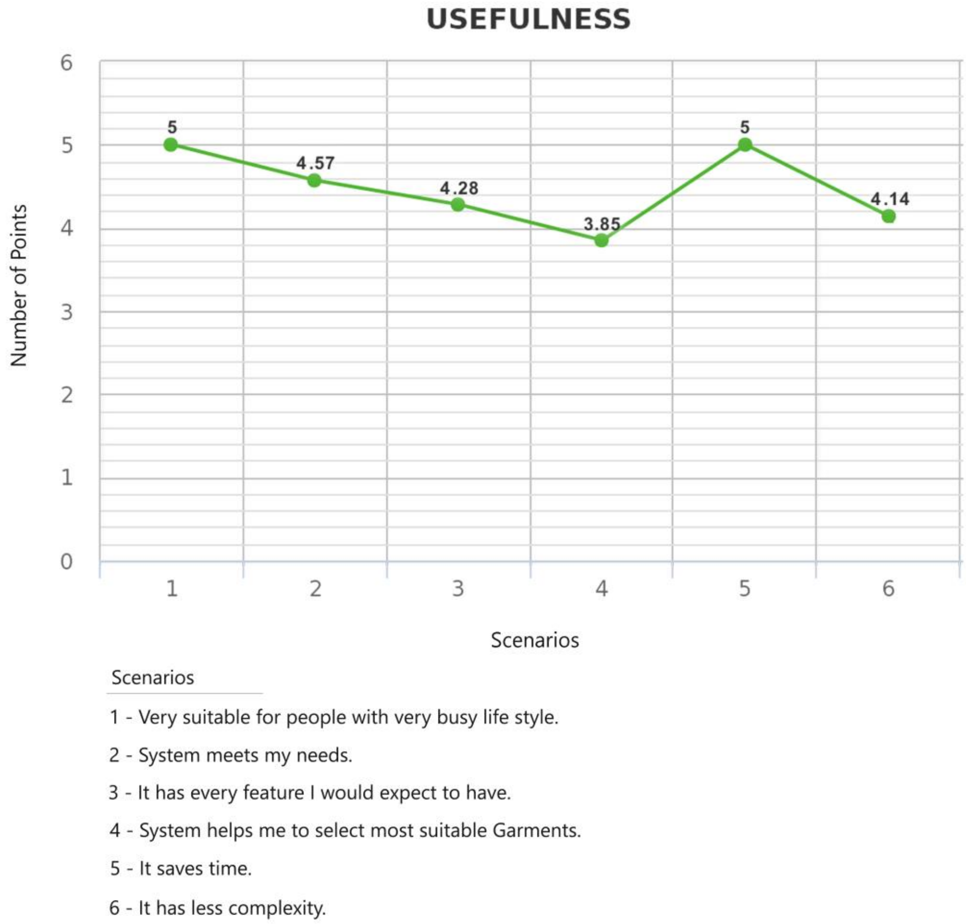

Figure 5.3 Usefulness

According to the Figure 5.3, it shows good average scores around and above point 4. That means that the virtual fitting room has full filled its usefulness requirement. Only scenario 4 has slightly less point below 4, because some users asked "how can I measure the size of the cloth that I want buy? Like whether is it small, medium, large…etc."

