

Implementation of Real Time Dress-up System based on Image Blending

Md. Zahangir Alom,
Dept. of CSE
BRAC University,
Dhaka, Bangladesh

Farazul Haque Bhuiyan
Dept of CSE
BRAC University,
Dhaka, Bangladesh

Hyo Jong Lee
Dept. of CSE
Chonbuk National University,
Jeonju, South Korea

ABSTRACT

In this paper the real time virtual dress-up system has been proposed. The proposed system consists of multiple tasks including extraction of different body parts, torso detection, resizing input dress images and dress up using blending and re-blending techniques over the subject. The coexistence of different clothing and cluttering backgrounds is the main difficulty for accurate body extraction in image. Haar classifier is applied for detecting face from input frames and geometrical information is used to extract different parts (like a face, a torso, and hands) of a body according to the face position in a frame. Due to the variability of human body, it is complicated to extract accurately. A novel dominant colors based segmentation method is proposed to tackle this problem. First, an image is segmented into uniform areas based on HSV color components. Then, dominant colors of the torso are adaptively selected using color probability model. The torso has been extracted based on the dominant colors to resize the input dress image to fit over the subject body as well as dress size prediction. Automatic dress blending points are calculated on human body using torso starting position and geometrical relationship with face region. A selected dress is scaled and rendered to fit with the subject's body even they move around. Some preprocessing and post processing techniques are used to make outputs more accurate and realistic.

Keywords

Torso detection; Body measurement; Real-time Clothing; dress-up system; dress color recognition; human-computer interaction.

1. INTRODUCTION

Until recently, retailers have encountered great difficulties while attempting to sell clothing items via the Internet. Although consumers would like to benefit from the savings and convenience of shopping online, they are unable to determine how the clothes will fit. Businesses would similarly like to increase the proportion of their online sales, which would necessitate fewer physical stores and diminish the possibility of losing profits due to local competitors and returned goods. Our goal is to provide a concept for real time system in details that is able to effortlessly try on countless pieces of clothing, without leaving the comfort of their own homes. Moreover, people can also try to wear good looking dress when they wanted to go out from their home to party or other places. People use mirrors everyday to see how they look and choose clothes they will put on for a day before leaving home. Also in clothing stores, many mirrors are located to help customers for making their decision to buy dress fitting well and looking dresses. In this sense, detail

concepts for real time dress up system can answer your questions about dress up as well as the size-fitness of dress without physical don and doff time. The needs for the real time virtual dress up system are obvious. Firstly, benefits for customers are to save don and doff time and estimate their body measurements easily for made-to-measure dress. Customers commonly try on many items and spend lots of time to don and doff to purchase dress. It is very inconvenient for them to take dress items they want to try on, go to a dressing room, take off and put on whenever they find attracting dress. Secondly, shop owners can save costs, because they do not need dressing rooms any more. Additionally, wasting clothes tried on by customers will be reduced.

2. RELATED WORKS

In this section, we are going to discuss on the background concepts and related works of this propose real time dress up system. Anthropometric measurement system, 3D scanner based dress up system and avatar based system are discussed in the following paragraph. Recently, one manual dress up system is proposed only for static model and dress images. They have used some supporting software's in their proposed system for process the dress images as well as model images manually. "Soft Scissors" is used for matting out the foreground of input images. To establish the necessary correspondences between the model and dress images, joint positions have been selected manually on the input image. Moreover, the skeleton is fixed on the model image manually. The system also required the user height, weight, top, and bottom size of model [1]. For real-time cloth simulation, Thalmann et al [2] introduced a hybrid approach, which segments dresses into pieces that are simulated by various algorithms, depending on how they are laid on the body surface and whether they stick or flow on it. It is more natural and easier for users to control their avatars via their body movements than via basic input devices such as keyboard, mouse and joystick. Commercially available motion capture equipment, however, is far too expensive for common use and cumbersome, requiring the user to wear 40-50 carefully positioned markers and skin-tight clothing. Chai and Hodgins [3] introduced an approach to performance animation that employs video cameras and a small set of retro-reflective markers to create a low-cost, easy-to-use system. Kim et al [4] proposed the new virtual reality platform magic mirror that is economic in development process and cost, flexible by contents and installation conditions. According to the magic mirror system; they used video based virtual world instead of 3D models and simple 2D motion tracking method. There are many traditional methods related to the clothing system and body measurements have been discussed in the following

papers [5, 6, 7, 8, 9 and 10]. [TC]² and its associated brand, Image Twin, are the world leaders in 3D body scanning hardware and software. [TC]² provides 3D body scanning solutions in apparel, virtual fashion, health, medical, gaming, and online virtual worlds applications[11, 12].

3. PROPOSED APPROACH

This proposed system consists with two main steps, which are registration and dress-up as shown in Fig 1. Firstly, the user loads video stream from computer file or live camera and dress images are loaded from database into system interface. Secondly, the following steps such as extract face region, segmentation, real time torso detection, dress up using blending dress pixels and re-blending those pixels respect to user position or movement in front of camera have been implemented. Finally, the system automatically dressed up over the subject body perfectly in real time. Relevant sections of this work are given in the following paragraph.

3.1 Face detection

Although many different algorithms exist to perform face detection, each has its own weaknesses and strengths. Some use flesh tones, some use contours, and other are even more

complex involving templates, neural networks, or filters. These algorithms suffer from the same problem; they are computationally expensive [13]. Analyzing the pixels for face detection is time consuming and difficult to accomplish because of the wide variations of shape and pigmentation within a human face. Pixels often require reanalysis for scaling and precision. Viola and Jones devised an algorithm, called Haar Classifiers, to rapidly detect any object including human faces using “AdaBoost” classifier cascades that are based on Haar-like features and not pixels [14, 15, and 16]. Face in the frame has been detected by using this method. According to HSV color component based segmentation algorithm I_{frame} input frame of video file or live video

stream from camera and f_w , f_h belongs to the width and height of input frames respectively. Individual values for H, S, and V component are select automatically by applying the holding of HSV from the input frame of the video. The new values of HSV components have been reassigned in the segmentation frame. Torso detection probabilistic model has been applied on the segmented frame from input frame.

