Information & Security

M. Caruana & J. G. Vella,
vol. 47, no. 3 (2020): 328-340
https://doi.org/10.11610/isij.4724

Published since 1998

ISSN 0861-5160 (print), ISSN 1314-2119 (online)

Research Article

3D Facial Reconstruction from 2D Portrait Imagery

***Matthew Caruana, Joseph G. Vella* (✉)**

Department of Computer Information Systems, University of Malta,
https://www.um.edu.mt/

ABSTRACT:

3D facial images are reconstructed from 2D portraits using regression trees for facial landmark alignment and 3D morphable models. Two generic regression trees were adopted, one being based on the widely used 68-landmark structure, and the other based on a 74-landmark structure. The FaceWarehouse dataset was used to create a novel 74-landmark regression tree and during the system's evaluation. The accuracy of the models generated was computed through the Root Mean Square, 75th Percentile and Arithmetic Mean comparison metrics. Two different datasets of 2D images were reconstructed. The evaluation results demonstrate that a higher level of accuracy and precision was attained from the models reconstructed using 68-landmark regression tree when compared to the 74 developed here. The accuracy produced by the 68-landmark regression tree applied to two sets was 85 % and 90 % as opposed to the 82 % and 83 % produced by the 74-landmark regression tree on the same model subsets; thus justifying its wide adoption.

ARTICLE INFO:

RECEIVED: 11 JUNE 2020

REVISED: 23 AUG 2020

ONLINE: 14 SEP 2020

KEYWORDS:

digital forensics, forensic facial reconstruction, landmark alignment

Introduction

Facial Reconstruction is a vast area and even though it is researched extensively, there is much scope for improvement. Facial Reconstruction is being used in Facial Rehabilitation after extensive face traumas,[1] and Facial Recognition in forensic investigations.[2] Moreover, it has other applications that may be of benefit to our society, namely, in locating missing people, where it could be useful to have 3D facial models to complement the available material to base their

✉ E-mail: joseph.g.vella@um.edu.mt

searching on. In academia and industry various approaches have been put forward in order to achieve the best results possible to meet these domain requirements; these approaches' main function is to compare an input portrait image that is in a 2D format to a set of 3D models and find a 3D model that is the better fit.[3] The production of a specific 3D model is generated by a set of algorithms that are applied to the input image. These ensure that the specific 3D model produced is devoid of any occultation (e.g. specs), disguise (e.g. beard), skin texture and facial expressions.

As described by Blanz and Vetter,[4] "Reconstructing the true 3D shape and texture of a face from a single image is an ill-posed problem." Consequently, this problem does not have a specific solution as it relies heavily on the input data provided. Nonetheless, the inclusion of datasets that contain images annotated with appropriate landmarks is beneficial for the best results to be extracted through the models generated. Furthermore, reconstruction of faces from images is currently considered to be a computationally heavy procedure. Finally, setting up and testing of the various available algorithms is required when accepting a wide array of input 2D images. Consequently, it is expected training dataset choice does affect the performance and results generated by the same algorithms.[5]

Aims & Objectives

The aim of this application is to address the computation of facial reconstruction and provide a platform that generates and evaluates a 3D model from a single 2D input image. To undertake reconstruction the system has to create regression trees fitting features extracted from the landmarks found within the 2D input image. The texture from the facial image is also extracted for its later application on the produced 3D model. To achieve a good regression, various datasets that contain portrait images and their annotated landmarks need to be available.

Moreover, the developed system needs an efficient and an effective approach to reconstruct a 3D model from an input 2D image of a human face. An important part of the system is to ensure tests are administered using various unit test suites during its development, and during its use any 3D model created from the 2D image is evaluated using 2D and 3D model comparison techniques.

Background & Related Work

Landmark Alignment

A crucial part of most facial reconstruction projects is the ability to identify and locate key human facial landmarks in a portrait image (see Figure 1). In most cases, having too many landmarks requires inordinate overall execution time for extracting them, however, having few landmarks might make the overall accuracy or quality of data yield subpar results.[6] Therefore, consideration must be undertaken with regards to the number of landmarks chosen to ensure that the best results possible are extracted.