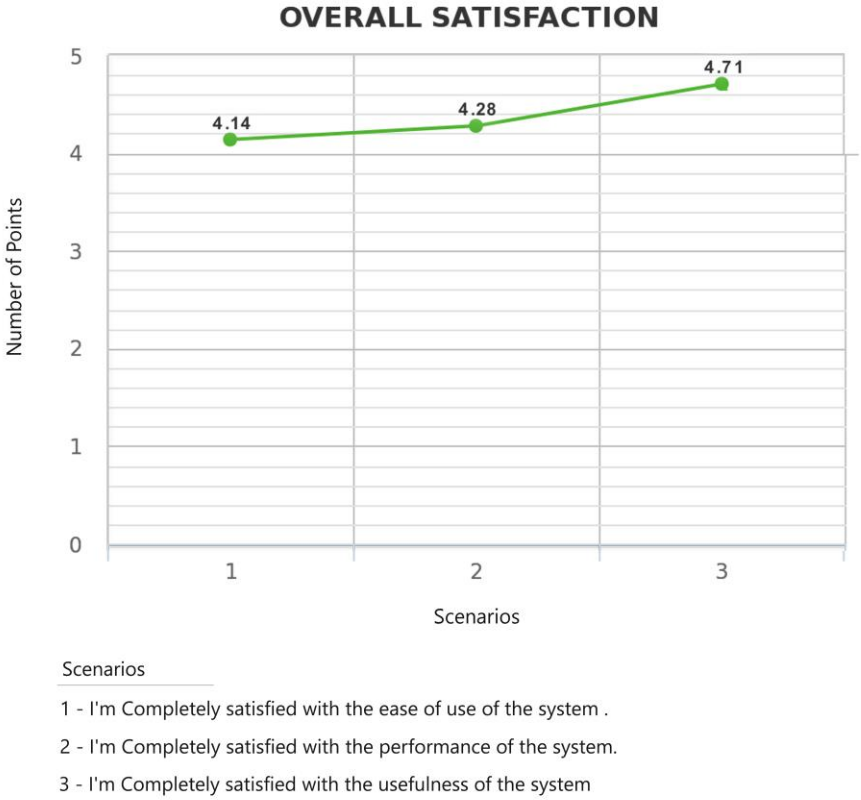

Figure 5.4 Overall Satisfaction

The overall system evaluation has depicted in Figure 5.4, and the results indicate all the evaluators were quite satisfied with the current Virtual Fitting Room system. Hence, the project objectives have successfully achieved within a limited boundary.

.

Chapter 6

CONCLUSION

6.1 Conclusion

This research has mainly forced on how to address traditional weak points, such as inability to try on without physically wearing, long waiting queues in front of fitting rooms, high return rates of garments...etc. The main objective of this research is to create a VFR with fast rendering that is capable of showing the physical appearance and the behavior of garments more realistically. Existing virtual try-on system researches can be based on either one of the following six variants namely,

1. Real Person / Kinect / Superimpose clothing
2. Real Person / Web Cam / Superimpose clothing
3. Avatar / Kinect / 3D Clothing
4. Avatar / Input Body Measures / 3D Clothing
5. Predefined Avatar / 3D Clothing
6. Photo / Superimpose Clothing

Most of the existing virtual try-on systems are based on non-customized pre-generated 3D avatars (trimirror [12], glamstorm [13]), 2D based Garments (Zugara [11]) or even photographs. Therefore, Most of the customers can't feel the realistic experience of a virtual try on-system.

In this research, planned to use the Microsoft Kinect sensor to create a VFR from the camera images in real time and planned improved rendering using different algorithms in order to provide more fast and realistic appearance. Compared to most of the virtual try-on system, specialty of the presented virtual fitting room is, it allows customers to try out multiple items at the same time (shirt, hat, eyeglasses, and bags) with a good performance and also provide the options to resize and reposition the superimposed garments as they wished. Another most important advantage of the presented system is, it does not require a complex setup. Instead, all you need a computer with a screen and a Kinect.

Evaluation of the presented VFR was conducted via seven independent evaluators. The Evaluation is divided into two main sections called Interface evaluation and functionality evaluation. Both of these sections had good evaluation results such as, for the overall system evaluation, it was rated to 87.53%. Thereby, the presented VFR has full filled its intended purpose to its targeted customers.

The identified few main limitations during the evaluation process were low quality of graphics, low quality of the camera can be listed, and hope to address these limitations in futures works of the research, in order to provide more promising and satisfactory experience to its target customers.

6.2 Limitations

Regarding limitations, while the process of development and evaluation, VFR had to be limited to several boundaries due to its massiveness with the given time constraint. Because it is an ongoing research area with huge trails and errors. For example, even superimposing a cloth on to a user can be considered as a research area. During this project implementation and evaluation had to face the following limitations.

