MAKING 3D EYEGLASSES TRY-ON PRACTICAL

Difei Tang[1], *Juyong Zhang*[1], *Ketan Tang*[2], *Lingfeng Xu*[2], *Lu Fang*[1]

[1]University of Science and Technology of China
[2]Hong Kong University of Science and Technology

ABSTRACT

In this paper, we propose a virtual 3D Eyeglasses Try-on (3DET) system, with efficient, realistic and real-time augmented performance. The 3DET system captures user's performance with the help of depth camera and renders the glasses properly on the video stream immediately after user's selection. The virtual eye-glasses follow with the motion (movement or rotation) of user's head simultaneously. The high efficiency of our 3DET system is achieved by simplifying the eyeglasses matching procedure, making use of the active appearance model (AAM) based face tracking algorithm. This is completely different from existing methods, which usually relies on eye detection. In addition, due to the exploiting of a generic 3D face model during tracking and displaying, the 3DET system can handle occlusion problem easily and render realistic glasses in videos effectively. Experimental results demonstrate that the proposed 3DET system is able to produce superior natural and smooth visual results with virtual glasses fitted on the users face at 30 fps with a high level of accuracy on common hardware.

***Index Terms*—** virtual try-on, face tracking, 3D face model, occlusion problem, augmented reality

1. INTRODUCTION

1.1. MOTIVATION

Virtual try-on for eye-glasses can provide a lot of convenience for consumers. Although taking off and picking up glasses seems easy to do, repeated work is always annoying and time consuming. Besides, it is hard to remember every style and choose one from them. Moreover, many consumers are short-sighted, they cannot see their looks clearly when trying on glasses frames without ophthalmic lens. Sunglasses prevent bright lights to come in to consumers' eyes, making consumers cannot see their looks clearly, too. Virtual try-on technique can help consumers get rid of all these problems to a large extend and contribute to online sales.

To meet consumers' demands, a practical virtual try-on system should be both efficient and effective. Algorithms including face tracking, glasses matching and occlusion problem solving etc should be considered. Currently, there are three main categories of eyeglasses virtual try-on techniques. The first one is based on 2D images, adding 2D glasses image onto user's facial image. Technically, this kind method is relatively simple, and the results are acceptable. In this way, it has been applied widely to commercial use, like online shopping already. However, it can only deal with frontal view, and cannot provide dynamic feedback for user's action. Another kind of methods combines 3D glasses model with 2D facial images. Taking 3D glasses models instead of 2D glasses image can help systems handle multiple views. That is to say, user can see his/her look from different angles. This kind of methods usually takes videos as input, and good tracking algorithm is needed. But when user rotates his/her head, some parts of the glasses are occluded. The occlusion problem needs to be solved. The third kind of methods not only uses 3D glasses model, but also reconstruct user's 3D head model in some way. Based on 3D face model, occlusion problem could be solved easily. However, the reconstruction of human face is challenging, and the texture always lacks sense of reality. So this kind of method is regarded as no effective enough in practical use. How to make virtual try-on into practical use is still a difficult problem.

1.2. RELATED WORK

Since trying virtual spectacles on consumers' faces has so much commercial potential, there are some related methods that have been deployed already. Previous methods are mostly based on 2D images or videos [1, 2, 3]. Users can upload their personal pictures and choose glasses they like, then the system will add the wanted pair of glasses onto their photos. [2] adopts a cascaded AdaBoost classifier to detect the eye corners and then fits the glasses image to the eye area using affine transform. This method is simple and effective, but can only handle frontal view. [3] uses videos as user's input and can control image parameters, but the system requires special hardwares, including 2 cameras, and can only run at 9.5 fps.

Mixed reality system in [4] adopts 3D glasses model instead of glasses images, and proposes a method for handling the occlusion problem. The solution for occlusion problem is based on several feature points which coordinates are known,

This work was supported in part by Natural Science Foundation of China (NSFC) under contract No. 61303151, 61331015 and 61390514.

