



A Statistical Model for Synthesis of Detailed Facial Geometry

Citation

Golovinskiy, Aleksey, Wojciech Matusik, Hanspeter Pfister, Szymon Rusinkiewicz, and Thomas Funkhouser. 2006. A statistical model for synthesis of detailed facial geometry. In Proceedings International Conference on Computer Graphics and Interactive Techniques, ACM SIGGRAPH 2006 Papers: July 30 - August 03, 2006, Boston, Massachusetts, 1025-1034. New York, N.Y.: ACM Press. Also published in ACM Transactions on Graphics 25(3): 1025-1034.

Published Version

doi:10.1145/1141911.1141988;doi:10.1145/1179352.1141988

Permanent link

<http://nrs.harvard.edu/urn-3:HUL.InstRepos:4101996>

Terms of Use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA>

Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. [Submit a story](#).

[Accessibility](#)

A Statistical Model for Synthesis of Detailed Facial Geometry

Aleksey Golovinskiy
Princeton University

Wojciech Matusik
MERL

Hanspeter Pfister
MERL

Szymon Rusinkiewicz
Princeton University

Thomas Funkhouser
Princeton University

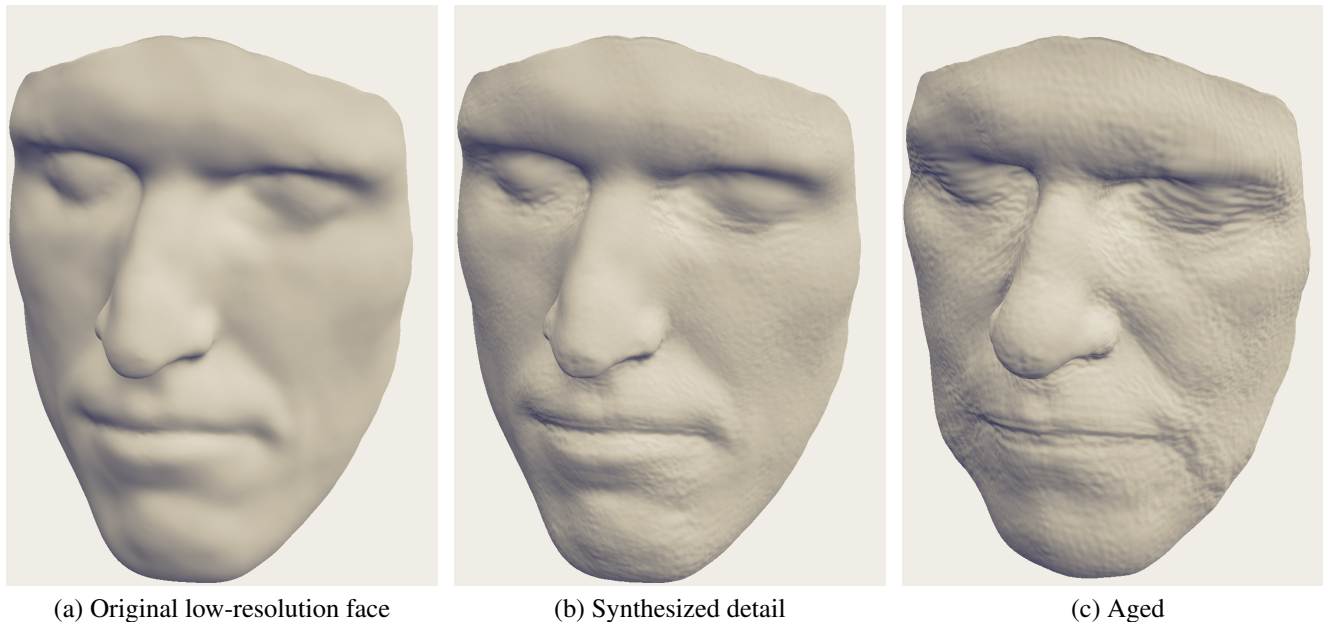


Figure 1: Synthesis of facial detail. (a) We begin with a low-resolution mesh obtained from a commercial scanner. (b) Then, we synthesize detail on it using statistics extracted from high-resolution meshes in our database. (c) Finally, we age the face by adjusting the statistics to match those of an elderly man. Note that all figures in this paper use Exaggerated Shading [Rusinkiewicz et al. 2006] to emphasize small geometric details.

Abstract

Detailed surface geometry contributes greatly to the visual realism of 3D face models. However, acquiring high-resolution face geometry is often tedious and expensive. Consequently, most face models used in games, virtual reality, or computer vision look unrealistically smooth. In this paper, we introduce a new statistical technique for the analysis and synthesis of small three-dimensional facial features, such as wrinkles and pores. We acquire high-resolution face geometry for people across a wide range of ages, genders, and races. For each scan, we separate the skin surface details from a smooth base mesh using displaced subdivision surfaces. Then, we analyze the resulting displacement maps using the texture analysis/synthesis framework of Heeger and Bergen, adapted to capture statistics that vary spatially across a face. Finally, we use the extracted statistics to synthesize plausible detail on face meshes of arbitrary subjects. We demonstrate the effectiveness of this method in several applications, including analysis of facial texture in subjects with different ages and genders, interpolation between high-resolution face scans, adding detail to low-resolution face scans, and adjusting the apparent age of faces. In all cases, we are able to re-produce fine geometric details consistent with those observed in high resolution scans.

