Journal of Systemics, Cybernetics and Informatics (2024) 22(2), 63-68
https://doi.org/10.54808/JSCI.22.02.63

Face Authentication by Constructing a 3D face image from 2D face images, and Encoding a Unique Identification

Menachem DOMB
Computer Science, Ashkelon Academy College
Ashkelon, 5265320, Israel

ABSTRACT

With the vast spread of automation, the significant demand for communication services, the addition of IoT to the Internet, and the corresponding increase and sophistication in malicious attacks utilizing system vulnerabilities to penetrate systems. Identification and authentication are heavily used in every access trial for electronic resources and communication networks. The traditional approach to coping with such challenges is using passwords, encryption, Secure ID, Firewall, Etc. More safety methods use Biometrics, which suffer from spoofing. The access control market is a fast-growing and highly volatile market that poses significant challenges for investors seeking to make secured decisions. As the market continues to evolve and become more mainstream, there is a growing demand for new identification and authentication technologies. This paper proposes expanding the number of features extracted from a 3D image, with unique features evolved during the generation of the 3D image of the prospect at the access control stage. Experiments support the proposed approach.
Keywords: Authentication, Image processing, 2D, 3D, Features.

1. INTRODUCTION

Face recognition is a type of biometric recognition technology that uses information about a person's facial features to identify them. The goal is to identify the vital facial landmarks to distinguish factors. It looks for the following features: histogram, color, template, structural, Haar, visual, pixel statistical, face image transformation coefficient, and face algebra features. Face recognition uses the Distance between the eyes, forehead to the chin, nose and mouth, eye socket depth, cheekbones shape, lips, ears, and chin contour. 3D face recognition inherits 2D face features and adds variant facial positions and expressions, introducing pose-invariant recognition, expression-invariant recognition, and occlusion-invariant recognition. One of the critical reasons why 3D face recognition outperforms 2D is its ability to perform pattern recognition. It captures and analyzes the Depth of the human face. While 2D methods rely on flat images only, 3D face recognition uses a 3D face model, which allows it to take advantage of the human face's geometry. It offers a potent shield against pitfalls like changing lighting conditions, diverse facial expressions, and varying head angles. Advanced algorithms conduct pattern analysis on the 3D facial data, ensuring robust recognition even amidst pose variations and shifts in illumination. This algorithmic prowess bolsters the system's capabilities, rendering it resilient against diverse environmental factors. 3D models significantly improve facial recognition results compared to 2D images. The system's adaptability and accuracy are unparalleled when combining this depth analysis with AI algorithms. While 2D might falter due to their striking resemblances, 3D would discern between their subtle facial depth differences, ensuring correct identification. The acquisition of 3D face samples involves infrared laser beams directed at the human face, drawing its shape features using triangulation methods to determine a precise map by calculating and grouping reflection points. Draw the surface shape. The system matches points acquired from different cameras and builds the precise 3D location of the matched point. The set of the matched points forms the 3D face. The most straightforward feature extraction is storing the entire face as a single feature vector in the database. In the feature-matching stage, the target face is compared with faces in the database using statistical classification functions and graph operators to extract the nose and eye parts and store these local features in the database. When a target face is inputted for recognition, it extracts the corresponding parts from the target faces and then searches the matched set of parts from the feature database.
2D face recognition will never have the accuracy needed for accurate Unsupervised Identity Verification and Authentication. This variability creates significant overlapping similarities between the 2D features of different humans and confuses the 2D algorithms, preventing them from achieving highly accurate FARs at usable FRRs. 3D technology has recently evolved in various fields, such as access control, health imaging, and topography. It provides an illusion of Depth and makes simulating actual reality possible. Humans and animals are endowed with eyes, right and left, comprising a cornea, retina, and pupil, functioning as a sophisticated camera. An image enters each pupil as a 2D image, transmitted by neurons to the brain, and combined into a 3D image having the depth effect. The element that gives the 3D effect is the Distance between the two pupils, approximately 6.35 cm. In birds, the Distance between the pupils is more significant, allowing them to see their 3D prey image from

