OVERLAY UPPER CLOTHING TEXTURES TO STILL IMAGES
BASED ON HUMAN POSE ESTIMATION

QUANG-SONG LE

Cao Thang Technical College

University of Science, VNU-HCM

song.lequang@gmail.com

MINH-TRIET TRAN

University of Science, VNU-HCM

tmtriet@fit.hcmus.edu.vn

The authors propose a system to enhance user's experience in virtual shopping with image processing techniques. Our proposed system allows a user to submit his or her upright photo taken by any regular camera and choose a product interactively, and then the system returns an output image, which is morphed with upper clothing textures corresponding to the human pose estimated from the given photo. Our approach is based on the result of estimating upper human pose to calculate similarity scores between pairs of image poses and our simple yet efficient method to scale a texture image of an upper clothing product to an appropriate size for morphing. From our study on various magazines and websites of fashion, we define 16 common types of human poses. We create a dataset with 390 photos taken in studios for many products in various poses which cover the 16 defined common poses to train and test our proposed virtual dressing room. Experiments with 12 volunteers to evaluate 150 result images from 50 user images in our system show that 89% of the overlaid output images of the system have good evaluation results. Our proposed system not only enhances users' experience but also helps users save time with a virtual dressing system before they decide to purchase upper clothing products.

Keywords. Articulated human pose estimation, virtual clothing, texture mapping, morphing.

1. Introduction

With the development of e-commerce, there are more and more online fashion shops. To enhance users' excitement and to attract their interests, different utilities and means of interaction have been integrated into online fashion websites. Customers can freely preview, vote, write comments, or share their favourite collections with friends via utilities of online fashion websites. Furthermore, one of the innovative features of an online fashion shop is to provide customers with the virtual dressing room where customers can virtually try on as many clothing products as possible.

A virtual dressing room can be more convenient for a customer than a physical in-store dressing room because that customer does not have to wait in a long queue especially during sale seasons. Furthermore, each customer might be advised to limit the number of items to try on in a real shop. Virtual dressing rooms are investigated to save time and bring more freedom to customers. In a virtual dressing room, customers can try as many items as they want, take photos, save them, and/or ask for

advice. This special feature can also provide relevant information to customers on items and help customers evaluate the fitness with advanced image processing techniques.

There have been various solutions to develop virtual dressing rooms. One typical approach is to use 3D virtual models to create meshes and coated meshes [1] [2] [3] [4]. In fact, creating 3D models from multiple 2D images taken from cameras is challenging. Kwang-Jin Choi and Hyeong-Seok-Ko [5] propose a clothing simulation model for 3D human characters with realistic motion. However, this method can only be performed on virtual environments for characters' clothing in games or animations.

Other approaches take advantages of depth camera technologies [6] [7], e.g. Kinects, to scan the 3D structure of a human body for virtual try-on applications. These approaches are heavily based on hardware devices and only suitable for big malls. An approach based on images is proposed in [8] [9]. This approach is difficult to implement due to these systems need more devices. Some approaches are similar to our proposal which processes the dressing by using input images and human poses [10] [11] [12]. But these methods have certain limitations such as the user must manually make marks and draw the skeletal structure of the image model.

In this paper, we propose a method morphing upper clothing texture based on upper human pose estimation – virtual dressing room. With our technique, we overcome the above limitations by simply allowing users use all types of input images which do not have the required specific and without requiring specialized equipment to take pictures.

Our propose system is based on upper human pose estimation [13] from a image submited by user and then morphing clothing product texture on this image. We conducted the survey to propose eligibility criteria and select common poses for dressing.

The main content of this paper is as follows. In Section 2 we summarize the related works. Section 3 we present our proposed system. Experiments are presented in Section 4. In the last section we present conclusions and proposing possible extensions to our work.

2.   Background and Related Works

2.1   Clothes Reconstruction

There are many virtual dressing applications rendering textured cloth meshes over a camera image. Obtaining that mesh is the key feature of such systems. Some approaches use existing CAD models[1]. When models are not available, they have to be reconstructed. Capturing clothes from a video sequence is a hard task. Many approaches use markers on the cloth for capturing, which makes them less suitable for our application. More recent approaches do not require markers [2] [3] [4]**.** However, all approaches that rely on point correspondences that are computed from the image data assume a certain texturedness of the garment. Kwang-Jin Choi and Hyeong-Seok-Ko [5] proposed a simulation model clothing on characters with realistic motion. The strength of this approach is the creation of the movement of the piece and lightly coat requires sensitivity reactions clothes. This approach can only be performed on virtual environments for animated character clothes.