Keywords: face modeling, texture synthesis

1 Introduction

Creating realistic models of human faces is an important problem in computer graphics. Face models are widely used in computer games, commercials, movies, and for avatars in virtual reality applications. The goal is to capture all aspects of a person’s face in a digital model – i.e., “Digital Face Cloning” [Pighin and Lewis 2005].

Ideally, a face model should be indistinguishable from a real face, easy to acquire, and intuitive to edit. However, digital face cloning remains a difficult task for several reasons. First, humans are very accustomed to looking at real faces and can easily spot artifacts in computer generated models. Second, capturing the high resolution geometry of a face is difficult and expensive. Finally, editing face models is still a time consuming and largely manual task, especially if changes to fine-scale details are required.

In this work we focus on a small but important source of realism in faces: geometry of small facial features, such as wrinkles and pores. As mentioned in Igarashi et al. [2005], wrinkles are folds of skin formed through the process of skin deformation, whereas pores are widely dilated orifices of glands that appear on the surface of skin. They are visible to the naked eye, and their appearance is very familiar to us. Small-scale variations in skin color and reflectance, typically caused by freckles and moles, are not the topic of this work.

Acquiring high-resolution face geometry with small features is a difficult, expensive, and time-consuming task. Commercial active or passive photometric stereo systems (e.g., EyeTronics, 3QTech) have fast acquisition times (in seconds) and have become affordable. However, they only capture large wrinkles and none of the important small geometric details that make skin look realistic. Laser scanning systems (e.g., CyberWare) may be able to capture the details, but they are expensive and require the subject to sit still for

tens of seconds, which is impractical for many applications. Moreover, the resulting 3D geometry has to be filtered and smoothed due to noise and motion artifacts. The most accurate method is to make a plaster mold of a face and to scan this mold using a precise laser range system (e.g., XYZRGB). However, not everybody can afford the considerable time and expense this process requires. In addition, the molding compound may lead to sagging of facial features [Pighin and Lewis 2005].

In this paper we present a statistical face model that makes it possible to extract, transfer, and synthesize small facial features. Our approach is based on the analysis of high-resolution face scans. Using a commercial 3D face scanner and a custom-built face scanning dome, we acquire high-resolution 3D face geometry. We use displaced subdivision surfaces [Lee et al. 2000] to separate facial detail (the displacement map) from the smooth underlying mesh (the subdivision surface). We then apply a novel tile-based extension of parametric texture analysis/synthesis [Heeger and Bergen 1995] to compile a spatially-varying statistical model for the displacement map. The model is used to synthesize fine geometric details (e.g., new wrinkles and pores) on the same or another base mesh. Our method of generating facial detail can be used in conjunction with methods such as the PCA-based analysis of Blanz and Vetter [1999], which are well suited for modeling coarse facial deformations.

This statistical model of facial detail, combined with our database of statistics for 149 subjects, opens ground for new applications. Chief among these is adding detail to a low resolution face (Figure 1b). This application allows anybody to create plausible, detailed face meshes from low resolution scans by synthesizing detailed displacement maps from statistics that we make publicly available. Second, users can interpolate between multiple faces without blurring high resolution details. Finally, they can age or de-age a face using statistics of older or younger subjects in the database (Figure 1c).

The main contributions of this work are:

- a statistical model of fine geometric facial features based on an analysis of high-resolution face scans;
- an extension of parametric texture analysis and synthesis methods to spatially-varying geometric detail;
- a database of detailed face statistics for a sample population that is made available to the research community;
- new applications, including introducing plausible detail to low-resolution face models and adjusting face scans according to age and gender.

1.1 Previous Work

There has been a wealth of research in capturing and modeling faces in computer graphics and computer vision. In this overview, we focus on the relevant work in statistical modeling and synthesis of face geometry.

Morphable Face Models: DeCarlo et al. [1998] used variational techniques to synthesize faces with some characteristic distances consistent with measured data. Because of the sparseness of the measured data compared to the high dimensionality of possible faces, the synthesized faces are not as plausible as those produced using a database of scans.

Blanz and Vetter [1999] were the first to study the space of faces. They use Principal Component Analysis (PCA) to generate a linear morphable face model from a database of face scans. This was extended by Vlasic et al. [2005], who used multi-linear face models to study and synthesize variations in faces along several axes, such as identity and expression. Morphable models have also been used in 3D face reconstruction from photographs [Blanz and Vetter 1999; Fuchs et al. 2005] or video [Vlasic et al. 2005]. These methods synthesize a range of plausible meshes.

However, current linear or locally-linear morphable models cannot be directly applied to analyzing and synthesizing high-resolution face models. The dimensionality (i.e., length of an eigenvector) of high-resolution face models is very large, and an unreasonable amount of data (i.e., number of eigenvectors) would be required to capture small facial details. In addition, during construction of the model, it would be difficult or impossible to find exact correspondences between high resolution details of all the input faces. Without correct correspondence, the weighted linear blending performed by these methods would blend small facial features, making the result implausibly smooth (see Figure 7a). We address the shortcomings of morphable models on meshes that are more than an order of magnitude larger than those used in Blanz and Vetter [1999].

