Evolutional Normal Maps: 3D Face Representations for 2D-3D Face Recognition, Face Modelling and Data Augmentation

Michael Danner[1,2], Thomas Weber[2], Patrik Huber[1,3], Muhammad Awais[1], Matthias Raetsch[2] and Josef Kittler[1]

[1]*Centre for Vision, Speech & Signal Processing, University of Surrey, Guildford, U.K.*

[2]*ViSiR, Reutlingen University, Reutlingen, Germany*

[3]*Department of Computer Science, University of York, York, U.K.*

Keywords: Deep Learning, Visual Understanding, Machine Vision, Pattern Recognition, 2D/3D Face Recognition, Local Describers, Normal-Vector-Map Representation.

Abstract: We address the problem of 3D face recognition based on either 3D sensor data, or on a 3D face reconstructed from a 2D face image. We focus on 3D shape representation in terms of a mesh of surface normal vectors. The first contribution of this work is an evaluation of eight different 3D face representations and their multiple combinations. An important contribution of the study is the proposed implementation, which allows these representations to be computed directly from 3D meshes, instead of point clouds. This enhances their computational efficiency. Motivated by the results of the comparative evaluation, we propose a 3D face shape descriptor, named Evolutional Normal Maps, that assimilates and optimises a subset of six of these approaches. The proposed shape descriptor can be modified and tuned to suit different tasks. It is used as input for a deep convolutional network for 3D face recognition. An extensive experimental evaluation using the Bosphorus 3D Face, CASIA 3D Face and JNU-3D Face datasets shows that, compared to the state of the art methods, the proposed approach is better in terms of both computational cost and recognition accuracy.

1 INTRODUCTION

Face recognition and matching are important technologies for many application scenarios, including identity verification, public security, human-computer interactions, person tracking and re-identification for process monitoring such as passenger progression through airports, secondary authentication for mobile devices, and for indexing into large multimedia archives of media and entertainment companies. During the last decade, face biometrics research has been dominated by 2D face recognition. This is primarily the consequence of the recent advent of deep neural networks, which can learn face image representations that are largely invariant to nuisance factors such as illumination and pose changes. However, since the first computer assisted face recognition system of Kanade (Kanade, 1977) half a century ago, there has always been interest in 3D face recognition as the ultimate target technology, that has the capacity to disentangle face skin texture from the 3D shape of this intrinsically 3D object, and use these two sources of biometric information in the most productive way.

The interest in 3D face recognition is evident not only from the continuing research on this topic in the literature (Kittler et al., 2005), but also from the advances in the development of 3D capture systems. 3D capable cameras based on binocular stereo vision and time of flight (ToF) technology are becoming more and more affordable. For example, Apple have been shipping user-facing depth cameras in their consumer mobile phones for a number of years now, and Sony have started developing next-generation 3D sensors with ToF technology. Clearly, the quality of images captured by such user devices is not comparable to the output of high-resolution 3D scanners and decreases rapidly with increasing distance, so that the working distance is typically just a few meters. However, if the lighting conditions are difficult, and if the subject is unconstrained in terms of facial expression and occlusions, then the recognition can be more effective than using purely 2D RGB images. Moreover, 3D sensing is also more robust to spoofing attacks.

Danner, M., Weber, T., Huber, P., Awais, M., Raetsch, M. and Kittler, J.
Evolutional Normal Maps: 3D Face Representations for 2D-3D Face Recognition, Face Modelling and Data Augmentation.
DOI: 10.5220/0010912000003124
In *Proceedings of the 17th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2022) - Volume 5: VISAPP*, pages 267-274
ISBN: 978-989-758-555-5; ISSN: 2184-4321

Our own interest in 3D face recognition is from the point of view of 3D assisted 2D face recognition. Sinha *et al.*(Sinha et al., 2006) showed that both photo-metric and shape cues are equally used by humans to recognise faces. The motivation for the 3D assisted 2D face recognition approach is the disentanglement of shape and texture, achieved by 3D face model fitting. The fitted 3D face should offer a better control and aggregation of the different sources of biometric information (shape and texture), as well as suppression of the effect of illumination, expression and pose. The 3D assisted 2D face recognition approach can also benefit from the availability of large 2D face databases, which are essential for effective machine learning.

