International Journal of Applied Engineering Research
ISSN 0973-4562 Volume 10, Number 8 (2015) pp. 20597-20612

http://www.ripublication.com

Ornament Dressing Room: A Virtual Dressing Room For Jewelry

Shridhar S Kini[1] and Suganya Raman[2]

1) Shridhar S Kini, 2nd year MTech, Computer Vision and Image Processing, CSE department, Amrita Vishwa Vidyapeetham, Coimbatore, India, 641112,
Email: shridharkini6@gmail.com
Ph:09944494787
2) Suganya Raman, Associate Professor, CSE department, Amrita VishwaVidyapeetham, Coimbatore, India, 641112,
Email: r_suganya@cb.amrita.edu

Abstract

The author proposes a system architecture for an Augmented Reality (AR) application “Ornament Dressing Room”. This paper describes a technique to identify the control points required for placing the jewellery virtually on the neck. The proposed system requires a user who is intend to give a trial to a jewellery, has to face a Kinect Sensor, and the system augments the jewellery over the user's neck. The proposed method makes use of both RGB as well as Depth stream available from Microsoft Kinect to identify the required control point. The system not only enhances the user experience but it will surely reduce the wear on and off time for every jewellery.

Keywords Virtual Dressing Room, Kinect sensor, Augmented Reality (AR), Depth Stream, RGB processing, Image Processing;

1. Introduction

Virtual Reality (VR) is an art of computer simulation by modifying the environment according to the response that we get from sensors. The modification in this context remains pre-calculated and stored. Whereas Augmented Reality (AR) is slightly different from VR, the modification of environment happens on live feed and nothing is pre-calculated. In Real Time, one has to process the feed and then augment the rendering content over the live feed.

For VR application we could have used the special purpose displays like Head Mounted Displays (HMD’s) where as in AR, we need to be able to do every augmentation on normal display systems, be it the mobile display or monitor or TV display. The Virtual Dressing Room is a computerised dressing room where virtual dresses are put over the user. Of course the Virtual Dressing Room is modified to Augmented Reality application in recent years because of need of the customers and joy that they get out of augmenting something on them rather animating on avatar.

But in this paper, the task of our system is to give the feel of wearing a jewellery on human body. In real sense the jewellery is getting augmented on human body in real time. As the orientation of user changes, the jewellery changes its orientation to fit the user orientation. It enables users to try jewellery to check, one or more of size fit or style, but virtually rather than physically. The system aims at “how the jewellery fits on me, and how it suits me”.

2. BACKGROUND AND RELATED WORKS

As the Ornamental Dressing Room concept is similar to Virtual Dressing Room (VDR), analysing few of the existing VDR system allowed us to decide the architecture for our system.

Young In Yeo [4], is an early dressing room for garments. Author discusses the different kinds of avatar simulation, placement of sensors for complete body control point tracking, template modelling for avatar, reconstruction of avatar model based on motion of the user. This white paper is a base to come up with architectural diagram for our proposed system.

Chang H.T [1], used two Microsoft Kinect sensor for acquiring the human complete body to analyse size of the garment that fit the user. Author shows the various joint that Microsoft Kinect will give to developer so that they can exploit it to reconstruct 3D model of user. Author also discuss two display system, one for actual rendering of garments (Dynamic Fitting Room display) and one more is for wardrobe screen which works as a shelf for user to select the garments from the shelf.

Stefan [3], discuss the advancement in virtual try on. They pinpointed the drawbacks of older way of identifying the rendering and reconstruction. They came up with solution to remove the computationally intensive motion capture, 3D reconstruction or modelling by introducing combined image based rendering of user and previously stored garments, which transfers the earlier appearance of garment to real time user matching input and recorded frames. They also used image-based Visual hull rendering and online registration methods.

The systems and techniques discussed in the above sections have lots of advantages as well as disadvantages. The Table (1) provides a summary of strengths and weaknesses of each technique/system.

Table 1. Comparison study of existing frame work.

Technique	Strengths	Weaknesses
Magic Mirror	Template modelling of avatars	Avatar based rendering - not an augmented reality based
Chang H T	Full body reconstruction	Uses two Kinect rather one
Stefan	Removed all computationally expensive methodology and arbitrary viewing angle is possible	Data set collection is cumbersome task. As single data set has to be captured by almost 10 cameras in all the angles simultaneously to reconstruct in arbitrary viewing angle

3. PROPOSED SYSTEM ARCHITECTURE

As the interest is only on jewellery augmentation rather than garments as a whole, we restructured the existing architecture as per our requirements by taking certain concepts from earlier methodologies discussed. The whole of dressing room concepts is based on a cycle known as Dressing cycle. Figure (1) shows the Dressing Cycle.

