A Review on Virtual Dressing Room for E-Shopping using Augmented Reality

A. A. Shaikh, P. S. Shinde, S. R. Singh, S. Chandra, R. A. Khan

Abstract— Augmented Reality combined with new algorithms and social media technologies have started a revolutionary shift away from the classic desktop paradigm and into the direction of intuitive, "natural interaction" where people interface with the technological world through hand gestures, speech and body language. The virtual dressing room will make use of Human Computer Interface and Augmented Reality and it will be used for online shopping. This will facilitate the shopping experience by letting customers to try-on apparel and/or mix and match accessories without being physically present in the retail shop. These platforms are not only powerful decision tools for the on-line shopper, but also contribute to the fun factor of in-store shopping. The system gets the data of custom body sizes to construct virtual fitting model through photos already uploaded. The model then tries on different costumes and the system shows the fitting effect. Augmented Reality Virtual Dressing room works by superimposing the model or picture of a garment or accessory within the live video feed of the customer. The super-imposed model or picture of the garment or accessory will then track the movements of the customer so it appears as if the customer is wearing the virtual item in the video-view. In addition, omnipresent social networking features allow sending photos or videos of the shopper wearing the apparel for quick feedback. The proposed project can achieve real-time, high-fidelity cloth simulation and provide encouraging online virtual fitting experiences.

Index Terms— Augmented Reality, Edge detection, Gesture recognition, Human-Computer Interaction, Information Kiosk, Motion tracking.

I. INTRODUCTION

Augmented Reality (AR) is a variation of *Virtual Environments* (VE), or Virtual Reality. VE technologies replace the real world completely with a synthetic environment. While submerged in the synthetic environment, the user cannot see the real world around him. In contrast, AR allows the user to see the real world, with virtual objects overlaid upon or composited with the real world. Therefore, AR add-ons reality rather than completely replacing it. Ideally, it would appear to the user that the virtual and real objects are harmonized in the same space.Augmented reality is an environment that includes both virtual reality and real-world elements. AR is a system that has the following three characteristics:

- They combine real and virtual.
- They are interactive in real time.
- They are registered in 3-D. [1]

Augmented reality is an interesting topic as combining real and virtual objects in 3-D can be very useful. Augmented Reality enhances a user's observation and interaction with the real world. With the help of the virtual object the user can get information about environment which he cannot detect with his own sense. The information obtained can be applied to the real world to make the task easier. Augmented reality is an interesting topic as it has vast application in fields such as medical visualization, maintenance and repair, annotation, robot path planning, entertainment, and military aircraft navigation and targeting. Using Augmented reality in *virtual dressing room for e-shopping* is very interesting as it would combine the real world with the virtual dresses superimposed on it. The necessary steps in the design of virtual Dressing room are as shown in fig. 1

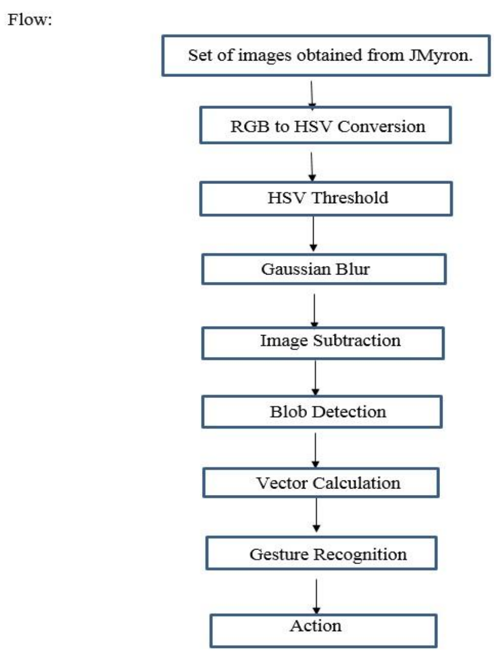

Fig. 1: Steps used in Virtual Dressing Room

Even though talking about Augmented Reality seems easy, but there are several problems which need consideration before an Ideal AR system can be developed. The most important problem which needs consideration is *REGISTRATION* which includes

Manuscript Received on November 2014.