In comparison with 2D, the progress in 3D face recognition has been hampered not only by expensive sensor hardware and lack of data for training, but also by the incongruence of 3D face representations in the form of 3D meshes with data structures enabling efficient processing by convolutional neural networks. However, this problem has recently been overcome by means of graph neural networks (Scarselli et al., 2009), and 3D face image remapping onto a 2D mage structure such as isomap or Laplacian map (Feng et al., 2018; Kittler et al., 2018). The additional problem of errors caused by reconstructing a 3D from its 2D projection has also been recently mitigated (Danner et al., 2019). Thus we are reaching the point when the vision of 3D assisted 2D face recognition is becoming realistic.

This paper is concerned with 3D face recognition in the context of 3D face biometrics per se, or as the ultimate step in 3D assisted 2D face recognition. Our approach involves mapping a 3D face mesh into 2D for CNN based matching. We confine our interest to the 3D shape information only, and investigate, how the face shape should be represented and in what form it should be provided to the neural network. We shall explore a number of alternatives to the raw 3D measurement information, and propose a novel representation, called *evolutional normal map*, which is shown to be very effective from the 3D face shape recognition point of view. We compare it with a number of existing representation methods, and demonstrate on several 3D databases, that it delivers impressive recognition accuracy.

The rest of the paper is organised as follows: Section 2 reviews the related work and recent approaches to 3D face representations and 3D assisted 2D face recognition. Section 3 discusses normal vector maps as an alternative to raw 3D measurement information and introduces our proposed Evolutional Normal Maps. Section 4 presents the results of an extensive evaluation of the proposed system and its comparison with the state of the art methods. We conduct experiments with Bosphorus 3D Face (Alyüz et al., 2008), CASIA 3D Face Database (Zhong et al., 2008) and JNU-3D dataset (Koppen et al., 2018) datasets to benchmark the 3D representations compared. Conclusions and future work are presented in section 5.

2 RELATED WORK

3D Face Recognition. A conventional 3D Face recognition approach comprises methods like 3D face landmarking, 3D face registration and facial feature extraction. The 3D face landmarking locates the geometric positions of reference points for the face. The 3D face registration aims to register 3D face scans in a coordinate system so that the adjustment of facial features can be carried out in a consistent manner. Extracting facial features means creating a distinctive face representation that should fully describe each 3D face scan. Kakadiaris *et al.* (Kakadiaris et al., 2017) already shows the effectiveness of a 3D-2D framework for face recognition. The main advantage is the more practical use than 3D-3D and higher accuracy for 2D-2D face recognition.

Face Reconstruction. Dou *et al.*proposed a way for Monocular 3D facial shape reconstruction from 2D facial images (Dou et al., 2017) in a more effective way than other approaches. Huber *et al.* (Huber et al., 2016) also showed the power of 3D Morphable Face Models in computer vision and the widely utilisable reconstruction of a 3D face from a single 2D image. These 3D Morphable Face Models can be used for pose estimation, analysis and recognition and also for facial landmark detection and tracking.

Depth Images and Point Clouds. Depth images, depth point clouds, or 3D meshes have emerged as an important tool and principle in biometrics and face recognition research. According to Kakadiaris *et al.* (Kakadiaris et al., 2017) the existing frameworks for face recognition vary across approaches (e.g. model-based, data-driven and perceptual) or facial data domains (e.g. images, point clouds, depth maps). The main benefits for the use of 3D facial scans such as depth maps are the insensitivity to ambient influences and the colour of the skin which could lead to missing details in 2D face images.

ToF- and Stereo-cameras. Structured-light RGB-D and Time-of-Flight (ToF) cameras are used in 3D perception tasks at close distances. While RGB-D cameras benefit from high resolution and frame rate, ToF-cameras have the capability to operate outdoors and perceive details (Alenyà et al., 2014).

Local Descriptors. For 3D face recognition tasks, local descriptors have been established due to their robustness in variations of illumination and facial expressions and achieved remarkable recognition rates. Much work is inspired by local descriptors like Radon transform (Jafari-Khouzani and Soltanian-Zadeh, 2005), Textons (Lazebnik et al., 2005) and Local Binary Patterns (LBP) (Ahonen et al., 2004). Thanks to the high performance and computationally low complexity and its flexibility to adapt, LBP has a huge number of improved and extended successors (Pietikäinen et al., 2011; Huang et al., 2006; Ouamane et al., 2017). High-order local pattern descriptors, like local derivative pattern on normal maps capture more detailed information by encoding various distinctive spatial relationships (Soltanpour and Wu, 2017; Soltanpour and Wu, 2019; Zhang et al., 2010). LBPs and other local descriptors will not be discussed in this work due to the huge number of variations and to the fact that this is a lossy conversion of 3D shape information. Further use in face modeling based on descriptors is therefore difficult. However, they can easily be adapted on the *evolutional normal maps* to further improve face recognition performance.