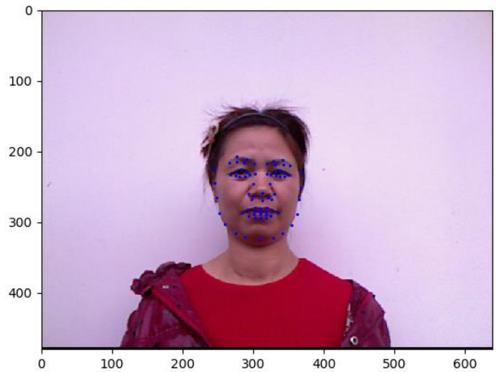


Figure 1: Results of Landmark Alignment.[7]

To capture the information shown in Table 1, regression trees, as proposed by Kazemi and Sullivan,[8] are applied to efficiently estimate the location of these facial landmarks. The regression tree takes the training set consisting of the images and their landmark positions to regress a model based on a set of hyper-parameters to be able to align the same landmark collection to the input image. Table 1 describes these hyper-parameters; to produce an acceptable regression tree one must balance speed, accuracy and model size and this is done through these parameters.

Active Shape Models[9] are often used instead of regression trees for the alignment of facial landmarks.

3D Model Generation

Throughout the various methods quoted in the literature, facial reconstruction projects make use of a number of 2D portrait images rather than a single image, hence making them even more computationally expensive processes for reconstructing a 3D model.[10]; [11]

3D morphable models as proposed by Blanz and Vetter,[4] make use of facial landmarks identifiable on a 2D image and to then have them applied to a reference 3D model based on a morphable model. A 3D morphable model is a database of 3D surfaces, where each vertex within a model corresponds to the same vertex of all the other surfaces within the database. While the morphable model can be a database of any structure, i.e. not only human faces, it is generally aimed towards the morphing of 3D human faces. With this approach, each morphable model (i.e. Surrey Face Model[12] and Basel Face Model[13]) makes use of Principle Component Analysis to learn from the database of 3D surfaces that the morphable models are based on.

For the deformations to be applied to the respective morphable model a mapping from the landmarks to their corresponding vertex in the reference 3D model is required.[12]

Table 1: Hyper-parameter Descriptions for Regression Trees[8]

Hyper-Parameter Name	Hyper-Parameter Description
Tree Depth	The total depth of the trees in each cascade.
Regularisation	By how much the algorithm will generalise on the training data.
Cascade Depth	The number of cascades that the model will be train with.
Feature Pool Size	The number of pixels that are used to generate features for the random trees.
Test Split Amount	The number of times that the training set is used during training.
Oversampling Amount	The amount of times the training data is augmented when training the regression tree.
Oversampling Translation Jitter	The amount of augmentation that is applied to the images in the dataset.

Other approaches that are attributed for model generation are Shape from Shading (SFS)[14] and the use of Convolutional Neural Networks (CNNs).[15]

Existing Systems

A number of existing projects tackle facial reconstruction from 2D imagery. Currently, these systems heavily rely on the inclusion of a number of input images from different angles, which in most cases is difficult to have. Moreover, many rely on landmark alignment by using regression trees with the 68 landmark structure,[12; 16; 17] meaning results are limited in showing the effects that the set of landmarks generated has on the overall accuracy of the model in relation to the input image.

Establishing which implementation provides the better results is difficult due to the tests being reported using different datasets, images, morphable models, and evaluation techniques.[18]

Datasets

In conjunction to the methodologies surrounding facial reconstruction, the choice of the dataset with regards to training and evaluation is an important one. The FaceWarehouse dataset,[7] maintained by Chen et al. has 150 different individuals and with each individual having 20 distinct images showcasing different poses from a frontal angle. Supplementing the images, FaceWarehouse also provides 74 facial landmarks per image, and a 3D model blend shape for that person's specific pose. The Celebrity Face Recognition dataset[19] is another dataset. It is collated by Prateek Mehta and focuses on 1100 celebrities and contains 800 thousand images of them. Both these datasets are available and have been extensively used here.

Design & Implementation

The system was developed using the Python development environment and availing of well-known packages. These include, OpenCV, dlib, and eos, that focus on image processing, landmark alignment, and 3D model morphing respectively. The system follows the structures and processes depicted in Figure 2.