Physical/Geometric Wrinkle Modeling: Some work has focused on directly modeling skin folding and physics [Wu et al. 1995; Wu et al. 1997; Boissieux et al. 2000]. However, these models are not easy to control, and do not produce results that can match high-resolution scans in plausibility.

Other work suggested modeling wrinkles geometrically [Bando et al. 2002; Larboulette and Cani 2004]. Such methods generally proceed by having the user draw a wrinkle field and select a modulating (cross-sectional) function. The wrinkle depth is then modulated as the base mesh deforms to conserve length. This allows user control, and is well-suited for long, deep wrinkles (e.g., across the forehead). However, it is difficult for the user to realistic sets of wrinkles, and these methods do not create pores and other fine scale skin features.

As such, these methods are well suited to complement our technique. They can be used to create long-range (and in some cases user-specified) wrinkle structures, and our technique can be used to adjust the results to match face detail statistics, adding pores and other fine-scale facial detail to the models.

1.2 Texture Synthesis

To analyze and synthesize skin detail we apply texture synthesis methods to 3D geometry displacements. The two main classes of texture synthesis methods are Markovian and parametric texture synthesis.

Markovian texture synthesis methods treat the texture image as a Markov random field. They typically build up an image patch by patch (or pixel by pixel) by searching the sample texture for a region whose neighborhood matches the neighborhood of the patch or pixel to be synthesized. This method was first proposed by Efros and Leung [1999]. Hertzmann et al. [2001] extended it for a number of applications, including a super-resolution filter, which created a higher-resolution image from a low-resolution one using a sample pair of low and high resolution images. Liu et al. [2001] applied similar ideas to the specific task of hallucinating detail on a lower-resolution facial image. Markovian methods have also been used for generation of facial geometry by Haro et al. [2001] to grow fine-scale normal maps from nickel-sized samples taken at different areas of the face.

Parametric methods extract a set of statistics from the sample texture. Synthesis starts with a noise image, and coerces it to match the statistics. The original method was proposed by Heeger and Bergen [1995], where the chosen statistics were histograms of a steerable pyramid of the image. Portilla and Simoncelli [2000] used a larger and more complex set of statistics to generate a greater variety of textures, but we found the simpler approach of Heeger and Bergen to be sufficient for our application. Matusik et al. [2005] interpolated between textures while preserving sharpness, similarly to how we interpolate between high-resolution face meshes. We augment this method with spatially varying texture statistics, as has been previously explored in the context of image segmentation (e.g., [Brox and Weickert 2006]).

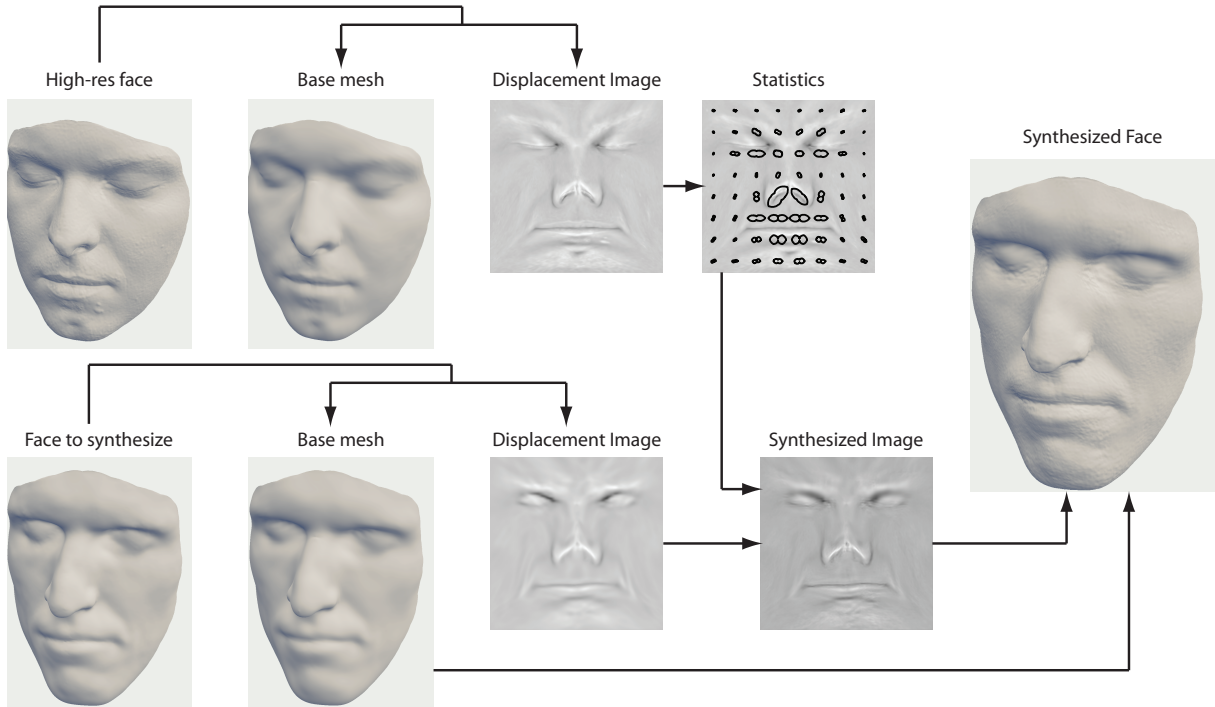


Figure 2: System overview. Faces are separated into smooth base meshes and detailed displacement images. Statistics are extracted from the displacement images. The image to be synthesized is coerced to match the desired statistics, and added back to the base mesh to create the new face mesh.