Normal-vector Map Representations. 3D polygonal surfaces are represented through their corresponding normal map, a bidimensional array which stores mesh normal vectors as the pixel's RGB components of a colour image. Abate *et al.*calculate the difference of two normal maps (Abate et al., 2005) and achieved remarkable results in face recognition based on these difference maps. In 3D face recognition, normal maps having x,y,z mapped to RGB for visualisation is a common task (Kakadiaris et al., 2006). Gilani *et al.* (Gilani and Mian, 2018; Gilani et al., 2018) used depth images and normal vectors with azimuth and elevation in the form of RGB images for training a deep convolutional neural network.

3 METHODOLOGY

The values of the normal map matrix are determined by a partial binary operation that maps the x and y coordinates into unit normal vectors. The horizontal and vertical coordinates define the resolution of the normal map and the range of this operation. In addition, the normal map matrix elements can be represented by other 2D parameterisations, such as the spherical, angle-based domain. The normal vectors can be computed from the positions of the 3D points and their neighbours while the surface is in a point cloud, a 3D mesh or a range image representation (Daoudi et al., 2013). The representation of the 3D face shape on a uniform square matrix that is needed for convolutional neural network, leads to the fact that all normal maps have the same resolution.

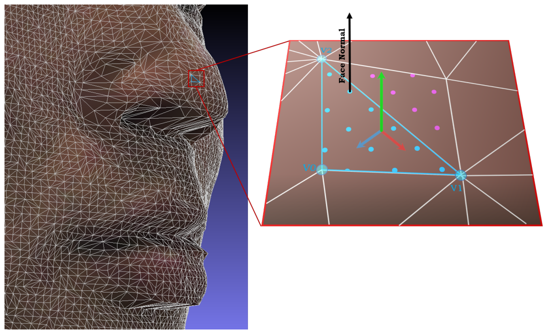

Figure 1: Each triangle is projected into the x,y-plane. the enclosed pixels are determined and the normal vector components are assigned to these pixels.

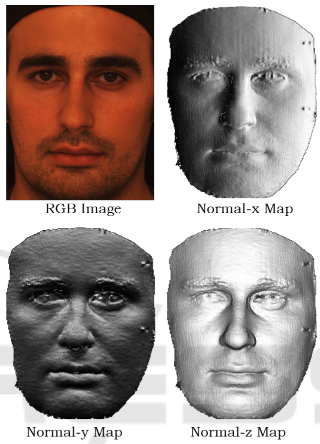

Figure 2: Portrait picture and normal maps created by conventional methods on Bosphorus 3D Face Database.

The surface normals contain more detailed and robust information compared to depth image for 3D data (Li et al., 2014). Our objective is to explore discriminative facial 3D representations in order to apply them to 3D facial machine vision tasks. We propose a refined feature extraction using the surface normals that provides richer and distinct information, as compared to the other 2D mesh representations.

3.1 Surface Normal Map

This work is inspired by recent algorithms (Mohammadzade and Hatzinakos, 2013; Li et al., 2014;

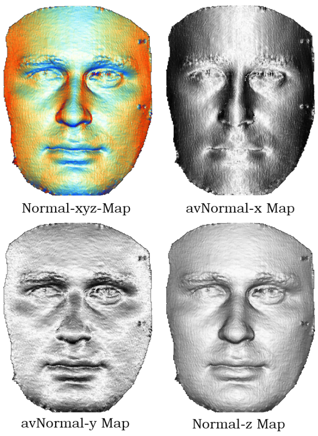

Figure 3: Evolutional normal maps on Bosphorus 3D Face Database.

Emambakhsh and Evans, 2017) in which surface normals have been applied for 3D face recognition. These works follow normal surface estimation methods presented by Klasing *et al.* (Klasing et al., 2009). Although the use of surface normals is a good practice for point clouds, it is not necessary in popular 3D Face Databases and registered scan scenarios. A 3D mesh consists of vertices and faces which can be used to calculate the normal vectors, instead of estimating.