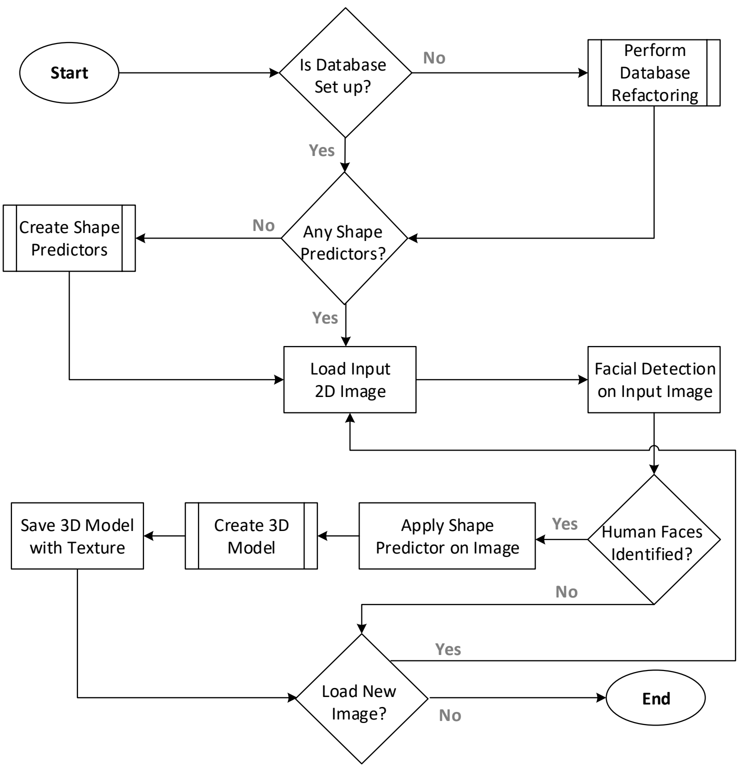


Figure 2: High Level System Processes for 3D Facial Reconstruction.

Landmark Alignment

In facial reconstruction project, landmark alignment is an essential process in order to transition from an input 2D image to a 3D model. It consists of two main phases:

The first phase involves the creation of the regression tree structure. Here the regression tree created is based on images and landmark structure provided in the FaceWarehouse dataset. As described previously, this process depends on the identification of the best set of hyper-parameters given in Table 1. This is achieved through an automated procedure that iteratively updates the hyper-parameter values in order to decrease the total amount of error produced during each testing phase. The hyper-parameters bounds and the optimal values found and adopted here are as shown in Table 2.

Table 2: Lower and Upper Bound and Optimal Hyper-Parameters Values for Regression Trees.

Hyper-Parameter Name	Lower-Bound	Upper-Bound	Optimal
Tree Depth	2	8	3
Regularisation	0.01	1	0.0134
Cascade Depth	10	25	25
Feature Pool Size	100	2000	578
Test Split Amount	20	250	249
Oversampling Amount	1	20	19
Oversampling Translation Jitter	0	0.3	0.1381

In the second phase, the input 2D image is used in conjunction to the optimised regression tree produced and then an alignment of the landmarks on that same image is undertaken. It is important to note that, the quantity and structure of the landmarks used in this phase is associated to the number and formation of the landmarks that the regression tree used for its training. As a result of this phase, the landmark structure is extracted and is used in reconstructing the final 3D model (see Figure 3.a).

Implementation of Facial Reconstruction

The process of reconstructing the 3D face is based on the 3D morphable model and is chosen because different types of face shape representations can be used. Also, this method does not require any additional information about the input image (e.g. illumination, face angles) which other methods are dependent on. Moreover, given the input required, the efficiency and accuracy are of a higher level compared to the other approaches as stated in, for example, Kazemi and Sullivan.[8]

The Surrey Face Model[12] is used to reconstruct the face of the 3D model. Since the Surrey Face Model focuses primarily on the frontal part of the face, it is deemed an adequate fit here. This is heavily reliant on setting up mappings between the 3D morphable model's vertices as once these features of the 3D Morphable Model are known, the landmarks of the face found within the input image are fitted onto the reference model. The fitting is done through an iterative process where the vertex and landmark pairs are compared based on the picture frame. Then, the appropriate adjustments of the vertex positions, as well as, the surrounding vertices positions are applied relative transformations to ensure that the general face structure of the model is contained whilst applying the transformations. By completing all iterations of the process, an untextured fitted shape model is constructed alongside the pose which dictates the angle of the face as shown in the input image (See Figure 3.a).