2.2 Using Kinect and HD Camera

S.Giovanni et al. [6] proposed Kinect and HD Camera Kinect devices allows a user stand in front of a screen, and select an item from the list of clothes to try on. A similar approach based on Kinects is demonstrated in another application [7]. This proposed approach is mainly based on the extraction of a user from video stream, alignment of models, and skin color detection. Authors use the 3D locations of the joints for positioning, scaling, and rotation in order to align 2D cloth models with the user. Then, authors apply skin color detection on video to handle the unwanted occlusions of the user and the model. Finally, the model is super imposed on the user in real time. These methods need to invest more hardware devices and to be deployed in a big shopping center.

2.3 Overlay clothes based on image

Anna and Peter Eisert Hilsmann [8] used a virtual mirror with a camera to take a picture for their system that allows a user change the color and texture while maintaining the distortion of posture wearing a T-shirt in human. Their method is limited to the color, texture and just simply displays the results for a particular T-shirt only. Hauswiesner et al. [9] used the approach based on an image by setting many cameras to try on clothes. They use available data sets recorded in the database and implement methods of tracking the user's current position, query position available in the database, then morph clothes the current user through images captured from the camera and recreate the 3D model.

2.4 Overlay clothes based on image and human pose

Ehara and Saito [**2**] propose a texture overlay method for a virtual clothing system. The deformation of the shirt's surface is estimated from its silhouette in an input image based on a pre-collected database of shapes of the same shirt. Their system needs homogeneous background and markers on shirts. Unfortunately, they do not take illumination into account. As an extension of the work [10], Hiroshi Tanaka et al. [11] describe properly the shape of the contour of the clothing. Therefore, the shape on the surface of the clothing is estimated correctly even when the hands occlude the clothing. In addition, the computation time is reduced by performing PCA and K-means method, which enables the system to work in real time.

The above methods still have the limitations, such as the user must manually make marks, create alpha map images, choose shape border in an image, and draw the skeletal structure of an image model.

3. Proposed Method

3.1 Overview our proposed method

Figure 1 illustrates our proposed method with the following steps. First, the system provides a user interface for a user to select and submit a user image to the system of virtual dressing room. Second, we apply the work of M. Eichner et al. [13] to estimate the upper human pose in the input image provided by the user. Details of the estimated upper human pose are presented in section 3.2. After human pose estimation, the system searches and compares the similarity between the user image's pose and the poses available in the data set. We perform the normalization parameters, define metrics

to determine the similarity and the selected pose closest users provide. Details are presented in the section 3.3. Next, in section 3.4 we present our method to calculate values r   elated to product scales and morphing product texture to user image. By the end of this section, we present our virtual dressing room application.
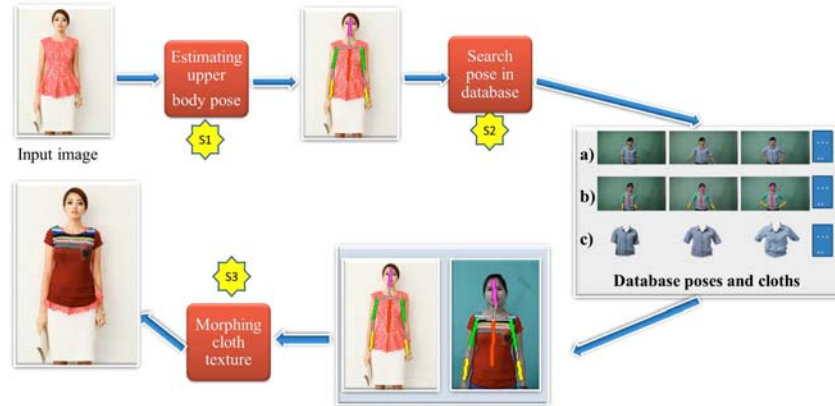