At first, we have to ensure the mesh is registered to the x,y-plane and define the space where the vertices are located. The vertices x,y-coordinates will be mapped to the resulting normal map determining a transformation. This transformation assigns the coordinates from the domain of the mesh to the domain of the normal matrix.

Then, for each polygon in the mesh, we calculate triangles, if necessary. Each triangle consists of three vertices points $P = \{p_i, p_j, p_k\}$ where $p_n \in \mathbb{R}^3$. Two vectors are determined with $v_0 = p_j - p_i$ and $v_1 = p_k - p_i$. Then, a unit normal vector $\hat{n}$ can be calculated by

$$\hat{n} = \frac{n}{||n||} = \frac{v_0 \times v_1}{||v_0||\,||v_1||\,|\sin\theta|} \tag{1}$$

where $\hat{n} = [\hat{n}_x, \hat{n}_y, \hat{n}_z]^T$.

Furthermore, each triangle is projected into the x,y-plane and the enclosed pixels are determined by giving consideration to the targeted image resolution. The process in described in Figure 1.

Azimuth and Elevation. Given the surface unit normal vector $\hat{n} = (\hat{n}_x, \hat{n}_y, \hat{n}_z)$ at a point, the azimuth angle α is defined as the angle between the positive x-axis and the projection of n to the x-y plane. The elevation angle ϕ is the angle between n and the vector's orthogonal projection onto the xy plane. The elevation angle is positive when going toward the positive z-axis from the xy plane.

$$\alpha = \tan^{-1}\left(\frac{\hat{n}_z}{\hat{n}_x}\right) \tag{2}$$

$$\phi = \tan^{-1}\left(\frac{\hat{n}_y}{\sqrt{(\hat{n}_x^2 + \hat{n}_z^2)}}\right) \tag{3}$$

For each pixel the normal vector component is stored in a 2D-matrix, resulting in five images for x, y, z, azimuth and elevation values. Optionally, the minimum and maximum of each matrix can be determined and the histogram values can be stretched to the full range.

Pixels in these matrices with no assigned values are considered as background and the background value is assigned to them, which is 0 by default. Finally, the matrices N_x, N_y, N_z are surface normal maps having respectively x, y and z dimensions.

3.2 Evolutional Normal Maps

We now have several methods for calculation and estimation of normals and the corresponding algorithms for depth map and multiple normal maps. The question is, can the normal maps and representations based on them still be improved without loss of information? Applying local descriptors lead to increased histogram information which is beneficial to face recognition but 3D shape information is lost. Indeed, based on the fact that increasing histogram information is helpful, the minimum and maximum values of the normal component matrices are used to stretch them to the full range. Additionally, on the matrices with normal component x and y, the absolute value is used which doubles the gradients from bright to dark. These techniques are loss-free and can be facilely reversed.

Another method to manipulate histogram values is used in inverting the values of the matrix to swap bright and dark grey tones. In the experimental evaluation, diverse operations are applied to prove this as a valuable operation on normal-maps.

It should be emphasised that the methods mentioned only represent part of the possibilities with improving normal maps. That is why we call the set of 3D representations 'evolutional' normal maps because there are lots of parameters to adjust the outcome and they can be flexibly adapted to many facial 3D machine vision tasks: (a) calculate N_x, N_y, N_z, N_a, N_e, Depth map or any combination of these. (b) Invert greyscale of the normal map. (c) Darken or lighten normal maps. (d) Calculate real

Figure 4: Example ENM set of images from Bosphorus 3D (left) and CASIA 3D (right): 1st row from left to right: Absolute-Values-Normal-X-map, Absolute-Values-Normal-Y-map, Normal-Z-map, Combination of avN_x, avN_y, N_z; 2nd row from left to right: Normal-X-map, Normal-Y-map, Normal-Z-map, Combination of N_x, N_y, N_z; 3rd row from left to right: depth map, Azimuth-Normal-map, Elevation-Normal-map, Combination of depth map, N_a, N_e.

or absolute normal values. (e) Adjust image resolution. (f) Horizontal and vertical rotation on the mesh for data augmentation. (g) Accessible for local describers.

