Building 3D Morphable Models from a Single Scan - Supplementary Material

Skylar Sutherland
Massachusetts Institute of Technology
77 Massachusetts Ave, Cambridge, MA 02139
skylar@mit.edu

Bernhard Egger
egger@mit.edu

Joshua Tenenbaum
jbt@mit.edu

1. Color-Correlated Asymmetric 3DMMs

In our main paper we build 3DMMs using Gaussian processes that include symmetry and color-channel correlation heuristics. To assess the effects of these heuristics individually, we can also build 3DMMs that include only one of these heuristics. Specifically, we experimented with constructing 3DMMs whose albedo models have correlated color channels but which lack symmetry. This enables us to compare the relative importance in an analysis-by-synthesis setting of the symmetry and color-correlation heuristics of our symmetric 3DMMs.

Our main paper defined albedo kernels $K^{\texttt{cor}}_{a,\texttt{xyz}}$ and $K^{\texttt{sym}}_{a,\texttt{rgb}}$. These kernels possess a color-channel correlation heuristic but lack an explicit symmetry heuristic ($K^{\texttt{sym}}_{a,\texttt{rgb}}$ is symmetric, but this is only because we assume that the reference face is symmetric). We may average these kernels to produce an albedo kernel $K^{\texttt{cor}}_{a} = 0.5(K^{\texttt{sym}}_{a,\texttt{rgb}} + K^{\texttt{cor}}_{a,\texttt{xyz}})$ which combines physical-space and RGB-space distance information. Samples from $K^{\texttt{cor}}_{a,\texttt{xyz}}$ and $K^{\texttt{cor}}_{a}$ are shown in Figure 1. By combining these albedo kernels with our shape kernel K_a, we can construct 3DMMs which possess a color-channel correlation heuristic but which lack an explicit symmetry heuristic. We list these 3DMMs in Table 1.

We repeat the face recognition experiment presented in Section 3.1.2 of our main paper with these 3DMMs (once again using the mean of the 2019 Basel Face Model [2] as our reference mesh). The results of this experiment are shown in Table 2, along with a copy of the results with the standard and symmetric 3DMMs that were shown in the main paper. Table 2 demonstrates that the color-correlated asymmetric 3DMMs perform comparably to the symmetric (and color-correlated) 3DMMs on faces with $15°$ and $30°$ yaw angles. On faces with $45°$ yaw angles, they are significantly worse, indicating that (unsurprisingly) a symmetry prior becomes more important as the yaw angle increases. Nevertheless, in general the color-correlated asymmetric 3DMMs perform quite well. This indicates that in an inverse graphics context the color-channel correlation heuristic is more important to our symmetric 3DMMs than the symmetry heuristic is, at least for input images with a low yaw angle.

name	shape kernel	albedo kernel
correlated-full	K_s	$K^{\texttt{cor}}_{a}$
correlated-RGB	K_s	$K^{\texttt{cor}}_{a,\texttt{rgb}}$
correlated-XYZ	K_s	$K^{\texttt{cor}}_{a,\texttt{xyz}}$

Table 1: Our Gaussian processes for modeling faces with color-correlated but asymmetric kernels.

angle	$15°$	$30°$	$45°$
probe id	140_16	130_16	080_16
standard-full	84.7	69.9	54.2
standard-RGB	76.3	57.8	28.9
standard-XYZ	77.1	62.7	35.7
correlated-full	92.4	**87.1**	66.7
correlated-RGB	71.5	58.6	37.8
correlated-XYZ	76.7	61.4	45.8
symmetric-full	**93.2**	85.9	**72.3**
symmetric-RGB	73.5	61.4	40.2
symmetric-XYZ	73.1	58.6	44.2

Table 2: Face recognition results on images from the Multi-PIE database [3]. Each column represents the accuracy for a set of probe images with a common yaw angle given in the first row. The second row gives the common ending of the IDs in the Multi-PIE dataset of the probe images with a given yaw angle. The gallery is constructed from images with a yaw angle of $0°$ (dataset IDs ending in 051_16).

2. Samples from Kernels

In Figure 1, we show samples from our various shape and albedo kernels, applied to the mean of the 2019 Basel Face Model [2]. While these samples are clearly non-naturalistic, this does not invalidate the results of Section 3 of our main paper. **We make no claim that our 3DMMs accurately model the distribution of human faces; rather, we claim**

that they are of sufficient quality to be useful in a machine vision context.