Figure 1 Overview of our proposed method

In order for our proposed system to be widely applied in practice and suitable for many different poses, we conduct a survey to choose suitable poses to prepare data for dressing. The image morphing can be deformed, do not reach the high quality of resolution and size, so, we surveyed the common poses in fashion shows or dressing models in fashion magazines and fashion e-commerce sites. Details of our survey are presented in Section 4. From the results of our survey, we choose 16 common poses that cover 98% of common possible poses. Figure 2 shows our process of building test data.
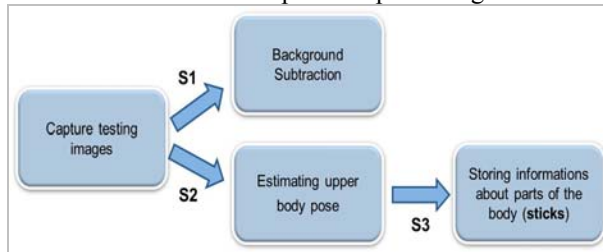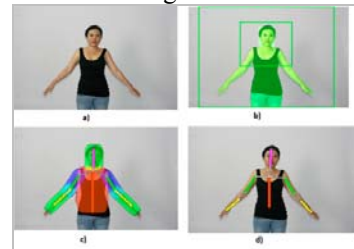

Figure 2 The process of building test data

Figure 3 Estimating upper body pose.

3.2   Step 1: Overview of upper human pose estimation in image

Before presenting the application of ***upper human pose estimation*** in our process, we present the overview of the method proposed by M. Eichner et al. [13]. Their work consists of the following steps:
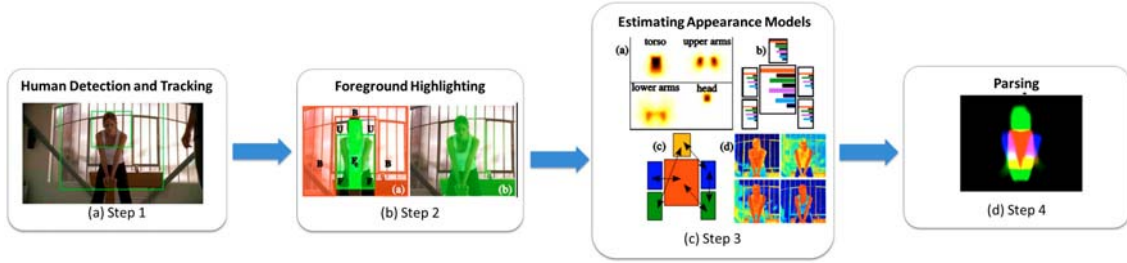
Figure 4 Model upper human pose estimation [13]

According to the model by M. Eichner et al., a user can submit any image taken by a regular camera. The output of the model is to determine the pose of human joints in the image.
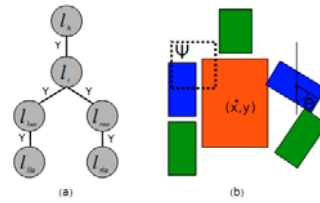

Figure 5 Pictorial Structure model. Each node represents a body part (h: head, t: torso, left/right upper/lower arms lua, rua, lla, rla). (a) The kinematic tree includes edges between every two body parts which are physically connected in the human body. (b) Cardboard representation where body parts are rectangular patches, parameterized by location (x, y) and orientation θ, connected by kinematic priors Ψ.

The general framework of ***pictorial structures*** for human pose estimation is as follows. A person's body parts are represented by a conditional random field. Typically, each part l_i is a rectangular image patch parameterized by the location (x, y), orientation θ, and scale s. The posterior of a configuration of parts L= $\{l_i\}$ given an image I, $\gamma(l_{head})$, $\gamma(l_{torso})$ requiring the orientation of the torso and head to be near-vertical. $\gamma(\cdot)$ gives the uniform probability to a few values of θ around vertical, and zero probability to other orientations. This reduces the search space for torso and head, thus improving the chances that they can be correctly estimated.

$$P(L \mid I) \propto \exp\left(\sum_{i,j \in E} \Psi(l_i - l_j) + \sum_i \Phi l_i + \gamma(l_{head}) + \gamma(l_{torso})\right) \qquad (1)$$

In [13], upper body Pictorial Structures model consists of 6 body parts, namely torso, head, upper and lower arms connected in a tree structure by the kinematic priors $\Psi(l_i, l_j)$ (c.f. Figure 5a). This method is based on the model proposed in [14] (c.f. Figure 5b) and extended with orientation priors described next in this section. We also reduce the spatial extent of the kinematic prior. Detailed steps are as follows:

Step 1: Human detection and tracking (Figure 4a). This is an important step in upper human pose estimation to limit the region of interest ROI in the image. They start with the approach of [15], where each examined window is subdivided into tiles described by Histogram of Oriented Gradients (HOG) and classified using a linear SVM. Next, they investigate the improvement brought by apart-based

hierarchical extension of [15] proposed in [16] (PBM). Finally they check whether complementing an upper-body PBM with a face detector [17] improves performance.