A key decision in designing our system was whether to use parametric or Markovian techniques. Markovian techniques have recently produced more impressive images, and have been shown to synthesize a larger range of textures. Instead, we chose to use a parametric texture model for the following reasons:

- A parametric model yields statistics for study. We can perform analysis, compare the statistics of groups, and gain some understanding of the detail we are synthesizing. This also allows for easier and more direct manipulation of statistics: it is simple to take mean statistics of a group of faces, use PCA, etc.
- Our parametric method respects detail existing at the start of synthesis and automatically adjusts to the resolution of the given target mesh. In contrast, example-based synthesis overwrites the image at whatever resolution it is being used.
- A parametric model allows for compression of data. The statistics we use are several orders of magnitude smaller than the original images. Example-based synthesis would require the images to be available in their entirety.
- The meshes and images used for this paper are restricted from being disseminated in their full resolution because of privacy concerns. A parametric model allows us to share statistics with other researchers to allow the synthesis and further study of high resolution faces. Markovian techniques, which require the full images, do not permit this.

We modified the method of Heeger and Bergen to the case of spatially-varying statistics, and reduced these statistics to a smaller set without sacrificing the quality of synthesized faces. We found this choice to strike a good balance between compactness and ease of use on the one hand, and quality of the synthesized meshes on the other.

2 System

Our system consists of an analysis stage, executed once for each scan in our database, and a number of application-driven synthesis

stages (Figure 2). Analysis begins with a high-resolution (500k polygons) scan of a face. This is reparameterized, and separated into a base mesh and a displacement image. The latter is broken into tiles, and statistics are computed separately for each tile. Synthesis uses these statistics to adjust a different face’s displacement image, and the result is combined with a base mesh to form a new face. Depending on the application, this sharpens high-frequency data lost due to interpolation, adds detail missing because of the low resolution of the scan, or adjusts detail to change the age of a face.

2.1 Data Acquisition

The first step of our process is to acquire high resolution face scans for a number of subjects. Each subject sits in a chair with a head rest to keep the head still during acquisition. We capture the complete face geometry using the commercial face-scanning system from 3QTech (www.3dmd.com). The output mesh contains 40k vertices and is manually cropped and cleaned. We then refine the mesh to about 700k vertices using Loop subdivision [Loop 1987].

The resulting mesh is too smooth to resolve fine facial details, and we capture these details using photometric stereo. The subject is surrounded by a geodesic dome with multiple cameras and LED lights, similar to the LightStage of Debevec et al. [Debevec et al. 2000; Debevec et al. 2002]. The system sequentially turns on each light while simultaneously capturing images from different viewpoints with 16 cameras. Using the image data, we refine the geometry and compute a high-resolution normal map using photometric stereo [Barsky and Petrou 2001]. Finally, we use the method of Nehab et al. [2005] to combine the high-resolution normals with the low-resolution geometry, accounting for any bias in the normal field. The result is a high-resolution (500k polygons) face mesh with approximately 0.5 mm. sample spacing and low noise (below 0.05 mm.), which accurately captures fine geometric details. We report the details about the acquisition system and its calibration procedure in [Weyrich et al. 2005].

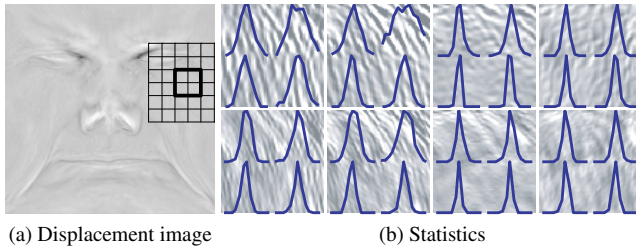


Figure 3: (a) The displacement image is divided into tiles. Filter responses and histograms of the outlined 2×2 section are shown in (b). All orientations and two scales are shown; tiles with more content have wider histograms.

2.2 Remeshing

The second stage of our process is to put the faces into per-vertex correspondence. To identify facial features, we have defined 21 feature points, which must be manually located on each face. With predefined connectivity, these form a “marker” mesh. Following the approach of Guskov et al. [2000], the marker mesh is subdivided and re-projected in the direction of its normals onto the original face scan several times, yielding successively more refined approximations of the original scan. Because the face meshes are smooth relative to the marker mesh, self-intersections do not occur.

A subtle issue is choosing the right subdivision strategy. If we use an interpolating subdivision scheme, marker vertices remain in place and the resulting meshes have relatively accurate per vertex correspondences. However, butterfly subdivision [Dyn et al. 1990] tends to pinch the mesh, and linear subdivision produces a parameterization that has discontinuities in its derivative. An approximating method, such as Loop subdivision, produces smoother parameterizations at the cost of moving vertices and making the correspondences worse. The choice of subdivision scheme then offers the tradeoff between a smooth parameterizations and better correspondences. Since the first several rounds of subdivision would move vertices the furthest under approximating schemes, we choose two linear subdivisions followed by two Loop subdivisions (projecting onto the original face after each subdivision).