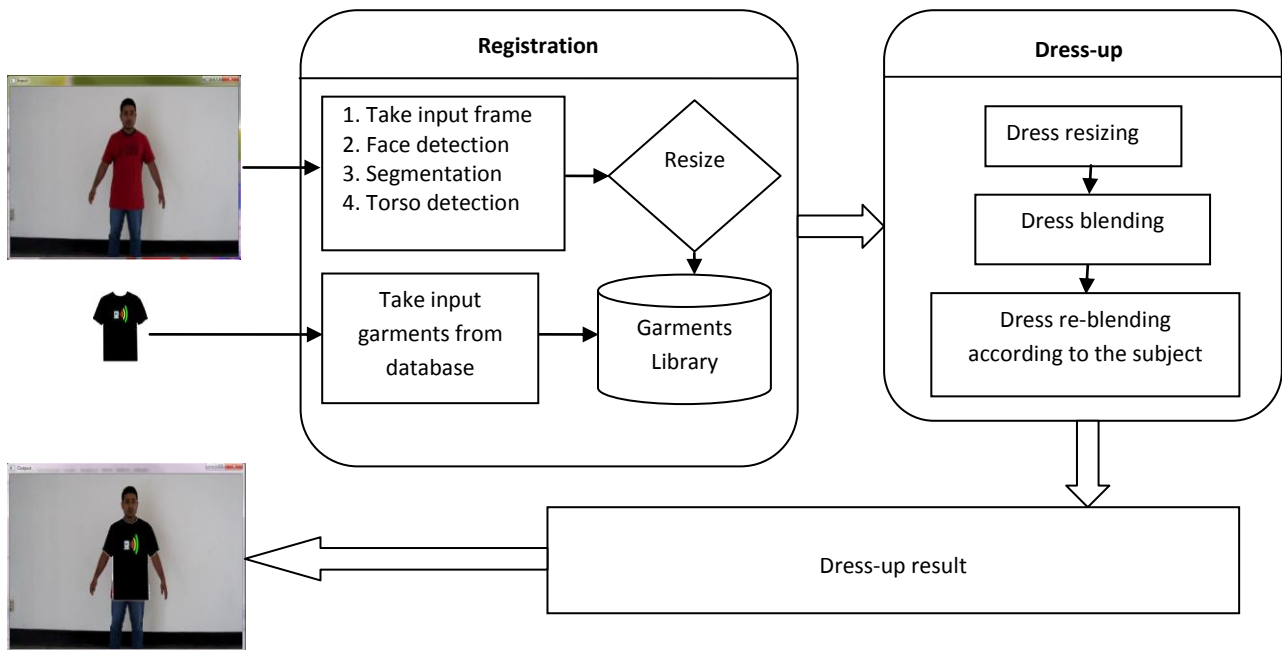


Fig 1: Proposed system overview: Our system has the two major steps, Registration and Dress-Up step. The user mostly registers input data at the Registration step, and the system automatically dresses up the model image in the dress at the Dress-Up step.

3.2 Real time torso detection

Torso detection is a very important part for body measurement. It plays an important role in human pose estimation, since torso connects most other body parts (arms, legs and head) together. However, locating of torso region in real time from video sequence is difficult due to different clothing, cluttering background and various poses. A number of schemes have been proposed for solving this problem for still image. Gang et al [17] utilize line segments to assemble torso shape, and Ren et al. [18] extract the torso region based on a pair of parallel line segments. In the exemplar-based approach [19], a lot of contour exemplars upon which the locations of joint were marked previously are stored. Then,

shape matching is used to transfer of the exemplars to test images. It is appropriate for images without cluttering backgrounds or complex clothing. When lots of line segments are observed near the torso due to complex backgrounds or clothing then the edge-based methods may not work well. The work of Moriet al [20], which is the most relevant to ours, is based on color image segmentation results produced by normalized cuts. Their work mainly focuses on baseball player in still images and some restricted cues employed in their method cannot be easily extended to handle real time images or video. In this paper, a novel dominant colors based method is proposed based on HSV color components. The segmentation of image is generated by pixel clustering of

HSV images component. Then, the dominant colors of the torso are selected by probability model based on HSV color

components.

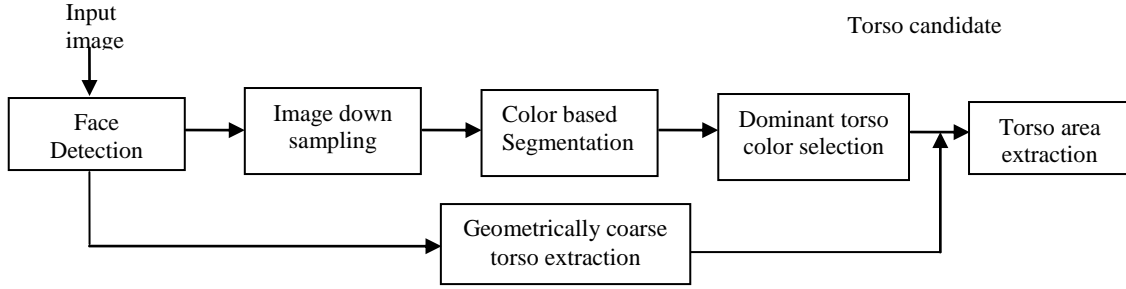


Fig 2: The framework of the proposed real time torso detection

There are two steps of color image segmentation procedure: pixel clustering and region merging. The details of these two steps are given below. Pixel clustering is belongs to human perception, the HSV components of each pixel are adopted to measure the color similarity, specially using H (hue) value. Using this color space, all the pixels with similar HSV components values are grouped together. The number of clusters is automatically decided by the average coarseness of the image [21]. K-means algorithm is used to cluster all the pixels values.

3.2.1 Merging of Regions

Many connected regions with the same cluster are small due to the cluttering background and illumination effect. To tackle this problem, the region merging process is followed as described below:

- By finding all the connected regions with the same cluster, select the largest region R_{\max} .
- If the area of R_{\max} is smaller than A_{\min} , consider as a noise region. Else, fill the holes in R_{\max} using the morphological operation.