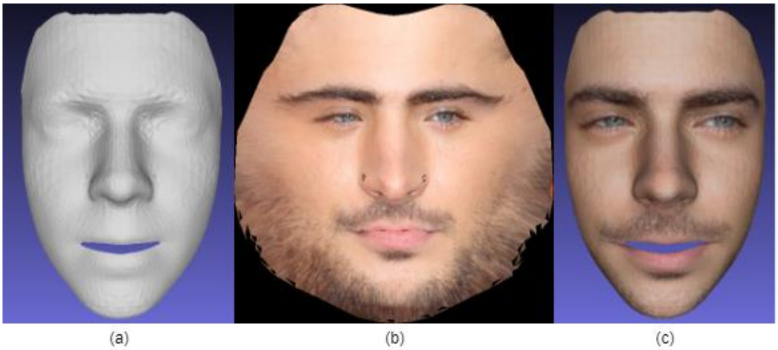


Figure 3: (a) Reconstructed 3D model; (b) Extracted Texture; (c) Textured 3D model.

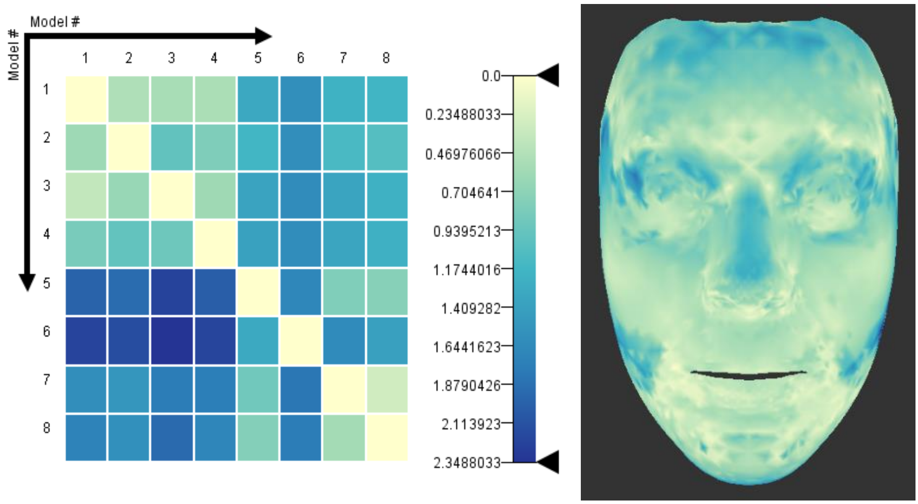
Evaluation Analysis

For the evaluation process, the Fidentis Analyst Software Application[20] is used to analyse the quality of the models generated. This was done using its Batch Processing feature which allows for all 3D models within the testing set to be evaluated using Fidentis Analyst's built in metrics to generate the similarity values of the generated models to one another. More specifically, basing the evaluation on the degree of similarity that is identified between 3D models that are of the same person from different images, while also being dissimilar from 3D models of different people.

The evaluation here used the Root Mean Square, 75th Percentile and the Arithmetic Mean metrics[21] with the value closest to zero is considered to be the closest model. In conjunction to this, Fidentis Analyst provides the ability to view the regions of the models that see a large amount of variance from the mean shape (See Figure 4.b). Also, heatmaps are generated to indicate the level of similarity that is exhibited from one model to the others (See Figure 4.a).

From the evaluation results, confusion matrices are constructed that enable expression of the level of accuracy and precision that is achieved within that particular set.

With regards to data used in the evaluation process, a total of four model sets were used and with each model set containing pairs of images of the same person. Two of model sets are constructed using a subset from the FaceWarehouse images dataset and with the regression tree 74-landmarks and as well as the 68-landmark regression tree widely used in various literature (these have been



(a) (b)

Figure 4: (a) Heat map Example (lighter is better); (b) Vertex variation model (lighter is of lesser variation).

Table 3: Fidentis[20] Settings Values.