4 EXPERIMENTAL EVALUATION

The datasets used in this work are CASIA, Bosphorus and JNU-3D. The CASIA 3D FaceV1 Database contains 4,674 scans of 123 subjects, where each subject is captured with more than 35 different expressions and poses. Bosphorus 3D Face Database comprises 105 identities and 4,666 scans and is popular because of its rich repertoire of expressions. Beside 3D face recognition, this database is often used for expression recognition and facial action unit detection. The JNU-3D data set consists of 774 3D faces and is used for augmenting the 3D database and for accuracy tests.

Evaluation Protocol. We evaluate the depth image and the normal maps on face recognition accuracy. For the benchmark training the normal map size is fixed to 300x300 since previous experiments showed this as an efficient resolution on recognition rate and training duration.

In total, we evaluate the accuracy with 16 different experiments shown in fig. 4 and fig. 5 . Therefore, as a run of the experiment we defined the basis, the "pure" depth map D. Followed up by the 8 different normal maps in x-dimension N_x and avN_x, y-dimension N_y and avN_y, z-dimension N_z and avN_z, azimuth angle N_a and the elevation angle N_e. Accordingly the dimension of the network input layer is (1, 300, 300).

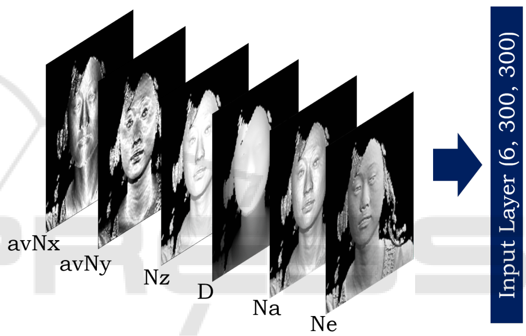

Figure 5: Multiple combinations of input images for the face recogniser from single channel grey scale images up to six channel $D + avN_{xy} + N_{zae}$

In the second run we evaluate the combination of avN_x, avN_y, avN_z, called avN_{xyz}, the combination of N_x, N_y, N_z, called N_{xyz} and the combination of D, N_a , N_e, called $D + N_{ae}$ also shown in the fig. 4. Accordingly the dimension of the network input layer is (3, 300, 300).

The final run includes the combination of the depth map and the five normal maps, called $D + N_{xyzae}$. Accordingly the dimension of the network input layer is (6, 300, 300). This input for the face recognition network is shown in fig. 5.

Each representation is trained on our face recognition network for 60 epochs on Bosphorus 3D Database, CASIA 3D and JNU-3D, respectively. On Bosphorus, all images are used, including facial expression and partial occlusion. The first neutral sample of each subject is used as the gallery and the remaining scans as the probe (neutral vs. all).

Since we want to evaluate the pure 3D representations we neither apply any data augmentation methods for the training data nor use local binary pattern

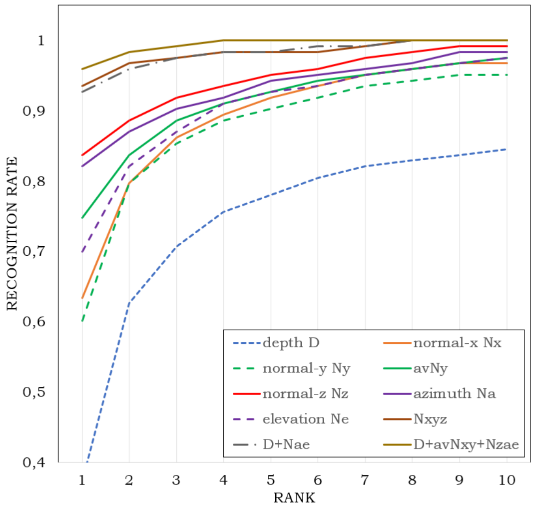

Figure 6: Recognition rate on CASIA 3D data set.

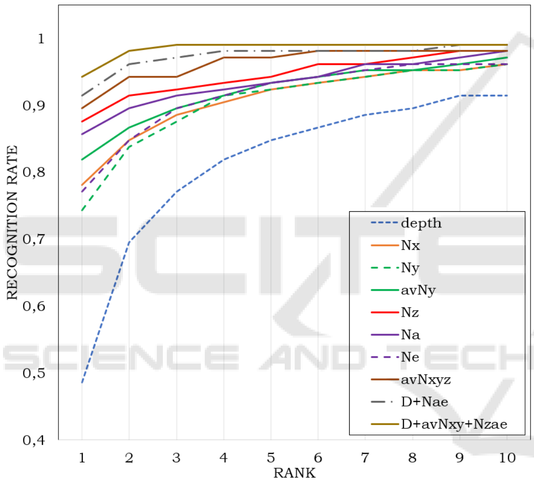

Figure 7: Recognition rate on Bosphorus 3D Face.

layers. Actually, the experiments show that with combining normal maps, we are able to achieve recognition rates equal to or better than previous work with lots of augmentation and adapted layers.

