HOW MANY OPTIFACES? A NEW EVALUATION METRIC FOR 3D FACE RECONSTRUCTION

Will Rowan[1], **Patrik Huber**[1], **Nick Pears**[1], **Andrew Keeling**[2]
[1]University of York, [2]University of Leeds

ABSTRACT

Three dimensional face reconstruction is a challenging problem, so much so that the mean face is highly competitive with recent learning-based approaches for 3D face reconstruction from 2D images. No other universal baselines for this task exist. We propose a novel baseline that selects a subset of face meshes, called OptiFaces, that minimise overall 3D reconstruction error. This is a universal approach to calculate dataset-specific metrics for 3D face reconstruction, offering intuitive new baselines for the interpretation of 3D reconstruction error.

1 INTRODUCTION

The accurate reconstruction of 3D facial models from 2D images is a fundamental task in computer vision, with applications ranging from animation to facial recognition systems. However, current approaches often do little better than the mean face of existing 3D face models (Sanyal et al., 2019). The mean face from the FLAME head model (Li et al., 2017) outperforms all pre-2019 approaches on the NoW benchmark (Sanyal et al., 2019). This demonstrates how little information these previous methods are able to extract from an input image beyond a generic prior on face shape and provides a meaningful and informative baseline to beat.

We extend this idea from a single reference face (i.e. the mean face) to a collection of N reference faces, called OptiFaces. We then consider the performance that can be achieved if we have a classifier that optimally performs face matching from an input image to the closest of these N faces. This enables intuitive judgements to be made about the degree of facial information learned by existing approaches. For example, if the error for 5 OptiFaces is state-of-the-art, one only needs to be able to consistently perform accurate face matching to 5 well-separated face shapes to achieve this performance. For face reconstruction benchmarks, where 3D shape is known for a given image, OptiFaces provide a new set of baselines to beat and help us better distinguish between existing methods.

Here we provide a formal definition of OptiFaces and a method for calculating a set of OptiFaces for a 3D face dataset. We apply this method to calculate OptiFaces from the Headspace dataset (Dai et al., 2019), and test these on the NoW validation set (Sanyal et al., 2019), providing a strong set of baselines for the state-of-the-art monocular 3D face reconstruction benchmark.

2 PROPOSED METHOD

We formulate the problem of finding OptiFaces as follows. Given a set of N face meshes, each represented by V vertices in a 3D space, our goal is to find a subset of X representative faces that minimise the total reconstruction error. This error is quantified by the mean squared error (MSE) between the vertices of the mesh when each of the N faces is compared with the closest one from the selected X faces. This error metric is chosen as it is easy to calculate for any set of faces in an unsupervised manner. Formally, the problem is defined as follows:

Let $F = \{f_1, f_2, ..., f_N\}$ be the set of all face meshes, where each face mesh f_i is represented by its vertices $V_i \in \mathbb{R}^{V\times 3}$ where V is the number of vertices and $V_{i,k}$ is the kth vertex in V_i.

The error function between two face meshes f_i and f_j may be defined as:

$$E(f_i, f_j) = \frac{1}{V}\sum_{k=1}^{V} \|V_{i,k} - V_{j,k}\|^2$$

The objective is then formulated as follows:

- Find a subset $S \subseteq F$ such that $|S| = X$.
- This subset S should minimise the total reconstruction error across all faces in F.
- The total error is expressed as:

$$\underset{S \subseteq F,\, |S|=X}{\text{argmin}} \sum_{f_i \in F} \min_{f_j \in S} E(f_i, f_j)$$

We define an **idealised discrete classifier** as a classifier that always selects the nearest OptiFace. The idealised discrete classifier, denoted as $C(f_i, S)$, selects the face mesh f_j from subset S that minimises the error function E for a given face mesh f_i. Formally, the classifier is defined as:

$$C(f_i, S) = \underset{f_j \in S}{\text{argmin}} E(f_i, f_j)$$

By calculating a representative set of faces from one dataset and then selecting the closest face for a given image using our idealised-classifier, we compute new error baselines for existing reconstruction benchmarks. We introduce a simple greedy algorithm to calculate the set of representative faces in Appendix A and implement an idealised discrete classifier for the NoW benchmark, which reports the face with the lowest NoW error for the given image input. We evaluate this on the NoW benchmark for 1, 5, and 10 OptiFaces, establishing strong new baselines for face reconstruction. Appendix B includes visualisations of the computed OptiFaces as meshes, within PCA space, and in comparison with the K-means algorithm.