Aleeza A. Shaikh, Department of Computer, Modern Education Society's College of Engineering ,University of Pune, Pune, India.
Prashant S. Shinde, Department of Computer, Modern Education Society's College of Engineering ,University of Pune, Pune, India.
Sandeep R. Singh, Department of Computer, Modern Education Society's College of Engineering ,University of Pune, Pune, India.
Swapnil Chandra, Department of Computer, Modern Education Society's College of Engineering ,University of Pune, Pune, India.
Prof. R. A. Khan, Department of Computer, Modern Education Society's College of Engineering ,University of Pune, Pune, India.

- *The Registration Problem*: The most basic problem which limits Augmented Reality application is the registration problem. The objects in real and virtual world must be aligned properly in all the aspects with respect to each other or the illusion of both the worlds coexisting would get negotiated. This problem gets enhanced due to a phenomenon known as Visual Capture [2] which is the tendency of the brain to believe what it sees rather than what it feels, hears, etc. When watching a television program, a viewer believes the sounds come from the mouths of the actors on the screen, even though they actually come from a speaker in the TV. Ventriloquism works because of visual capture. Similarly, a user might believe that her hand is where the virtual hand is drawn, rather than where her real hand actually is, because of visual capture. This effect increases the amount of registration error users can tolerate in Virtual Environment systems. Thus Registration is a topic of continuing research.
- *Static errors:* This class of error includes errors which are constant throughout the system. The four main sources of static errors are:
  - Optical distortion [3].
  - Errors in the tracking system
  - Mechanical misalignments
  - Incorrect viewing parameters (e.g., field of view, tracker-to-eye position and orientation, inter pupillary distance)

➢ *Dynamic errors*: This class of error includes errors which are caused due to system lag or delay. The end-to-end system delay is defined as the time difference between the moment that the tracking system measures the position and orientation of the viewpoint to the moment when the generated images corresponding to that position and orientation appear in the displays. The delays in the tracking subsystem, the communication delays, the time it takes the scene generator to draw the appropriate images in the frame buffers, and the scan out time from the frame buffer to the displays all contribute to end-to-end lag. Methods used to reduce dynamic registration fall under four main categories:

- Reduce system lag [4]
- Reduce apparent lag
- Match temporal streams (with video-based systems)
- Predict future locations

II. OBTAINING THE IMAGE

The images/video feed used in the virtual Dressing room are obtained from the JMyron library which stores the images/ video feed from the webcam and also provides an interface to use these images for further processing. JMyron library is used because it is user friendly and provides easy access and function to handle these images.

III. RGB TO HSV CONVERSION

RGB to HSV conversion is a very necessary step in virtual Dressing room because RGB color information is noisier than the HSV information. HSV, which stands for hue saturation value, separates luma, or the image intensity from Chroma or the color information. Hence we convert the RGB image input given to the system for image processing to an HSV image. Each of the HSV components is assigned a separate value which combines to acquire the original color in RGB. So it becomes necessary to convert RGB image to HSV image to apply further processing.

IV. HSV COLOUR THRESHOLDING

HCI exists all around us in our day to day lives. It is usually realized by computers' physical controls that are its hardware, namely, the mouse, keyboard or touchscreen. Along with HCI, image processing is another major activity used for communicating with computers. Image processing is analyzing and manipulating images with a computer with the result being that the image maybe altered in some way or it might be a report based on analysis of the image. Once the HSV value is calculated then standard deviation is calculated using which, the ranges of the three color components are computed and the image is thresholded. The threshold value is set from the hue histogram. We obtain a black and white image which contains distinct regions (black or white). Hence, the image processing effectively detects the distinct regions in an image which makes it easy to implement gesture recognition in the other algorithms. However, this algorithm has a few drawbacks for processing of shiny surfaces and brown colored hair in the images, barring which this method is reliable and leads fairly to accurate results. In our project of virtual Dressing room we will be following the same steps of image processing and HCI used in this paper and obtain an accurate skin detection process which will then be used for further processes like superposition of the clothes and gesture recognition.