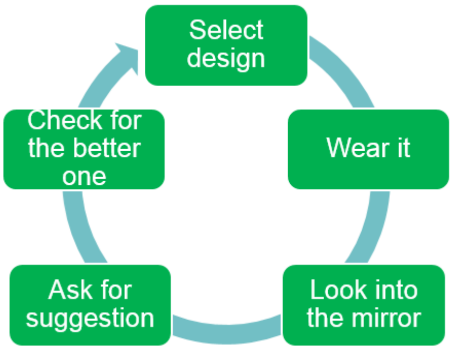

Figure 1. Dressing Cycle

Goal of this application is to recreate all the blocks available in dressing cycle by AR Technology to ease the user in jewellery selection process. The 1st, 2nd and 3rd process are crucial processes that we would like to model from the system.

Data Set Details: The data set consists of jewelry images taken from 8MP resolution camera along with the mannequin. Reason for keeping the jewelry over mannequin is, to give the proper physics and shape to the jewelry. These images are preprocessed using photo editing software's. The irrelevant portions of the jewelries are etched out manually using the software. Since mannequin and background portions will not contribute during rendering, these pixel are white washed. The pre-processed images are shown in Figure (3). The metadata of each jewelry is obtained and stored in a file. The metadata includes the end point of the jewelry, its name, its scaling factor by which scaling has to be done while rendering.

Figure 2. Raw Data Set

Figure 3. Pre-processed data set

The Figure (4) shows the Architecture diagram of our proposed method.

The proposed system consists of 5 modules namely Acquisition, Depth-Processing, RGB-Processing, Control Point identification and tracking, and Visualization and Rendering. Each module will be detailed in the upcoming sections.

3.1 ACQUISITION

We have used Microsoft Kinect as acquisition device which captures RGB video as well as Depth Stream. All the device drivers has to be installed prior to acquisition. The library packages from Microsoft SDK as well as OpenNI 2.0 has been installed on the computer. The Depth Stream and RGB streams together used for segmentation and control point identification whereas RGB video is also used for visualization. The User has to stand 400mm away from the Kinect Sensor so that depth measurement is not blocked by its Blind zone.

3.2 DEPTH-PROCESSING

This unit is responsible for the getting boundary of user using Depth stream alone. The boundary obtained from depth stream is far better compared to the boundary obtained from RGB stream. This module has 4 sub-modules namely Depth

Segmentation, Segmentation Refinement, Boundary Extraction and Connected Component Analysis.

3.2.1 DEPTH SEGMENTATION

The task of Segmentation is to separate the regions which are crucial for us from non-region of interest. The algorithm that we have used is kmeans clustering [5] method for Depth Segmentation, which would cluster the pixel of foreground and background. The background pixels contribute to the background or non-region of interest, whereas foreground corresponds to the user. So the number of classes required for segmentation remains two. The result of Depth Segmentation is shown in Figure (5).