- Repeat step a) until all the pixels is allocated to a certain region.

The parameters $A_{\min} = \alpha S_f$ where S_f is the area of the face, is a weighting factor and is set to 0.05 in our experiment.

3.2.2 Extraction of coarse torso region

There is a stable relationship between the ratio of the width of the torso and the head. We use the anthropometric data to get the coarse torso region, which is depicted as a rectangle with the center of the top at the position of the neck, as shown in Fig 5(b) [22]. The torso color probability is composed of the region probability and the location probability. The region probability presents the contribution of each region to the torso, and the location probability refers the probability for each pixel belonging to the torso.

3.2.2 Region Probability

Suppose there are L regions in total with more than one pixel inside the coarse region, represented by R_i , $i = 1, 2, \dots, L$. We have denoted the number of the pixels inside the region (see the rectangle in Fig 5 (c) for example) of R_i as T_i , the area

of region R_i as S_i , and the area of the coarse region as S_p . Then the region probability of $P(R_i)$ is defined as follows:

$$P(R_i) = \frac{T_i}{S_i} \left(\frac{T_i}{S_p} \right)^\alpha$$

The parameter α belongs to (0, 1), and is set to 0.5 experimentally here.

3.2.3 Location Probability

The probability of a pixel (x, y) is represented by $P_{x,y}$. Generally, the pixels near the neck are more likely to be the torso pixels and a general Gaussian distribution can properly describe for probability $P_{x,y}$, i.e.

$$P_{x,y} = \exp \left(- \left| \frac{\rho_{x,y}}{C} \right|^\beta \right)$$

where, (x, y) represent the image coordinates, and $\rho_{x,y}$ represents the Euclidean distance from pixel (x, y) to the neck. The parameters C and β are experimentally set as W_t and 4. W_t refers the width of the coarse region.

3.2.4 Selection of Dominant Torso Colors

The contribution of each color in the predicted region based on the torso color probability model is defined as follows:

$$N(k) = \sum_{i \in C(k)} \left(P_i \sum_{(x,y) \in R_i} P_{x,y} \right), \quad k = 1, \dots, K$$

where, K is the number of clusters in the predicted region, and C_k is the set of the regions belonging to the k -th cluster. Dominant torso colors can be extracted according to $N(k)$. The largest values of clusters are selected as the dominant torso colors.

3.2.5 Torso Detection

The regions with the most possible torso colors are estimated for the torso candidate (m). A single torso candidate consists of two components: the region probability L_{region} and the

edge probability L_{edge} [23]. For a single test image, we extract edges based on the results of color image segmentation. For each pixel of the lower half of the estimated candidate's contour, the closest edge point is calculated based on Euclidean distance; the edge probability measurement is then defined as:

$$L_{edge} = \exp(-\lambda_{edge} \bar{D})$$

where, the average shortest distance is \bar{D} , and λ_{edge} refers a constant and the value of λ_{edge} is set to 4 here. The measure of integrated probability is given by: $P(I|m) = L_{region} \times L_{edge}$. The final torso estimation procedure is as follows:

- By finding the regions with the most likely torso color. If the above area is less than S_{min} , merge the segments with the next most likely torso color.
- The combined regions based on torso candidate have extracted and calculate the probability again.
- Check with S_{max} if the combined area is less than that; merge the connected region to make a larger one, and then return to b). The parameters $S_{min} = \beta S_f$, $S_{max} = \gamma S_f$. where β and γ are set to 0.5 and 5 respectively in our experiments. The corresponding candidate is near to the true torso is belongs to the larger likelihood.
- Finally, torso width (t_w), height (t_h) and starting top right position (t_{rx}, t_{ry}) have been calculated.

3.3 Dress color recognition

This paper has discussed on real time automatic dress color recognition for shirt, pants and cap. The program detects person's face then a small region below of detected face is of the dress like shirt, T-shirt. Pant position is calculated by establishing geometrical relationship with face and torso position. Shirt or pant color have been recognized accurately using propose method. This can be performed even without a complex method of determining the exact contour region of a person's dress, since all of us know that the color of a region below human face, as opposed to the color of their entire shirt or body after that below of shirt position we can easily get pant color. This proposed system is also able to recognize of dress color for more than one person even they changed dresses in front of camera lively. Pant color is recognized by the same way of shirt color recognition with child torso of pant region. The algorithm for detecting dress color is given below:

Table 1. Algorithm for dress color recognition

Calculate width and height of input frame $I_{frame}(f_w, f_h)$
Initialize the child rectangular torso region.
Initialize the color confidence=1.0f
1. Take rectangular child torso from middle of the detected torso.
2. Re-assign confidence value, since a smaller region, confidence is : 0.7f.
3. Calculate width and height of child torso $I_{child_torso}(t_{cw}, t_{ch})$
4. for $y = 1 \dots t_{ch}$
for $x = 1 \dots t_{cw}$
4.1 Get the HSV pixel component (H, S, and V)
4.2 Determine color type of HSV pixel
4.3 Set HSV pixels according to color type
4.4 Keep count of these colors
End
End
5. Calculate total number of pixels in the child torso ($t_{cw} * t_{ch}$)
6. Count maximum number of pixels as v.
7. p can be defined as:
$p = (confidence) * ((v * 100) / (t_{cw} * t_{ch}))$
8. Show p (%) as final result.

Finally, the color confidence is represented in percentage. In some times peoples (or shoppers) would like to choice the same shirt, T-shirt and pant with different color that they already worn. In that particular case, system needs not to blend the new dress image over the subject body. The recognized dress colors are replaced with desired texture color. Therefore, people would be seemed that they wear desire dresses.

3.4 Automatic dress size recognition

Dress size of the subject has been recognized automatically in this proposed system. The final torso size has been used to predict the size of the dress of the subject. Fitting area has considered according to the Fig 3. The red oval area for the subject and there is a camera in front of the subject. The distance between the front camera and fitting area is around 3 meters in average has been considered. About 20 persons picture have been capture by the camera from the front of side of the subject. The heights of the people are different regarding to their age structure between 8 to 40 years. The experimental result for dress size recognition is given in Table 3.