V. GAUSSIAN BLUR

A Gaussian blur (also known as Gaussian smoothing) is the result of blurring an image by a Gaussian function. This is a very important step in image processing as it reduces image noise and also reduces unwanted detail. After applying Gaussian Blur the image appears as it is viewed through a translucent screen. Mathematically, applying a Gaussian blur to an image is the same as convolving the image with a Gaussian function. Since the Fourier transform of a Gaussian is another Gaussian, applying a Gaussian blur has the effect of reducing the image's high-frequency components; a Gaussian blur is thus a low pass filter. The formula for Gaussian blur for a two dimensional image is

$$G(x,y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}}$$

where x is the distance from the origin in the horizontal axis, y is the distance from the origin in the vertical axis, and σ is the standard deviation of the Gaussian distribution. Values from this distribution are used to build a convolution matrix which is applied to the original image. Each pixel's new value is set to a weighted average of that pixel's neighborhood. The original pixel's value receives the heaviest weight (having the highest Gaussian value) and neighboring pixels receive smaller weights as their distance to the original pixel increases. This results in a blur that preserves boundaries and edges better than other, more uniform blurring filters. The image obtained after applying Gaussian blur is as shown in fig. 3

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication Pvt. Ltd.

Fig. 2: Normal Image

Fig. 3: Gaussian Blurred Image

VI. IMAGE SUBTRACTION

Image subtraction is a process in which the numeric value of whole image is subtracted from another image. This is done for detecting changes between two images. [5] This result of changes can be used to tell if something in the image moved. It is majorly used in fields like astrophotography to assist with the computerized search for asteroids. This is mostly done if the image in question is a part of a video stream. Image subtraction provides important cues for numerous applications in computer vision, for example video surveillance tracking or human poses and gesture estimation.

In our project we use this technique to detect the movement of the user for one of the two reasons – to map the apparel on to his body perfectly and to recognise his hand gesture for carrying out required operations. There are various methods to implement this procedure. Some of them are frame differencing, mean filter, running Gaussian average, and background mixture models. The first algorithm gives us the best results with minimum complexity as compared to later ones. Hence we are using Frame Differencing method in our project. The subtraction of two images is performed straightforwardly in a single pass. The values of output pixel are given by:

$$Q(i,j) = P_1(i,j) - P_2(i,j)$$

Or if the operator computes absolute differences between the two input images then:

$$Q = |P_1(i,j) - P_2(i,j)|$$

Or if it is simply needs to subtract a constant value C from a single image then:

$$Q = P_1(i,j) - C$$

If the pixel values in the input images are actually vectors rather than scalar values (e.g. for color images) then the individual components (e.g. red, blue and green components) are simply subtracted separately to produce the output value. [5]

VII. BLOB DETECTION

A blob is a region of a digital image in which some properties are constant or vary within a limited range of values. Blob detectors have two classes: differential methods, which are based on derivatives of the function with respect to position, and methods based on local extrema, which are based on finding the local maxima and minima of the function. In early work, blob detection was used to obtain regions of interest for further processing. These regions could signal the presence of objects or parts of objects in the image domain with application to object recognition and/or object tracking. There are many algorithms for finding out colored blobs from an image. Some of them are Laplacian of Gaussian (LoG), Difference of Gaussians (DoG), Determinant of Hessian (DoH), maximally stable extremal regions, etc. There’s also Lindeberg's watershed-based algorithm for grey-level blob detection. In our project we first convert the RGB colored image into HSV image and then we threshold that image. The threshold value is chosen by the user itself according to the color of his palm. Hence we don’t use any of the above mentioned algorithms. Our algorithm is based on local extrema method. The algorithm used by us is explained below. A vector table is constructed first say V(x1, x2, y) to store the coordinates of blob. The image is scanned row wise until a black pixel is found, that is also the reason that we are storing a single value of y coordinate in the vector table. Its x value is stored in a vector table as x1. The scan then continues until the last black pixel is found in the same row. Its x and y values are stored as x2 and y respectively. Then the scan continues for the remaining rows till the last row. If the algorithm previously made two vectors for two different blobs which are connected below, then a single vector is maintained by merging the two vectors values. Then the min and max values of x and y coordinates are taken out of the vector table and the blob is detected. This is depicted in Fig. 4

VIII. VECTOR CALCULATION

Generally in vector calculation the pixels which are important for recognizing the gesture are calculated and it is used to find the center of gravity of the image so that the proper meaning of the gesture can be extracted.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication Pvt. Ltd.