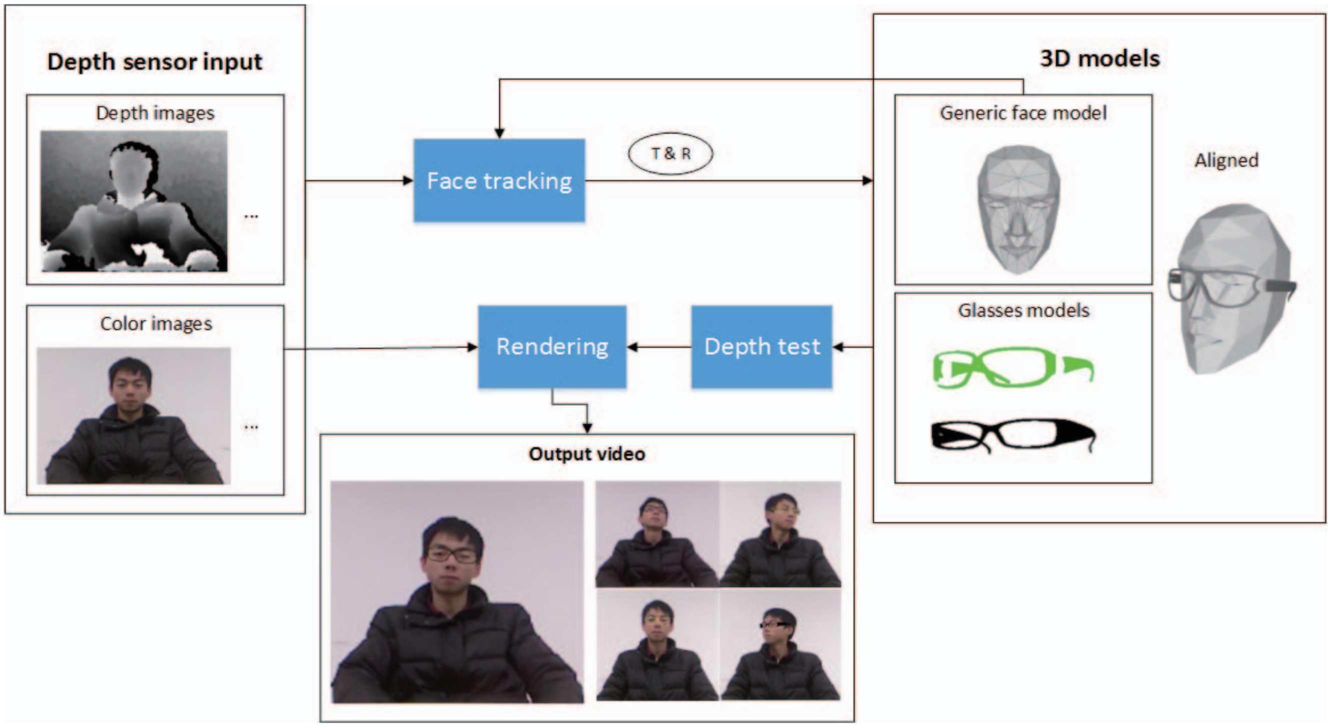


Fig. 1. The pipeline of proposed 3DET system.

and compares other points' coordinates with theirs. Since feature points can only present the geometry of eyeglasses model itself, the result can not be very precise, and some parts can be seen through transparent lens. This system also requires a special mirror system for display. [5] concentrates on human-centric design of glasses based on 3D glasses models, and presents an augmented reality system for glasses try on. However, the tracking procedure of the head movement in this system needs a 3D scan of user's head and after that user has to wait 5 to 10 minutes until a training procedure is done.

The last category of existing methods is based on user's 3D head model [6]. It relies on a 3D head reconstruction, and then fits 3D glasses model to 3D head model and renders the final results. The drawback is that the process for reconstructing a realistic 3D head model with satisfying texture is challenging and time-consuming, and the output cannot correspond to user's action, hence user's experience is degraded.

1.3. OUR CONTRIBUTION

In this paper, we present a 3D virtual try-on system based on face tracking algorithm and augmented reality techniques, which can adding virtual glasses onto video streams efficiently. Compared to previous works, our system makes use of a generic 3D face model, Candide-3[7], to both track user's face and render the final results without occlusion problem. Since the geometry of eyeglasses is simple and rigid, a generic model can provide enough information about human faces. In this way, we get rid of challenging 3D reconstruction problem, making our system very efficient. It employs an efficient face tracker and can reach real-time rate. Besides, the virtual glasses are added to real live video scenes, so it looks natural and realistic, which is an big advantage of augmented reality systems [8].