Setting Name	Value
Method	Surface (ICP)
ICP Metric	Vertex to Vertex
Use Scale	True
Error Rate	0.05
Max Iterations	15
Average Meshes	8
Comparison	Nearest Neighbour Distance

named FW – 74-RT and FW – 68-RT respectively). The other two model sets are constructed using a subset from the Celebrity Face Recognition under the same regression trees previously mentioned (named Celeb – 74-RT and Celeb – 68-RT respectively). To create a uniform evaluation of the system, the settings for Fidentis, as stated in Table 3, are used during all evaluations.

Results and Evaluation

In this section the evaluation process of the results acquired through the use of the developed system are given. To ensure that different conditions are met within the system, numerous tests have been executed so as to better analyse the results acquired under such circumstances.

Evaluation of 74-landmark Based Reconstruction

Every model within each training set exhibiting the models generated using the 74-landmark regression tree was compared to the other models within that same training set for an aggregate of 192 total comparisons. Upon computing the Root Mean Square, 75^{th} Percentile, and the Arithmetic Mean, the accuracy and precision was deduced on that specific testing set.

As can be seen in Table 4 and Table 5, both evaluation sets provide similar results with the ability to identify models of the same person. From these results, the accuracy and precision values are, 82% and 65% for the FaceWarehouse subset using the 74-landmark regression tree, and 83% and 69% accuracy and precision for the Celebrity subset using the 74-landmark regression tree.

Table 4: Results of FaceWarehouse Subset using 74-landmark Regression Tree.

	Expected: False	**Expected: True**
Actual: False	True Negative $\frac{127}{192} = 66\%$	False Negative $\frac{17}{192} = 9\%$
Actual: True	False Positive $\frac{17}{192} = 9\%$	True Positive $\frac{31}{192} = 16\%$

Table 5: Results of Celebrity Subset using 74-landmark Regression Tree.

	Expected: False	**Expected: True**
Actual: False	True Negative $\frac{129}{192} = 67\%$	False Negative $\frac{15}{192} = 8\%$
Actual: True	False Positive $\frac{15}{192} = 8\%$	True Positive $\frac{33}{192} = 17\%$

Evaluation of 68-landmark Based Reconstruction

Similar to the first two sets administered, the models constructed using the 68-landmark regression tree undergo the same comparison techniques.

From the information gathered through Table 6 and

Table 7, similar results are identified. The accuracy and precision acquired for the FaceWarehouse subset using the 68-landmark regression tree were 85% and 71% respectively. Meanwhile, the accuracy and precision acquired from the Celebrity subset using the same regression tree were 90% and 79% respectively.

Table 6: Results of FaceWarehouse Subset using 68-landmark Regression Tree.

	Expected: False	Expected: True
Actual: False	True Negative $\frac{130}{192} = 68\%$	False Negative $\frac{14}{192} = 7\%$
Actual: True	False Positive $\frac{14}{192} = 7\%$	True Positive $\frac{34}{192} = 18\%$

Table 7: Results of Celebrity Subset using 68-landmark Regression Tree.

	Expected: False	Expected: True
Actual: False	True Negative $\frac{134}{192} = 70\%$	False Negative $\frac{10}{192} = 5\%$
Actual: True	False Positive $\frac{10}{192} = 5\%$	True Positive $\frac{38}{192} = 20\%$

Comparison of Results

Comparing the results acquired from all the model sets, the accuracy and precision of each of the model sets were investigated individually.

Table 8 provides the accuracy and precision values that each model set provided through the confusion matrices. From this information, two observations are possible. The two most accurate and precise model sets in identifying the images of the same people are generated using the widely used 68-landmark regression tree. Also, this tells us that a more determining factor in the accuracy and precision is the regression tree used rather than the dataset used even though both generated an acceptable level of results.

Summary of Results

From the evaluation undertaken it can be concluded that the process of facial reconstruction is dependent on the regression tree used during the landmark alignment phase.

Table 8: Accuracy and Precision Results of each model set.

Name of Model Set	Accuracy	Precision
FW – 74-RT	82%	645 %
FW – 68-RT	85%	71%
Celeb – 74-RT	83%	69%
Celeb – 68-RT	90%	79%

Therefore, this implementation has functioned well when constructing a new regression tree based on the FaceWarehouse dataset with an accuracy of 82% and 83% on the FaceWarehouse data subset and the Celebrity Face Recognition data subset respectively. However, when comparing the results acquired from using the industry standard 68-landmark regression tree, to system's implementation of the regression tree, the 68-landmark regression tree produced an accuracy of 85% and 90% on the same model sets. This underlines the robustness and effectiveness of the 68-landmark regression tree when compared to the 74-landmark regression tree.