3. Additional Bird and Fish Models

In our paper we presented bird and fish 3DMMs created analogously to the standard-full and symmetric-full 3DMMs, i.e. with K_s and K_a, and with $K_s^{\mathtt{sym}}$ and $K_a^{\mathtt{sym}}$, respectively. We can also define similar bird and fish 3DMMs using albedo kernels that only rely on physical distance; i.e., using $K_{a,\mathtt{xyz}}$ and $K_{a,\mathtt{xyz}}^{\mathtt{sym}}$ instead of K_a and $K_a^{\mathtt{sym}}$. This produces two new 3DMMs for each reference mesh, which for space reasons are listed as listed as "XYZ standard" and "XYZ symmetric". Figure 2 shows samples from these two bird 3DMMs, as well as reconstructions produced with these models of the bird images that were used in the main paper. Figure 3 shows analogous samples and reconstructions for the two new physical-distance-based fish 3DMMs. The results are close to those produced in the main paper; this is unsurprising given that the 3D mesh used to build these 3DMMs does not include complex coloration, and instead has near-piecewise-constant albedo.

In both Figure 3 and in the main paper, we obtain our input natural fish images from Wikipedia[1].

4. Additional Registration Results

Figures 4, 5, 6, and 7 offer a variety of quantitative metrics of the shape error of the registered meshes produced in our paper's registration tasks. As in the main paper, in all figures the "shape and albedo" option refers to meshes registered using both shape and albedo information in the MCMC method, while the "shape only" option refers to meshes registered using only shape information in the MCMC method. These figures do not take into account the stability of the reconstruction or the albedo error.

We estimate shape error through Hausdorff distance (Figures 4 and 6) and chamfer distance (Figures 5 and 7) between either the vertices of the registered meshes and the corresponding scans (Figures 4 and 5) or a sparse set of landmarks on the registered meshes and corresponding scans (Figures 6 and 7). We obtained landmark information by using the landmark annotations given in [5] for each of the 10 input meshes and the landmark annotations provided with the 2019 Basel Face Model [2] for the registered meshes (since these have the same topology as the 2019 Basel Face Model).

Figures 4 and 5 demonstrate that including albedo information along with shape information slightly increases the shape reconstruction error. As noted in the main text, this is to be expected; the shape-only reconstruction is optimized to produce the lowest shape error possible, whereas the reconstruction produced using both shape and albedo is also optimized to produce a low albedo error, and by definition cannot have a lower shape error than the reconstruction with the minimum shape error. However, the increase in shape error is not very large. Furthermore, Figures 6 and 7 demonstrate that including albedo in registration does not significantly affect the shape error of landmarks. This suggests that the incorporation of albedo information does not reduce the registration quality of the *important aspects* of face shape.

[1]The links are:
https://en.wikipedia.org/wiki/File:Acanthurus_achilles1.jpg,
https://en.wikipedia.org/wiki/File:Acanthurus_dussumieri.jpg, and
https://en.wikipedia.org/wiki/File:Paracanthurus_hepatus_(Regal_Tang).jpg

5. Specificity, Generalization, and Compactness

Figure 8 shows plots of the specificity, generalization, and compactness [6] of our 3DMMs and the 2017 Basel Face Model [2]; specifically, it shows the specificity and generalization of the shape and albedo models of each 3DMM as a function of the number of principal components included. We compare the 2017 Basel Face Model ("BFM 2017") and versions of the standard-full ("standard"), symmetric-full ("symmetric"), and correlated-full ("correlated") models built using the mean of the 2017 Basel Face Model as template. We use as our dataset the ten scans provided with the 2009 Basel Face Model [5]. We also include the symmetric-x models, where x is a scan ID number (`001`, `002`, `006`, `014`, `017`, `022`, `052`, `053`, `293`, or `323`); for these models we exclude the scan used to build the model. We report results averaged across the symmetric-x models as "single-scan". We measure specificity and generalization using 1, 2, 5, 10, 20, 50, 100, and all (199) principal components. We indicate the specificity and generalization of the mean of the 2017 Basel Face Model, considered as a 3DMM with zero principal components, with a black line.

We may observe that, for all numbers of principal components, the generalization of our 3DMMs' shape models is comparable to that of the 2017 Basel Face Model, while the generalization of our 3DMMs' albedo models is in fact superior to that of the 2017 Basel Face Model. The specificity of our 3DMMs' shape and albedo models, is, of course, inferior to that of the 2017 Basel Face Model. This is unavoidable as our models' were constructed using far less data than the 2017 Basel Face Model. We may additionally observe that our single-scan models perform comparably to the standard-full model across all conditions.