3 EXPERIMENTS

We use the FLAME-registered faces (Zielonka et al., 2022) of the Headspace dataset (Dai et al., 2019) as a database of 1,211 face shapes. We calculate OptiFaces for this using the approach outlined in Section 2. We then implement an idealised discrete classifier for the NoW validation set, using this to compute errors for 1, 5, and 10 OptiFaces.

Method	Median	Mean	Standard Deviation
Deep3D (Deng et al., 2019)	1.286	1.864	2.361
DECA (detail)	1.19	1.469	1.249
DECA (Feng et al., 2021)	1.178	1.464	1.253
AlbedoGAN (detail)	0.95	1.173	0.987
MICA (Zielonka et al., 2022)	0.913	1.130	0.948
AlbedoGAN (Rai et al., 2023)	0.903	1.122	0.957
1 OptiFace	1.527	1.859	1.524
5 OptiFaces	1.144	1.436	1.231
10 OptiFaces	1.125	1.416	1.226

Table 1: Reconstruction error (mm) on the validation set of the NoW benchmark (Sanyal et al., 2019) in non-metrical reconstruction. Comparison results are presented from Rai et al. (2023).

With just 5 OptiFaces, we beat all pre-2022 face reconstruction methods. More precisely, all methods before that period are outperformed by an approach that can select the optimal face from just 5 unique face shapes. Headspace and the NoW benchmark are independent datasets, collected in different parts of the world. This shows that OptiFaces generalise across distributions of face shape.

4 CONCLUSION

We present OptiFaces, a novel baseline for 3D reconstruction techniques that selects face meshes that optimally minimise reconstruction error. We demonstrate that a small number of OptiFaces offer a valuable baseline on the NoW benchmark and a novel way to evaluate reconstruction performance by connecting it to the associated classification problem. OptiFaces are easy to compute and provide a novel set of dataset-specific metrics for 3D face reconstruction, giving a more meaningful interpretation of 3D shape reconstruction error.

URM Statement

The authors acknowledge that at least one key author of this work meets the URM criteria of ICLR 2024 Tiny Papers Track.

References

Paolo Cignoni, Marco Callieri, Massimiliano Corsini, Matteo Dellepiane, Fabio Ganovelli, Guido Ranzuglia, et al. Meshlab: an open-source mesh processing tool. In *Eurographics Italian chapter conference*, volume 2008, pp. 129–136. Salerno, Italy, 2008.

Hang Dai, Nick Pears, William Smith, and Christian Duncan. Statistical modeling of craniofacial shape and texture. *International Journal of Computer Vision*, 128(2):547–571, Nov 2019. ISSN 1573-1405. doi: 10.1007/s11263-019-01260-7. URL `https://doi.org/10.1007/s11263-019-01260-7`.

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. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, pp. 285–295, 2019.

Yao Feng, Haiwen Feng, Michael J Black, and Timo Bolkart. Learning an animatable detailed 3d face model from in-the-wild images. *ACM Transactions on Graphics (ToG)*, 40(4):1–13, 2021.

Tianye Li, Timo Bolkart, Michael J Black, Hao Li, and Javier Romero. Learning a model of facial shape and expression from 4d scans. *ACM Trans. Graph.*, 36(6):194:1–194:17, 2017.

F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

Aashish Rai, Hiresh Gupta, Ayush Pandey, Francisco Vicente Carrasco, Shingo Jason Takagi, Amaury Aubel, Daeil Kim, Aayush Prakash, and Fernando De la Torre. Towards realistic generative 3d face models. *arXiv preprint arXiv:2304.12483*, 2023.

Soubhik Sanyal, Timo Bolkart, Haiwen Feng, and Michael Black. Learning to regress 3d face shape and expression from an image without 3d supervision. In *Proceedings IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, June 2019.

Wojciech Zielonka, Timo Bolkart, and Justus Thies. Towards metrical reconstruction of human faces. In *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XIII*, pp. 250–269. Springer, 2022.