Step 2: Foreground highlighting (Figure 4b)**.** After step 1, in order to remove the background image, the pixels is not necessary in human pose estimation in expanding window, reducing the search space to improve performance, M. Eichner et al. has done foreground highlighting in expanding window by using GrabCut.

Step 3: M. Eichner et al. describe a mechanism for estimating good image and person-specific appearance models from a single image based on two observations: (i) certain body parts have rather stable location w.r.t. the detection window; (ii) often a person's body parts share similar appearance. This mechanism is then used to compute appearance models specific to new instances found in step 1 (Figure 4c). **Step 4**: Parsing: An articulated pose is estimated by running inference (Eq. (3)) with person-specific appearance models (computed in step 3) and generic appearance models (edges). The area to be parsed is restricted to the region output of foreground highlighting. The output of the parsing stage is the posterior marginal distribution $P_i(x, y, \theta)$ for every body part of each person detected in the image.

We utilize the algorithm for ***upper human pose estimation*** [13] into to our process. From an original image provided by a user (Figure 3a), the system performs the detection and tracking human face and shoulders (small blue rectangles in the Figure 3b), then this rectangle will expand to cover the entire upper body, the next step, the system will foreground highlighting. In Figure 3c, posterior foreground probabilities are then used to derive a color soft-segmentation of the image for each body part. Figure 3d is the result of the upper human pose estimation.

3.3   Step2: Normalization and definition measurement of similarities between the poses

Output of the upper human pose estimation [13] is described as follows: sticks (:, n) = **[x_1 y_1 x_2 y_2]** where each stick representing each part consisting 2 points with coordinates **(x_1, y_1)** and **(x_2, y_2), n** represents the upper of the body, 1 – torso, 2 – left upper arm, 3 – right upper arm, 4 – left lower arm, 5 – right lower arm, 6 – head. (c.f. Figure 6)
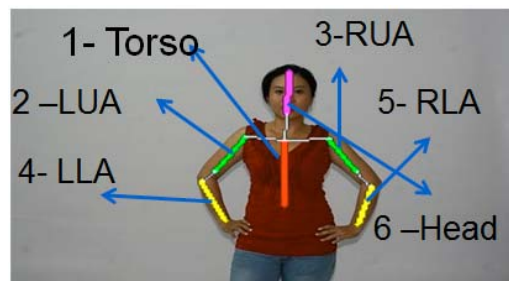

Figure 6 Output sticks

According to the author's algorithm: Stick Coordinates defined from the left on the original image, so they depend on the origin of the image and can change when scale or crop. From that, ***we proposal features consist of 6 vectors***: Torso, Left upper arm, Right upper arm, Left lower arm, Right lower arm

and Head (Figure7). ***Advantages of these are independent of the size and position of the person appearing in the photo.***
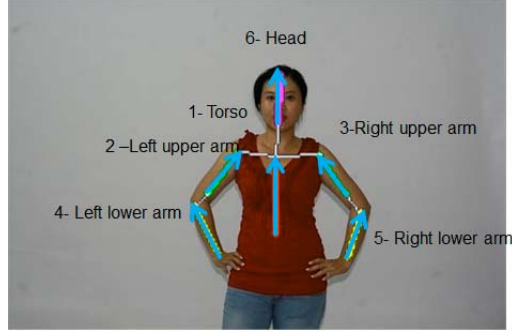

Figure 7 Proposed vectors

To calculate the similarity between poses, first, we normalize vectors of sticks. From 2 point of a stick of a part, we normalized to original coordinates ***O (0, 0)*** of the image as follows: Set start and end of a stick are ***point$_A$ (x_1, y_1)***, ***point$_B$ (x_2, y_2)***, so, after normalized we have a ***point$_{Norm}$ (x_2-x_1, y_2-y_1)***. Applying this method, the system will normalize all stick of parts to the corresponding point. To make the comparison of similarity between test image and training images in the database, we normalized the sticks of test image and calculate the similarity metric between the test image to each image in the data set and find image corresponding best.

We use the Euclidean (Eq. (2)) metric to perform the calculation similarity between the test image pose and the pose of the images in the data set. The system will loop over all upper body to calculate similarity via Euclid formula:

$$d(p,q) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + ... + (p_n - q_n)^2} = \sqrt{\sum_{i=1}^{n}(p_i - q_i)^2} \qquad (2)$$

p and ***q*** are 2 point sticks have been 2 normalized corresponding of 2 parts between test image and each image in the data set. From these measurements, the system will select the best metric to use.