huge distances. 3D cameras imitate the same concept of two parallel lenses with about 6.5 cm between their centers. The camera takes two pictures simultaneously with both lenses, and a 3D image is generated with embedded software. 3D images are used in medicine, security, and smart home applications. This paper elaborates on a simplified, affordable, real-time, and calibratable method of transforming 2D images to 3D, generating a secured image ID with a unique embedded secret. We propose a system that connects two standard cameras mounted on an ordinary bar, allowing the calibration of the Distance between the lenses of the two mounted cameras. Hence, the user can determine the camera's position, achieving vision of distant objects in 3D online. We propose adding new features to the standard 3D image features, such as the Distance between the camera lenses and between the cameras and the subject and distortion. We describe the required mechanical setup with calibration capabilities, real-time synchronization between the two cameras, and the composition process of building the 3D image by discovering interesting points common to the two 2D images and applying a stitching mechanism to design the desired secured 3D image.

2. LITERATURE REVIEW

Many methods for face recognition have been published and classified by various classes, such as local, holistic, and hybrid, or using face image features or the whole facial features with ML. They are evaluated based on a range of factors: robustness, accuracy, complexity, discrimination, database structure, and supervised and unsupervised learning. We review articles on the main subjects relevant to our research. We first present papers approving that 3D recognition is better than 2D recognition due to added features, geometry, and better security resistance. The first stage in our proposal is 3D construction from two 2D images. Therefore, we outline several methods for transforming 2D to 3D. Since this work is about 3D Authentication, we outline 3D face recognition methods.

3D face recognition is more accurate than 2D face recognition:

The advantage of 3D is that it has enhanced features and recognition systems' efficiency, contributing to its accuracy in results. The FAR and ERR in 3D-based recognition systems are lower than those of 2D recognition systems. Many surveys covering historical and current face recognition systems show that 3D face recognition systems are more accurate than 2D systems[1a]. Jing, Y et al. [1] comprehensively explain 3D face recognition, attributing face geometry and deep learning as the main drive for new developments. The use of AI methods helps minimize the number of features and speed up the recognition process, making it suitable for near real-time applications, such as our proposal. ZHD Eng et al. [2] claim that 3D face recognition is more accurate than 2D due to the geometric features 3D has and justifies our approach of utilizing 2D images to compose a reliable 3D image. Authors claim this is different when comparing 2D and 3D inverted faces due to the lack of 3D stereoscopic effects that influence face recognition during holistic processing but not during featural processing.

Constructing 3D images from 2D images: An approach of transforming 2D images to 3D [3, 4], utilizing low-level linear features, neglecting extracted planar facades, or higher-order features. However, this approach is biased towards large elements such as buildings and does not fit face images. Xin Wen et al. [5] suggest reconstructing a 3D shape from a 2D image by capturing the 2D semantic features and reconstructing them through a 3D decoder of various forms, such as voxel, point cloud, and mesh. Although we get a 3D image, it needs to be completed and consistent concerning the feature richness, which is a crucial foundation of our proposal. Another attempt to generate a 3D from a 2D image is described in [6]; they introduce an automatic extension of Generative adversarial networks (GANs) images to become 3D aware, near 3D but not yet a complete 3D as it still lacking the 3D geometry and thus insufficient for our purpose. In [7], they highlight two issues in the [6] proposal: the function has a local perspective, needs to include the global view, and complicates optimization difficulty. To cope with these problems, they propose explicitly learning a structural and textural representation. It learns a feature volume representing the underlying structure and then converts it to a feature field. The feature field is accumulated into a 2D feature map as the textural representation, followed by a neural renderer for appearance synthesis, providing independent control of the Shape and the appearance. Haochen Wang et al. [8] repurpose aggregating 2D scores at multiple camera viewpoints into a 3D score by using 2D images to predict a vector field of gradients, activate a chain rule on these gradients, and circulate the score of a diffusion model via the Jacobian of a differentiable renderer. This process requires 2D images for the training stage to apply an ML complex mechanism, which is different from our goal of a fast, simple, affordable, and based on a firm concept such as mimicking the human eye. To speed up the process, [9] propose I2P-MAE and acquire superior 3D representations from 2D pre-trained models through Image-to-Point Masked Autoencoders. [10] Blanze et al. outline the construction of 3D face images from single images, building a morphable face model and recovering domain knowledge about face variations by applying pattern classification methods. Gallucci et al. [11] suggest using a synthetic replica and present the generation of synthetic 3D models using machine learning and injecting a common reference 3D template into every scan, bringing all scans in full correspondence.