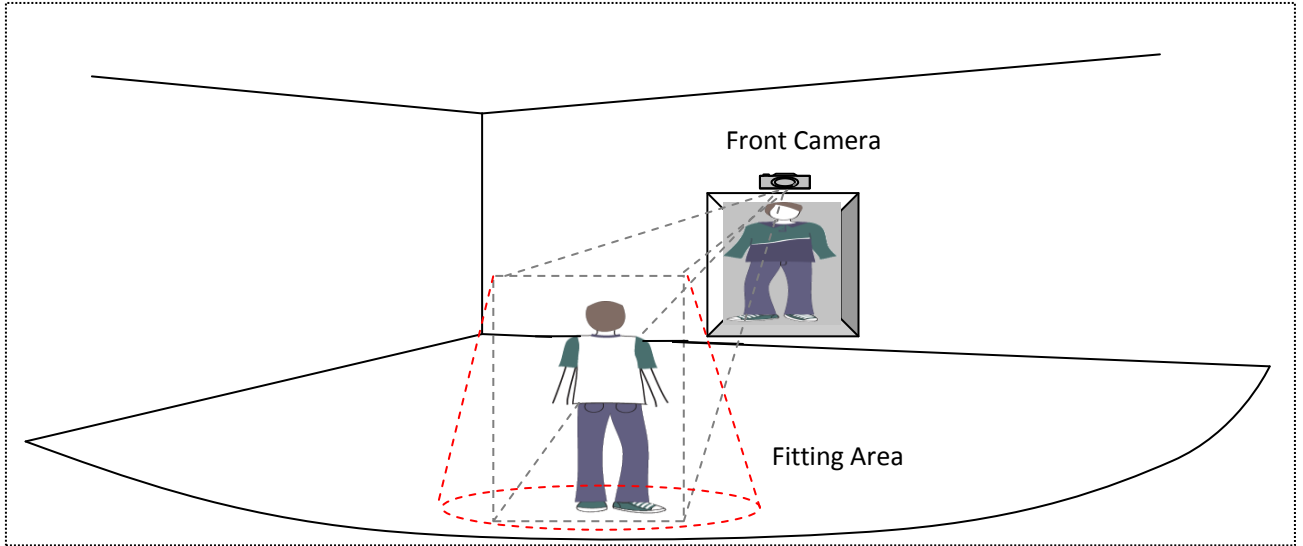


Fig 3: The concept of dress size recognition

3.5 Real time dress-up

The main issues to implement the virtual dress up system consist for real time video stream or video from input files and dress. The basic guideline is that the real time dress up system would blend the input dress images over the subject body and respond users interaction in real-time. The initial starting points is required for blending and starting points calculation procedure for cap, shirt or t-shirt, pant and female dresses are as follows: (fc_{sx}, fc_{sy}) represents the starting point of face region that has been detected by Haar classifier, fc_h and fc_w represents detected face height and width respectively. And (c_{sx}, c_{sy}) represents the starting blending point of cap, c_h and c_w are height and weight of cap respectively.

$$\text{where, } c_h = 0.7 * fc_h, \quad c_w = fc_w + 5$$

$$c_{sx} = fc_{sx} \text{ and } c_{sy} = fc_{sy} - c_h$$

The torso starting position (t_{sx}, t_{sy}) has been calculated for torso position, width and height.

$$t_{sx} = fc_{sx} - fc_w, \text{ and } t_{sy} = fc_{sy} + fc_h$$

For calculating blending starting point for shirt or t-shirt as follows: s_w and s_h are considered primary width and height of shirt or t-shirt respectively that we have calculated from geometrical relationship with face width and height as $s_w = 2.85 * fc_w$, and $s_h = 4.3 * fc_h$. Torso width (t_w) and height (t_h) are calculated using dominant color based segmentation method. Final width (s_{fw}), height (s_{fh}) and starting point for blending (s_{sx}, s_{sy}) for shirt or t-shirt has calculated according to the following equations:

$$s_{fw} = \frac{s_w + t_w}{2}, \quad s_{fh} = \frac{s_h + t_h}{2}$$

$$\text{and } s_{sx} = \frac{t_{rx} + t_{sx}}{2}, \quad s_{sy} = \frac{t_{ry} + t_{sy}}{2}$$

Consequently, we have calculated the starting position (p_{sx}, p_{sy}) of pant for blending, width ($P_w = 0.89 * s_{fw} = 2.5 * fc_w$) and height ($P_h = 1.2 * s_{fh} = 5.2 * fc_h$) for pant. All the parameters are calculated from geometrical relationship with detected face and torso position. Where,

$$P_{sx} = s_{sx} + 0.03 * s_{fw}$$

$$\text{and } P_{sy} = s_{sy} + s_{fh} - 0.25 * s_{fh}$$

3.5.1 Dress-up Algorithm

The proposed system integrates with various images processing and computer vision techniques. In the previous section, the procedures of starting position calculation for blending or re-blending have been discussed. Starting location for blending refers from where the image must be put for dress up accurately. S and D are used as blending coefficients for frame and dress images respectively. When overlaying is occurred, it is obviously replacing pixel values of frame by the pixel values of input dress image. These values are computed by multiplying S with the pixel values at the source and adding to D multiplying with the pixel values of input dress image. As we have considered the database dress images background are white so the values of pixels gather than 250 of input dress image belong to the input frame pixels. To do this, we must iterate over the entire dress image to be overlaid. According to blending algorithm in Table 2, I_{frame} refers input frame of input video file or live video stream from camera. f_w, f_h are belongs to the width and height of input frames respectively. I_{dress} refers input dress image from dress database. d_w, d_h are belongs to the width and height

of input dress respectively. p_{fxy} , p_{dxy} , and p_{mxy} refers frame, dress image and merge image pixels respectively. Algorithm for dress-up with blending and re-blending techniques is given below.