This gives us a coarse control mesh using which we can define a scalar displacement image that captures the remaining face detail following Lee et al. [2000]. Specifically, we subdivide this mesh three times with Loop subdivision (without re-projecting). This yields a smooth, fine mesh we refer to as the base mesh. We project the base mesh onto the original face along the normals of the base mesh, and define the displacement image by the length of the displacement at each vertex. To map this to an image, we start with the marker mesh mapped in a pre-defined manner to a rectangle, and follow the sequence of subdivisions in the rectangle. We chose to represent the displacement images with 1024×1024 samples. One such displacement image is shown in Figure 3a.

2.3 Extraction of Statistics

In our next step, we aim to build a statistical model of the fine detail in the facial displacement maps. Our goal is to represent the displacements with enough accuracy to retain wrinkles and pores in a compact model suitable for synthesis of details on new faces.

Our method is an extension of texture synthesis techniques commonly used for images. Following Heeger and Bergen [1995], we extract histograms of a sample texture’s steerable pyramid [Simoncelli and Freeman 1995] to collect the texture’s statistics at several scales and orientations. Direct application of previous methods would define a set of global statistics for each face, which are not immediately useful for our problem since the statistics of facial detail vary spatially. We make the intuitive modification of taking statistics of image tiles to capture the spatial variation. Specifically, we decompose the images into 256 tiles in a 16×16 grid and build

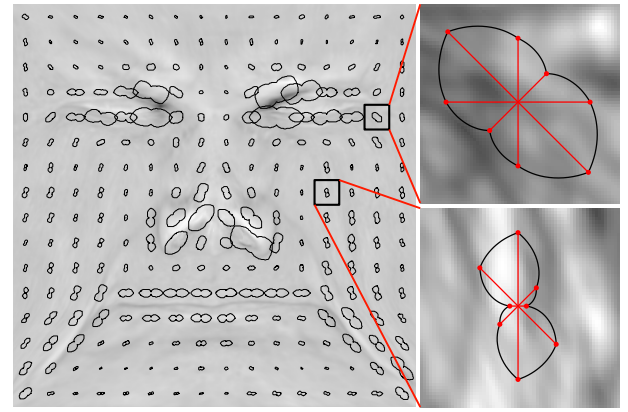


Figure 4: Visualization of statistics of an old face for one scale: circles at each tile center are expanded in each filter direction by the standard deviation of the filter response within the tile. These statistics capture the spatially local directionality and amount of detail at a given scale.

steerable pyramids with 4 scales and 4 orientations for each tile. We also consider the high-pass residue, but not the low-pass residue, which we take to be part of the base mesh. This produces 17 filter outputs. Sample histograms and filter outputs for two scales are shown in Figure 3b.

Storing, analyzing, interpolating, and visualizing these histograms is cumbersome, since they contain a lot of data. However, we observe that the main difference between the histograms in the same tile for different faces is their width. Following this observation, we approximate each histogram by the standard deviation of the pixels in the tile. This approximation allows significant compression of the data. The reduced statistics of a face then contain a scalar for each tile in each filter response: $17 \times 16 \times 16 = 4,352$ scalars, compared with $128 \times 17 \times 16 \times 16 = 557,056$ scalars in the histograms (if we use 128 bins), and $1024 \times 1024 = 1,048,576$ scalars in the original image. We have confirmed empirically that faces synthesized from these reduced statistics are visually indistinguishable from those synthesized with the full set of histograms.

This reduced set of statistics not only decreases the cost of storage and analysis, but also allows for easier visualization and better understanding of how the statistics vary across a face and across populations of faces. For example, for each scale and tile, we can draw the standard deviations for all filter directions as a circle expanded in each direction by the standard deviation computed for that direction. Figure 4 shows such a visualization for the second scale of the pyramid (512×512 pixels) for the face in Figure 5a.

2.4 Synthesis

The final step is to use these statistics to synthesize facial detail. Heeger and Bergen [1995] accomplish this as follows. The sample texture is expanded into its steerable pyramid. The texture to be synthesized is initialized with noise, and is also expanded. Then, the histograms of each filter of the synthesized texture are matched to those of the sample texture, and the pyramid of the synthesized texture is collapsed, and expanded again. Since the steerable pyramid forms an overcomplete basis, collapsing and expanding the pyramid will change the filter outputs if they have been adjusted independently. However, repeating the procedure for several iterations has been found to lead to convergence.