The rest of this paper is organized as follows: Section 2 presents an overview of our 3DET system as well as some technique details. Experimental results will be presented in Section 3. Section 4 concludes the paper and discusses future work.

2. PROPOSED 3DET SYSTEM

2.1. SYSTEM OVERVIEW

The pipeline of our 3DET system is illustrated in Figure 1, where the color images together with the corresponding depth maps are captured by the Kinect Sensor in real-time [1]. According to the input images, the area of user's face could be detected and tracked. An Active Appearance Model (AAM) - based tracking algorithm is used to capture the rotation and translation information of user's head. Then we apply this transformation to the 3D glasses model to follow user's movement. The glasses models are aligned with a face model in preparation. At last, we render the glasses model and blend

[1] It should be noted that our 3DET system is not limited to Kinect sensor, other depth sensors can be used for capturing input information as well.

it with the original color images. In addition, we propose to handle the occlusion problem using depth test during afore mentioned procedure. Consequently, the output video is the user fitted in virtual eyeglasses, taking advantage of both the geometry information provided by generic face model and the natural scenes provided by real video streams. In the following subsections, we will elaborate our 3DET system from the points of ”Efficiency” and ”Effectiveness” respectively.

2.2. EFFICIENCY OF OUR SYSTEM

To achieve a real-time response to user’s movement, an efficient and robust face tracker is essential. Inspired by [9], our 3DET system exploits an AAM based face tracking method, converting eyeglasses fitting problem into a robust tracking procedure without laborious face reconstruction. The face tracker is based on active appearance models [10], and integrates two constraints to make matching and tracking results smooth and accurate even in cluttered background. Additionally, the tracker is able to accomplish superior efficient face tracking with 50 frames per second.

Active appearance model is a kind of generative models like Morphable Models and Active Blobs. Linear in both shape and appearance but nonlinear in terms of pixel intensities, AAM is very suitable for facial representation. AAM assumes that a shape is described by N feature points, $s = [\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_N]$, and a shape can be represented as a mean shape s_0 plus a linear combination of n shape basis s_i:

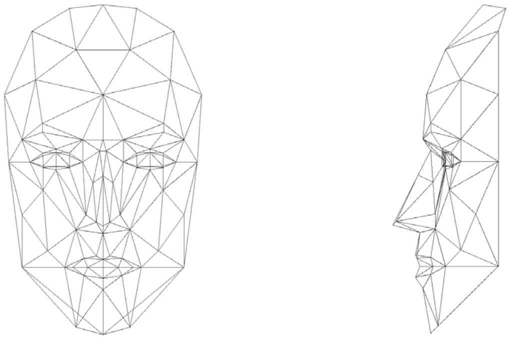
$$s(\mathbf{w}) = s_0 + \sum_{i=1}^{n} p_i s_i, \tag{1}$$

where $\{p_i\}$ are the shape parameters, and will be denoted as $\mathbf{p}$.

Here we use a generic face model Candide-3 as the mean shape. Candide-3 is a parameterised face mask specifically developed for model-based coding of human faces. This face model includes 113 vertices and 168 surfaces, and has been widely used in face coding and animating[7], as shown in Figure 2.

The appearance A in AAM is defined as the image patch enclosed by the mean shape s_0. Similar to the shape, A can be presented by a mean appearance plus a linear combination of appearance basis:

$$A = A_0 + \sum_{i=1}^{n} \lambda_i s_i. \tag{2}$$



(a) Frontal view of Candide-3 (b) Profile of Candide-3

Fig. 2. Candide-3 face model used in tracking and rendering procedure.

To locate the shape on an observed image I, AAM aims to find the optimal p_i and λ_i to minimize the difference between the warpedback appearance $I(W(\mathbf{p}))$ and the synthesized appearance A_λ:

$$\begin{aligned} E_a(\mathbf{p}, \lambda) &= \|A_\lambda - I(W(\mathbf{p}))\|_2 \\ &= \sum_{x \in s_0} [A_0(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i A_i(\mathbf{x}) - I(W(\mathbf{x}; \mathbf{p}))]^2, \end{aligned} \tag{3}$$

where $W(\mathbf{x}; \mathbf{p})$ is a warping function to map point $\mathbf{x}$ in the model to its corresponding image pixel.