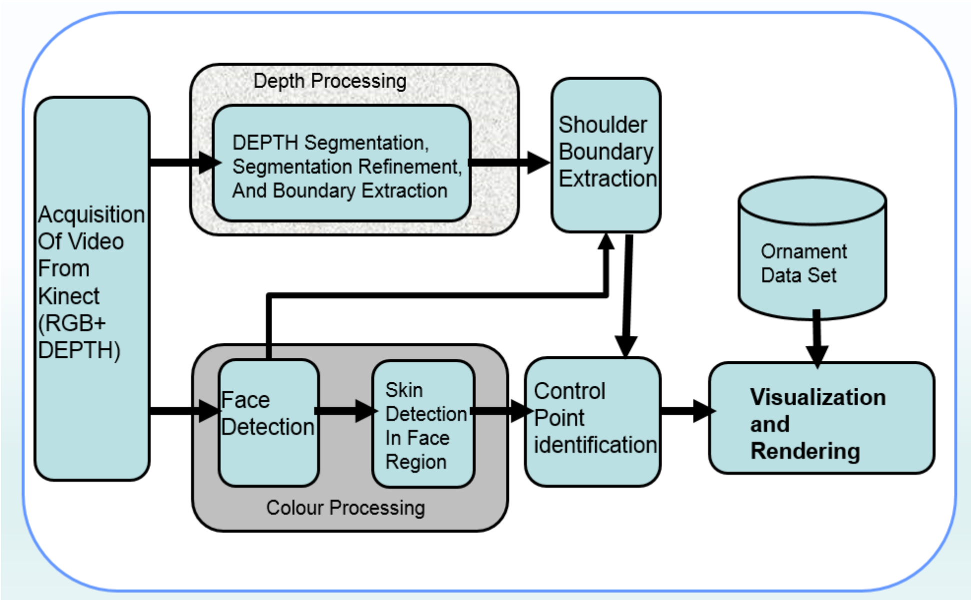

Figure 4. Architectural Block Diagram

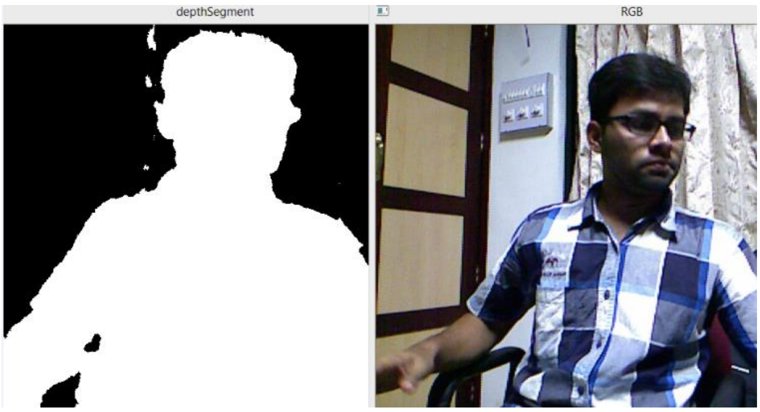

Figure 5. Result of Segmentation

3.2.2 Segmentation Refinement

It is observed that, pixels near the boundary of foreground and background are extrapolated as foreground, shown in figure (7). It is because of un-synchronization between IR transmitter and IR receiver of Kinect. So, in order to nullify this effect, we made it as 3 class problem rather than two class clustering, where each region will be background, foreground and error region. Once we obtain 3 clusters, the smallest cluster, which is error cluster is grouped to background cluster so that the end product would be again two class problem. This task of achieving two class from three class is termed as Segmentation Refinement. The result of Segmentation Refinement is shown in Figure (7). The region which is shown in red colour for Figure (6) is the error region. In Figure (7), error region is grouped to background.

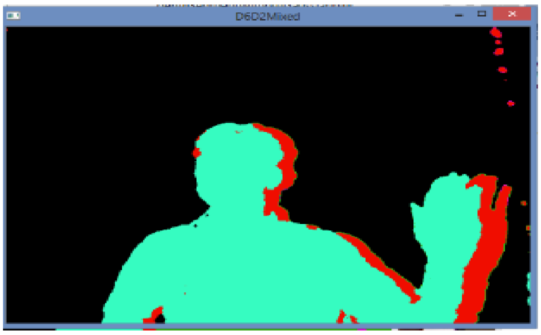

Figure 6. Extrapolation of error pixel

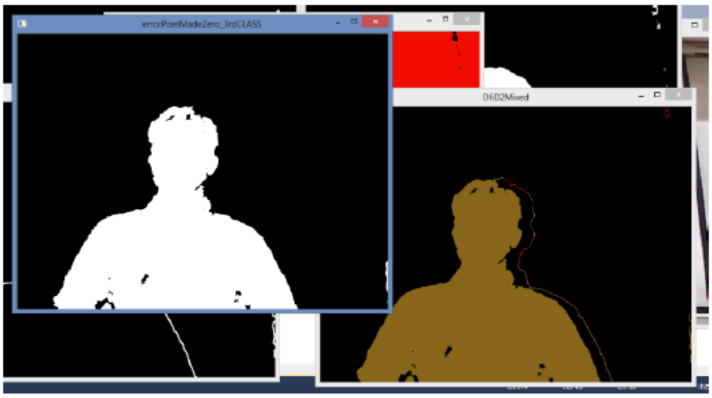

Figure 7. Result of Segmentation Refinement

3.2.3 Boundary Extraction

Once the foreground object is available which is free from all the error pixels, the boundary of foreground is extracted by Morphological Boundary Extraction Method [6], which is obtained by subtracting eroded image with original image.

$$\text{Boundary}(I) = I - \text{Erode}(I) \quad (1)$$

Where,
I – Foreground Image pixels which are in white colour
Erode – Morphological Erosion operation on Binary Image
The result of Boundary Extraction is shown in Figure (8)