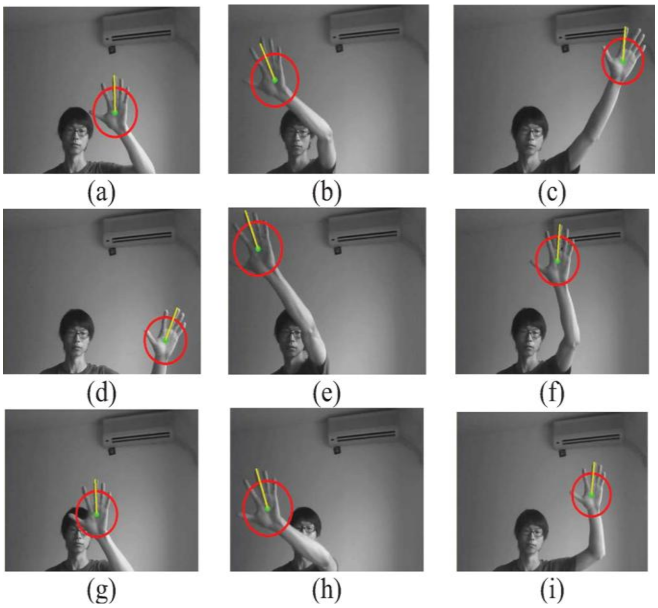

Fig. 4: Blob Detection Examples

IX. GESTURE RECOGNITION

Gestures are expressive and meaningful body motion which involves movements of different parts such as hands, arms, fingers, face, head or body with a motto of conveying meaningful information and interacting with the environment. Gesture recognition is a technique which derives meaning from the gesture made by the human, using his arms, face, hands, head and/or body. Gesture recognition is most important for designing an intelligent and effective HCI (Human Computer interface). In gesture recognition the gesture made by the user are recognized by the receiver and accordingly the task is performed. Gesture Recognition is rather a complex and the most important task in HCI. The Gesture Recognition cost is high mainly due to the tracking process. The drawbacks of the Gesture Recognition system get drastically worse in an uncontrolled context such as outdoor, where it is difficult to calibrate the environment, control lighting and limit the operating range to facilitate Gesture Recognition. Various approaches to handle gesture recognition [7], ranging from mathematical models based on hidden Markov chains [8] to tools or approaches based on soft computing [9]. In addition to the theoretical aspects, any practical implementation of gesture recognition typically requires the use of different imaging and tracking devices or gadgets. These include instrumented gloves, body suits, and marker based optical tracking. Traditional 2-D keyboard-, pen-, and mouse-oriented graphical user interfaces are often not suitable for working in virtual environments. Rather than using different imaging and tracking devices or gadgets which includes instrumented gloves, marker based optical tracking, devices that sense body (e.g., hand, head) position and orientation, direction of gaze, speech and sound, facial expression, galvanic skin response, and other aspects of human behaviour or state can be used to model communication between a human and the environment. A gesture can be static (where the use shows a particular pose or gesture), dynamic (where the user changes the environment rapidly) or a combination of static and dynamic. A gesture can get affected by the prior and/or the subsequent gestures. Also the gestures change from time to time, person to person and culture to culture, like shaking hands in western culture and saying Namaste in Indian culture. Typically, the meaning of a gesture [10] can be dependent on the following:

- *Spatial information*: where it occurs;
- *Pathic information*: the path it takes;
- *Symbolic information*: the sign it makes;
- *Affective information*: its emotional quality.