Considering temporal matching error between I_{t-1} and I_t:

$$E_t = \sum_{j \in \Omega_t} \sum_{x \in R^j} [A_{t-1}(x)/\bar{g}^j_{t-1} - I_t(W(\mathbf{x}; \mathbf{p_t}))/\bar{g}^j_t(\mathbf{p}_t)]^2, \tag{4}$$

where Ω_t is a set of feature points, R^j is the corresponding local patch of the j-th feature point, $\bar{g}^j_t$ is the average intensity of the j-th patches of frame t. This constraint makes the feature points in the next frame to be consistent at the pre-matched position, assuring the tracking procedure to be smooth.

While in cluttered backgrounds, the background edges may influence the matching of face outline. Skin color is effective in handling this problem. [9] segments the face region using an adaptive color model, encouraging shape points to be located inside the face region:

$$E_c = \sum_{k=1}^{K} I_D(W(\mathbf{x_k}; \mathbf{p}))^2. \tag{5}$$

The total cost function to be minimized is:

$$E = E_a + \omega_t E_t + \omega_c E_c, \tag{6}$$

where E_a denotes the difference between warped-back appearance and the synthesized appearance, E_t denotes temporal matching constraint, E_c denotes face segmentation constraint, and ω_t, ω_c control the strength of each constraint respectively. By minimizing this cost function, shape parameters $\mathbf{p}$ and the location of outline points $\mathbf{x_k}$ can be found, and the movement can be easily obtained correspondingly.

In fact, Kinect SDK can provide the coordinates of detected face, and the transformation matrices of it, including translation matrix T and rotation R. The interface is exploited in 3DET system. While other existing systems usually adopt feature-based detection and tracking methods to find eye corners or eye center, and estimate an affine transform for head pose[2] according to the tracking results. Compared with these feature-based tracking, AAM takes both the shape and appearance information of the whole face into consideration rather than only some features, and can track human face more accurately and stably with little jitter, and 3D head's movement can be known directly.

2.3. EFFECTIVENESS OF OUR SYSTEM

Our 3DET system is also very effective. On one hand, the system is augmented by video streams of real scene rather than adopting all 3D models. The dynamic output interacted with user's movement in real-time provides a sense of reality. On the other hand, 3DET system can handle with occlusion problem very well. When user's movement is captured by tracking, applying the transformation to 3D glasses model can make glasses model (denoted as GL) fit the face:

$$GL_{current} = R * GL_{previous} + T. \tag{7}$$

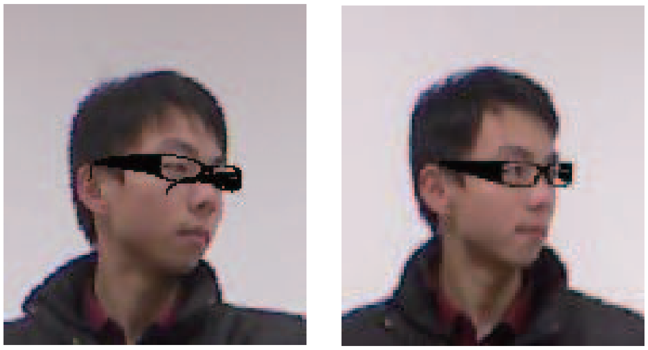
However, some parts of glasses occluded by user's head should not be seen. Figure 3 (a) shows the unreal scene caused by occlusion. To solve the problem, depth information is needed in rendering. However, considering that the geometry of eyeglasses model is simple and the structure of them is rigid, 3DET system makes use of the geometry of the same generic 3D face model (denoted as GF) employed in tracking procedure. By comparing the depth information of face model and glasses model, the occlusion problem could be solved easily.