6. Qualitative Reconstructions

Figures 9, 10, 11, and 12 provide additional qualitative reconstruction results. Figures 9 and 10 present qualitative reconstructions (in frontal and side views, respectively) of images from the Labeled Faces in the Wild dataset

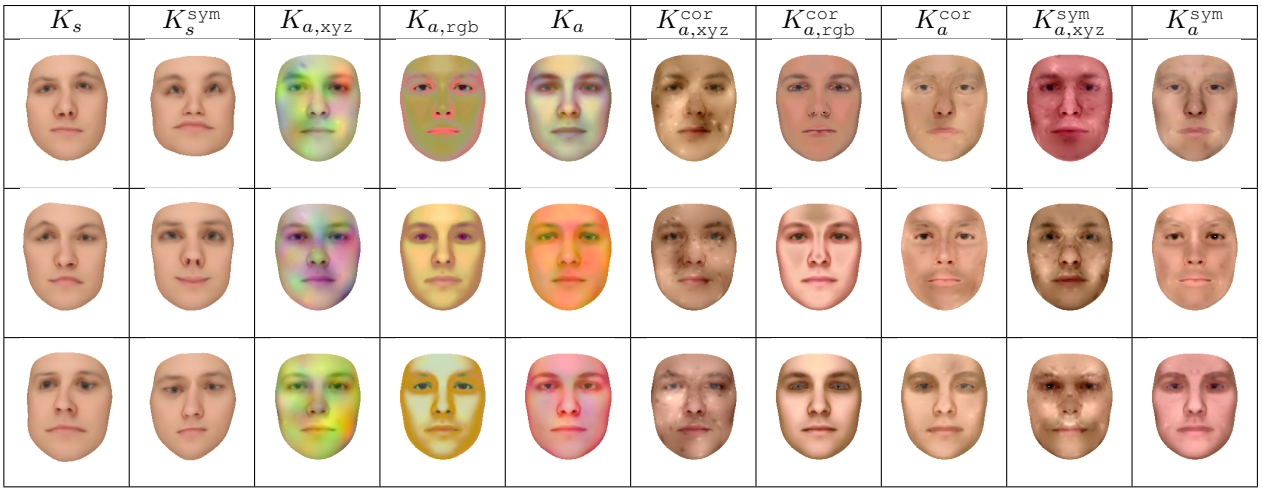

Figure 1: Three random samples from each of the shape and albedo kernels applied to the mean of the 2019 Basel Face Model [2] and rendered under ambient illumination. The first two columns are the two shape kernels, while the remaining eight columns are the albedo kernels.

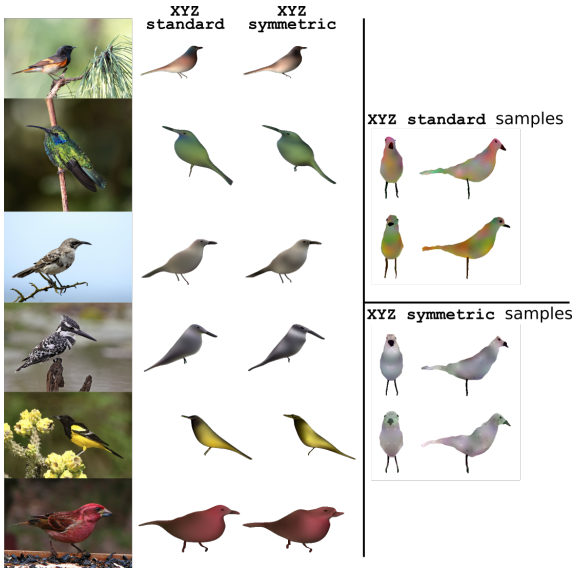

Figure 2: On the left: the reconstructions produced by the two bird models built using only physical distance information on six images taken from the Caltech-UCSD Birds 200 dataset [7]. On the right: samples from these models, shown in frontal and side views.

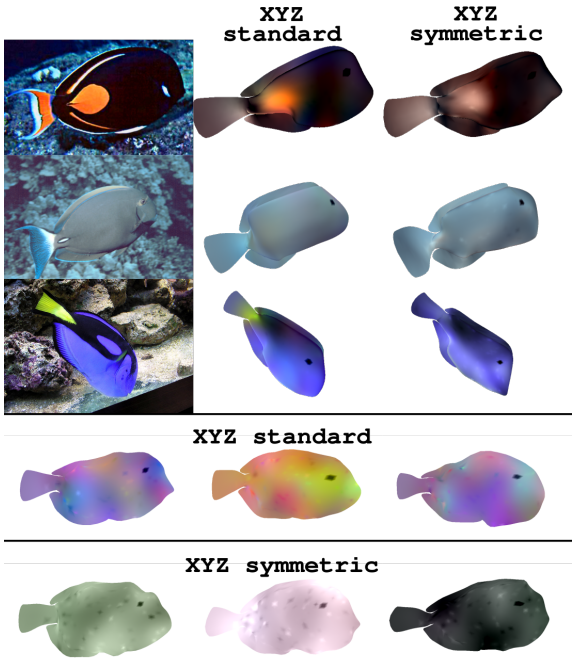

Figure 3: On the top: the reconstructions produced by the two fish models built using only physical distance information on three natural images of fish. On the bottom: samples from these models, shown in side views.