This procedure needs to be modified to use our reduced set of spatially varying statistics. The histogram-matching step is replaced with matching standard deviations by scaling the range of values in each tile. Each tile is grown to the centers of its neighboring tiles, so that each pixel is covered by its four neighboring tiles (except on the perimeter, where pixels are covered by one tile at the corners and by two tiles elsewhere). The matching of standard deviations

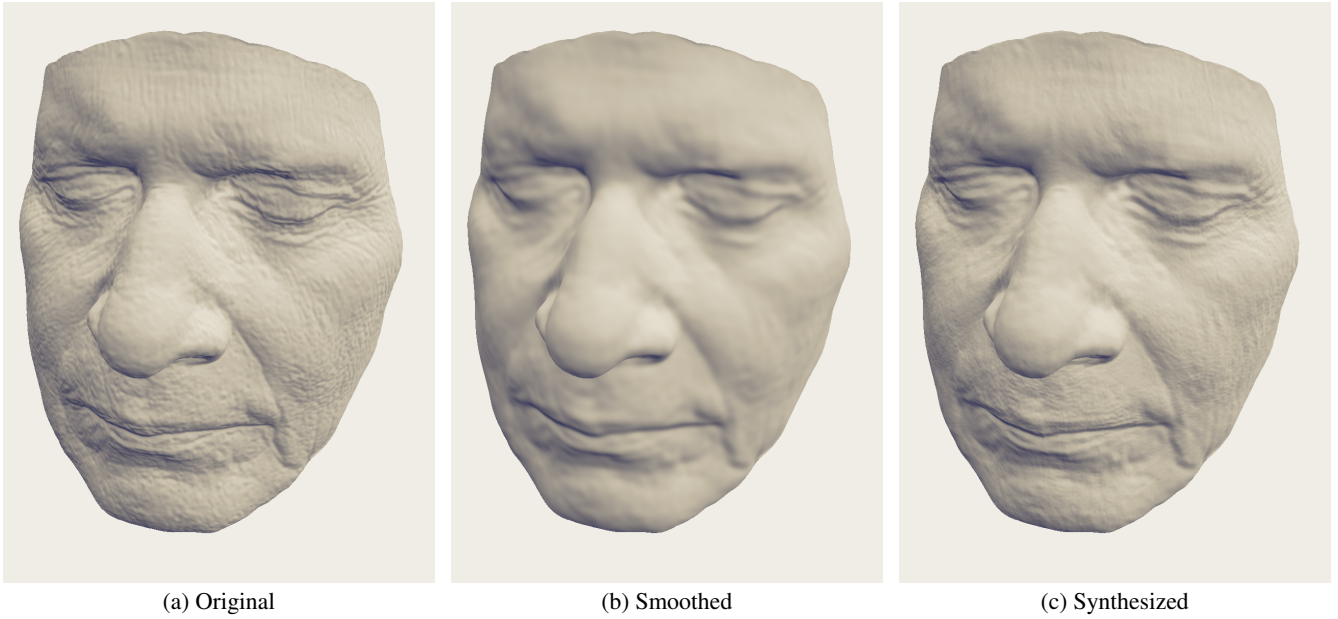


Figure 5: To evaluate the plausibility of our analysis/synthesis methods, we begin with a high-resolution face (a) and apply smoothing (b). We then re-synthesize a face by using only the statistics from (a). Though the synthesized details differ from the original face, they appear qualitatively similar.

is done separately for each (expanded) tile, using its own statistics, resulting in several synthesis results available at each pixel. The final image is assembled from the individually synthesized tiles, with bilinear weighting used to control the interpolation.

Note that adjusting standard deviations in this manner does not end with the synthesized tiles having the same standard deviation as the target tiles. If, however, this step is repeated several times, the deviation of the synthesized tiles converges to the desired deviation. In practice, performing this matching iteratively does not result in a mesh visually distinguishable from that synthesized with only one matching step per iteration.

Parametric texture synthesis usually begins with a noise image. Instead, for our applications, we begin synthesis with an existing displacement image. In this case, iterative matching of statistics does not add new detail, but modifies existing detail with properly oriented and scaled sharpening and blurring. If the starting image has insufficient detail, we add noise to it. We use Gaussian noise, and our experiences suggest that similarly simple noise models (e.g., Perlin noise [Perlin 1985]) lead to the same results. We must be careful to add enough noise to cover possible scanner noise and meshing artifacts, but not so much that the amount of noise overwhelms existing detail.

As an empirical validation of this approach, and to illustrate its strengths and limitations, we attempt to re-create the details of a high resolution face that has been smoothed by coercing it to match its original statistics. We show the result in Figure 5. Note that the synthesized wrinkles at corners of the eyes, and the forehead, are oriented correctly. Moreover, the eyelids and crease of the mouth have been sharpened. The synthesis is faithful to detail existing in the starting mesh: it sharpens existing wrinkles. But, the newly created wrinkles are not as elongated or correlated as they are on the original mesh.

These observations point to several limitations of the synthesis process, which become less prominent with increasing resolution of the starting mesh. Completely new detail created in this process can only come from the added noise. As such, it is limited by the noise model and basis functions of the filters. With white noise and a steerable pyramid, this does not produce long, correlated structures. Also, in the absence of detail, this method struggles with

deterministic facial details, such as creases on the mouth or eyelids. When such a crease is insufficiently sharp, the synthesis may instead sharpen a similarly oriented indentation that happens to be in the same tile. Despite these drawbacks, we find that our method creates plausible faces for several applications, as shown in the following section.