All these proposals indeed generate 3D images but for different purposes and input. However, no one proposal suggests a construction process that can be accurately calibrated. Calibration ability is a critical component in our proposal, as we use the ability to adjust the inter-camera distance to construct a unique 3D image, different from another 3D image constructed from a different inter-camera distance. We managed to find a way to develop an adjustable mechanism by adopting the Harris algorithm

[17] for identifying “interesting points” and using them for stitching the input 2D images.

3D face recognition methods (3DFR): Menghan et al. [12] provide a comprehensive 3DFR survey describing the standard and up-to-date methods, including feature extraction, classification, and disadvantages such as pose, illumination, expression variations, self-occlusion, and spoofing attack. [13] proposes to start with enhancing the image quality, apply a deep-learning (DL) based distance ternary search, extract key features, and execute the MIT-CBCL face recognition using Texas 3D face databases, obtaining 99.31% accuracy. W. Yang et al. [14] proposed injecting the head pose and facial expression variation between video frames into a face image to learn 3D face shapes and then extracting and reversing the injected variation to reconstruct the face image to its original state. During training, the model learns to decouple the pose and expression properties for performing cycle-consistent face reconstruction. Chen, G.Y [15] introduces selective denoising with block-matching and 3D filtering (BM3D), computes filter faces, and extracts the HOG features from the extracted feature maps, achieving the correct classification rate (98.4%) for the Extended Yale Face dataset B and comparable results (100%) for the CMU-PIE datasets. For hyperspectral face recognition, the method achieves a perfect classification rate (100%) for the PolyU-HSFD and the CMU-HSFD dataset and dataset. Qi Wang et al. [16] claim that the point clouds lack detailed textures, causing facial features to be affected when expression or head pose changes. Therefore, it proposes a unique face recognition network comprising an operator based on a local feature descriptor and a feature enhancement mechanism to enhance the discrimination of facial features.

3. THE MATCHING PROCESS

Standard Identification and Authorization systems are done in two processes, Enrollment and Authentication, as described in Figure 1. The person's picture is taken at the Enrollment step, and then its features are extracted and saved in the database. Later, when the person tries to access a secured resource, his picture is taken, and its features are extracted and matched against any of the features in the database by the Feature Matcher. If the extracted features match any features set in the database, the access is approved; otherwise, it is denied. In our proposal, we change the matching process to a 3D image matching instead of a 2D matching process, where the core change is in the enhanced features associated with a 3D image Vs. a 2D image, and in addition, several more features representing the measurements of the transformation from 2D to 3D, such as the Distance between lenses and the Distance between the cameras and the prospect, at the image taking moment. The 3D feature values are unique to each Distance between the camera lenses or the Distance between the cameras and the prospect. This difference is the secret hidden behind the 3d image features. This attribute makes the 3D comparison dependable with incredibly low FRR and ERR measurements.