Development limitations,

- Clothing without animations and less details.
- Unable to fully wrap the cloths on to user, instead currently showing cloths just in front of the user.
- Low resolution RGB camera.

Evaluation limitation,

- Limited to 7 evaluators.
- Could not evaluate with commercial products. Due to several parameters. E.g. for some systems you have to pay a subscription fee to use it, some systems uses cloud based approaches with high end servers…etc.

6.3 Future work

Regarding future work, the VFR can be enhanced in some points that were identified during the evaluation process. One of the main concerns was lack of graphics and details in garments. But it can be improved using good 3D modelled clothes. In order to do that planned to use nvidia's APEX Clothing [40] or Verlet algorithm [41] based on the success rate. The APEX tool can easily use to re-create the required cloth models for the VFR, because this tool allows creating quick and dynamic clothing with following features.

- Full control over clothing settings and behavior
- Support self-collision effects (bone collision and A-collision)
- Ease of scalability
- Ability to control Level of details in cloths
- Provides support for single-layered and thick clothing

The next main concern is how to fully wrap a cloth model on to a user's body. Fitting 3D clothing model on to the user's body is an important research topic in the virtual try on systems. In the process of 3D cloth model fitting, researchers still find varies difficulties while superimposing clothing on to the user. Because the physical garment modeling is considerably complicated and time consuming task. In order to rectify this, some of the stable solutions like interactive operation for the mesh models (e.g. point cloud) [43] can be used. But still, too many interactive selections from the point cloud data affect the efficiency [42] of the clothing fitting while too fewer selections cause the accuracy of the fitting. So this area will be considered in future works.

Another main concern is the low-resolution video stream of Kinect's RGB camera. Which is max to 1024x768 pixels. In order avoid this drawback there were few researchers [25] conducted with external cameras to capture the video stream. But the problem is coming with calibration. Because have to map depth data from the Kinect with the video stream from the external camera. But now it has more simplified with the introduction of Kinect sensor V2. Because it has resolutions up to 1920×1080 pixel .Thereby problem of the low-resolution video stream will not be a much problem in future developments.

References

[1] "OpenNI 2 Downloads and Documentation | The Structure Sensor", *Structure.io*, 2017. [Online]. Available: http://structure.io/openni.

[2] "PrimeSense", *En.wikipedia.org*, 2017. [Online]. Available: https://en.wikipedia.org/wiki/PrimeSense.

[3] "Willow Garage", *Willowgarage.com*, 2017. [Online]. Available: http://www.willowgarage.com/.

[4] "Cite a Website - Cite This For Me", *Sidekick.co.il*, 2017. [Online]. Available: http://www.sidekick.co.il/.

[5] "OpenNI", *En.wikipedia.org*, 2017. [Online]. Available: https://en.wikipedia.org/wiki/OpenNI.

[6] "Bm.com", *Bodymetrics.com*, 2017. [Online]. Available: http://www.bodymetrics.com/.

[7] "3D Virtual fitting dressing room / mirror", *Fitnect.hu*, 2017. [Online]. Available: http://www.fitnect.hu/.

[8] "Imagine That Technologies", *Imaginethattechnologies.com*, 2017. [Online]. Available: http://imaginethattechnologies.com/.

[9] "Virtual Try on - Augmented reality solutions | TryLive by Total Immersion", *Trylive.com*, 2017. [Online]. Available: http://www.trylive.com/.

[10]"TUKATECH", *TUKATECH*, 2017. [Online]. Available: http://www.tukatech.com/.

[11]"Augmented Reality and Natural User Interface Experiences", *Zugara*, 2017. [Online]. Available: http://zugara.com/.

[12]"- TriMirror Virtual Fitting Room", *Trimirror.com*, 2017. [Online]. Available: http://www.trimirror.com/en/.

[13]2017. [Online]. Available: http://glamstorm.com/.

[14] F. Isikdogan and G. Kara,"A Real Time Virtual Dressing Room Application using Kinect",2012

[15] "YCbCr", *En.wikipedia.org*, 2017. [Online]. Available: http://en.wikipedia.org/wiki/YCbCr.