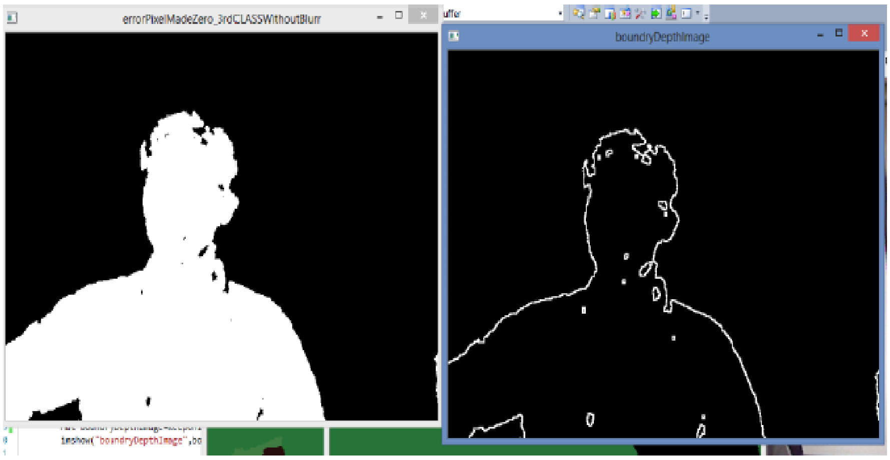

Figure 8. Result of Boundary Extraction

3.2.4 Connected Component Analysis

It is observed that once after obtaining the boundary, we may get many disconnected groups with smaller boundaries on the foreground object. This is due to segmentation error. To remove this small non-connected component from the boundary of user, we need to go for Connected Component Analysis [7] to identify all connected component and choose the biggest connected component which belongs to the exact boundary of the user. There are many techniques for connected component analysis, but one we choose is Connected Component Analysis by Two Pass method [7].The result of Connected Component Analysis is shown in Figure (9).

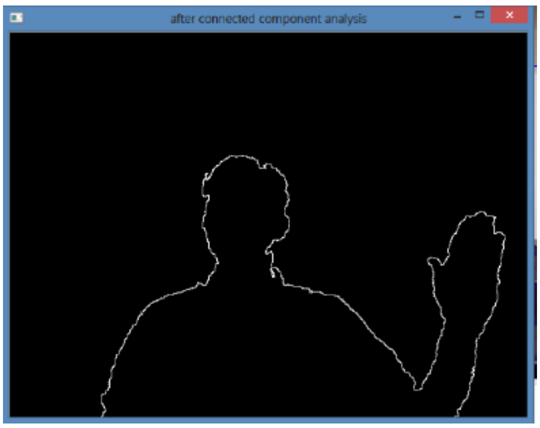

Figure 9. Result after Connected Component Analysis

Thus after depth processing stage, the boundary that we obtained is ready to be given to the control point identification where we can place the jewellery selectively over these points.

3.3 RGB PROCESSING

In this stage we are going to make use of the RGB frame from video in order to detect the Face, which intern gives the Shoulder ROI.

3.3.1 FACE DETECTION

We have incorporated Viola-Jones [8] method for face detection. The result of it gives the region of interest (ROI) of Face in the RGB image. The Viola-Jones method uses Haar Basis Functions of size 24X24 for extracting features from the greyscale images. To speed up the process of computation of these features, they brought up Integral Image concept, which will take constant amount of time to calculate the feature for an image. Once the features has been identified, the number of features turned to be more than actual number of pixel, in order to remove the weak features a modified version of AdaBoost classifier was created which selects only the features which may contribute in face detection. Later, they have used classifiers in cascade form so that speed of the detectors can be increased by removing unwanted non-face regions from the frame in every classifier that it pass through. By doing so the complex and crucial calculation are done only at the facial region rather than non-facial region. The face detection result is shown in figure (11).

3.3.2 SKIN DETECTION

Once the Face ROI is available from previous method, apply a skin detection algorithm to identify the skin pixels in the face. We have used two methods cascaded to identify skin pixels in-order to get good accuracy of skin detection. The first one is the Ratio based method [11] by Ghazali et, al and the second one is Kovac model of rules [12].

$$0.0 \leq \frac{(R-G)}{(R+G)} \leq 0.5 \quad \text{and} \quad 0.0 \leq \frac{(B)}{(R+G)} \leq 0.5 \qquad [11]$$

$$R > 95 \ \&\ G > 40 \ \&\ B > 20 \ \&\ \{max(R, G, B) - min(R, G, B)\} > 15$$

$$\&\ |R - G| > 15 \ \&\ R > G \ \&\ R > B \qquad [12]$$

The pixels which are termed as skin pixel in both the method, those pixels are selected as skin pixel or else made it as non-skin pixel. Thus obtained skin pixels will not form a complete face. So, to remove small-small disconnected skin regions as well as non-skin regions, a Connected Component Analysis [7] is done, where bigger connected component resembles the actual skin of the Face. Further, Dilation followed by Erosion is applied to fill the hole in skin region as well as to smooth the boundary of the skin region. The result of these process is shown in figure (10).