3.4   Step 3: Uniform scale product image and overlay cloth texture to image

After comparing and calculating the similarity between test image and the images in the data set, the system will select product image with best metric to morphing. After selecting the product which removes background, requirements we need to know the scale, position of the human upper body of test images submitted by users to morphing this product image to test image. We proceed morphing as follow (Figure 8):
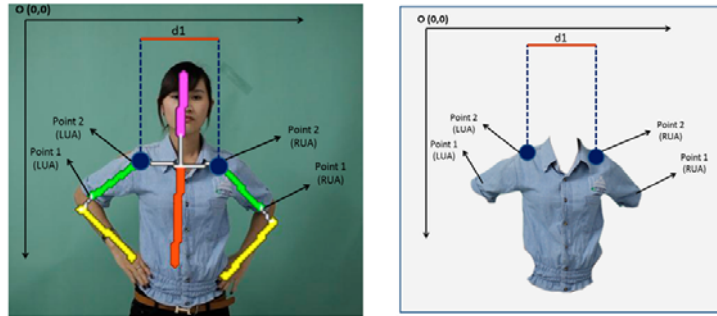
Figure 8 Calculation uniform scale of product image

In the process upper human pose estimation for image product dataset, besides removing the background image; our system will save pose sticks for each image. Based on these stick to the system can be mapped onto the product images and calculated the following: first, we calculate original distance d_1 between two points of stick LUA (Left Upper Arms) and stick RUA (Right Upper Arms) on the product image (the background was separated): $d_1 = x_{pt2\ (RUA)} - x_{pt2\ (LUA)}$, original product image **scale $(s_1) = h/w$**, similarly, we will perform the calculations distance d_2 between 2 stick on the image test submitted by user was previously estimated, then, we calculate new size of product image **w'(new width)= w* d_2 / d_1, h' (new height)= w' * s_1**. Next, we determine the location on test image to morphing overlay texture image product, this position coordinates **$x = x'_{pt2\ (LUA)} + d_2/2 - w'/2$, $y = y_{pt1}$ (Head) + Abs ($y_{pt1\ (Head)}$ - $y_{pt2\ (Torso)}$)/4** ($x'_{pt2\ (LUA)}$ is X coordinate of LUA, **y** coordinate of point 1 **Head** stick on test image). After determined the location and scale of product image from test image, the system will precede morphing product image to the correct location on the test image. (Figure 9)

Figure 9 Calculation process

3.5 *Virtual dressing room application*

We built virtual dressing room for our method. The application consists of two parts, the left side to upload the users' images and the right side is a list of products to try. When users click on any product to the right, the system will follow the steps suggested above and morphing to user image on left.

Figure 10 Virtual dressing room application

3.6 The process of developing test data

We built test data according to the steps described in Figure 2. First, we shooting 350 pictures in the studio for more products with various poses cover 16 postures. After shooting, in **step 1**: We did background subtraction from these images to obtain the product images. The pictures of products are used for the morphing texture as described in step 3 (see Figure 11).

Figure 11 Product with background subtraction

Figure 12 Upper human pose estimation

In **step 2** of this process, we proceed to upper human pose estimation [13]. After running this step, we obtained sticks of parts of body (see Figure 12). In **step 3**, we store these stick and the name of the corresponding image to calculation similarity. Based on, we search in database to choose the closest corresponding products to morphing.

4. Experiments

4.1 Experiment 1: Survey and analysis to select the appropriate poses while performing the operation fashion shows and dressing

The goal of this section is to identify common poses for our system. We conduct the survey on 10 fashion magazines and fashion e-commerce sites with 269 samples (see **Table 1**)

No	Website	# samples	No	Website	# samples
1	Magazine: tapchithoitrangtre.com.vn	25	6	Magazine: vmode.vn	32
2	Magazine: dep.com.vn	22	7	e-commerce: sendo.vn	27
3	Magazine: 24h.com.vn	16	8	e-commerce: vnwiz.com	20
4	Magazine: tapchi365.com	34	9	e-commerce: yes24.vn	27
5	Magazine: thoitrangso.vn	30	10	e-commerce: zalora.vn	37
Total			269 samples		

Table 1 Number of samples from magazines and fashion e-commerce websites

After the statistical process, we identify that there are 16 main poses that cover approximately 98% poses from fashion magazines and e-commerce websites (see Table 2).