Table 2. Algorithm for blending the dress

Take input frame and calculate width and height of input

frame $I_{frame}(f_w, f_h)$

Calculate width and height of dress image $I_{dress}(d_w, d_h)$

Initialize variables s_x, s_y as starting point for dress mapping

Initialize variables $S[4] = \{0.0, 0.0, 0.0, 0.0\}$,

$D[4] = \{1.0, 1.0, 1.0, 1.0\}$ as blending coefficients

Initialize variables $p_{fxy}[4] = p_{dxy}[4] = p_{mxy}[4] = 0$

```

1.  for  $x = 1$  ..... to .....  $d_w$ 
        if  $(x + s_x \geq f_w)$  Continue;
2.  for  $y = 1$  ..... to .....  $d_h$ 
        if  $(y + s_y \geq f_h)$  Continue;
         $p_{fxy} = \text{GetPixel}(I_{frame}, y + s_y, x + s_x)$ ;
         $p_{dxy} = \text{GetPixel}(I_{dress}, y, x)$ ;
        for  $i = 1$  ..... to ..... 4
            if  $(p_{dxy}.val[i] > 250)$ 
                 $p_{mxy}.val[i] = p_{fxy}.val[i]$ 
            else
                 $p_{mxy}.val[i] = S.val[i] * p_{fxy}.val[i] + D.val[i] * p_{dxy}.val[i]$ 
        End
         $\text{SetPixel}(I_{frame}, y + s_y, x + s_x, p_{mxy})$ 
    End
End

```

The same blending and re-blending procedure for pant and cap images respect their blending starting position have been applied. Starting position of pant is calculated with a geometrical relationship between torso and face position. And cap blending position is calculated according to the position of face in an input frame. Above proposed algorithm or function is useful since the “cvAddWeighted” does not have location parameter. Especially, when overlay input image is much smaller compared to source image.

$$I_d(x, y) = \alpha * I_1(x, y) + \beta * I_2(x, y) + \lambda \quad \text{OpenCV}$$

provides cvAddWeighted ($I_1, \alpha, I_2, \beta, \lambda, I_d$). The function computes the weighted sum of two arrays. The overview of mentioned function parameters: $I_1(x, y)$ the first source array, α - weight for the first array elements,

$I_2(x, y)$ - the second source array; must have the same size and same type as $I_1(x, y)$, β weight for the second array elements, λ - scalar, added to each sum, $I_d(x, y)$ - The destination array; it will have the same size and same type as $I_1(x, y)$ and $I_2(x, y)$. According to this function, user should provide same size of input images. As there is no starting position for overlay images, so it would be started from (0, 0) position automatically. Therefore, this function is not useful for this proposed system because of its positioning problem. On the hand, OpenGL provided one function name glTexSubImage2D for texture mapping with one image over another. The main limitation of this function for dress up system is starting position for texture mapping on the body surface. According to OpenGL, starting position of the function is from down left position on the texture image. Therefore, it is very difficult to adjust position the dress image texture over the subject body texture.

3.6 Result and discussion

This proposed method has been tested on indoor video files from computer as well as real time video. In this experiment, we have considered the frame rate 30 frame/second for a 640X480 video on Intel® Core™ 2Due CPU E7200 @ 2.53GHz (2CPUs) desktop with 4096MB RAM, which can meet the real-time requirement. The canon IXUS 110IS digital camera is used to capture video. On the other hand for live video streaming, PCM-101 is used which is with 35Mega Pixel CMOS Image Sensor MAX 45Mega Pixel, active pixel size 8mmx8mm, and frame rate: VGA (15fps), Cif (30 fps), view angle: 52 and memory recommend 64Mb. This proposed system has been implemented in Visual C++ using Intel's Open Computer Vision Library (OpenCV).

Table 3. Dress size recognition

Height in inches	Distance from camera to subject in meters	Average height of Torso(pixels)	Size of T-shirt/ Shirt in centre meters
48~58	3	108	90~95
59~65	3	125	100~110
65~+	3	137	110~+

3.6.1 Discussion on Database

The image database is a collection of digitized images. It is maintained primarily to support research in image processing, image analysis, and computer vision. The database is divided into volumes based on the basic character of the dress images. The proposed system database contains more than hundreds various sizes images such as average 100x100 pixels for cap images, average 100x200 pixels for pant images, average 200x300 pixels for shirt or T-shirt images and finally average 260x400 pixels for female dresses. Most of the database images have been collected from websites specially [25]. Some of experimental database images are given here as example:



Fig 4. Example images in database

3.6.2 Advantages of this Proposed System

The previous proposed anthropometric measurement based system is not realistic as well as user friendly. 3D digital scanner based proposes dress up system is more complex, costly and require more space to setup. Avatar based virtual dress up system is not able to show the original user images with dress, they created user cartoon as a model and dress up over the cartoon. Actually, avatar based system is simulation and modeling technique. Recently, one manual dress up system is proposed only for static model and dress images [1]. They are used some supporting software's in their proposed system to process the input dress images as well as model images manually. To establish the necessary correspondences between the model and dress images, they selected the joint positions on the input image manually. Moreover, the skeleton is fixed on the model image manually. The system also required the user height, weight, top, and bottom size. For doing all the mentioned steps, the proposed image based system required processing time around 17 minutes on average that shown in Table 4. As this proposed system is real-time system, so user could visualize their dressed up output lively on the computer screen.

Table 4. Time complexity of existing system

Experiment no	Purposes	Require time (m)
01	Photo shoots	3
02	Alpha Map (Soft Scissors)	3
03	Alpha Map (Hand Made)	10
04	Skeleton Setting	1

Additionally, this proposed system is pose invariant in the sense because user could take any type of posed dress image as input image and if user gives the same pose in front of camera lively then the input dress will be automatically adjust over the subject.