[4] produced using all the 3DMMs constructed using the mean of the 2019 Basel Face Model [2]. Figure 11 presents qualitative reconstructions of the same images produced using 3DMMs built from the scans included with the 2009 Basel Face Model [5]. These reconstructions are significantly lower-quality, because a significant portion of the shape of the template mesh is preserved during the MCMC process. Figure 12 presents qualitative reconstructions of different images from the Labeled Faces in the Wild dataset that contain significant occlusion. Figure 12 includes both

3D reconstructions as well inferred occlusion masks. Figure 12's results were produced using the occlusion-aware MCMC method described in [1].

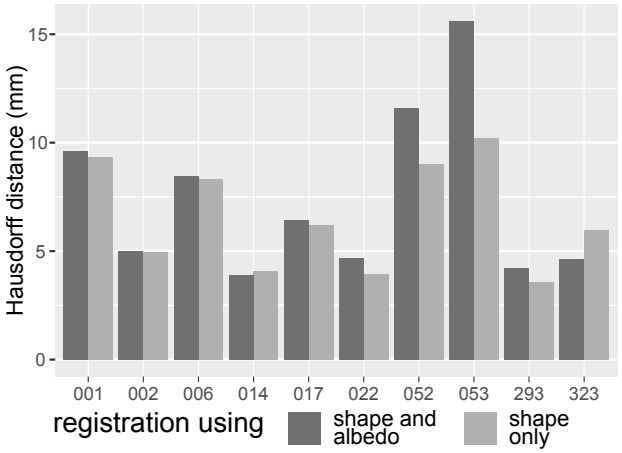

Figure 4: The Hausdorff distance between the vertices of each of the registered meshes and the vertices of the corresponding face scan.

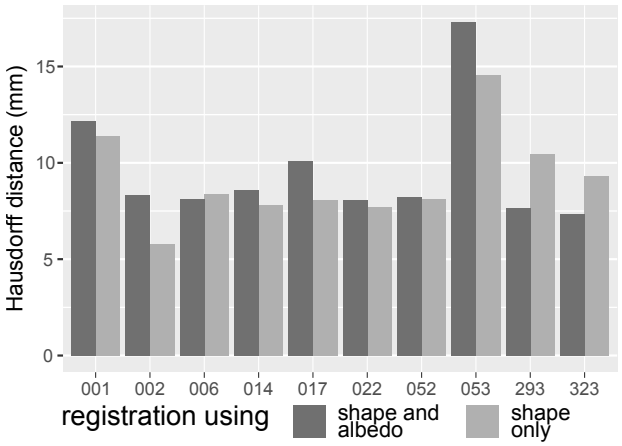

Figure 6: The Hausdorff distance between the landmarks of each of the registered meshes and the landmarks of the corresponding face scan.

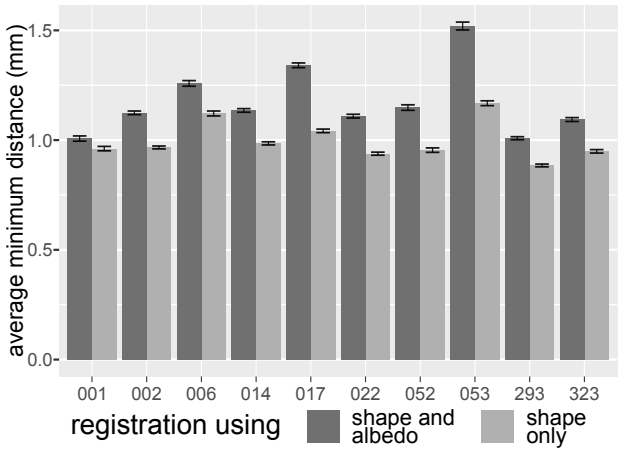

Figure 5: The average distance between each vertex in each of the registered meshes and the closest point in the corresponding face scan, with error bars (± 1.96 standard error).

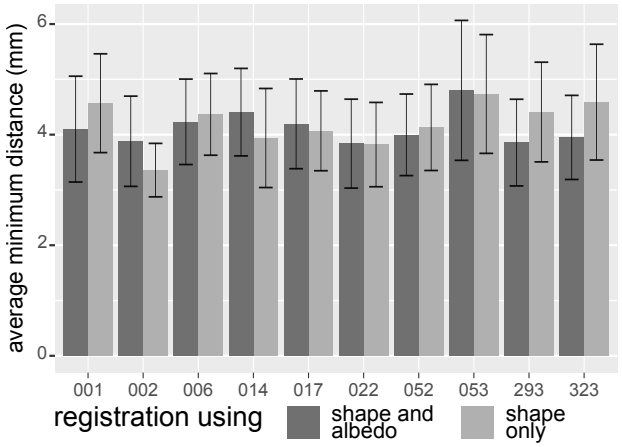

Figure 7: The average distance between each landmark in each of the registered meshes and the closest landmark in the corresponding face scan, with error bars (± 1.96 standard error).