Pose	Image	%	Pose	Image	%	Pose	Image	%
Pose 1		2%	Pose 2		3%	Pose 3		4%
Pose 4		5%	Pose 5		6%	Pose 6		12%
Pose 7		2%	Pose 8		4%	Pose 9		14%
Pose 10		3%	Pose 11		5%	Pose 12		9%
Pose 13		11%	Pose 14		12%	Pose 15		4%
Pose 16		2%	**Others**		**2%**			

Table 2 Statistics for each pose in collection

Based on the statistics from Table 2, we decide to choose the 16 main poses to support fitting for users. Synthetic poses are shown in Figure 13.

Figure 13 List of poses supported by our system

4.2 Experiment 2: Evaluate the accuracy of the method proposed by M. Eichner et. al [13] to our data set

The data preparation process is illustrated in **Figure 17**. Our test data was taken in our studio **Figure 17**a) with the poses supported as described in section 4.1. Our data were taken on 3 models, with a total of 350 samples taken (**Figure 17**b). The proposed morphing method based on pose estimation system should be able to support any type of top patterns (T-shirts, shirts, short sleeves, long sleeves ...). After the shooting, we did background subtraction for products (Figure 13c) and also ran the algorithm by Eichner, M *et al* [13] proposed (**Figure 17**d). After running the algorithm [1], we obtained the stick (**Figure 17**e), we store these results to calculate similarity and morphing texture cloth. Through our statistic, experimental result of the algorithm proposed by M. Eichner *et al* [13] with our test data is 96, 45%. However, some samples when applying the method of [13] are wrong pose estimation, see Figure 18.

Wrong pose

Figure 14 Experimental evaluation of the accuracy Eichner, M in our data set

4.3 Experiment 3: Applying actual fitting to image provided by the users – User Evaluation

We using 50 actual recorded images provided by users and then we use these images to put into our process. For each image is provided, test try on 2 to 4 products, after that we using 150 images to evaluate. We invite 12 volunteers participate with our review criteria as follows:

Criterion	Answer	Weight
Whether the upper clothing texture covers the torso part?	Y/N	1.35
Whether the upper clothing texture covers the left upper arm part?	Y/N	0.85
Whether the upper clothing texture covers the right upper arm part?	Y/N	0.85
Whether the upper clothing texture covers the left lower arm part?	Y/N	0.55
Whether the upper clothing texture covers the right lower arm part?	Y/N	0.55

Table 3 Criteria for volunteers to evaluate our system

Average evaluation scores for each result image among the 150 images are demonstrated in Figure 15 and Figure 16.
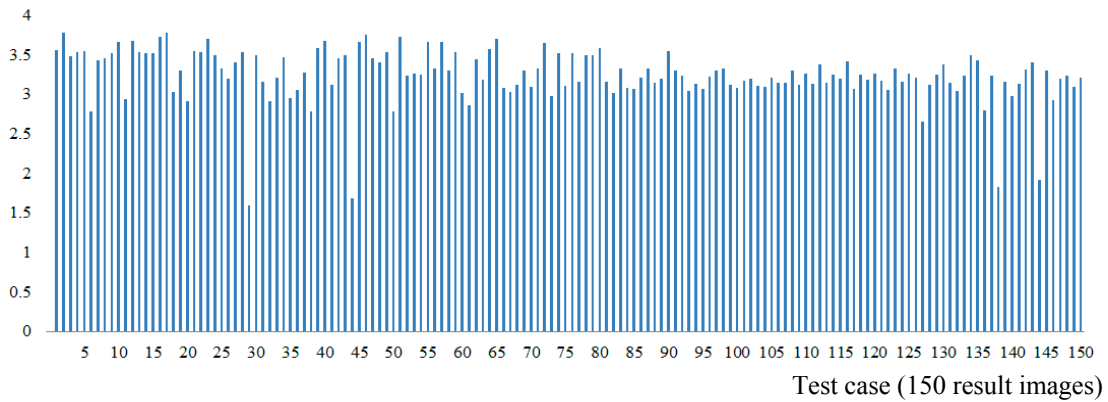

Test case (150 result images)