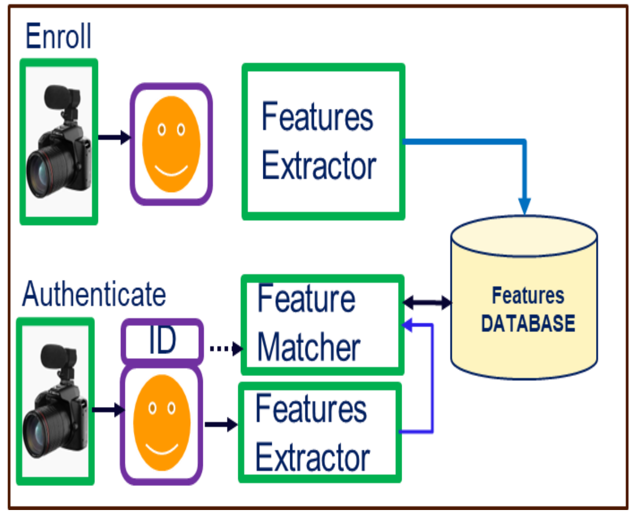

Figure 1: Enrollment and Feature Macher

Figure 2 depicts the three-stage transformation of a pair of 2D images into one 3D image. The processes are designated by the numbers 1, 2, and 3, while the outcome of each process is designated by the letters a, b, and c. In stage 1, the two cameras are synchronized and simultaneously shot to create identical images of the person from two different angles. This stage is performed by a process developed in C++. The algorithm starts by reading the two images, validating their completeness and correct synchronization. Finally, it saves the images as R.jpg and L.jpg, ready for the next stage, merging them into a 3D image. In stage 2, the two 2D images are merged to generate one 3D image, as depicted in Figure 3. The 3D construction process was developed in Python in three steps. For the construction of a 3D image, we consider several alternatives. Step 1: Scan the images to identify their "interesting points" using the Multi-Scale Harris algorithm [17]. It has been proven to be most stable against rotation and noise and insensitive to scale change. The algorithm consists of two main stages. In stage one, candidate interesting points are found for each scale level, and in stage two, the stability of each point is measured, and based on the measure, the final "interesting point" is selected from candidates. Step 2 identifies common "interesting points" to the two images and applies the Image "stitching algorithm" to combine the two images into one panoramic image. Step 3: We use the "Epipolar Line" algorithm [18] to combine the images into one 3D image. Multiple-view geometry relates the camera points in 3D, and the corresponding observation is called the Epipolar geometry of a stereo pair. Stage 3 features are extracted from the 3D image using a standard method. The following features were added: Distance between the two cameras and from the subject and % distortion.

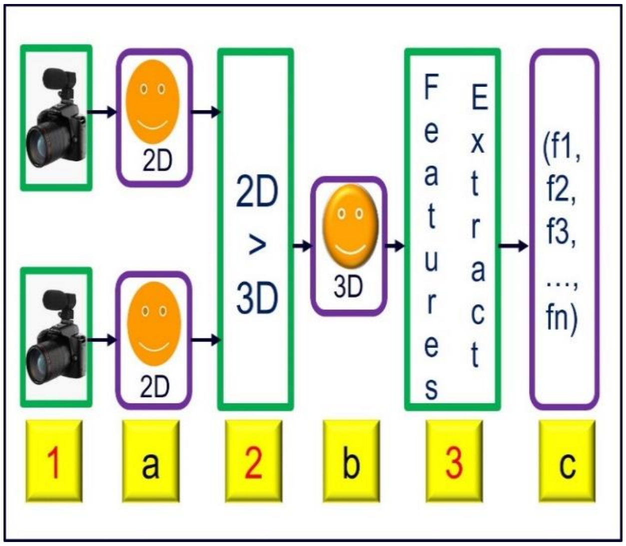

Figure 2: The proposed new extended matching process

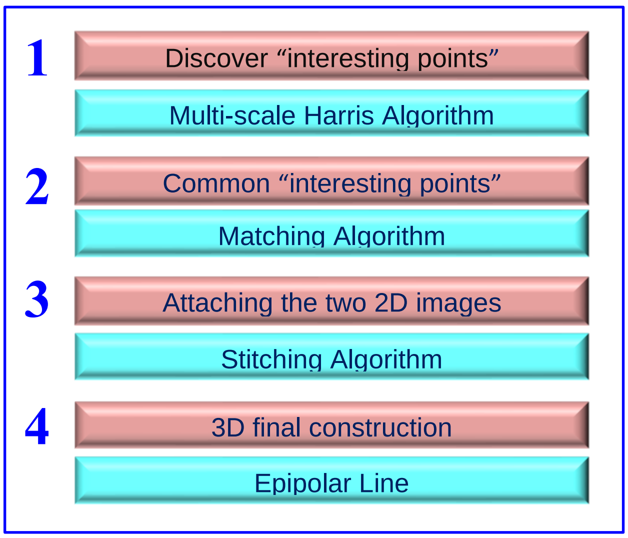

Figure 3: The process of constructing 3D image

Before executing the transformation process, the system is set by the following attributes: a. The Distance between the two cameras. b. The camera's location defines the exact distance from the subject. c. The feature-extract algorithms are to be used in stage 3. These setup parameters will be added to the features list of the 3D image.