A OptiFace Optimiser

To calculate our error metric for N OptiFaces, we first must calculate a set of OptiFaces from a set of face shapes as in Section A.1 and then evaluate these representative faces on a face reconstruction benchmark in Section A.2. The following section provides additional details on the practical implementation of these steps.

A.1 Selecting Representative Faces

We employ a greedy algorithm for iteratively building a set of OptiFaces. This algorithm assumes all input faces follow the same topology, a condition we meet by using a database of faces already registered to the FLAME head model. It can be applied to any set of faces, provided they are registered to a common face topology, which can be completed as a pre-processing step if necessary. The pseudo-code for our proposed OptiFace Optimiser is presented in Algorithm 1.

We apply the OptiFace Optimiser to the Headspace dataset (Dai et al., 2019). Headspace is chosen due it being one of the largest available datasets of 3D face shape, including a wide range of age and strong gender balance. This ensures that we are able to select representative faces that cover the wide natural variety of human face shape. On a single GTX 1080, the computation times for this algorithm are 1 minute for 1 OptiFace, 33 minutes for 5 OptiFaces, and 94 minutes for 10 OptiFaces.

Algorithm 1 Greedy algorithm for OptiFace calculation.

Require: Faces F, Desired number of OptiFaces N
Ensure: Subset of selected OptiFaces S, Indices of selected OptiFaces I

Calculate the mean face mean_face from F
Initialize S with the face closest to mean_face
Initialize I with the index of this face
while $|S| < N$ **do**
    Set min_error to infinity
    for each face f in F not in S **do**
        Temporarily add f to S
        Calculate reconstruction error E for F given S
        if $E <$ min_error **then**
            Update min_error to E
            Set best_face to f
        end if
    end for
    Add best_face to S
    Add the index of best_face to I
end while
return S, I

A.2 Idealised Discrete Classifier: Evaluating OptiFaces on a Benchmark

Now that we have a set of representative faces, we can evaluate these using an implementation of the aforementioned idealised discrete classifier. This algorithm is designed to evaluate the effectiveness of a set of OptiFaces in face reconstruction tasks, particularly within the context of benchmarks. It processes a dataset consisting of image pairs alongside their corresponding ground truth 3D meshes, aiming to determine the OptiFace that most closely approximates each ground truth mesh. Psuedocode for our discrete idealised classifier is presented in Algorithm 2.

We perform evaluation on the NoW benchmark (Sanyal et al., 2019), a standard benchmark for 3D face reconstruction from a single image. We selected it for its diversity in individuals, environments, and capture settings—ranging from neutral and expressions to selfie and occlusions. These uncontrolled settings more accurately represent real-world conditions, making NoW a challenging benchmark where our shape-only baselines enable us to better distinguish existing approaches.

The summary statistics computed by the algorithm provide insights into the overall efficacy of the OptiFaces in reconstructing faces from the benchmark dataset. These statistics serve as a benchmark to evaluate and compare different sets of representative faces or methods used for face reconstruction.

Algorithm 2 Benchmark-Based Idealised Classifier for OptiFace Selection

Require: Benchmark dataset D containing pairs of images and ground truth meshes, Set of representative OptiFaces S
Ensure: Summary statistics of errors for each image in D with the best matching OptiFace

```
function BenchmarkClassifier(D, S)
    Initialize an empty list errors to store errors for each image
    for each pair (image, gt_mesh) in D do
        Initialize min_error to infinity
        Initialize selected_face to null
        for each OptiFace f_j in S do
            Calculate error E between gt_mesh and f_j
            if E < min_error then
                Update min_error to E
                Update selected_face to f_j
            end if
        end for
        Add min_error to the list errors
    end for
    Compute summary statistics from errors
    return Summary statistics
end function
```

B Visualisation of OptiFaces

In figure 1, we present a visualisation of the first 10 OptiFaces calculated using the OptiFace Optimiser and evaluated on the NoW benchmark. We show the first two principal components of shape from the FLAME head model, alongside all other faces in the Headspace dataset. This effectively demonstrates the extensive coverage our OptiFaces provide for the two principal components of face shape that capture the most variation, out of a total of 300. The OptiFaces are numbered in the order they are calculated.