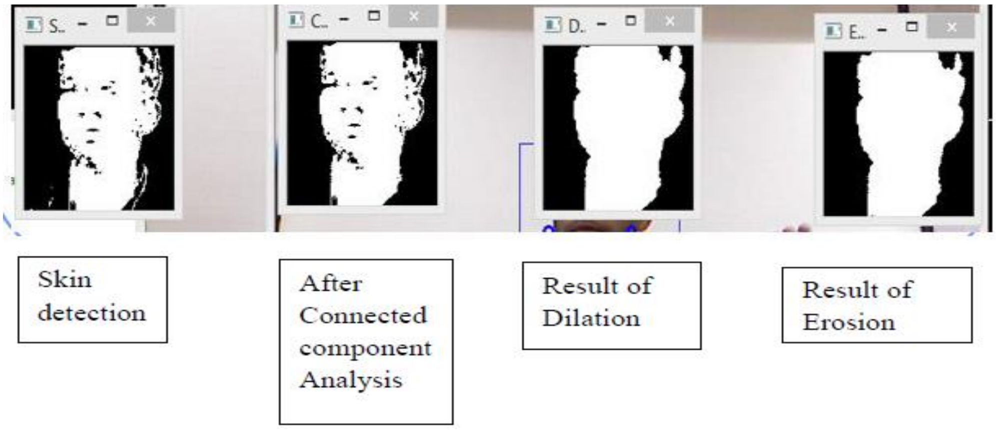

Figure 10. Result of Skin Detection.

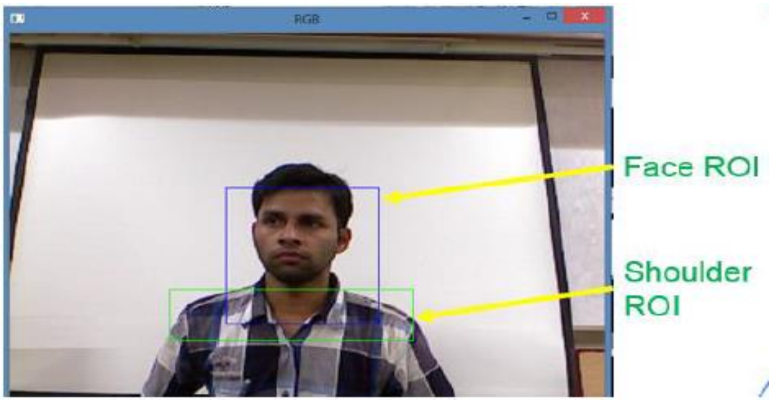

Figure 11. ROI for Face and Shoulder

3.3.3 Shoulder ROI

The knowledge of Face ROI can be used to identify the presence of shoulder of the user. As human beings are so regular and symmetric, identifying shoulder for one individual remains same for the all. Though we are not going to use any grey level information from the shoulder ROI, it will be used in further processes.

3.4 Control Point Identification

The most crucial part in the whole process of Dressing room is control point for the placement of Jewellery. Control point identification section is divided into 4 sub section. Each subunits is explained below.

3.4.1 Collection of Shoulder Points in Boundary image

The knowledge of shoulder ROI is utilized to segment the Boundary obtained in section 3.2. The boundary points are segregated as two shoulders. This segregation is easily identifiable as they form two disjoint sets always.

3.4.2 LINE FITTING FOR SHOULDER POINTS

Each shoulder points are given to Line Fitting algorithm by Least Square Method [14]. The Least Square Method is simplified to simple correlation sort of equations by the author [14] in their text book. The step by step process is explained below.

Consider set of ordered pairs (x_1, y_1), (x_2, y_2),..., (x_n, y_n) (for single shoulder alone is taken)

Step1: Calculate the mean of the *x*-values and the mean of the *y*-values.

$$\overline{X} = \frac{\sum_{i=1}^{n} x_i}{n}, \quad \overline{Y} = \frac{\sum_{i=1}^{n} y_i}{n}$$

Step2: Use following formula to get the slope of the line of best fit

$$m = \frac{\sum_{i=1}^{n} (x_i - \overline{X})(y_i - \overline{Y})}{\sum_{i=1}^{n} (x_i - \overline{X})^2}$$

Step3: Compute the *y*-intercept of the line by using the formula

$$b = \overline{Y} - m\overline{X}$$

Step4: Use the slope *m* and the *y*-intercept *b* to form the equation of the line. The process is shown in figure (12).

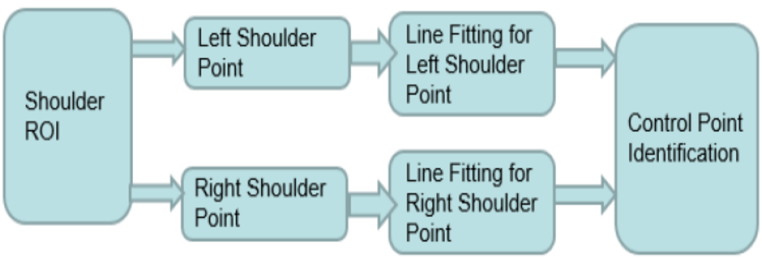

Figure 12. Best Fit Line For shoulder points

The result of Best Fit Lines to the shoulder point is shown in figure (13).