Conclusions & Future Work

The main aim of this study was to implement a system for Facial Reconstruction with and heavy emphasis on reporting its efficacy.

The resultant prototype provides an approach that manages to extract 3D facial models from a single 2D portrait image of a person. This was achieved through the implementation of a regression tree that aligns a predefined set of landmarks onto the input image and morphs a 3D morphable model into the desired shape based on those landmarks.

When comparing the results of the models generated under this system, the 68-landmark regression tree outperformed the 74-landmark regression tree by achieving an accuracy of 85% and 90% as opposed to the 82% and 83% accuracy for the same test sets. Hence, accuracy is dependent on the regression tree used during the landmark alignment phase of this system.

Python and the availability of third-party libraries and software applications made the development process manageable. Lastly, FaceWarehouse offered an effective dataset to enable a cohesive and relatively complete prototype.

Further enhancements are possible and a few include:

- The generalisation of landmarks despite the occurrence of occlusions that are visible within a portrait image.
- The ability to execute a wider range of hyper-parameter search to ensure the shape predictor is taught in the best conditions.
- The inclusion of more dedicated 3D model comparison algorithms to compare and contrast the resultant models produced and how each model relates from one another.

References

[1] Gary Storey, Richard Jiang, and Ahmed Bouridane, "Role for 2D Image Generated 3D Face Models in the Rehabilitation of Facial Palsy," *Healthcare Technology Letters* 4, no. 4 (Aug 2017): 145-148. https://doi.org/10.1049/htl.2017.0023.

[2] Dalong Jiang, Yuxiao Hu, Shuicheng Yan, Lei Zhang, Hongjiang Zhang, and Wen Gao, "Efficient 3D Reconstruction for Face Recognition," *Pattern Recognition* 38, no. 6 (June 2005): 787–798, https://doi.org/10.1016/j.patcog.2004.11.004.

[3] Dang-hui Liu, Lan-sun Shen, and Kin-man Lam, "Image synthesis and face recognition based on 3d face model and illumination model," in *Proceedings of the First international conference on Advances in Natural Computation - Volume Part II (ICNC'05)*, Springer-Verlag, 2005, pp. 7-11, https://doi.org/10.1007/11539117_2.

[4] Volker Blanz and Thomas Vetter, "A morphable model for the synthesis of 3D faces," in *Proceedings of the 26th annual conference on Computer graphics and interactive techniques (SIGGRAPH '99)*, ACM Press/Addison-Wesley Publishing Co., USA, 1999, pp. 187–194, https://doi.org/10.1145/311535.311556.

[5] Aaron Jackson, Adrian Bulat, Vasileios Argyriou, and Georgios Tzimiropoulos, "Large Pose 3D Face Reconstruction from a Single Image via Direct Volumetric CNN Regression," *IEEE International Conference on Computer Vision, ICCV 2017, Italy, October 22-29,* 2017, pp. 1031-1039, https://doi.org/10.1109/ICCV.2017.117.

[6] Christos Sagonas, Epameinondas Antonakos, Georgios Tzimiropoulos, Stefanos Zafeiriou, and Maja Pantic, "300 Faces In-The-Wild Challenge: Database and Results," *Image and Vision Computing* 47, (2016): 3-18, https://doi.org/10.1016/j.imavis.2016.01.002.

[7] Chen Cao, Yanlin Weng, Shun Zhou, Yiying Tong, and Kun Zhou, "FaceWarehouse: A 3D Facial Expression Database for Visual Computing," *IEEE Transactions on Visualization and Computer Graphics* 20, no. 3 (2014): 413-425, https://doi.org/10.1109/TVCG.2013.249.

[8] Vahid Kazemi and Josephine Sullivan, "One millisecond face alignment with an ensemble of regression trees," *IEEE Conference on Computer Vision and Pattern Recognition, Ohio USA*, 2014, pp. 1867-1874, https://doi.org/10.1109/CVPR.2014.241.

