Sunglasses Virtual Try-on

***Abstract*—** **Virtual try- on is a technology that allows people to nearly check the appearance when placed over them of accessories, makeup, haircut, hair color, clothes etc. The virtual pass- on presents numerous advantages over real pass- on, it speeds up the process furnishing the possibility to test multiple products without the need to reach a store. This project deals with a virtual pass- on web operation specific for eyeglasses and sunglasses that can be fluently used by simply taking a picture of a face and opting the asked frames. The pass- on process is performed on a 3D face from the input video (or webcam access if needed) allowing the stoner to see the virtual face and spectacles from different shoes. The pass- on process is completely automated and doesn't bear the stoner to give anything additional but the selection of the sunglasses to test.**

***Keywords*—Pose landmarks, Pose Estimation, Computer Vision, Augmented reality, Virtual try-on**

1. INTRODUCTION

The latest in fashion is a virtual try-on system that uses augmented reality to let users virtually try on clothes, accessories, makeup, hairstyles, and more. These systems come in a variety of formats, including web apps, mobile applications, and actual stations for use in retail settings. The virtual try-on has numerous benefits over the real try-on, including speeding up the procedure and enabling product testing without having to physically visit a store; it also enables the possibility to try things that might not be available in the store.

Another benefit is the ability to quickly and simply check the try-on results from various angles, giving a look at aspects that are typically not possible. The ability to see the outcome of the try-on session is a feature that is helpful while trying on glasses; this is especially helpful when the user has vision issues and finds it challenging to assess how the glasses will look without lenses.



In this work, we suggest an online application for trying on eyeglasses and sunglasses that can be utilized by simply snapping a picture of the user's face and choosing the desired frames. The try-on procedure uses real-time input so the user may view the simulated face and glasses from various angles. The user only needs to choose the glasses frames they want to try on for the try-on process to be totally automated.

We give a quick overview of the state of the art for both the virtual try-on and locating the user's facial locations for the positioning of frames in Section II. The solutions chosen for the frame placement and the glasses try-on procedure are explained in Section III. The application's user experience is described in Section IV. Lastly, in Section V, we analyze the project's current state, its shortcomings, and potential improvements.

1. RELATED WORKS

In this work, we analyze the advantages and disadvantages of the various methods with a focus on virtual try-ons for eyeglasses and sunglasses. As our try-on is based on frame placement on a real time input, we quickly go over the various approaches, focusing on those that only require one face image.

1. *Virtual try-on:* In response to growing demand from commercial businesses, a number of virtual try-on applications have been created in recent years. These solutions, which are typically offered as mobile or online applications, are designed to let a prospective buyer digitally try on some goods offered by a retailer or manufacturer.

Fig. 1: Virtual Try-on

According to our information, the possibly most well-liked virtual try-on options for makeup and face accessories are:

* Ditto’s Virtual try-on — The program is designed specifically for 3D eyewear, paying close attention to real geometry dimensions. The application can suggest the best-looking glasses by measuring the user's face measurement and drawing from a library of eyewear models. The glasses are rendered after the try-on process in several frames of a film that was shot, each with a different facial orientation. For the try-on procedure, the user is required to take a brief video while rotating his face horizontally and to place an object the size of a credit card on his forehead. This object is used by the program to estimate the user's face size. The user must adhere to precise instructions for the try-on to function as intended.
* XLabz’s Glassify — an application that renders several eyeglass and sunglass models using a single front face photograph. The software then tries to fit the glasses on the face, and the user can manually change the glasses' position and size as part of the try-on process. The user is then asked to select the face shape that best matches his or her own image. Moreover, the program only displays the front portion of the glasses frame and only front face photographs are compatible with it.
* Perfect Corp’s YouCam Makeup — Their virtual try-on system is primarily designed for makeup, while it may also be used for accessories, hair colour, and hairstyles (e.g. jewels, glasses). The framework has a number of features for facial recognition and tracking, augmented reality, and an AI-powered beauty advisor. Several of the mobile application's try-on capabilities operate in real time on the live camera video stream. However, there are certain drawbacks. For example, when the portrayed face is not in a perfect front view, the glasses try-on typically fails to accurately estimate location, size, and orientation because it only renders the front frame and ignores the temples.

The majority of these solutions are offered as stand-alone programs and frameworks that can be connected to already-in place services and commercial platforms. The bulk of these apps enable for interaction with social media and e-commerce sites, enabling users to share their virtual try-on experiences and make purchases. For simplicity of comparison, Table I summarizes the key characteristics of virtual try-on systems.

Furthermore, Google has recently introduced a set of func- tions for face detection, 3D reconstruction and augmented reality in the ARCore SDK currently available on a limited list of mobile devices.