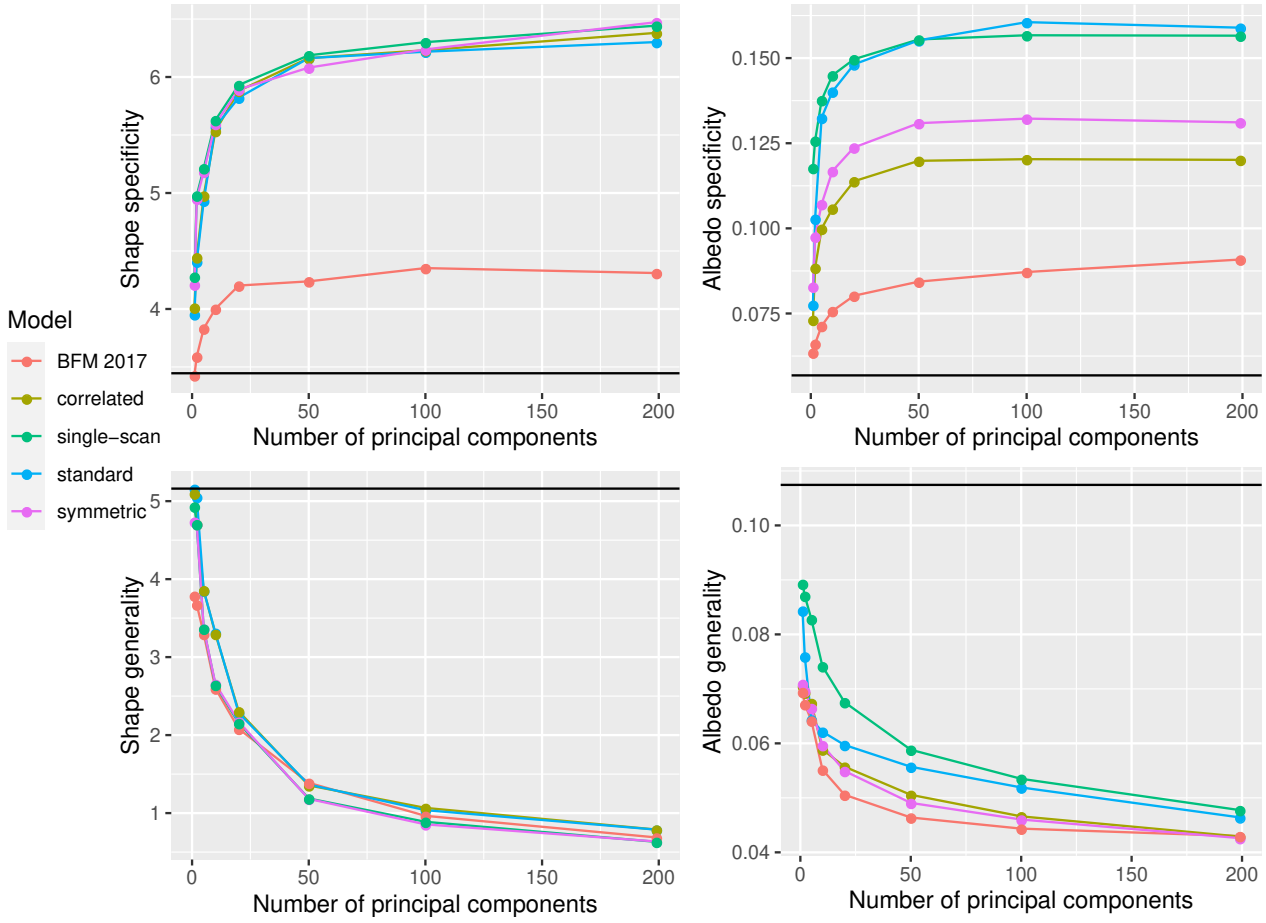

Figure 8: A plot of the specificity, generalization, and compactness [6] of our 3DMMs' relative to the 2017 Basel Face Model [2]. "standard", "correlated", and "symmetric" refer to versions of the standard-full, correlated-full, and symmetric-full models built using the mean of the 2017 Basel Face Model, while the "single-scan" results are an average of the performance of the various symmetric-x models. The scans included with the 2009 Basel Face Model [5] were used as a dataset; for the symmetric-x 3DMMs, the scan used to construct the 3DMM was excluded. See Section 5 for more details.