Secondary evaluation, using CASIA 3D Database, is used under same conditions. The first scan of each subject is considered as gallery and finally the JNU-3D data set is used for accuracy tests on small networks.

Results. To illustrate recognition efficiency, the Cumulative Match Characteristics (CMC) curves on CASIA 3D are presented in fig. 6. Results from table 1 show our ENM achieves the best results. High detection rates in CASIA and Bosphorus are already scored without data augmentation or the use of local binary patterns or similar descriptors. Table 2 presents the

Table 1: Rank 1 recognition rate [%] in results on Bosphorus and CASIA 3D.

	Bosphorus	CASIA
depth map	48.6	37.4
N_x	78.1	63.4
N_y	74.3	60.2
N_z	87.6	83.7
N_a	85.7	82.1
N_e	77.1	69.9
$D+N_{ae}$	91.4	92.7
ENM avN_y	81.9	74.8
ENM avN_{xyz}	89.5	93.5
ENM 6-layer	**94.3**	**95.9**

Table 2: Rank 1 recognition rate [%] results on Bosphorus and CASIA 3D with augmented data sets.

	Bosphorus	CASIA
GoogleNet[1]RGB	63.4	85.9
Resnet152[2]RGB	7.1	52.9
VGG-Face[3]RGB	96.4	94.1
GoogleNet 3D	26.8	50.8
Resnet152 3D	3.8	25.3
VGG-Face 3D	48.1	72.0
$D+N_{ae}$ FR3DNet[4]	**100.0**	99.7
ENM avN_{xyz}	100.0	97.3
ENM 6-layer	**100.0**	**99.8**

performance of the proposed method in recognising the 1^{st} rank of the 3D faces by training with augmented data. It achieves the best performance for CASIA 3D faces and measures up to FR3DNet on Bosphorus 3D with 100% rank 1 recognition rate. The superiority of proposed ENM is also reflected in fig. 7.

5 CONCLUSIONS AND FUTURE WORK

We addressed the problem of 3D face shape representation with a focus on 2D normal maps. These maps are various functions of the 3D surface normals, that are defined on a 3D face mesh and mapped systematically onto a 2D image. We showed that these normal maps can be computed very efficiently from the 3D shape mesh.

An extensive comparative evaluation of multiple variants of these normal maps and their combinations has been carried out using face recognition accuracy as a measure of their effectiveness. Motivated by the results of the comparative study, we proposed a 3D face shape descriptor, referred to as Evolutional Normal Maps, that assimilates, modifies and optimises a subset composed of six of these normal map representations. The proposed descriptor is extensively evaluated on three benchmarking 3D face datasets with very promising results. The descriptor is computationally efficient and most importantly, it outperforms the state of the art methods in 3D face recognition.

The proposed Evolutional Normal Maps have many potential applications, apart from 3D face recognition. For instance, can can be used for data augmentation and in generative adversarial networks to render and develop synthetic 3D data. The future plans also include their use for 3D face reconstruction and 3D face analysis for emotion, pose and age estimation.

ACKNOWLEDGEMENTS

Part of the research in this paper uses the CASIA-3D FaceV1 collected by the Chinese Academy of Sciences Institute of Automation (CASIA).

This work is partially supported by a grant of the BMWi ZIM program, no. ZF4029424HB9.

REFERENCES

Abate, A. F., Nappi, M., Ricciardi, S., and Sabatino, G. (2005). Fast 3d face recognition based on normal map. In *Proceedings of the 2005 International Conference on Image Processing, ICIP 2005, Genoa, Italy, September 11-14, 2005*, pages 946–949. IEEE.

Ahonen, T., Hadid, A., and Pietikäinen, M. (2004). Face recognition with local binary patterns. In Pajdla, T. and Matas, J., editors, *Computer Vision - ECCV 2004, 8th European Conference on Computer Vision, Prague, Czech Republic, May 11-14, 2004. Proceedings, Part I*, volume 3021 of *Lecture Notes in Computer Science*, pages 469–481. Springer.