[16] S. Saha, P. Ganguly, and S. Chaudhury, "Vision Based Human Pose Estimation for Virtual Cloth Fitting," *Proceedings of the 2014 Indian Conference on Computer Vision Graphics and Image Processing - ICVGIP '14*, 2014.

[17] R. Li, K. Zou, X. Xu, Y. Li, and Z. Li, “Research of Interactive 3D Virtual Fitting Room on Web Environment,” *2011 Fourth International Symposium on Computational Intelligence and Design*, 2011.

[18] M. Yuan, I. R. Khan, F. Farbiz, S. Yao, A. Niswar, and M.-H. Foo, “A Mixed Reality Virtual Clothes Try-On System,” *IEEE Transactions on Multimedia,* vol. 15, no. 8, pp. 1958–1968, 2013.

[19] S. Milborrow and F. Nicolls, “Locating facial features with an extended active shape model,” *in Proc. Eur. Conf. Comput. Vis.*, 2008,pp. 504–513.

[20] S. Hauswiesner, M. Straka, and G. Reitmayr, “Virtual Try-On through Image-Based Rendering*,” IEEE Transactions on Visualization and Computer Graphics*, vol. 19, no. 9, pp. 1552–1565, 2013.

[21] T. Capin and I. Pandzic, “A dead-reckoning algorithm for virtual human figures,” *Proceedings of IEEE 1997 Annual International Symposium on Virtual Reality*.

[22] I. Pachoulakis, “Augmented Reality Platforms for Virtual Fitting Rooms,” *The International journal of Multimedia & Its Applications*, vol. 4, no. 4, pp. 35–46, 2012.

[23] U. Gültepe and U. Güdükbay, “Real-time virtual fitting with body measurement and motion smoothing,” *Computers & Graphics*, vol. 43, pp. 31–43, 2014.

[24] "Unleash the power of Kinect for Windows SDK! – EternalCoding", *Eternalcoding.com,* 2017. [Online]. Available: https://www.eternalcoding.com/?p=1663.

[25] S. Giovanni, Y. C. Choi, J. Huang, E. T. Khoo, and K. Yin, “Virtual Try-On Using Kinect and HD Camera,” *Motion in Games Lecture Notes in Computer Science*, pp. 55–65, 2012.

[26] Z. Zhang, “Microsoft Kinect Sensor and Its Effect,” *IEEE Multimedia*, vol. 19, no. 2, pp. 4–10, 2012.

[27] P. C. Kucharski, D. Sielski, K. Grudzień, W. Kozakiewicz, M. Basiuras, K. Greif, J. Santorek, and L. Babout, “Comparative analysis of multitouch interactive surfaces,” *Proceedings of the 2017 Federated Conference on Computer Science and Information Systems*, 2017.

[28] Immersivetouch.com. (2017). *ImmersiveTouch Inc.*. [online] Available at: http://www.immersivetouch.com/.

[29] H. Gonzalez-Jorge, B. Riveiro, E. Vazquez-Fernandez, J. Martínez-Sánchez, and P. Arias, “Metrological evaluation of Microsoft Kinect and Asus Xtion sensors,” *Measurement*, vol. 46, no. 6, pp. 1800–1806, 2013.

[30] Y. Zhao and C. Jiang, “Online Virtual Fitting Room Based on a Local Cluster,” *2008 International Workshop on Education Technology and Training & 2008 International Workshop on Geoscience and Remote Sensing*, 2008.

[31] B. Sobota, L. Jacho, S. Korecko, and K. Nogradiova, “On the way to virtual training system based on human body movements,” *2015 IEEE 13th International Scientific Conference on Informatics*, 2015.

[32] Y. Yu, C. Zhou, L. Huang, and Z. Yu, “A Moving Target Detection Algorithm Based on the Dynamic Background,” *2009 International Conference on Computational Intelligence and Software Engineering*, 2009.

[33] H. Boujut, M. Ourir, and T. Zaharia, “A fully automatic framework for building 3D animated avatars,” *2014 IEEE Fourth International Conference on Consumer Electronics Berlin (ICCE-Berlin)*, 2014.