In figure 2, we compare the OptiFaces calculated using the k-means algorithm with those derived from the OptiFace Optimiser. It is evident that the OptiFace Optimiser selects more representative faces from the Headspace dataset for the first two principal components of face shape, thereby covering a wider range of potential face shapes. To perform this analysis, we applied the k-means algorithm to all Headspace faces within the PCA space of the FLAME head model. We used the implementation from scikit-learn (Pedregosa et al., 2011), setting a fixed seed of 10 and running for 100 iterations. The centroids obtained from this process were then selected as our OptiFaces.

Figure 3 displays the ordered meshes of all 10 OptiFaces, which are also illustrated in figures 1 and 2 and evaluated in table 1. All images are rendered to a common scale using MeshLab (Cignoni et al., 2008) with a field of view (FOV) of 60. The ordering of the meshes demonstrates how the OptiFace Optimiser selects representative faces to minimise reconstruction error.

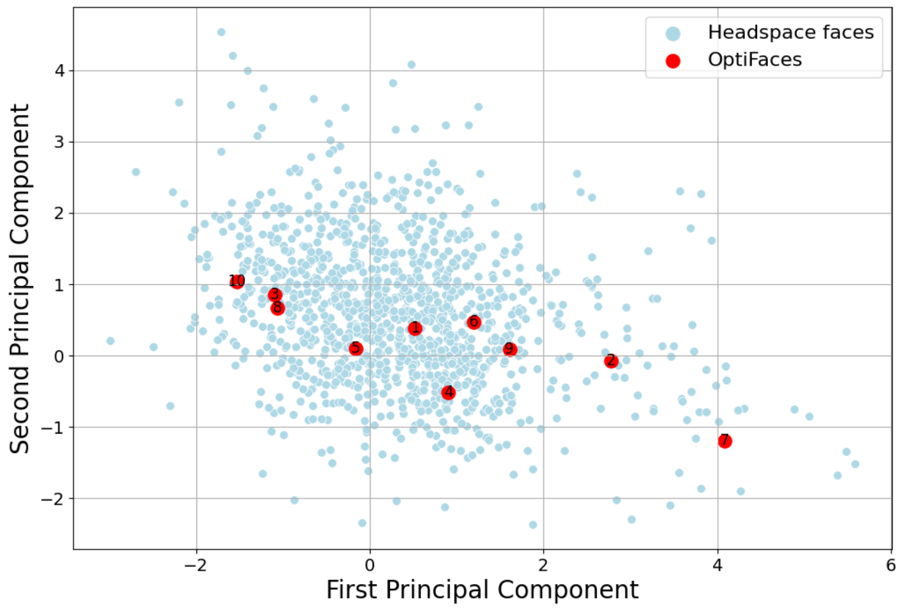


Figure 1: A visualisation of the first two principal components of FLAME shape for the first 10 OptiFaces calculated using the Headspace dataset. OptiFaces are numbered in the order they are calculated.