Since the eyeglasses model has been registered with the generic face model in preparation, we regard those two models as a whole, and apply transformation matrices to both of them:

$$GT = GF + GL, \tag{8}$$

$$GT_{current} = R * GT_{previous} + T. \tag{9}$$

Then we can render the virtual glasses' texture and blend it with real color images. In rendering process, only glasses model's texture is wanted, so the generic face model is set to be transparent. In addition, the relative position of these 2 models is considered in rendering process. We employ depth test to decide which part should be rendered: for each triangle T in the model GT, calculate the z-depth value of T at every pixel(x,y), then choose the triangle with smallest z-depth value to render. In this way, an glasses image I_g without occlusion problem is produced. Then I_g is added to image I:



(a) Occlusion problem (b) Occlusion is handled

Fig. 3. Illustration of occlusion problem.

$$I_{result}(x,y) = \begin{cases} I_g(x,y); & \text{if } I_g(x,y) \neq 0 \\ I(x,y); & \text{otherwise.} \end{cases} \tag{10}$$

Figure 3 (b) shows the result of 3DET system with occlusion problem dealt. Other systems, like system in [2] can only do frontal view, and [4] simply judges occlusion from some feature points. These methods will cause a feeling of not realism in some cases. For example, when in side views, some parts of the arm should not be removed because they can be seen through lens. Compared with them, any occlusion part from any angles and views can be handled correctively in 3DET system.

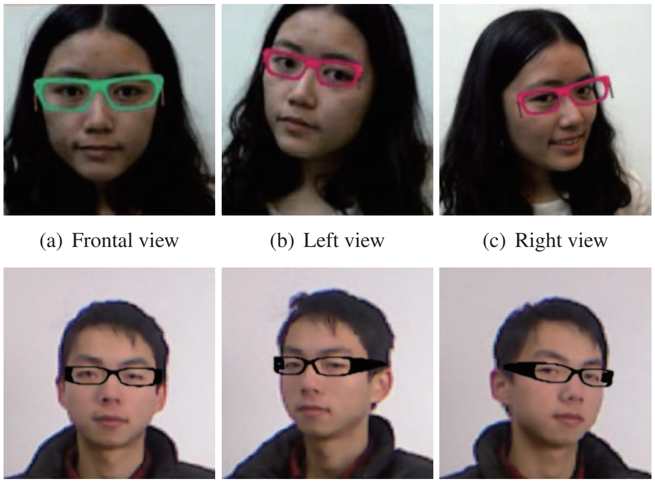
3. EXPERIMENTAL RESULTS AND DISCUSSIONS

3.1. Data acquisition and preparation

In our experiment, a Kinect sensor is used as the input depth camera. It is capable of capturing both color and depth information of a scene at 30 fps. These images are used as the original input of our system. The Candide-3 model [7] is used as a generic face model in our system for tracking and rendering. Eye-glasses models can be added to our system by simply aligning them with the face model. Our system runs on computer with an Intel Core i3 CPU @3.40 GHz at 30 frames per second.

3.2. Results and discussions

Figure 4 shows some results given by 3DET system. The virtual glasses are added to the user's face automatically following the user's movement during the try-on process. Users do not feel significant delay. From Figure 4 (a) - (e), we can see that 3DET system can track user's head pose very well, and



(a) Frontal view (b) Left view (c) Right view

(d) Frontal view (e) Left view (f) Right view

Fig. 5. Comparison with eyeglasses fitting system in [2]. (a) - (c) are results of [2], and (d) - (g) are results of 3DET system.