3.6.3 Limitations of Proposed Dress up System

Although the current system yields good results, there is still in room for improvement. The main limitation is that the proposed system is not pose invariant in real sense. Another reason, some parts of worn cloths is appeared in the system outputs images. If the body direction or appearance is different respect to the input dress images then system show those kinds of results. Moreover, after dressing up over the exiting dresses should be disappeared of the uncovered part of worn dress. The real body skin color should be appeared but our system does not provide that kind of result. In the future, we would try to implement this concept for recovering this limitation. The presented torso detection using video frame is quite robust and gives reasonably accurate results, but it has still some limitation. The use of color-distribution histograms to distinguish between the clothing and rest of the image means that torso detection will fail in images where the colors of the shirt and background are very similar, or if the user wears clothing that is too close to their skin color. The torso detection also give less-than optimal results in cases where the user's shirt contains multiple dominant colors, or is sharply divided into different-colored regions by shadows.

3.6.4 Performance Evaluation

To evaluate the performance of torso detection algorithm, we set up an image database with single and multiple people in a single image. We have also analysis the performance of our proposed torso extraction method for around 30 persons. The center point of the torso is the most important. Center position $P(x, y)$ of the torso has been considered. Ground truth, position of the center position P_0 of the torso is marked manually for each person in single photo. We have applied the proposed torso detection method on the input images and got a new center point according to the detected rectangle. The distance between the center position of the detected torso and the true one is measured by D_p , where D_p is the Euclidean distance between true center position and the detected torso center, and it is normalized by the width of face. For the most likely candidate of the 80% of photos, D_p is less than 0.2 and 90% of photos D_p is less than 0.5. Actually, this kind of variation of the distance of the center points for different types of pose in the considered images.

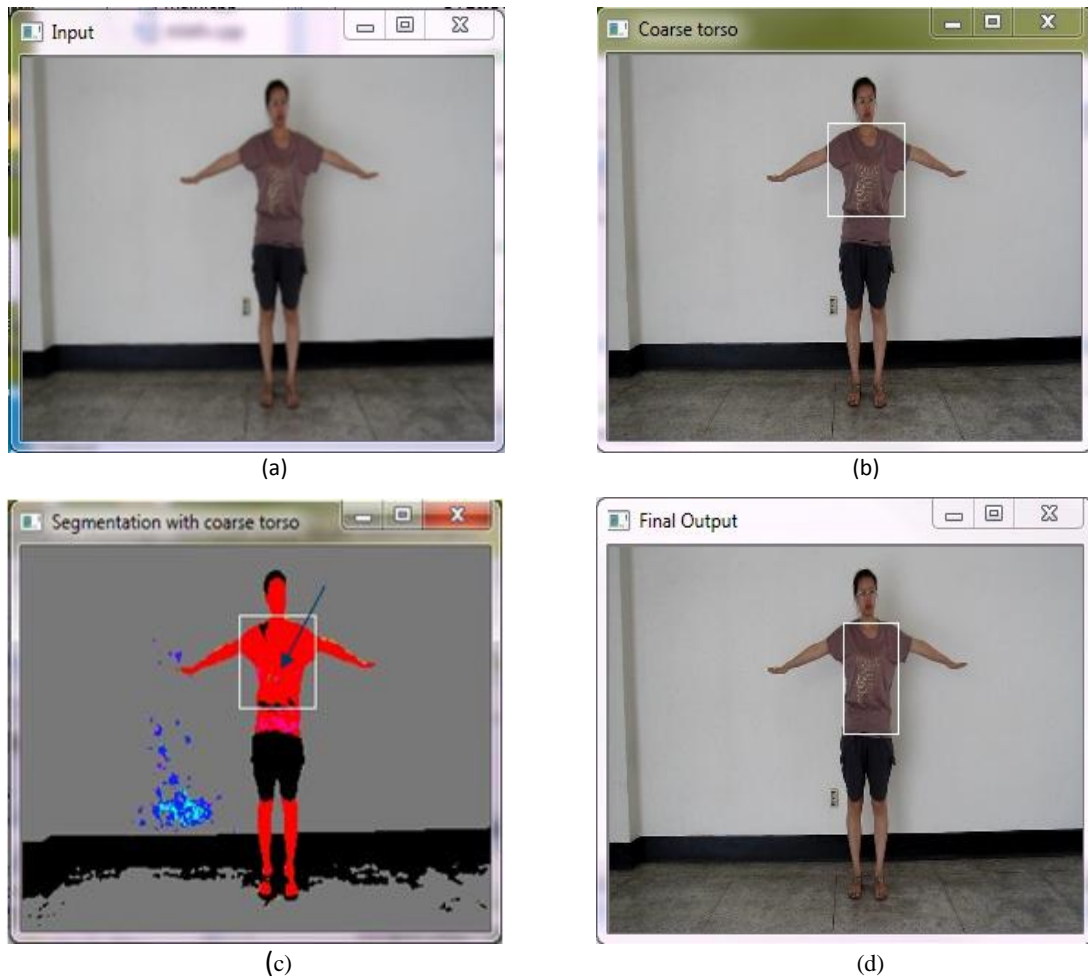


Fig 5: The whole process of the torso detection(a) Input frame; (b) the correspondence frame with the white rectangle is the boundary of the coarse torso region; (c) the pixel regions in coarse predicted torso regions; (d) the white rectangle is the most likely torso candidate extracted based on the most probability color.