3 Applications

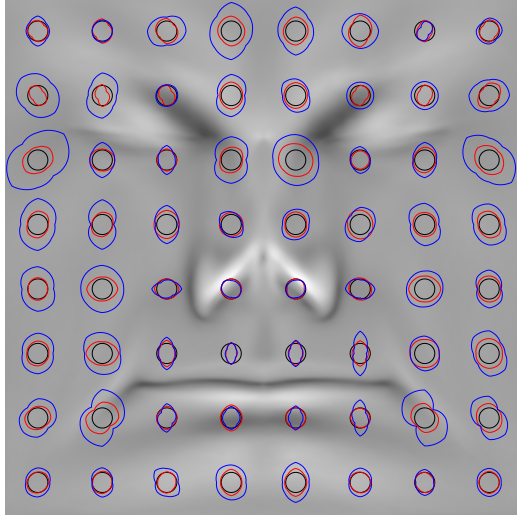
Our statistical model of detailed face geometry is useful for a range of applications. The statistics allow analysis of facial detail, for example, to track changes between groups of faces. They also allow synthesis of new faces for applications such as sharpness-preserving interpolation, adding detail to a low resolution mesh, and aging.

3.1 Analysis of Facial Detail

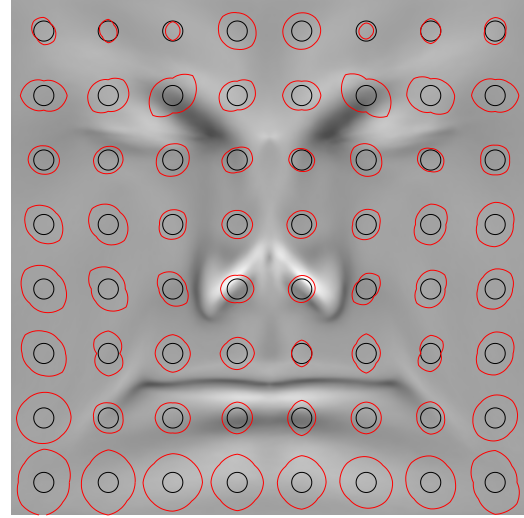
As a first application, we consider analysis and visualization of facial detail. This may be useful, for example, for classification of face scans. We wish to gain insight into how facial detail changes with personal characteristics by comparing statistics between groups of faces. To visualize the differences between groups, we normalize the statistics of each group to the group with the smallest amount of content, and compare the mean statistics on a tile-by-tile basis. We follow this approach to study the effects of age and gender.

Age To study the effect of age, we compare three groups of males aged 20-30 (21 subjects), 35-45 (17 subjects), and 50-60 (5 subjects). The group statistics are shown in Figure 6a, colored with black, red, and blue respectively. We see that wrinkles develop more from the second age group to the third than from the first to the second. This suggests that after the age of 45, the amount of roughness on skin increases more rapidly. These images also suggest that around that age, more directional, permanent wrinkles develop around the corners of the eyes, the mouth, and some areas on the cheeks and forehead.

Gender To investigate how facial detail changes with gender, we compare in Figure 6b 20-30 year-old women (10 subjects, in black) to males of the same age group (21 subjects, in red). The change of



(a) Variation across age



(b) Variation across gender

Figure 6: Comparison of statistics between groups, normalized to the first group. (a) Age (black, red and blue for groups of increasing age). (b) Gender (females in black and males in red).

high frequency content from females to males is different in character from that of the change between varying age groups. Males have more high frequency content, but the change, for this age group, is relatively uniform and not as directional. In addition, males have much more content around the chin and lower cheeks. Although none of the scanned subjects had facial hair, this is likely indicative of stubble and hair pores on the male subjects.

3.2 Interpolation

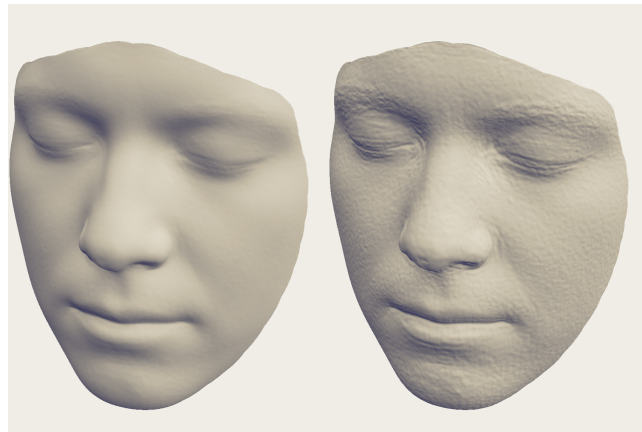
There are a number of scenarios in which it may be useful to interpolate between faces. A user interface for synthesizing new faces, for example, may present the user with faces from a data set, have her define a set of weights, and return a face interpolated from the input faces with the given weights. Alternatively, linear models (e.g., [Blanz and Vetter 1999]) could synthesize a face as a weighted sum of a large number of input faces. However, interpolation be-

tween a large number of faces blurs detail, leading to implausibly smooth faces. For example, Figure 7a shows an average of 14 faces produced with a linear model. Note that the fine details of the skin geometry are not visible.

Representing our statistics for each face as a vector of numbers, we can define algebraic operations such as interpolation on them. To prevent the blurring of detail that occurs with interpolation, we would like to synthesize a face with statistics matching the interpolation of the input face statistics, which we call the target statistics. We augment algebraic interpolation of mesh vertices with a detail-adjustment step that coerces the result to match the target statistics.