Figure 15 Average evaluation score for each result image

Good (score > 3): 133 images (89%)
Fair (2 ≤ score < 3): 12 images (8%)
Average (score < 2): 4 images (3%)
Average score: 487.1/150 = 3.2
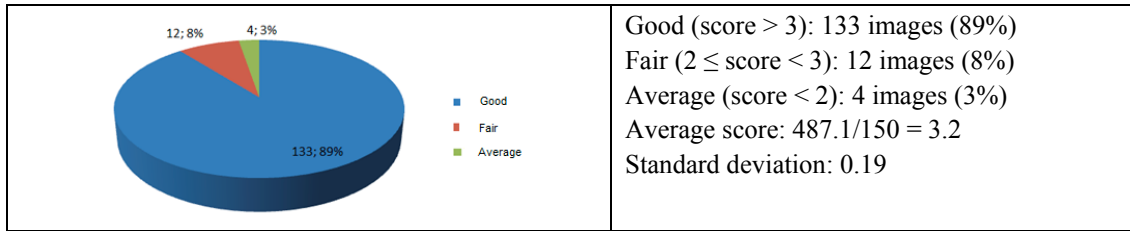Standard deviation: 0.19

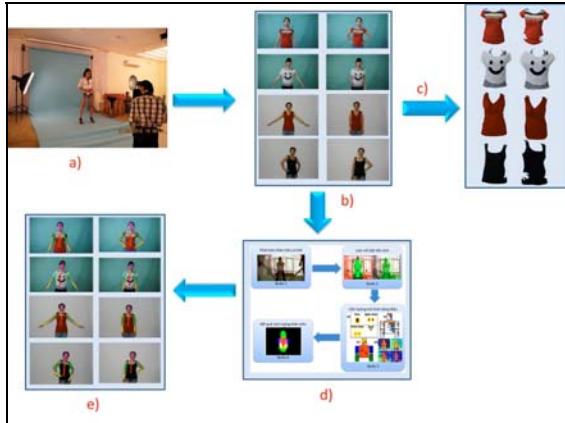Figure 16 Statistic on 150 images

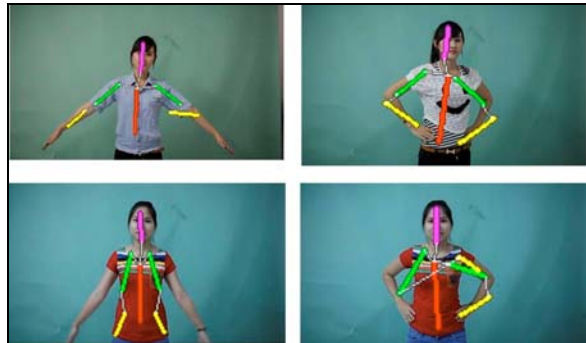
Figure 17 Data preparation process

Figure 18 Wrong sample return by [13]

For a user image with upright pose, lifted hands, the fitting results are evaluated as Good, consistent with the appearance of users. Product images covering the entire body position, shoulders of the user in the image. See **Table 4**.

Original user images	Result image using our proposed method	Evaluation
		Good
		Good

Table 4 Product image morphing experiment and evaluation of a user with upright poses

For the user's hands from both sides, the fitting result is also good reviews, though the body and hands on the products when users shift morphing 8-10 pixels (c.f.Table 5).

Original user images		Result image using our proposed method		Evaluate
				Good

Table 5 Product image morphing experiment and evaluation of the user's hands both sides

For the user's pose has the angle from the vertical axis of about 5 to 10 degrees, the product is now morphed slightly 10 to 15 degrees vertically. However, in some cases, as the estimated pose from the work [13] is wrong, the morphing results are not as good as we expect (see Table 6).

Original user images		Result image using our proposed method		Evaluation
				Fair, Average
				Bad

Table 6 Image morphing experiment & evaluation products of the two users as compared to the vertical deflection angle of 10 degrees from 5

Some results of our approach for bad results. The main reason of this limitation is dependent on human pose estimates of the algorithm [13]. In such case, the pose estimated between 2 points of shoulders is error and our system cannot calculate the correct distance d_1 and the morphing result are led to the wrong location and size.

Figure 19 Samples of incorrect results

5. Conclusions

In this paper we proposed method to overlay cloth to user image based upper human pose estimation [13] – virtual dressing room. We proposed 6 feature vectors from the original stick represent one human pose, normalize, and compare similarity between the poses and using Euclid to find corresponding pose in the data set. Proposed estimating the rate coefficients horizontally, vertically and position based initially stick to overlay textures of upper clothing products to user images. Experimental and identified 16 common pose 98% compared with other pose to support. Collect, organize and build cloth data on 390 images of 4 model for 12 sets cover 16 pose to try. Experimental and run the algorithm defined by Eichner, M et al [16] with our data set, the results achieved 96.45%. Tested over 50 image provided by the user and the result is shown in the previous section. We have also built virtual dressing room application to illustrate our method proposed.