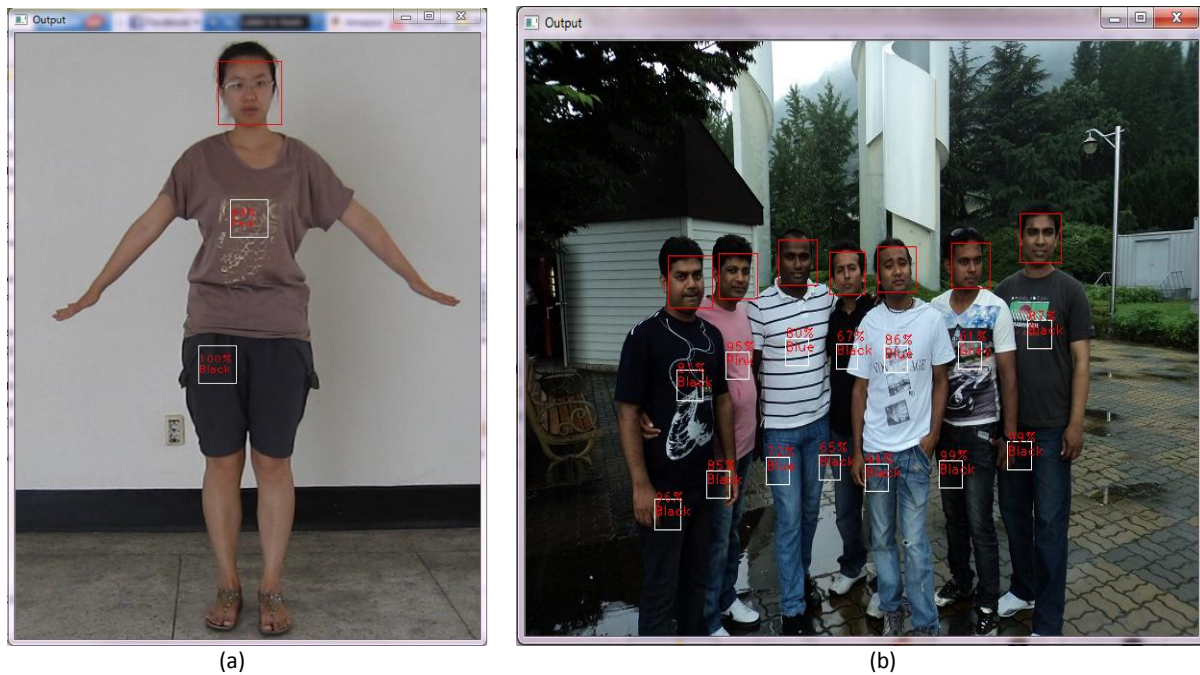


Fig 6: Dress color recognition (a) Dress color recognition with percentage for single person (b) System output for multiple people

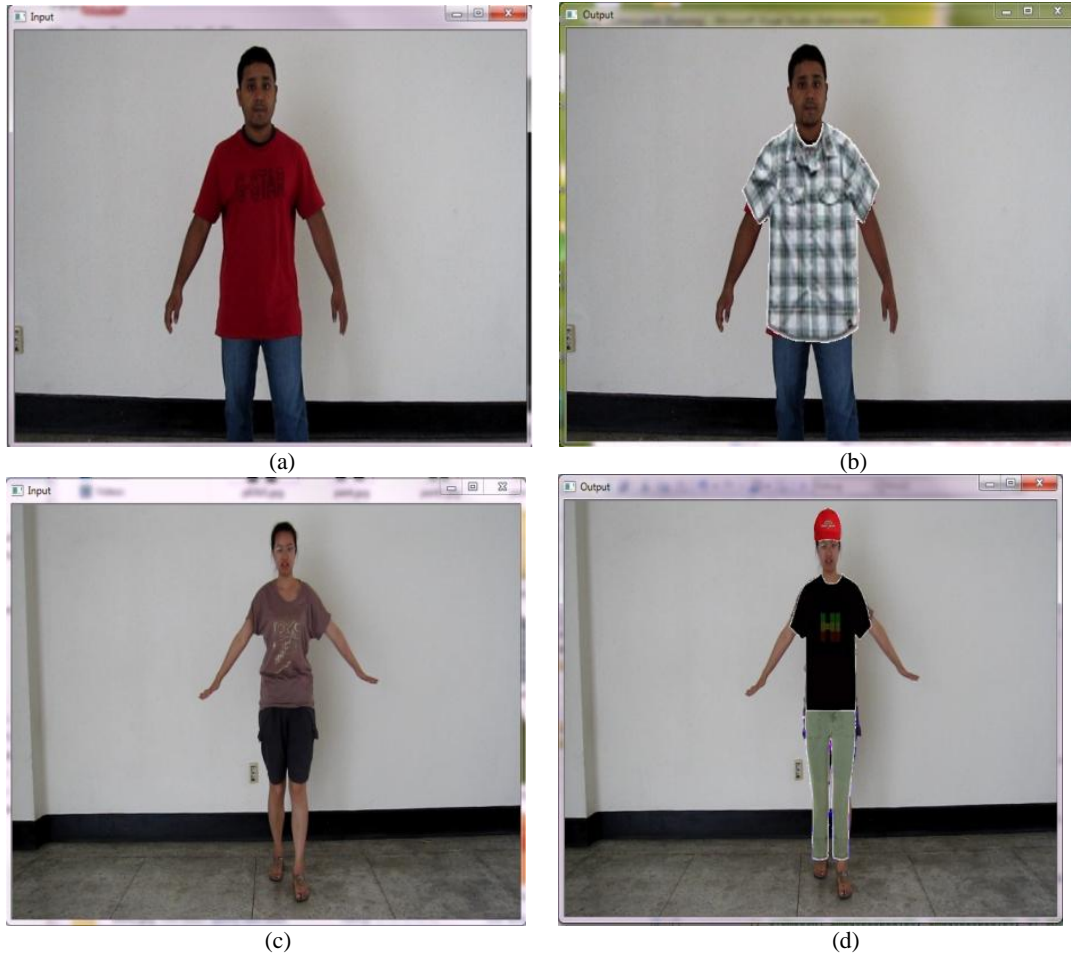


Fig 7: Correspondence input and output frames (a) Input frame (b) Dressed up output with T-shirt , (c) Input frame (d) System output with Cap, T-shirt and Pant



Fig 8. Proposed system outputs for female with different dress images.

A video stream with single person is considered from our lab database to evaluate the performance of mentioned main two steps of this proposed system. For this case, we have experimented with input video around 220 frames with different types of pose in video. For calculating standard deviation of processing time, 20 frames have been skipped during the calculation. The standard deviation of require time for each step like face detection, segmentation or torso

detection for (640x480) size frames have been calculated. Now standard deviation of require time of this propose real time dress-up system for individual input dress image as well as combine input dress images is given below. Shirt or T-shirt, cap, pant and female dresses are considered for calculating standard deviation respect input frames as processing time. According to this system, standard deviation for a complete dress up with shirt and pant is 1.452 seconds. Moreover, the

dress color detection process only takes a few milliseconds to execute (depending on the image size).The require time for shirt/T-shirt(200X300), pant(100X200), cap(80x80) and female dress(260X400) are 0.952, 0.5, 0.25 and 1.55 seconds respectively. Therefore, the total interaction time as standard deviation of this proposed system is 2.485 seconds including face detection, segmentation, torso detection and dress up with shirt or t-shirt.