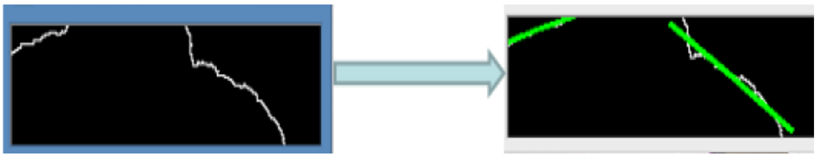

Figure 13. Result of Best Fit Line

3.4.3 INTERSECTION POINT IDENTIFICATION

The obtained line from previous step and skin region obtained in section 3.3 will intersect at one particular point. And this point is a point where neck joins the shoulder. This point is called as Control point here after. Now the task here is to take all the point on this line and to check to see that it intersects or not. To obtain all the points on the line, we can make use of line equations. But it takes too much iterations to identify the point. So approach that we used is the Binary Search Pattern, where search region halves every time to find suitable region, such that the middle point of halved line falls on boundary of the skin and non-skin.

For every middle point (M) that we obtained from binary search method, we consider the 5X5 matrix around it and calculate its sum. And it follows that,

If sum < 255 then,
M is not at boundary (non-skin region)
Else is sum > 255*24 then,
M is not at boundary (skin region)
Else
M is at boundary

Thus we obtained control point for both shoulders.

The Process of Control point Identification is shown in Figure (14).

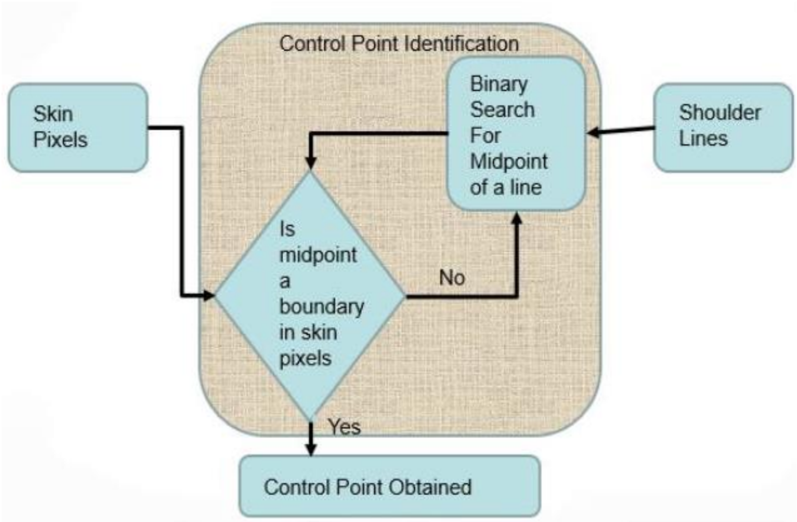

Figure 14. Control Point Identification

3.4.4 Control Point Tracking
The obtained control point has to be tracked in the subsequent frames so that whole process is not repeated again. If this control point is missed in any frame, then we need to re-calculate its position. Because we have missed the control point because of either tracking error or user might have undergone significant orientation change.

For tracking we can go for Gradient level tracking instead of grey level tracking. The reason is, slight variation in grey level due to illumination results in identifying the patch near control point is difficult. Whereas the gradient of this patch remains almost constant even if grey level varies a little bit. For this reason, we incorporate Lucas-Kanade-Thomasi point tracker [9] algorithm. If in case there is a larger motion of control point, then these point can we obtained by Laplacian Pyramid [10], followed by Lucas-Kanade-Thomasi point tracker.

3.5 Visualization and Rendering
Once control point is obtained, we need to render the jewellery by suitably placing its end point with the control point. The end point of jewellery and its scaling factors are retrieved from the Meta data that we created during the Data Set collection phase. The jewellery is scaled according to its scaling factor, and then jewellery image is superimposed over the user's neck by replacing the pixels of jewellery with user's body pixels so that it looks like an Augmented Reality application.

3.6 Flow Chart
The flow chart of our implementation is shown in the figure (15). The initial check for availability of Control Point is crucial, as we can ignore the tedious process of finding it again.

4. Experimental Result
4.1 System Requirements
We have implemented our proposed system with Microsoft Kinect as the acquisition system. Our algorithm runs on Microsoft Windows 8 with 4 GB RAM powered by Intel's Core i5 of 2.4GHz. We have used open source library OpenCV-2.4.9 to implement all image processing concepts. For acquisition from Kinect, OpenNI-2 has been used.

4.2 Result
We could able to get the control point accurately near the region where neck and shoulder joins. The augmentation process could have been assisted with 3D graphics of jewelry rather than plane jewelry images. Though this system is a early bird to the Industry Standard Project, improving this project with 3D modelling would have given better visualization to the real time environment.

Figure (16) shows the control point identified by our algorithm. Green circle represents the start, end and middle point of the shoulder line. The red circles represents the middle point identified by Binary Search Technic. 2 Red circles in the figure (16) tells us that it took 2 iterations of binary search to find out the exact control point. The Blue Circle represents the Control point identified in the intersection region.