Alenyà, G., Foix, S., and Torras, C. (2014). Using tof and RGBD cameras for 3d robot perception and manipulation in human environments. *Intelligent Service Robotics*, 7(4):211–220.

Alyüz, N., Gökberk, B., Dibeklioglu, H., Savran, A., Salah, A. A., Akarun, L., and Sankur, B. (2008). 3d face recognition benchmarks on the bosphorus database with focus on facial expressions. In Schouten, B. A. M., Juul, N. C., Drygajlo, A., and Tistarelli, M., editors, *Biometrics and Identity Management, First European Workshop, BIOID 2008, Roskilde, Denmark, May 7-9, 2008. Revised Selected Papers*, volume 5372 of *Lecture Notes in Computer Science*, pages 57–66. Springer.

Danner, M., Raetsch, M., Huber, P., Awais, M., Feng, Z., and Kittler, J. (2019). Texture-based 3D face recognition using deep neural networks for unconstrained human-machine interaction. In *Proceedings of the 15th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications VISAPP 2020*. SCITEPRESS.

Daoudi, M., Srivastava, A., and Veltkamp, R. (2013). *3D Face Modeling, Analysis and Recognition*. Wiley.

Dou, P., Shah, S. K., and Kakadiaris, I. A. (2017). End-to-end 3d face reconstruction with deep neural networks. In *2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, July 21-26, 2017*, pages 1503–1512. IEEE Computer Society.

Emambakhsh, M. and Evans, A. N. (2017). Nasal patches and curves for expression-robust 3d face recognition. *IEEE Trans. Pattern Anal. Mach. Intell.*, 39(5):995–1007.

Feng, Z.-H., Huber, P., Kittler, J., Hancock, P., Wu, X.-J., Zhao, Q., Koppen, P., and Rätsch, M. (2018). Evaluation of dense 3D reconstruction from 2D face images in the wild. In *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*, pages 780–786. IEEE.

Gilani, S. Z. and Mian, A. (2018). Learning from millions of 3d scans for large-scale 3d face recognition. In *2018 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2018, Salt Lake City, UT, USA, June 18-22, 2018*, pages 1896–1905. IEEE Computer Society.

Gilani, S. Z., Mian, A. S., Shafait, F., and Reid, I. (2018). Dense 3d face correspondence. *IEEE Trans. Pattern Anal. Mach. Intell.*, 40(7):1584–1598.

He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27-30, 2016*, pages 770–778. IEEE Computer Society.

Huang, Y., Wang, Y., and Tan, T. (2006). Combining statistics of geometrical and correlative features for 3d face recognition. In Chantler, M. J., Fisher, R. B., and Trucco, E., editors, *Proceedings of the British Machine Vision Conference 2006, Edinburgh, UK, September 4-7, 2006*, pages 879–888. British Machine Vision Association.

Huber, P., Hu, G., Tena, J. R., Mortazavian, P., Koppen, W. P., Christmas, W. J., Rätsch, M., and Kittler, J. (2016). A multiresolution 3D morphable face model and fitting framework. In Magnenat-Thalmann, N., Richard, P., Linsen, L., Telea, A., Battiato, S., Imai, F. H., and Braz, J., editors, *Proceedings of the 11th Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2016) - Volume 4: VISAPP, Rome, Italy, February 27-29, 2016*, pages 79–86. SciTePress.

Jafari-Khouzani, K. and Soltanian-Zadeh, H. (2005). Radon transform orientation estimation for rotation invariant texture analysis. *IEEE Trans. Pattern Anal. Mach. Intell.*, 27(6):1004–1008.

Kakadiaris, I. A., Passalis, G., Toderici, G., Murtuza, M. N., and Theoharis, T. (2006). 3d face recognition. In

Chantler, M. J., Fisher, R. B., and Trucco, E., editors, *Proceedings of the British Machine Vision Conference 2006, Edinburgh, UK, September 4-7, 2006*, pages 869–878. British Machine Vision Association.

Kakadiaris, I. A., Toderici, G., Evangelopoulos, G., Passalis, G., Chu, D., Zhao, X., Shah, S. K., and Theoharis, T. (2017). 3d-2d face recognition with pose and illumination normalization. *Comput. Vis. Image Underst.*, 154:137–151.