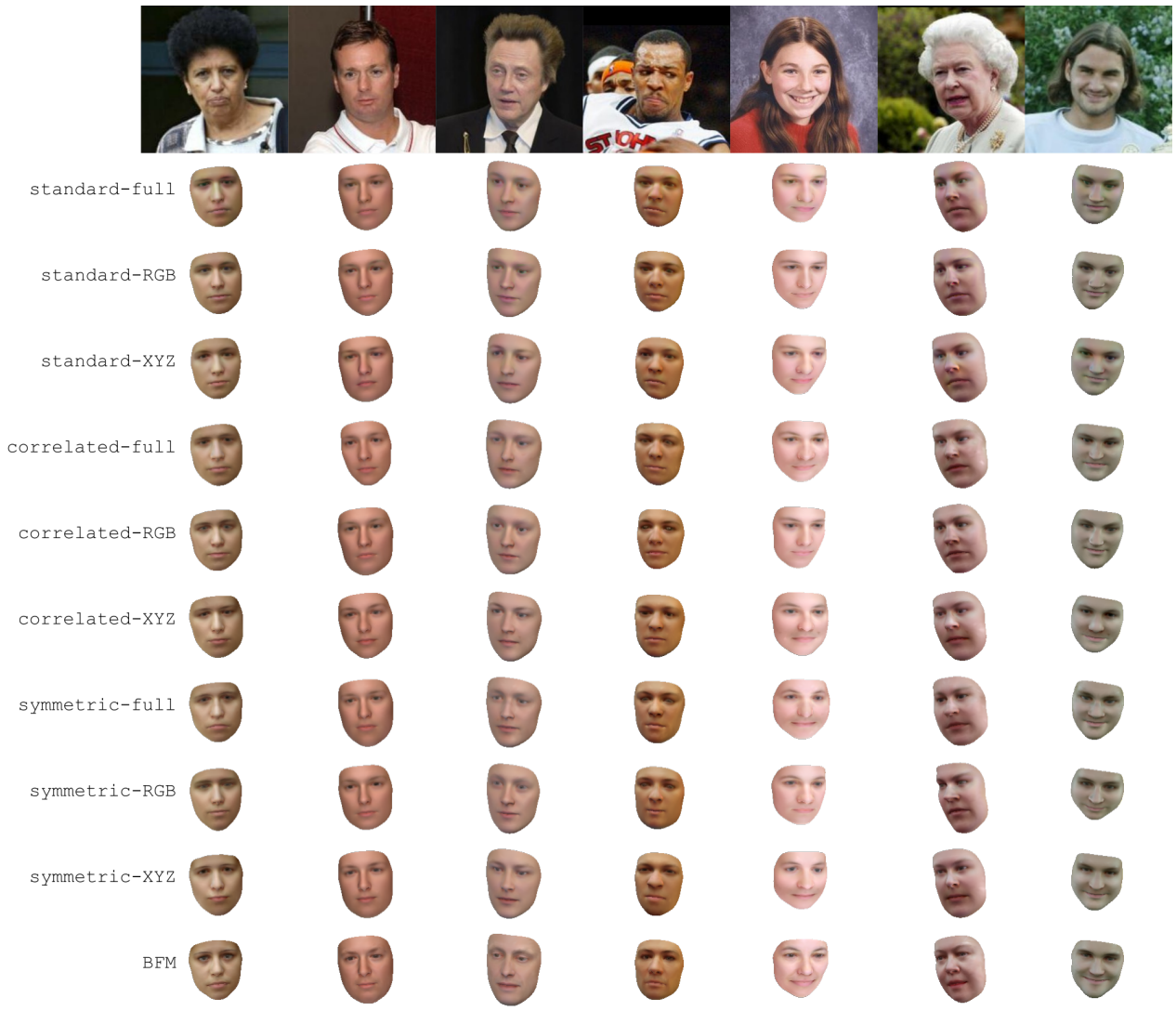

Figure 9: The face reconstructions produced from all the 3DMMs built using the mean of the 2019 Basel Face Model [2] on natural images from the Labeled Faces in the Wild dataset [4], as well as the reconstructions produced using the 2019 Basel Face Model itself ("BFM").