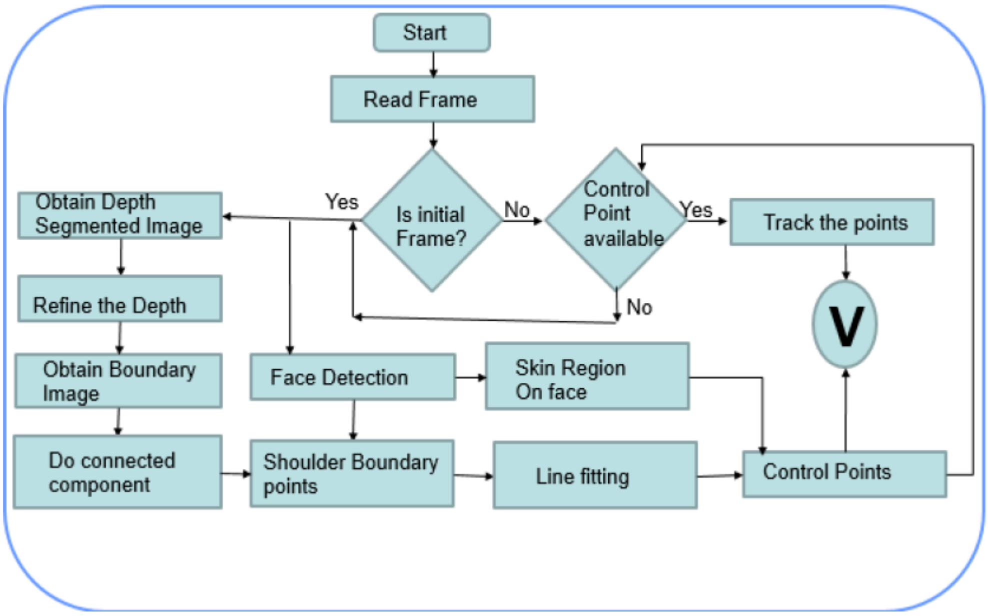

Figure 15. Flow Chart.

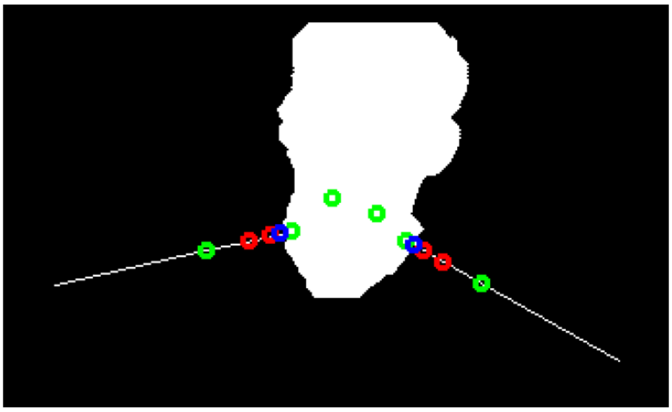

Figure 16. Result of Control Point identification

Figure (17) shows the control point on the user's neck. The circle in sky blue colour represents the control point where we can keep the jewellery.

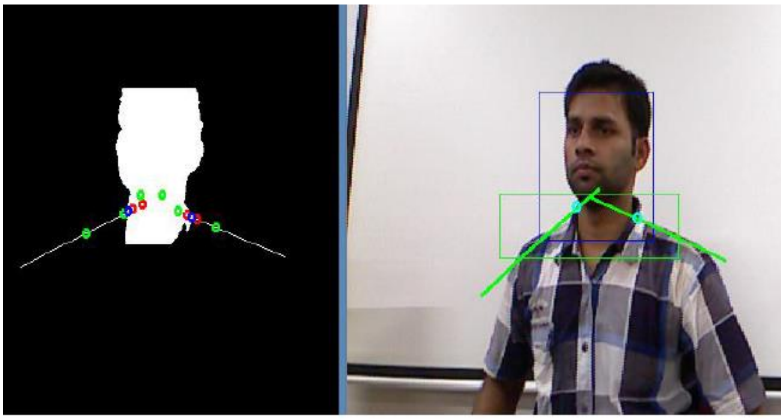

Figure 17. Control Point on User

Figure 18 shows the actual rendering of jewellery over the user.