Kanade, T. (1977). Computer recognition of human faces. *Interdisciplinary Systems Research*, 47.

Kittler, J., Hilton, A., Hamouz, M., and Illingworth, J. (2005). 3d assisted face recognition: A survey of 3d imaging, modelling and recognition approachest. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR Workshops 2005, San Diego, CA, USA, 21-23 September, 2005*, page 114. IEEE Computer Society.

Kittler, J., Koppen, P., Kopp, P., Huber, P., and Rätsch, M. (2018). Conformal mapping of a 3D face representation onto a 2D image for CNN based face recognition. In *2018 International Conference on Biometrics, ICB 2018, Gold Coast, Australia, February 20-23, 2018*, pages 124–131. IEEE.

Klasing, K., Althoff, D., Wollherr, D., and Buss, M. (2009). Comparison of surface normal estimation methods for range sensing applications. In *2009 IEEE International Conference on Robotics and Automation, ICRA 2009, Kobe, Japan, May 12-17, 2009*, pages 3206–3211. IEEE.

Koppen, W. P., Feng, Z., Kittler, J., Awais, M., Christmas, W. J., Wu, X., and Yin, H. (2018). Gaussian mixture 3d morphable face model. *Pattern Recognit.*, 74:617–628.

Lazebnik, S., Schmid, C., and Ponce, J. (2005). A sparse texture representation using local affine regions. *IEEE Trans. Pattern Anal. Mach. Intell.*, 27(8):1265–1278.

Li, H., Huang, D., Morvan, J., Chen, L., and Wang, Y. (2014). Expression-robust 3d face recognition via weighted sparse representation of multi-scale and multi-component local normal patterns. *Neurocomputing*, 133:179–193.

Mohammadzade, H. and Hatzinakos, D. (2013). Iterative closest normal point for 3d face recognition. *IEEE Trans. Pattern Anal. Mach. Intell.*, 35(2):381–397.

Ouamane, A., Boutellaa, E., Bengherabi, M., Taleb-Ahmed, A., and Hadid, A. (2017). A novel statistical and multiscale local binary feature for 2d and 3d face verification. *Comput. Electr. Eng.*, 62:68–80.

Parkhi, O. M., Vedaldi, A., and Zisserman, A. (2015). Deep face recognition. In Xie, X., Jones, M. W., and Tam, G. K. L., editors, *Proceedings of the British Machine Vision Conference 2015, BMVC 2015, Swansea, UK, September 7-10, 2015*, pages 41.1–41.12. BMVA Press.

Pietikäinen, M., Hadid, A., Zhao, G., and Ahonen, T. (2011). *Computer Vision Using Local Binary Patterns*, volume 40 of *Computational Imaging and Vision*. Springer.

Scarselli, F., Gori, M., Tsoi, A. C., Hagenbuchner, M., and Monfardini, G. (2009). The graph neural network model. *IEEE Trans. Neural Networks*, 20(1):61–80.

Sinha, P., Balas, B. J., Ostrovsky, Y., and Russell, R. (2006). Face recognition by humans: Nineteen results all computer vision researchers should know about. *Proceedings of the IEEE*, 94(11):1948–1962.

Soltanpour, S. and Wu, Q. J. (2019). Weighted extreme sparse classifier and local derivative pattern for 3d face recognition. *IEEE Trans. Image Process.*, 28(6):3020–3033.

Soltanpour, S. and Wu, Q. M. J. (2017). High-order local normal derivative pattern (LNDP) for 3d face recognition. In *2017 IEEE International Conference on Image Processing, ICIP 2017, Beijing, China, September 17-20, 2017*, pages 2811–2815. IEEE.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S. E., Anguelov, D., Erhan, D., Vanhoucke, V., and Rabinovich, A. (2015). Going deeper with convolutions. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2015, Boston, MA, USA, June 7-12, 2015*, pages 1–9. IEEE Computer Society.

Zhang, B., Gao, Y., Zhao, S., and Liu, J. (2010). Local derivative pattern versus local binary pattern: Face recognition with high-order local pattern descriptor. *IEEE Trans. Image Process.*, 19(2):533–544.

Zhong, C., Sun, Z., and Tan, T. (2008). Learning efficient codes for 3d face recognition. In *Proceedings of the International Conference on Image Processing, ICIP 2008, October 12-15, 2008, San Diego, California, USA*, pages 1928–1931. IEEE.