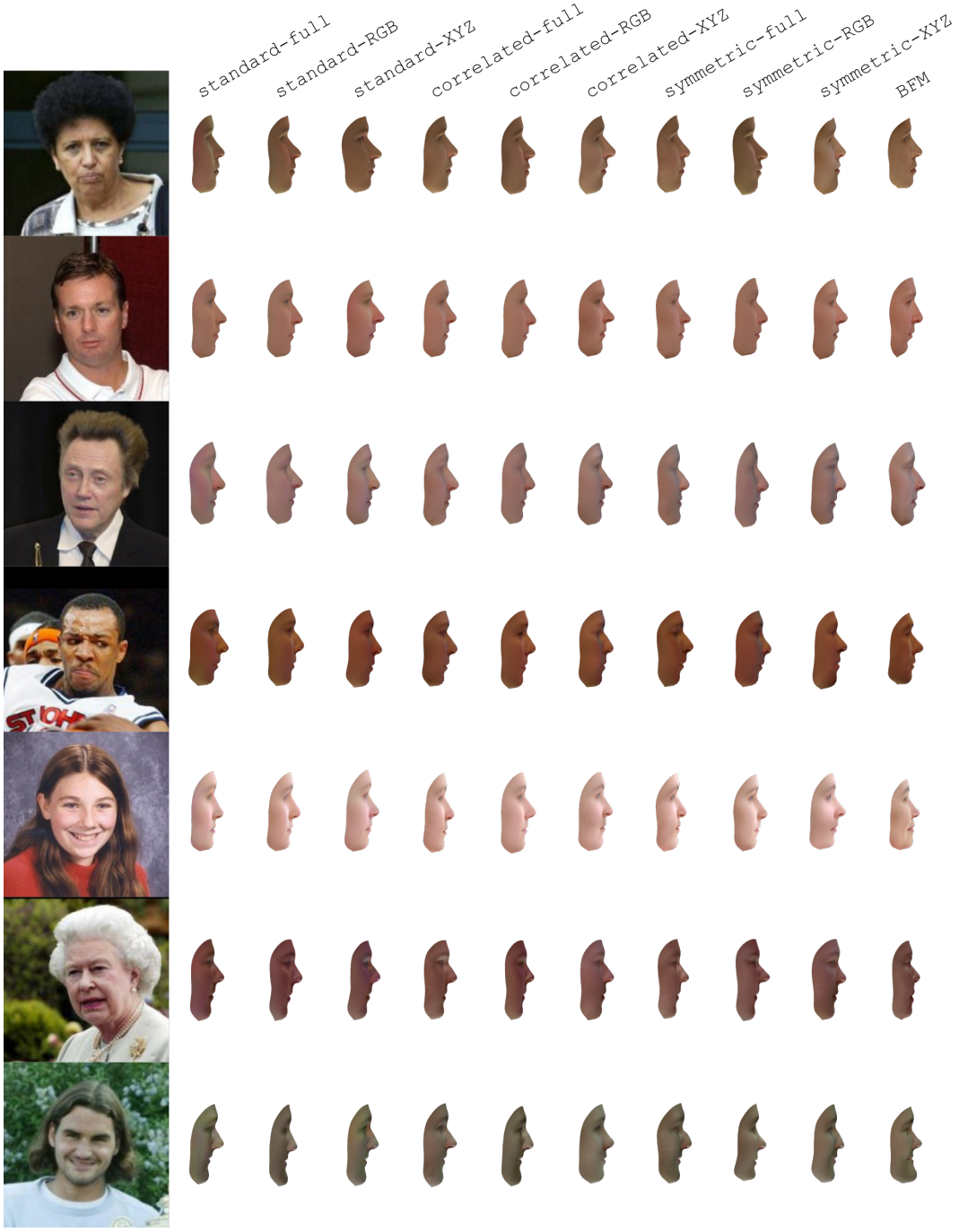

Figure 10: Side views of the reconstructions presented in Figure 9.

Figure 11: The face reconstructions produced from all 3DMMs built using the symmetric-full kernel and scans included with the 2009 Basel Face Model [5] on natural images from the Labeled Faces in the Wild dataset [4].