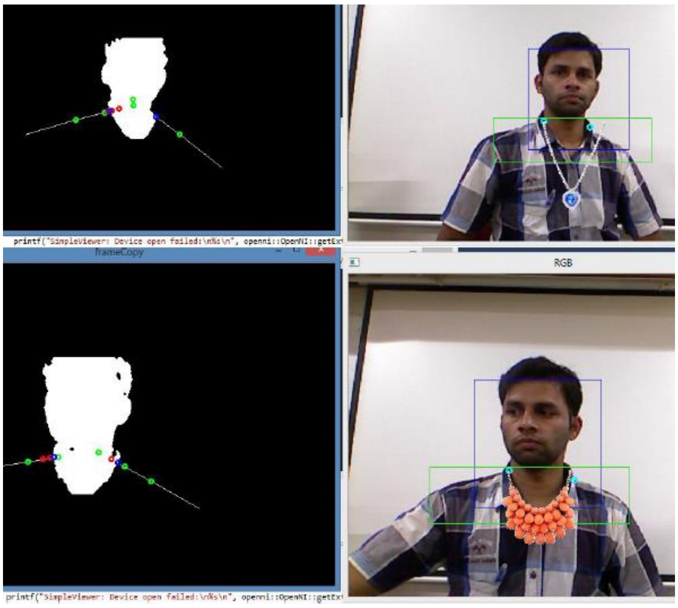

Figure 18. Jewel Augmented on User

4.3 RESULT ANALYSIS

The efficiency at which we got the control point is extremely satisfied. If in case skin detection fails, we can incorporate the Gaussian Mixture Model (GMM) to do the skin detection. The mixing of Depth and RGB can be exploited to get many other features such as the distance of the user from the Kinect so that proper scaling can be done while rendering of the jewel. The current system that we implemented runs at 3 FPS, which is not as reasonable as other systems are.

4.4 Future Enhancement

The application would have made more interactive and effective by incorporating 3D modelling, Orientation adaptation, and gesture control and so on.

As we can make it out from the flow chart that, there exists a process parallelism. For example the RGB processing and Depth Processing could have been paralleled and later they could have been combined to get the points.

The GPU could be used to accelerate the code level parallelism which could intern boost the Frames per Second (FPS). Process like Erosion, Dilation and Connected Component Analysis has high amount of code level parallelism inherit into it. The reason for such a parallelism is that, we are working on pixel domain not on frequency domain.

4.5 Conclusion

We have implemented a basic or minimal version of virtual fitting room. Our efforts will be there to improve this project to the next level. Suggestions for this project is always open to anyone who feels this project has potential to perform.

References

[1] Chang H.T, Li Y.W, Chen H.T, Feng S.Y, Chien T.T, "A Dynamic Fitting Room Based on Microsoft Kinect and Augmented Reality Technologies", *15th, International conference Human-computer interaction, 8007; 177-185*

[2] Quang-Song Le, Minh-Triet Tran, "Overlay Upper Clothing Textures to Still Images based on Human Pose Estimation", *Journal of Mobile Multimedia, vol. 10, no. 1&2 (2014) 043-057*

[3] Stefan Hauswiesner, Matthias Straka, and Gerhard Reitmayr, "Virtual Try-On Through Image-based Rendering", *IEEE Transactions on Visualization and Computer Graphics, vol. 19, no. 9, September 2013*

[4] I. Kim, H. Lee, H. Kim, "Magic mirror: A new VR platform design and its applications", *ACM SIGCHI International conference on Advances in computer entertainment technology 2004*

[5] Tapas Kanungo, David M. Mount, Nathan S. Netanyahu, Christine D. Piatko, Ruth Silverman, and Angela Y. Wu, "An Efficient k-Means Clustering Algorithm: Analysis and Implementation", IEEE *Transactions On Pattern Analysis And Machine Intelligence, Vol. 24, No. 7, July 2002*

[6] A.M.Raid, W.M.Khedr, M.A.El-Dosuky and Mona Aoud, "Image Restoration Based On Morphological Operations", *International Journal of Computer Science, Engineering and Information Technology (IJCSEIT), Vol. 4, No.3, June 2014*

[7] Akmal Rakhmadi, Nur Zuraifah Syazrah Othman, Abdullah Bade, Mohd Shafry Mohd Rahim and Ismail Mat Amin, "Connected Component Labeling Using Components Neighbors-Scan Labeling Approach", *Journal of Computer Science 6 (10): 1099-1107, 2010*

[8] Paul Viola and Michael J. Jones, “Robust Real-Time Face Detection”, *International Journal of Computer Vision 57(2), 137–154, 2004*

[9] Carlo Tomasi and Takeo Kanade. “Detection and Tracking of Point Features” *Carnegie Mellon University Technical Report CMU-CS-91-132*, April 1991.

[10] Peter J. Burt, Member And Edward H. Adelson “The Laplacian Pyramid as a Compact Image Code”, *IEEE Transactions On Communications, Vol. Com-3l, No. 4, April 1983*

[11] Ghazali Osman, Muhammad Suzuri Hitam and Mohd Nasir Ismail, “Enhanced Skin Colour Classifier Using Rgb Ratio Model”. *International Journal on Soft Computing (IJSC) Vol.3, No.4, November 2012*

[12] J. Kovaec P. Peer and F. Solina, ” Human skin colour clustering for face detection”, *International Conference on Computer as a Tool*, 22-24 September 2003, vol.2, pp. 144-148