1. *3D face reconstruction:* The challenge of 3D reconstruction has been addressed in a variety of ways throughout the years, typically needing specialized gear or numerous photos. We wish to streamline the acquisition and reconstruction process in order to give the user a simple-to-use application for virtual try-on.

TABLE I: Comparing with other try-on’s

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Methods | Input | Output | no | extra | input | 3D | glasses |
| Ditto  Glassify | video  image | images  image | —  — | | C  — | | |
| YouCam  **Ours** | image  video | image  video | C  — | | —  C | | |

We decided to concentrate on frame placement from a single video since it would enable the system to work with virtually any face image without necessitating the user's compliance with onerous instructions in order to gather input for the try-on session.

Here we provide a brief analysis of the most common methods for try-on with a single video.

* 3DMM fitting based methods — The manipulation of a 3D morphable model (3DMM), first presented by, is a well-liked approach to the face reconstruction problem from a single view. Several popular techniques, like, look for local features on the pictures that match to locations on the 3DMM. The 3DMM coefficients are then regressed using this correspondence to create a 3D face mesh that resembles the one in the image when they are applied to the model. CNNs are used in more recent techniques, as to regress the 3DMM coefficients. The whole face 3D model is always available with this technique, even when parts of the input image are obscured; these parts are recreated using the geometry of the 3DMM model. The primary drawback of these techniques is that the reconstruction frequently resembles the original 3DMM model too closely; characteristic qualities that are typically determined by minute details of the face geometry are lost during the reconstruction process, proposes a contemporary technique based on 3DMM models and asserts that it can produce superior outcomes to earlier efforts.
* Shape regression based methods — In order to obtain more accurate reconstructions than those based on 3DMM, various strategies were devised. The method put out by is interesting; its main idea is to simply translate the pixels from the input image to the entire 3D face structure in voxel space using volumetric CNN regression. The output face form is not constrained to a face model space, which is the key benefit. Another method is that suggested previously. In this method, the facial mesh is directly regressed from the input image without the need of a voxel space. As a result, the latter approach produces more accurate and thorough reconstructions while simultaneously being lighter and quicker to implement.

1. PROPOSED METHOD

The currently available solutions for face accessories virtual try-on present some limitations that we discussed previous Section. With the proposed method we want to define a solution that allows the user to try different glasses. To accomplish this task,

To accomplish this task, we have created a virtual sunglasses try-on using computer vision. We will first detect the human body pose and then overlay a sunglass on the person based on specific points. Pose estimation uses computer vision to determine the body's configuration out of an image or video. The variety of applications that really can profit from this technology is the reason for its significance.

Human pose estimate pinpoints body key locations to precisely identify a person's postures from an image. Either 3D or 2D estimations are carried out.

There are two fundamental steps in the primary process of estimating human pose:

a) Localizing important joints and locations on the human body

b) assembling those joints into a legitimate human stance configuration.

Finding each of the important human locations is the primary goal of the first stage. Eyes, Head, Shoulder, Arm, Hand, Knee, and Ankle, for instance. The second stage, which establishes the pairwise relationships between body components, involves arranging those joints into legitimate human pose configurations.



Fig 1: Pose tracking full body landmarks

If For this purpose a module called cvzone.Posemodule was used which facilitates the execution of AI and image processing operations. It primarily makes use of the OpenCV and Mediapipe libraries. The Media Pipe Pose framework for significantly high attitude tracking infers 33 3-dimensional landmarks on the entire human body from RGB video sequence input. Whereas other cutting-edge approaches generally rely on robust operating systems for inferencing, this methodology beats them and produces excellent results

in real-time.

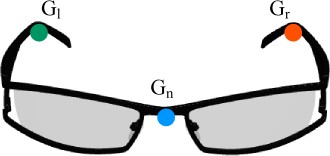
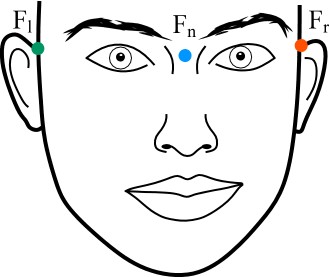
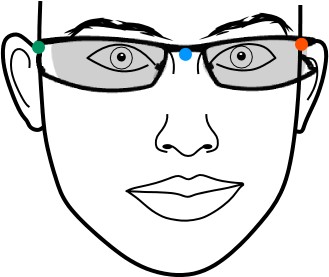
Fig, 2: Landmarks obtained using Pose Detector module

The terminal portion of the preceding graphic contains a list of pose markers. The following are the elements of each landmark: The image height and width, respectively, normalized the landmark coordinates x and y to [0.0, 1.0].