4. EXPERIMENT

The experiment is divided into two consecutive stages: 1). constructing a 3D image from two 2D identical images taken from various locations. 2). testing of the identification and authentication accuracy. Constructing the 3D image out of 2D images: This experiment aims to prove the feasibility and quality of our merging process of two 2D images into one 3D image. We started the experiment with large non-human objects. We photographed the object simultaneously from three cameras, two 2D cameras and one 3D camera. We applied our 2D to 3D transformation, compared the resulting 3D image with the 3D image from the third camera, and found the images similar, proving our proposal's accuracy. After applying some distortions to avoid face identification, we repeated the experiment with a human face. We used two standard remote-control stills cameras mounted on an adjustable camera platform sliding bar system, as depicted in Figure 4. An additional 3D camera is used. All three cameras are activated simultaneously to ensure proper construction of the 3D image and be able to compare it to the 3D image taken by the 3D camera. More details appear later herein. Due to privacy constraints, we took pictures of family and friends of various ages. Figures 5a and 5b show the two input 2D objects, and 5c is the resulting 3D image. We repeated the same experiment, analyzed all the intermediate results, and found the same results, proving that the process is accurate, dependable, and robust. We continued testing with more real faces and got comparable results. The features of image 4c contain the additional features of the Distance between the camera and the Distance between the cameras and the subject, which proves the feasibility of the physical setup to merge two images into one 3D image. We proceeded with the experiment with twelve adults. We first took a picture of their face and transformed it into a 3D image with our system, extracted its features, and saved them in the database. We asked them to come again and go through the same process with the same setup and Distance from the camera. We compared each 2D image feature to the 3D features in the database, and as expected, there was no match. We then extracted the 3D features, compared them to the database features, and found a match. We asked the person to repeat the process, but the Distance from the cameras was lower than the one in the database. We repeated the same for the following nine people. In two cases, the system ended with no match, while it is the same person due to the very tight numbers and hence sensitive to the exact measures. It is managed by setting a threshold to allow a small grace margin. Further testing is planned when we continue this project. On top is a mounted computer, which controls the camera synchronization, collects the images from the two cameras, and processes them to get the desired 3D image. Below the computer are two cameras mounted on the horizontal pole, which have a mechanism to adjust the distance between them. The cameras are connected to the computer system at the top. The system controls the Distance between the cameras, schedules the simultaneous shooting of both cameras, accepts the 2D images, merges them into a 3D image, and saves the extended features of the 3D image. Our proposed system of constructing the 3D image is feasible, dependable, robust, affordable, and easy to implement, and hence, appropriate to be used as a platform for 3D face authentication.

Assessing the feasibility and accuracy of our proposal towards the identification and authentication: We started by setting up the camera stand in a solid location, setting the inter-distance cameras to 6.5 cm, and determining the person's distance of 150 cm from the cameras. We asked the person to stand at the designated location. Photograph the person by the two cameras simultaneously, transmit the two images to the on-top mounted computer, which activates the 3D contraction application, generates a 3D image, and finally extracts the features from the 3D image, add to the features list three more features, the inter-camera Distance of 6.5 cm, the 100 cm distance between the person and the cameras, and the distortion of 4%. Store the collected features in the database with the persons' ID and demographic data. The registration of the person is complete. In the next step, we authenticate the same person. We asked the person to repeat the registration photography and get a current 3D image and its features list. We then compared the current 3D features to those stored in the database, and they match. We then changed the person's Distance from the camera to 150 cm, constructing the 3D image and features. We found differences in the generic 3D features and the three extended features. We repeated the test but changed the inter-cameras Distance to 6.8cm, and again, it did not match the generic and the extended features, meaning that changing the photography setup impacts the 3D image and the corresponding features.