In the future, we will implement a method to extract the product image to separate parts (hands, body, etc.) according to the mask layer based on the catalog of product. By using part of the product image, we will perform morphing to image a more accurate.

References
1.   Protopsaltou,D., Luible,C., Arevalo-Poizat,M., and Magnenat-Thalmann,N., A Body and Garment Creation Method for an Internet based Virtual Ftting Room. in Proc. Computer Graphics International 2002 (CGI '02), Springer, 105–122, 2002.
2.   Furukawa, Y. and Ponce, J. , Carved Visual Hulls for Image-based Modeling. in Int. J. Comput. Vision 81 (January), 53–67, 2009.
3.   Bradley, D., Popa, T., Sheffer, A., Heidrich, W., and Boubekeur,T., Markerless Garment Capture. in ACM SIGGRAPH 2008, ACM, New York, NY, USA, SIGGRAPH '08, 2008.
4.   Hasler, N., Asbach, M., Rosenhahn, B., Ohm, J.R., and Seidel,H.P., Physically based Tracking of Cloth.in 11th International Fall Workshop on Vision, Modeling, and Visualiza-tion 2006 (VMV 2006), IOS, Aachen, Germany, L. Kobbelt,T. Kuhlen, T. Aach, and R. Westermann, Eds., European As-sociation for Computer Graphics (Eurographics), 49–56, 2006.
5.   Choi, K.J. and Ko, H.S., Stable but Responsive Cloth. in ACM Transactions on Graphics, SIGGRAPH 2002 vol. 21 (2002).

6. Giovanni, S., Choi, Y.C., Huang, J., Khoo, E.T., and Yin,K.K., Virtual Try-On Using Kinect and HD Camera. in Motion in Games - 5th International Conference, MIG 2012, Rennes, France, November 15-17, 2012. Proceedings 2012, 2012.

7. Isikdogan, F. and Kara, G., A Real Time Virtual Dressing Room Application Using Kinect. 2012.

8. Hilsmann, A. and Eisert, P., Tracking and Retexturing Cloth for Real-Time Virtual Clothing Applications. in In Proc. of Mirage, 94–105, 2009

9. Hauswiesner, S., Straka, M., and Reitmayr, G., Image-based Clothes Transfer. in Proc. of International Symposium on Mixed and Augmented Reality (ISMAR), 2011

10. Ehara, J. and Saito, H., Texture Overlay for Virtual Clothing Based on PCA of Silhouettes. in 5th IEEE and ACM International Symposium on Mixed and Augmented Reality, ISMAR 2006, 139–142, Santa Barbara, USA (2006).

11. Tanaka, H. and Saito, H., Texture Overlay onto Flexible Object with PCA of Silhouettes and K-Means Method for Search into Database. In Proceedings of the IAPR Conference on Machine Vision Applications, Yokohama, JAPAN, 2009.

12. Yoon, J.C, Lee, I.K., and Kang, H., Image-based Dress-up System. in Proceedings of the 5th International Conference on Ubiquitous Information Management and Communication, ICUIMC 2011, Seoul, Republic of Korea, February 21 - 23, 2011.

13. Eichner, M., Marin-Jimenez, M.Zisserman, A., and Ferrari, V., 2D Articulated Human Pose Estimation and Retrieval in (Almost) Unconstrained Still Images. in International Journal of Computer Vision, ; Vol. 99, No. 2, pp. 190-214, September 2012.

14. Ramanan, D., Learning to Parse Images of Articulated Bodies. in NIPS, 2006.

15. Dalal, N., Histogram of Oriented Gradients for Human Detection. in IEEE Computer Society Conference on Computer Vision and Pattern Recognition CVPR 2005, 886–893, 2005.

16. Felzenszwalb, P., McAllester, D. and Ramanan, D., A Discriminatively Trained, Multiscale, Deformable Part Model. in IEEE Conference on Computer Vision and Pattern Recognition CVPR 2008, 1-8, 2008

17. Viola, P., Rapid Object Detection Using a Boosted Cascade of Simple Features. In Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition CVPR 2001, 511–518, 2001.