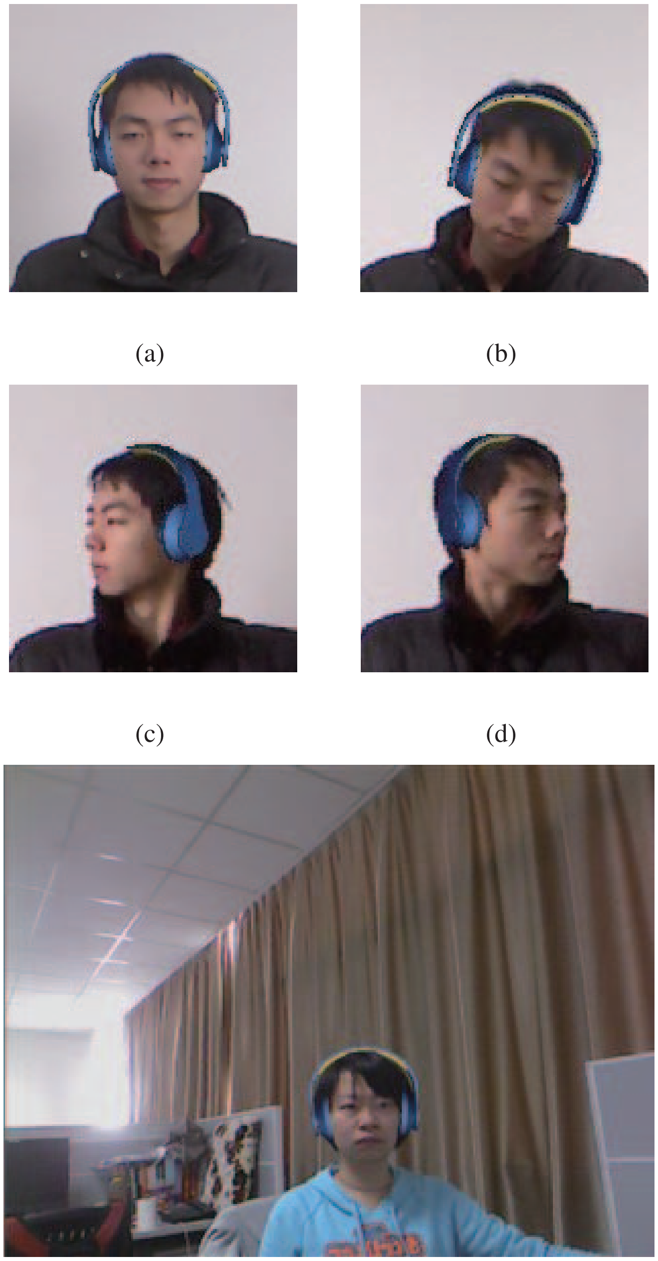
Figure 4 (f) shows that 3DET system is robust to the distance between the user and the camera. Thus our system provides users a large range of freedom of moving in a wide space. Figure 4 (h) - (i) show the try-on results for other spectacles with extreme angles, demonstrating our system's effectiveness.

In addition, a comparison with virtual try-on system in [2] is conducted. As depicted in Figure 5, system in [2] merely shows glasses without temples, such that the rotation of users head is significantly limited. While our system can handle with much larger rotation angles with the help of occlusion problem solved. More results and comparisons can be viewed on our video demo, demonstrating the stability and robustness of 3DET system.

Moreover, the proposed framework can be easily applied to virtual try-on for other facial accessories, like earrings, headphones, masks etc. Figure 6 (a) - (d) show some results of extending 3DET system on earphone try-on. Our system can work in cluttered background as well, as shown in Figure 6 (e). The results demonstrate that our system presents a natural view from different angles. Furthermore, the idea of taking advantages of both real scene and coarse 3D model's geometry can be extended to other tracking and displaying systems, helping some applications get rid of refined 3D model reconstruction.

4. CONCLUSIONS AND FUTURE WORKS

In this paper, we propose an effective augmented system 3DET for virtual glasses try-on. This system makes use of a generic 3D face model for face tracking and rendering. Our system does not need 3D reconstruction or online learning, thus is very efficient and suitable for commercial use. Besides, a robust face tracker is involved, ensuring the stability



(a) (b)

(c) (d)

(e)

Fig. 6. 3DET system's performance for other facial accessories. (a) - (d) show earphone try-on with different angles, and (e) shows the performance under cluttered background.