The lower the value of z, the nearer the key point is to the camera. z: This reflects the landmark depth by using the height at the midpoint of hips as origin. Z's magnitude is almost measured on the same scale as x.

Visibility: A number between [0.0, 1.0] reflecting the likelihood that the landmark will be seen in the picture. MediaPipe is operating flawlessly.

Following that, we gained access to all of landmarks 3 and 6's x and y values. The initial width of the glasses to be overlaid was specified as the difference in distance between the coordinates. Nevertheless, this is the improper strategy because the user won't be in a single location the entire time, which will cause the size and coordinates in which the glasses are to be placed to change. In order to determine this, we assessed the 83-pixel width of the typical user's ears and the 16-pixel distance between the landmarks 3 and 6 in relation to each other. To calculate the width of the glass that will be overlaid, this ratio was multiplied by the separation between the landmarks. As the starting point for the ear, the width value was used to determine the coordinates of the glass positions. The width was then multiplied by the glasses' height to width ratio, which was determined to be 0.78 because the glasses' height and width were measured at 140 pixels and 180 pixels, respectively.

* + 1. Overlaid Glasses landmarks (b) Face landmarks (c) Final Output

Fig. 3: Landmarks for overlaying glasses.

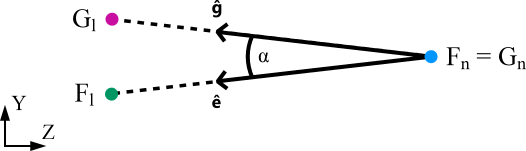


Fig. 4: Angle between ear and landmarks.

By this time, the user had the glasses successfully superimposed. In order to alter the input, or to try on different pairs of glasses, a virtual button image was positioned in the corner of both sides. As the user extended their hand in the direction of the button, the images changed and the new pairs of glasses were superimposed for try-on. The landmarks 15 and 16, representing the left and right wrists, were used to accomplish this. The coordinate values were in the 800s in the left corner and the 200s in the right. Hence, the following images were overlay whenever the 15th and 16th landmark values are lower than 900 and 300, respectively.

For a better User-Experience, when the user extends their hand towards any button on the application, the button is ringed in green in the right and left directions, respectively, and when the button is fully circled, the following images are superimposed.

TABLE II: Preliminary usability study results.

|  |  |
| --- | --- |
| Feature | Average rating (range 1-5) |
| Glasses fitting precision | 4.14 |
| Realism of the 3D view | 3.89 |
| Usefulness of the 3D view | 4.86 |
| Ease of use | 4.93 |
| Favorable to virtual try-on | 4.29 |

Twenty persons participated in a preliminary usability research, the results of which are shown in Table. Each user was instructed to use the application to take a picture of himself and conduct a virtual try-on session. At the conclusion of the session, each participant was asked to rate the application's usefulness, usability, convenience of use, and enthusiasm in using it if it were made publically available.  Most of them praised the application's simplicity of use and said it was very helpful to have the option to view the three - dimensional model from various angles. If combined with online stores, some consumers asserted that they'd possibly prefer this option over physical try-on. Yet, the biggest problem was the time necessary for the installation of lens in the user's face in proper position.

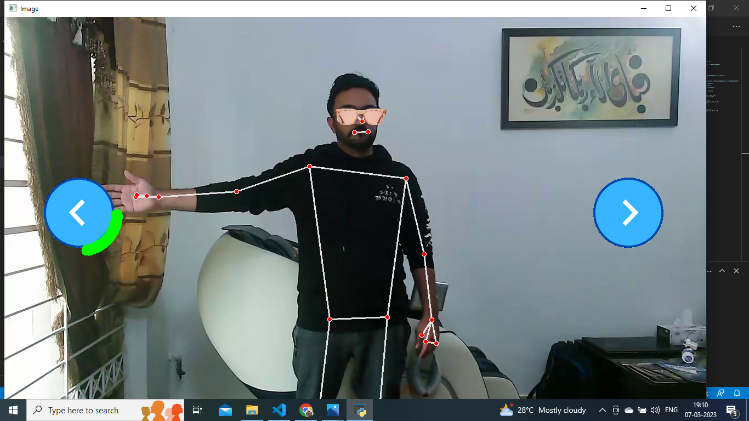
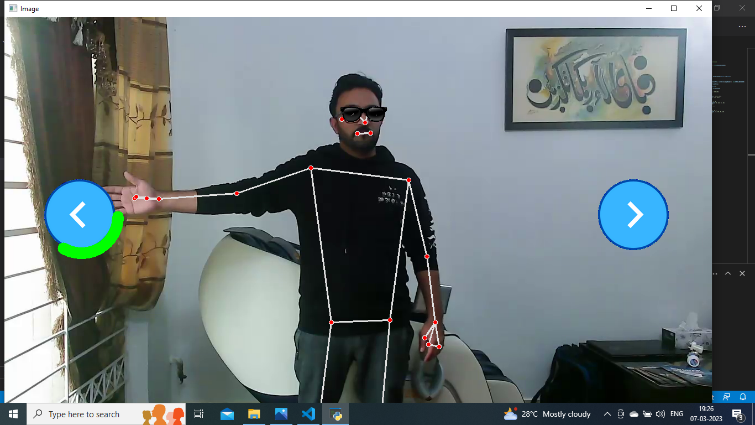
The other exciting feature relates to the proportions of eyeglasses; while some users want that genuine measurements be automatically taken into consideration, others would rather have the option to alter size and position to suit their preferences. Another widely held belief is that the geometry and texture of glasses need to be improved. Integration with social media and online retailers is another requirement that is frequently made.