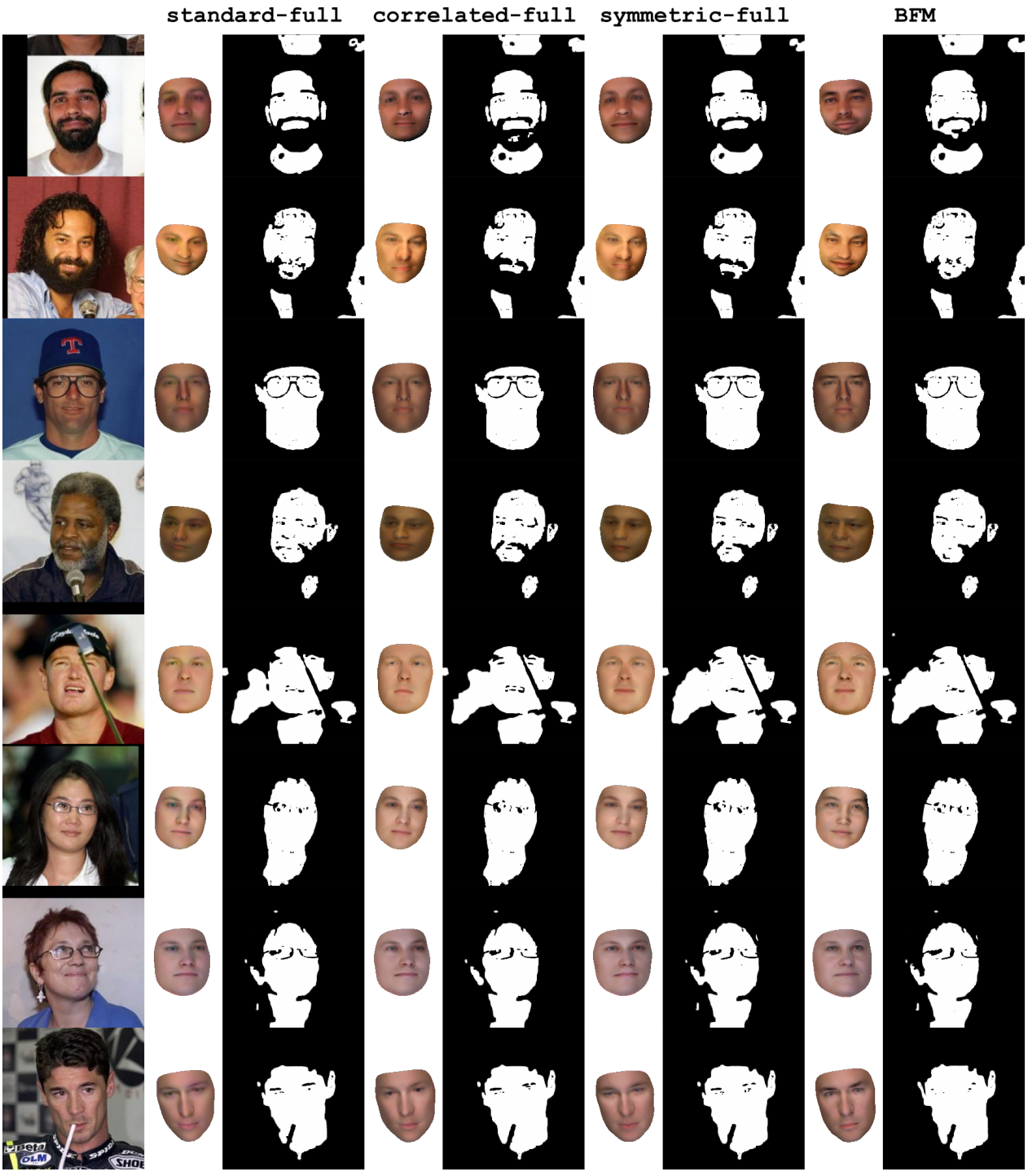

Figure 12: The face reconstructions produced by our standard-full, correlated-full, and symmetric-full models, as well as the 2019 Basel Face Model [2] ("BFM"), on images from the Labeled Faces in the Wild dataset [4], produced using the occlusion-aware MCMC method described in [1]. Both the segmentation masks and face reconstructions were inferred purely with top-down inference.

References

[1] Bernhard Egger, Sandro Schönborn, Andreas Schneider, Adam Kortylewski, Andreas Morel-Forster, Clemens Blumer, and Thomas Vetter. Occlusion-aware 3d morphable models and an illumination prior for face image analysis. *International Journal of Computer Vision*, 126(12):1269–1287, 2018. 4, 9

[2] Thomas Gerig, Andreas Morel-Forster, Clemens Blumer, Bernhard Egger, Marcel Lüthi, Sandro Schönborn, and Thomas Vetter. Morphable face models—an open framework. In *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*, pages 75–82. IEEE, 2018. 1, 2, 3, 5, 6, 9

[3] Ralph Gross, Iain Matthews, Jeffrey Cohn, Takeo Kanade, and Simon Baker. Multi-PIE. *Image and Vision Computing*, 28(5):807–813, May 2010. 1

[4] Gary B Huang, Marwan Mattar, Tamara Berg, and Eric Learned-Miller. Labeled faces in the wild: A database for studying face recognition in unconstrained environments. 2008. 3, 6, 8, 9

[5] Pascal Paysan, Reinhard Knothe, Brian Amberg, Sami Romdhani, and Thomas Vetter. A 3D face model for pose and illumination invariant face recognition. In *2009 Sixth IEEE International Conference on Advanced Video and Signal Based Surveillance*, pages 296–301. IEEE, 2009. 2, 3, 5, 8

[6] Martin A Styner, Kumar T Rajamani, Lutz-Peter Nolte, Gabriel Zsemlye, Gábor Székely, Christopher J Taylor, and Rhodri H Davies. Evaluation of 3d correspondence methods for model building. In *Biennial International Conference on Information Processing in Medical Imaging*, pages 63–75. Springer, 2003. 2, 5

[7] Catherine Wah, Steve Branson, Peter Welinder, Pietro Perona, and Serge J. Belongie. The Caltech-UCSD Birds-200-2011 dataset. 2011. 3