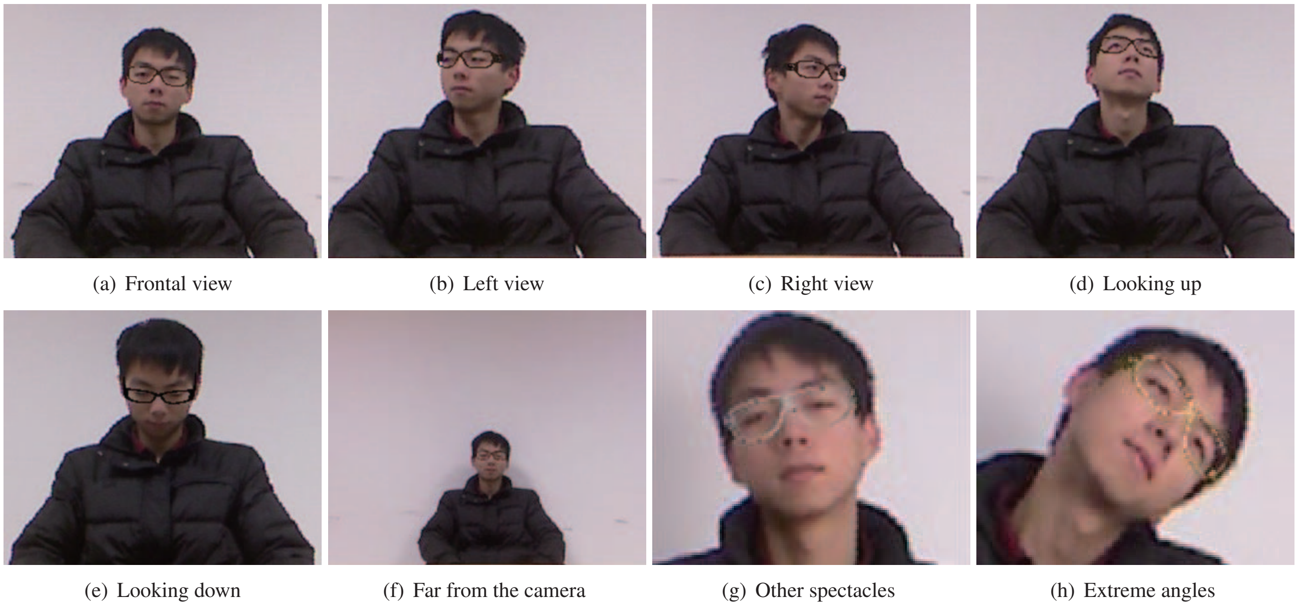

(a) Frontal view (b) Left view (c) Right view (d) Looking up (e) Looking down (f) Far from the camera (g) Other spectacles (h) Extreme angles

Fig. 4. Some virtual try-on results of 3DET system.

and efficiency of system's performance. The results provided by our system is also very natural and realistic. Furthermore, 3DET system has a potential to be extended to other facial accessories' virtual try-on applications easily.

Our system can be further improved from the following aspects: First, since the lighting in the scene can influence glasses' appearance, environmental issues could be taken into consideration to obtain better performance. Second, the quality of output data is based on the depth camera and 3D glasses models, so if the input data have a better resolution, the results will be better, too. Some super-resolution methods may be helpful for improving resolution when input source is poor.

5. REFERENCES

[1] Juan Li and Jie Yang, "Eyeglasses try-on based on improved poisson equations," *International Conference on Multimedia Technology (ICMT)*, pp. 3058–3061, 2011.

[2] Wan-Yu Huang, Chaur-Heh Hsieh, and Jeng-Sheng Yeh, "Vision-based virtual eyeglasses fitting system," *IEEE International Symposium on Consumer Electronics (ISCE)*, pp. 45–46, 2013.

[3] Oscar Déniz, Modesto Castrillón, Javier Lorenzo, Luis Antón, Mario Hernandez, and Gloria Bueno, "Computer vision based eyewear selector," *Journal of Zhejiang University SCIENCE C*, vol. 11, no. 2, pp. 79–91, 2010.

[4] Miaolong Yuan, Ishtiaq Rasool Khan, Farzam Farbiz, Arthur Niswar, and Zhiyong Huang, "A mixed reality system for virtual glasses try-on," *Proceedings of the 10th International Conference on Virtual Reality Continuum and Its Applications in Industry*, pp. 363–366, 2011.

[5] Szu-Hao Huang, Yu-I Yang, and Chih-Hsing Chu, "Human-centric design personalization of 3d glasses frame in markerless augmented reality," *Advanced Engineering Informatics*, vol. 26, no. 1, pp. 35–45, 2012.

[6] Arthur Niswar, Ishtiaq Rasool Khan, and Farzam Farbiz, "Virtual try-on of eyeglasses using 3d model of the head," *Proceedings of the 10th International Conference on Virtual Reality Continuum and Its Applications in Industry*, pp. 435–438, 2011.

[7] Jörgen Ahlberg, "Candide-3-an updated parameterised face," 2001.

[8] Ville Valjus, Sari Jarvinen, and Johannes Peltola, "Web-based augmented reality video streaming for marketing," *IEEE International Conference on Multimedia and Expo Workshops (ICMEW)*, pp. 331–336, 2012.

[9] Mingcai Zhou, Lin Liang, Jian Sun, and Yangsheng Wang, "Aam based face tracking with temporal matching and face segmentation," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 701–708, 2010.

[10] Iain Matthews and Simon Baker, "Active appearance models revisited," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 135–164, 2004.