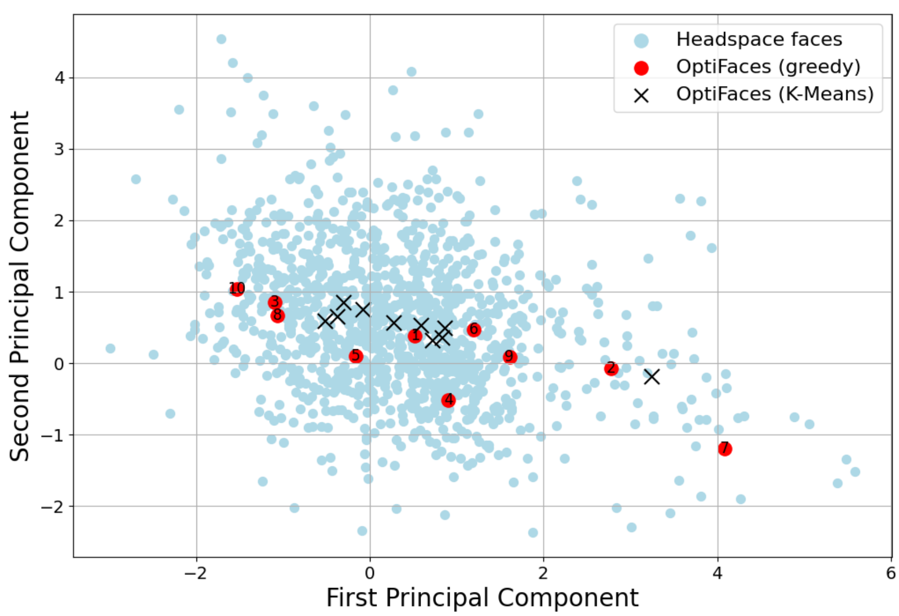
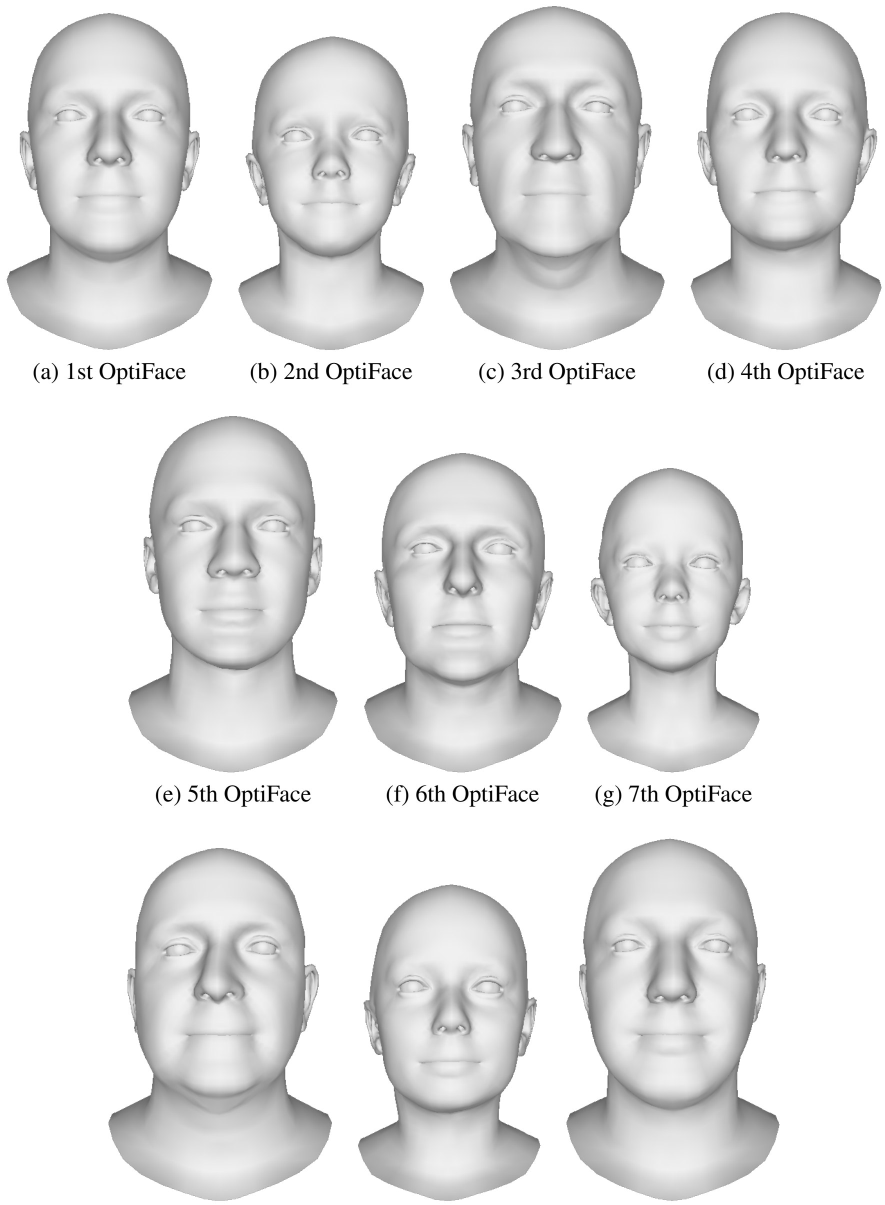


Figure 2: A comparison of the OptiFaces computed using the proposed OptiFace Optimiser and a k-means implementation.



(a) 1st OptiFace (b) 2nd OptiFace (c) 3rd OptiFace (d) 4th OptiFace

(e) 5th OptiFace (f) 6th OptiFace (g) 7th OptiFace

(h) 8th OptiFace (i) 9th OptiFace (j) 10th OptiFace

Figure 3: OptiFace Heads Visualisation for the first 10 OptiFaces calculated using the OptiFace Optimiser