4. CONCLUSION AND FUTURE WORKS

This paper has been presented an efficient and affordable method for real time dress up system. The method consists of two major steps: image registration and dress up. There exist many benefits from this proposed real time dress up systems for customers, shop owners and companies, such as removing don and doff time, space saving and reducing wasting cloth tried on. Additionally, the previous existing methods such as image processing based system required more time around 17 minutes for each person, 3D scanner based system has high costs and under physical space issues, Contrary to previous works, this proposed real time system only require 2.485 seconds as standard deviation of 200 frames skipped by 20 during process for dress up of single person with shirt or T-shirt. Moreover, it does not require the physical space and it is much easier to use. Therefore, system has used an automatic scaling function to resize the input dress images respect to torso size lively for blending the shirt or T-shirt accurately over subject body. The real time dress up system can be brought into home and used for on-line dress shopping. Current on-line shopping malls are mostly picture-based and customers are unsure of accurate sizing of dress. This virtual dress up system can solve the sizing problem by having virtual trying on dress instead of physical one. It also makes people easier to choose dress perfectly within a short time. Finally, experimental results are demonstrated this proposed method is accurate and reliable to solve the promising and challenging real-time automatic dress up system

5. REFERENCES

- [1] Jong-Chul Yoon, In Know Lee and Henry kang published one paper on this topic named "Image-based Dress-up System" in ICUIM- 2011.
- [2] Nadia Magnenat-Thalmann, H. Seo, F. Cordier,"Automatic modeling of virtual humans and body clothing", Proc. 3-D Digital imaging and modeling, IEEE Computer Society Press 2003
- [3] J. Chai, Jessica K. Hodgins,"Performance animation from low-dimensional control signals", SIGGRAPH 2005
- [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] Karla Peavy Simmons, Body measurement techniques: A comparison of three-dimensional body scanning and physical anthropometric methods, thesis for the degree of Ph.D in North Carolina State University, 2001.
- [6] Paquette, S., Brantley, J.D., Corner, B.D., Li, P., Oliver, T., Automated extraction of anthropometric data from 3D images, U.S. Army Soldier Biological & Chemical Command, Natick, MA, 1998.
- [7] Paquette, S., 3D scanning in apparel design and human engineering, IEEE Computer Graphics and Applications, 16 (5), 11-15, September, 1996.
- [8] Hwang, Su-Jeong, Three dimensional body scanning systems with potential for use in the apparel industry, thesis for the degree of Ph.D in North Carolina State University, 2001.
- [9] Hein A.M. Daanen, G. Jeroen van de Water, Whole body scanners, Displays 19(1998) 111- 120.
- [10] 3D Scanning methodologies for internationally compatible anthropometric databases, draft international standard ISO/DIS 20685, 2004.
- [11] General requirements for establishing an anthropometric database, draft international standard ISO/DIS 15535, 2001.
- [12] XinJuan Zhu, Xin Zhang, Jing Qi, Communion system of 3D anthropometric databases based on internet, China Inventing Patent, ZL200510096300.5.
- [13] K. Jeong, S. Jang, J. Chae, G. Cho, and G. Salvendy. 2008. Use of Decision Support for Clothing Products on the Web Results in No Difference in Perception of Tactile Sensation Than Actually Touching the Material, International Journal of Human-Computer Interaction, 24(8), 794-808.
- [14] Sun-Young Lee, Jong-Chul Yoon and In-Kwon Lee, 2007, Cartoon-like Video Generation Using Physical Motion Analysis, NPAR 2007 Poster, 2007.
- [15] Kenji Hara, Ko Nishino and Katsushi Ikeuchi, 2003, Determining Reflectance and Light Position from a Single Image Without Distant Illumination Assumption, ICCV, Proceedings of the Ninth IEEE International Conference on Computer Vision, Vol 2, p. 560.
- [16] Dmitri Bitouk, Neeraj Kumar, Samreen Dhillon, Peter Belhumeur and Shree K. Nayar, 2008, Face Swapping : Automatically replacing faces in photographs, International Conference on Computer Graphics and Interactive Techniques, ACM SIGGRAPH 2008, Article No. 39, p. 1-8.
- [17] G. Hua, M.H. Yang, and Y. Wu, Learning to estimate human pose with data. driven belief propagation, Proceedings of international Conference on Computer Vision and Pattern recognition, San Diego, 2005, pp.747-754.
- [18] X. Ren, A.C. Berg, J. Malik, Recovering human body configurations using pair wise constraints between parts, proceedings of International Conference on Computer Vision, eijing, 2005, pp.8
- [19] G. Mori and J. Malik, Estimating human body configurations using shape context matching, In European Conference on Computer Vision, Copenhagen, 2002, pp. 666–680.
- [20] G. Mori, X. Ren, A. Efros, and J. Malik, Recovering human body configurations: Combining segmentation and recognition, In Proc. IEEE Conf. on Computer Vision and pattern Recognition, Washington, 2004, pp.326–333.
- [21] P. Viola, M. Jones, Robust real-time face detection, Int. J. computer Vision, 2004, 57(2): pp.137-154.

- [22] Q.X. Ye, W. Gao, W.Q. Wang, and T.J. Huang, A color image segmentation algorithm by using color and spatial information. *Journal of Software*, 2004, 15(4): pp. 522-530.
- [23] NIST. Anthrokids - anthropometric data of children, <http://ovrt.nist.gov/projects/anthrokids/>, 1977.
- [24] M. W. Lee and I. Cohen, Proposal maps driven MCMC for estimating human body pose in static images. In *Proc. IEEE conf. on Computer Vision and Pattern Recognition*, Washington, 2004, pp. 334–341.
- [25] Most of the database images have been collected from : <http://www.abercrombie.com>