IV. CONCLUSION

For convenience, a user interface is offered in the form of an application. Once the program is activated, all the user needs to do is stand in front of it. The software automatically determines the form of the user's face and positions the picture of the sunglasses accordingly. The user can then insert their palm through the virtual button to try-on the next pair of glasses. The user is free to wander about and look at things from various perspectives.

A completely functional application for virtual try-on of eyeglass frames is the result of the work completed. However, there are several restrictions: the created mesh does not include the hair and is missing on the top and back, and the texture is frequently poor in quality. These factors, along with the opinions generated by the usability study, may influence future changes. For simplicity of use and improved device interaction, the online interface can also be turned into a mobile application.

V. REFERENCES

1. D. Technologies, “Ditto website,” [https://www.ditto.com/,](http://www.ditto.com/) accessed: May 31, 2019.
2. X. Technologies, “Glassify android application,” https://play.google. com/store/apps/details?id=com.xlabz.glassify, accessed: May 31, 2019.
3. P. Corp, “Youcam makeup,” https:[//www.perfectcorp.com/app/ymk,](http://www.perfectcorp.com/app/ymk) ac- cessed: May 31, 2019.
4. A. Ablavatski and I. Grishchenko, “Real-time ar self-expression with machine learning,” https://ai.googleblog.com/2019/03/ real-time-ar-self-expression-with.html, 2019, accessed: May 31, 2019.
5. 3dMD, “3dmdface system,” <http://www.3dmd.com/static-3dmd> systems/#face, accessed: May 31, 2019.
6. J. Roth, Y. Tong, and X. Liu, “Adaptive 3d face reconstruction from unconstrained photo collections,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 4197–4206.
7. M. Piotraschke and V. Blanz, “Automated 3d face reconstruction from multiple images using quality measures,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 3418–3427.
8. S. Bianco, G. Ciocca, and D. Marelli, “Evaluating the performance of structure from motion pipelines,” *Journal of Imaging*, vol. 4, no. 8, 2018.
9. V. Blanz, T. Vetter *et al.*, “A morphable model for the synthesis of 3d faces.” in *Siggraph*, vol. 99, 1999, pp. 187–194.
10. P. Huber, P. Kopp, W. Christmas, M. Ra¨tsch, and J. Kittler, “Real-time 3d face fitting and texture fusion on in-the-wild videos,” *IEEE Signal Processing Letters*, vol. 24, no. 4, pp. 437–441, 2017.
11. A. Tuan Tran, T. Hassner, I. Masi, and G. Medioni, “Regressing robust and discriminative 3d morphable models with a very deep neural network,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 5163–5172.
12. A. Ranjan, T. Bolkart, S. Sanyal, and M. J. Black, “Generating 3d faces using convolutional mesh autoencoders,” in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 704–720.
13. A. S. Jackson, A. Bulat, V. Argyriou, and G. Tzimiropoulos, “Large pose 3d face reconstruction from a single image via direct volumetric cnn regression,” in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 1031–1039.
14. Y. Feng, F. Wu, X. Shao, Y. Wang, and X. Zhou, “Joint 3d face recon- struction and dense alignment with position map regression network,” in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 534–551.
15. R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker, “Multi-pie,”

*Image and Vision Computing*, vol. 28, no. 5, pp. 807 – 813, 2010.

1. Y. Guo, L. Zhang, Y. Hu, X. He, and J. Gao, “Ms-celeb-1m: A dataset and benchmark for large-scale face recognition,” in *Computer Vision – ECCV 2016*, 2016, pp. 87–102.
2. Fig. 4: Examples of glasses fitted on 3D faces reconstructed from some images of the MS-Celeb-1M dataset [16]. Glasses models from CadNav and Blend Swap.

Fig. 5: Glasses overlaid on the user