During the experiment, we encountered the following limitations: 1) Our approach associates the uniqueness of multi-encoding with the capability to change the inter-camera distance. However, this calibration is limited and requires the recommended sliding range. 2) the setup must be precise, which is a challenge for daily operations.

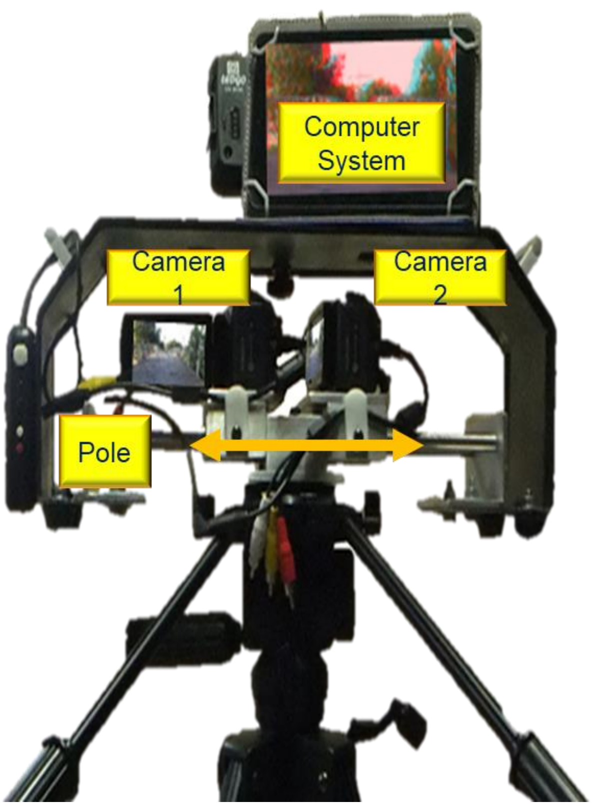

Figure 4: The physical setup

Figure 5: From 2D (a, b) to 3D © images

5. CONCLUSIONS

This work described a simple, affordable, solid, and calibratable construction of a 3D image using two 2D images. The generated 3D image is encoded and used for face-authentication. We performed a feasibility test using a reasonable number of cases. We plan a comprehensive test using about two hundred people's faces who will agree to take part in our experiment. We expanded the authentication process accordingly. In the future, we consider expanding the 3D construction process by adding more cameras (3,4), activated simultaneously, and exploring the contribution of the added cameras to the authentication accuracy and long-term operations.

6. REFERENCES

[1a] Ali, W., Wenhong, T., Din, S. U., Desire, I., & Khan, A. A. (2021). Classical and modern face recognition approaches A complete review. Multimedia Tools and Applications, 80(3), 4825-4880. Doi: https://doi.org/10.1007/s11042-020-09850-1

[1] Jing Y., Lu X. & Gao S. 3D face recognition: A comprehensive survey in 2022. Comp. Visual Media 9, 657–685 (2023). https://doi.org/10.1007/s41095-022-0317-1

[2] ZHD Eng, Y.Y. Yick, Y Guo, H. Xu, M. Reiner, T.J. Cham, S.H.A. Chen, 3D faces are recognized more accurately and faster than 2D faces, but with similar inversion effects, Vision Research, Volume 138, 2017, Pages 78-85, ISSN 0042-6989, https://doi.org/10.1016/j.visres.2017. 06.