[34] S.-I. Park, J. Lee, D. Yoo, and B. Goo, “An approach to the reduction of time delay in a networked performance,” *Proceedings of the Asia-Pacific Advanced Network*, vol. 36, p. 92, 2013.

[35] Msdn.microsoft.com. (2018). *Skeletal Joint Smoothing White Paper*. [online] Available at: https://msdn.microsoft.com/en-us/library/jj131429.aspx [Accessed 31 Jan. 2018].

[36] M. Kuo, S. Collinson, M. Jager, and J. Morris, “Fast accurate rendering,” *2008 13th Asia-Pacific Computer Systems Architecture Conference*, 2008.

[37] M. A. Livingston, J. Sebastian, Z. Ai, and J. W. Decker, “Performance measurements for the Microsoft Kinect skeleton,” *2012 IEEE Virtual Reality (VR)*, 2012.

[38] P. Cerveri, A. Pedotti, and G. Ferrigno, “Robust recovery of human motion from video using Kalman filters and virtual humans,” *Human Movement Science*, vol. 22, no. 3, pp. 377–404, 2003.

[39] Msdn.microsoft.com. (2018). *The MVVM Pattern*. [online] Available at: https://msdn.microsoft.com/en-us/library/hh848246.aspx [Accessed 10 Feb. 2018].

[40] Docs.nvidia.com. (2018). *Clothing Tool — NVIDIA APEX Documentation*. [online] Available at: http://docs.nvidia.com/gameworks/content/gameworkslibrary/physx/apexsdk/APEX_Clothing/ClothingTool.html [Accessed 10 Feb. 2018].

[41] H. Li and W. Wang, “A new numerical integration method using trapezoid formula for 3D virtual clothes animation,” *2017 2nd International Conference on Image, Vision and Computing (ICIVC)*, 2017.

[42] Z. Li, X. Jin, B. Barsky, and J. Liu, “3D clothing fitting based on the geometric feature matching,” *2009 11th IEEE International Conference on Computer-Aided Design and Computer Graphics*, 2009.

[43] K. Sugimoto, R. A. Cohen, D. Tian, and A. Vetro, “Trends in efficient representation of 3D point clouds,” *2017 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, 2017.

[44] H. Kato and M. Billinghurst, “Marker tracking and HMD calibration for a video-based augmented reality conferencing system,” *Proceedings 2nd IEEE and ACM International Workshop on Augmented Reality (IWAR99)*.

[45] K. Essmaeel, L. Gallo, E. Damiani, G. D. Pietro, and A. Dipanda, “Temporal Denoising of Kinect Depth Data,” *2012 Eighth International Conference on Signal Image Technology and Internet Based Systems*, 2012.

[46] H. Saito, S. Baba, M. Kimura, S. Vedula, and T. Kanade, “Appearance-based virtual view generation of temporally-varying events from multi-camera images in the 3D room,” *Second International Conference on 3-D Digital Imaging and Modeling (Cat. No.PR00062)*.

[47] G. Cheung, S. Baker, and T. Kanade, “Visual hull alignment and refinement across time: a 3D reconstruction algorithm combining shape-from-silhouette with stereo,” *2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings.*

[48] A. M. Andrew, “Multiple View Geometry in Computer Vision20012Richard Hartley, Andrew Zisserman. Multiple View Geometry in Computer Vision. Cambridge: Cambridge University Press 2000. xvi 607 pp., ISBN: 0-521-62304-9 hardback, £60.00,” *Kybernetes*, vol. 30, no. 9/10, pp. 1333–1341, 2004.

[49] “3D scanning | laser scanner strip light scanner DAVID,” *CNC-STEP.com - CNC Machine Manufacturer*, 23-Jan-2018. [Online]. Available: https://www.cnc-step.com/3d-scanning/. [Accessed: 01-Mar-2018].

[50] K. Khoshelham and S. O. Elberink, *Sensors (Basel, Switzerland)*, 2012. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3304120/. [Accessed: 02-Mar-2018].