In Gesture recognition generally sensitive features are extracted from the region where it can be distinguished and extracted easily from a normalized image. Often the dynamic images obtained from tracking device are used to generate suitable features. The location and intensity are very important for recognizing the gesture. To determine all the aspects correctly we should sense and track the body position, angle, rotation and movement accurately and dynamically. For this purpose either a sensing device (like magnetic field tracker, gloves, cameras or computer vision technique) can be attached to the user. The choice of the device depends on the requirements including range of motion, accuracy, resolution, latency, user comfort, and cost. The basic Gesture Recognition algorithm is outlined as follows:

- Colour system conversion from RGB to HSV.
- Estimation of similarity measures between model and input regions.
- Thresholding similarity measures.
- Noise removal and dilation.
- Detection of Gesture candidate regions (i.e. Blob Detection).
- Selection of Gesture region.

The gesture recognition algorithm can be outlined as follows:

- Detect body for boot-strapping the tracker.
- Recognize the starting body shape, and initialize tracker with its template.
- While body is in view repeat
- Track the body and output encoded motion information until shape change is detected.
- Recognize the new shape and initialize the tracker with template of the recognized shape.
- Using hidden Markov models (HMMs), find the gesture, which gives the maximum probability of occurrence of observation sequence composed of shape templates and motion information.

X. ACTION

Once all the above necessary steps have been implemented the last step is to merge the real world image with the virtual dress. In the action part the virtual dresses are augmented on the real world image and displayed to the user.

XI. CONCLUSION

We conclude by saying that, the Virtual Dressing Room using Augmented Reality would hence prove to make e-shopping for the customers more efficient and thus be beneficial to the e-shopping websites as well. By implementing each of the above steps accurately one can efficiently build the Virtual Dressing Room application which will prove to be a huge success and take online shopping altogether to a new level.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication Pvt. Ltd.

XII. ACKNOWLEDGMENT

We would like to take this opportunity to express our deepest gratitude to all those who have supported and helped us with every tiny detail related to our project. We are highly indebted to Prof. R. A. Khan for her guidance and constant supervision as well as for providing necessary information regarding the project and also for her support in completing the project. Furthermore, we would like to acknowledge the crucial role of the University of Pune for providing us with the privilege to undergo this enriching experience. In conclusion, we would like to thank our parents and colleagues for their kind co-operation and investing their time, and willingly helping us out with their abilities to help make this project a reality.

REFERENCES

[1] Ronald Azuma, "A Survey of Augmented Reality," In Presence: Teleoperators and Virtual Environments 6, 4 (August 1997), 355-385.
[2] Welch, Robert B, "*Perceptual Modification: Adapting to Altered Sensory Environments,"* Academic Press (1978). ISBN 0-12-741850-4.
[3] Deering, Michael, "High Resolution Virtual Reality," *Proceedings of SIGGRAPH '92 (Chicago, IL, 26-31 July 1992). In Computer Graphics 26, 2 (July 1992), 195-202.*
[4] Foley, James D., Andries van Dam, Steven K. Feiner, and John F. Hughes, Computer Graphics: Principles and Practice (2nd edition). Addison-Wesley (1990).
[5] HIPR2 homepage at The University of Edinburgh School of Informatics,http://en.wikipedia.org/wiki/Image_subtraction#cite_note-1
[6] R. Gonzales and R. Woods, *Digital Image Processing*, Addison Wesley, 1992, pp 47 - 51, 185 - 187.
[7] V. I. Pavlovic, R. Sharma, and T. S. Huang, "Visual interpretation of hand gestures for human computer interaction," *IEEE Trans. Pattern Anal. Mach. Intell., vol. 19, no. 7, pp. 677–695, Jul. 1997.*
[8] L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proc. IEEE, vol. 77, no. 2, pp. 257–285,Feb. 1989.*
[9] S. Mitra and T. Acharya, Data Mining: Multimedia, Soft Computing, and Bioinformatics. New York: Wiley, 2003.
[10] S. Mitra and T. Acharya, "Gesture Recognition: A Survey," *IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS—PART C: APPLICATIONS AND REVIEWS, VOL. 37, NO. 3, MAY 2007.*