[9] Tim Cootes, "An Introduction to Active Shape Models," in *Image processing and analysis*, ed. R.Baldock and J.Graham (Oxford: University Press, 2000), 223-248.

[10] A. Vinay Kumar, V. V. Ram Prasad, K. M. Bhurchandi, Vishal Satpute, Lizy Pious, and S. Kar, "Dense Reconstruction of 3D Human Face Using 5 Images and No Reference Model," *2017 4th International Conference on Control, Decision and Information Technologies (CoDIT),* 2017, pp. 1185–90. https://doi.org/10.1109/CoDIT.2017.8125644.

[11] Marcel Piotraschke and Volker Blanz, "Automated 3D Face Reconstruction from Multiple Images Using Quality Measures," 2016 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* 2016, pp. 3418-3427, https://doi.org/10.1109/CVPR.2016.372.

[12] Patrik Huber, Guosheng Hu, Rafael Tena, Pouria Mortazavian, Willem P. Koppen, William J. Christmas, Matthias Rätsch, and Josef Kittler, "A Multiresolution 3D Morphable Face Model and Fitting Framework," *Proceedings of the 11th Joint*

Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications SciTePress, 2016, pp. 79–86, DOI: https://doi.org/10.5220/0005669500790086.

13 Thomas Gerig, Andreas Morel-Forster, Clemens Blumer, Bernard Egger, Marcel Luthi, Sandro Schoenborn, and Thomas Vetter, "Morphable Face Models - An Open Framework," *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*, 2018, pp. 75–82. DOI: https://doi.org/10.1109/FG.2018.00021.

14 Emmanuel Prados and Olivier Faugeras, "Shape From Shading," in *Handbook of Mathematical Models in Computer Vision,* ed., Nikos Paragios, Yunmei Chen, and Olivier Faugeras (Boston, MA: Springer US, 2006), 375-388, https://doi.org/10.1007/0-387-28831-7_23.

15 Keiron O'Shea and Ryan Nash, "An Introduction to Convolutional Neural Networks," *Computing Research Repository (CoRR)*, 2015, pp. 1-11.

16 Yu Deng, Jiaolong Yang, Sicheng Xu, Dong Chen, Yunde Jia, and Xin Tong, "Accurate 3D Face Reconstruction With Weakly-Supervised Learning: From Single Image to Image Set," *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW),* 2019, pp. 285–95, https://doi.org/10.1109/CVPRW.2019.00038.

17 Tal Hassner, "Viewing Real-World Faces in 3D," *2013 IEEE International Conference on Computer Vision*, Sydney, Australia, 2013, pp. 3607-3614, https://10.1109/ICCV.2013.448.

18 Luo Jiang, Juyong Zhang, Bailin Deng, Hao Li, and Ligang Liu, "3D Face Reconstruction with Geometry Details from a Single Image," *IEEE Transactions on Image Processing* 27, no. 10 (2018): 4756-4770, https://doi.org/10.1109/TIP.2018.2845697.

19 Prateek Mehta, "Celebrity Face Recognition Dataset," Last modified 24th October, 2017, https://github.com/prateekmehta59/Celebrity-Face-Recognition-Dataset.

20 Igor Chalas, Petra Urbanova, Mikolas Jurda, Zuzana Ferkova, Katarina Furmanova, Barbora Kozikova, and Jiri Sochor, "FIDENTIS Analyst 1.3.2," 2015.

21 Zuzana Ferková, "Comparison and Analysis of Multiple 3D Shapes," MSc Thesis (Brno, Czech Republic: Masarykova Univerzita, 2016.

About the Authors

Matthew **Caruana** is a Software Engineer with interests in the area of 3D graphics, image processing, and digital forensics and has a Bachelor's Degree in Software Development from the University of Malta. Matthew has also participated in an artificial intelligence competition held by the Malta Information Technology Agency, developing an informative game highlighting the issues surrounding global warming.

Joseph G **Vella** lectures and researches in the areas of database technology at the University of Malta. His first degree was a BSc in Mathematics and Computing (UM) and a doctoral degree from the University of Sheffield Engineering Faculty. Dr Vella has participated in numerous projects at national and EU levels mainly dealing with data integration and consolidation for data warehousing, data science, and cloud services.