[51] “Time-of-flight camera,” *Wikipedia*, 01-Mar-2018. [Online]. Available: https://en.wikipedia.org/wiki/Time-of-flight_camera. [Accessed: 01-Mar-2018].

[52] R. Horaud, M. Hansard, G. Evangelidis, and C. Ménier, “An overview of depth cameras and range scanners based on time-of-flight technologies,” *Machine Vision and Applications*, vol. 27, no. 7, pp. 1005–1020, 2016.

[53] S. Agarwal, N. Snavely, I. Simon, S. M. Seitz, and R. Szeliski, “Building Rome in a day,” *2009 IEEE 12th International Conference on Computer Vision*, 2009.

[54] A. Hilsmann and P. Eisert, “Tracking and Retexturing Cloth for Real-Time Virtual Clothing Applications,” *Computer Vision/Computer Graphics CollaborationTechniques Lecture Notes in Computer Science*, pp. 94–105, 2009.

[55] Larry Li, “Time-of-Flight Camera – An Introduction,” *Texas Instruments* . Jan-2014.

Appendix A - Before System Evaluation Questionnaire

How often do you go shopping? (Choose one)	
1) Once per week	
2) Once per month	
3) More than once per month	
4) Less than once per month	
5) As necessary	
When do you plan to go shopping? (Choose at least one)	
1) Weekend	
2) Public holiday	
3) Weekday	
4) As necessary	
How much average time do you spend on each shopping day? (Choose one)	
1) Less than one hour	
2) Between 1 to 3 hour	
3) Half of day	
4) Full day	
5) As necessary	
How much average time do you spend on each shopping day? (Choose one)	
1) Less than one hour	
2) More Than one hour	
3) Half of day	
4) Full day	
5) As necessary	
How many clothes do you try on average on each shopping day? (Choose one)	
1) Less than 5	
2) 5 to 10	
3) 11 to 15	
4) More than 15	

List the most time consuming activity during shopping?
……..

Have you ever heard or tried out a virtual try on system? (Choose at least one)	
1) Yes	
2) No	
3) I have no idea about virtual try on system	

Appendix B – Interface evaluation

What do you think about the interface		1	2	3	4	5	
	Terrible						Wonderful
	Difficult						Easy
	Dull						Interesting
	Rigid						Flexible
Screen organization and learning		1	2	3	4	5	
1) Sequence of screens	Confusing						Clear
2) Feature categorization	Confusing						Clear
3) Self-Learning	Difficult						Easy
4) Performing tasks is straightforward	Never						Always

Which functionalities did you tried	Yes	No
Change Gender		
Take a photo		
Remove Clothes		
Sound on/off		
Resize or Move up/down cloths		

List the most Negative aspects
List the most Positive aspects

Appendix C – Functionality Evaluation

Ease of Use		1	2	3	4	5	
5) Easy to use.	Disagree						Agree
6) User friendly.	Disagree						Agree
7) Possible to use with fewer instructions.	Disagree						Agree
8) Requires the fewest steps possible to accomplish what I want to do with it.	Disagree						Agree
9) Possible to use in day to day life.	Disagree						Agree
10) Will it be easy to use by general users?	Disagree						Agree
Performance		1	2	3	4	5	
1) Performance of the system is acceptable.	Disagree						Agree
2) Graphics and details of the garments are acceptable	Disagree						Agree
3) System works smoothly.	Disagree						Agree
4) System in a state that can be improved for commercial use.	Disagree						Agree
5) Possible to try out different garments very quickly.	Disagree						Agree
Usefulness		1	2	3	4	5	
1) Very suitable for people with very busy life style.	Disagree						Agree
2) System meets my needs.	Disagree						Agree
3) It has every feature I would expect to have.	Disagree						Agree
4) System helps me to select most suitable Garments	Disagree						Agree
5) It saves time.	Disagree						Agree
6) It has less complexity	Disagree						Agree
Overall Satisfaction		1	2	3	4	5	
1) I'm Completely satisfy with the ease of use of the system	Disagree						Agree
2) I'm Completely satisfy with the performance of the system							

3) I’m Completely satisfy with the usefulness of the system							