[3] Alidoost, F., Arefi, H., Tombari, F. 2D Image-To-3D Model: Knowledge-Based 3D Building Reconstruction (3DBR) Using Single Aerial Images and Convolutional Neural Networks (CNNs). Remote Sens. 2019, 11, 2219. https://doi.org/10.3390/rs11192219

[4] Lingyun Liu, Ioannis Stamos, A systematic approach for 2D-image to 3D-range registration in urban environments, Computer Vision and Image Understanding, Volume 116, Issue 1, 2012, Pages 25-37, ISSN 1077-3142, https://doi.org/10.1016/j.cviu.2011.07.009

[5] Xin Wen, J. Zhou, Yu-Shen Liu, Hua Su, Zhen Dong, Zhizhong Han; 3D Shape Reconstruction From 2D Images with Disentangled Attribute Flow, IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 3803-3813

[6] Zhao, X., Ma, F., Güera, D., Ren, Z., Schwing, A.G., Colburn, A. (2022). Generative Multiplane Images: Making a 2D GAN 3D-Aware. In: Avidan, S., Brostow, G., Cissé, M., Farinella, G.M., Hassner, T. (eds) Computer Vision – ECCV 2022. ECCV 2022. Lecture Notes in Computer Science, vol 13665. Springer, Cham. https://doi.org/10.1007/978-3-031-20065-6_2.

[7] Yinghao Xu, S. Peng, Ceyuan Yang, Yujun Shen, Bolei Zhou; 3D-Aware Image Synthesis via Learning Structural and Textural Representations IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022, pp. 18430-18439

[8] Haochen Wang, Xiaodan Du, Jiahao Li, Raymond A. Yeh, Greg Shakhnarovich; Score Jacobian Chaining: Lifting Pretrained 2D Diffusion Models for 3D Generation, IEEE/CVF, Computer Vision and Pattern Recognition (CVPR), 2023, pp. 12619-12629

[9] Renrui Zhang, Liuhui Wang, Yu Qiao, Peng Gao, Hongsheng Li; Learning 3D Representations From 2D Pre-Trained Models via Image-to-Point Masked Autoencoders, IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2023, pp. 21769-21780

[10] Volker Blanz, Thomas Vetter, A Morphable Model For The Synthesis Of 3D Faces, Seminal Graphics Papers: Pushing the Boundaries, Volume 2August 2023Article No.:1, 8Pages 157–164' ACM, https://doi.org/10.1145/3596711.3596730

[11] Gallucci, A.; Znamenskiy, D.; Long, Y.; Pezzotti, N.; Petkovic, M. Generating High-Resolution 3D Faces and Bodies Using VQ-VAE-2 with PixelSNAIL Networks on 2D Representations. Sensors 2023, 23, 1168. https://doi.org/10.3390/s23031168

[12] Menghan Li, Bin Huang, Guohui Tian, A comprehensive survey on 3D face recognition methods, Engineering Applications of Artificial Intelligence, Volume 110, 2022, 104669, ISSN 0952-1976, https://www.sciencedirect.com/science/article/pii/S0952197622000057, https://doi.org/10.1016/j.engappai.2022.104669

[13] Bhavani, SA, Karthikeyan, C. Robust 3D face recognition in unconstrained environment using Distance based ternary search Siamese network. Multimed Tools Appl (2023). https://doi.org/10.1007/s11042-023-17545-6

[14] W. Yang, Y. Zhao, B. Yang, and J. Shen, "Learning 3D Face Reconstruction from the Cycle-Consistency of Dynamic Faces," in IEEE Transactions on Multimedia, DOI: 10.1109/TMM.2023.3322895

[15] Chen, G.Y., Krzyzak, A. Face recognition via selective denoising, filter faces and hog features. SIViP (2023). https://doi.org/10.1007/s11760-023-02769-8

[16] Qi Wang, Hang Lei, and Weizhong Qian, Point CNN:3D Face Recognition with Local Feature Descriptor and Feature Enhancement Mechanism, Sensors 2023, https://doi.org/10.3390/s23187715, 6-9-2023.

[17] Wang, Guang Xue, et al. A New Multi-Scale Harris Interesting Point Detector. Applied Mechanics and Materials, vol. 602–605, Trans Tech Publications, Ltd., Aug. 2014, pp. 1950–1955. Doi: 10.4028/www.scientific.net/amm.602-605.1950.

[18] Hata, Kenji, and Silvio Savarese. CS231A Course Notes 3: Epipolar Geometry." 18c.)/K. Hata, S. Savarese. 2018. http://web.Stanford.edu/class/cs231a/course notes/03-epipolar-eometry.