Each of the input faces is remeshed to yield a base mesh and a displacement image from which statistics are calculated. The base mesh of the output face comes from the interpolation of the base meshes of input faces, and remains unchanged. The displacement image of the output face is initialized with the interpolation of the input displacement images. The initial image does not match the target statistics, and we coerce it to match them, resulting in the output displacement image. Applying the output displacement image to the base mesh yields the synthesized face.

Figure 7b uses this method to sharpen the result of Figure 7a. This sharpens the creases of the mouth, lips, and eyelids. It adds slight vertical indentations to the lips, without smoothing them horizontally. It also creates pores on the cheeks, chin, and forehead.



(a) Average face

(b) Adjusted statistics

Figure 7: Interpolation between 14 faces aged 30 to 35. (a) The arithmetic average looks implausibly smooth. (b) The average with adjusted statistics has sharper creases on mouth and eyelids, and pores on cheeks, chin, and forehead.

3.3 Adding Detail

Low-resolution meshes can be produced from a variety of sources. Such meshes can come from a commercial scanner (as in Figure 8a), can be created manually, or can be synthesized using a linear model from a set of input meshes. On the other hand, high resolution meshes are difficult and expensive to obtain. It is useful to be able to add plausible high-resolution detail to a low-resolution face without having to obtain high-resolution meshes. One might wish, given a database of face images, to select a face to which to adjust statistics. Alternatively, it may be convenient to adjust the low-resolution mesh to the mean statistics of an age group.

Our framework allows the synthesis of detail on a low resolution mesh in a straightforward manner. We start with the displacement image of the low-resolution mesh, adjust it to match target statistics,

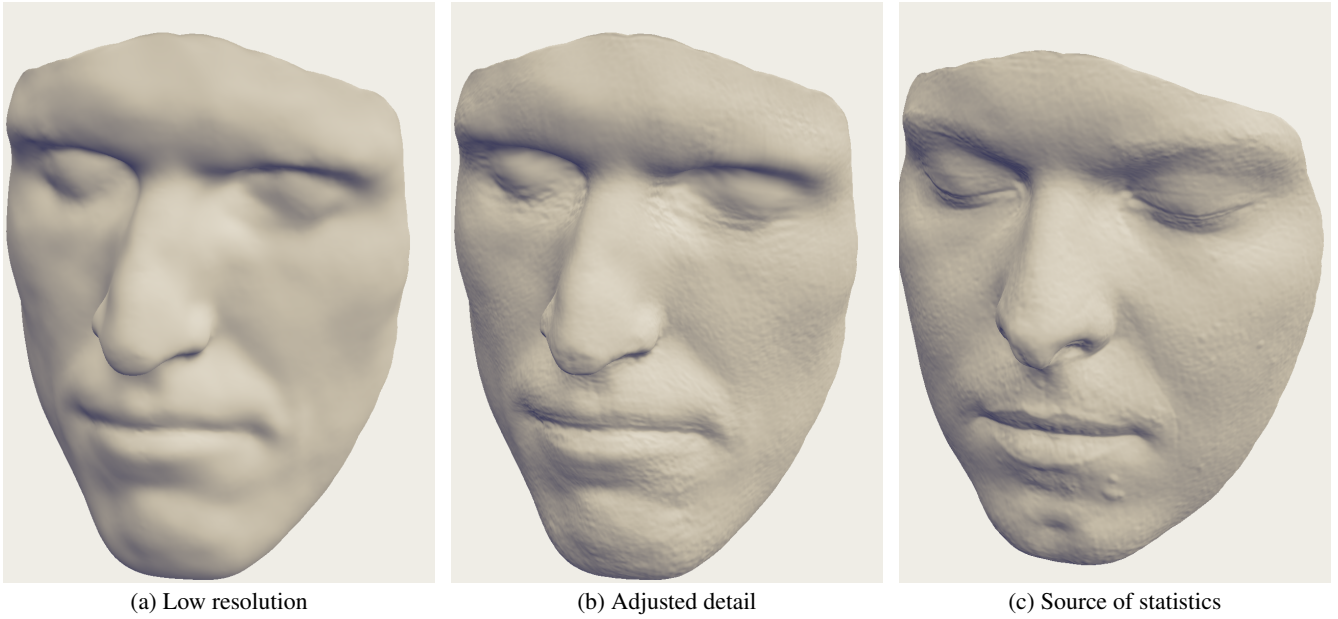


Figure 8: Adding detail to a face. We start with a low-resolution face obtained from a scanner (a), and synthesize detail on it (b) to match the statistics of the high-resolution face (c). Creases in the mouth and eyelids were sharpened, and pores of similar amplitude, distribution, and orientation as in the target face were added.

and add it back to the base mesh. This process inherently takes advantage of the available level of detail in the low resolution mesh; therefore, a more accurate starting mesh will yield a more plausible face.

Figure 8a shows a low resolution face acquired with a commercial 3D scanner by 3QTech (www.3dmd.com). Figure 8b shows the result of matching the statistics of this face to those of a high resolution face shown in Figure 8c. The synthesis keeps the features of the original subject, while matching the scale, amount, and orientation of high-resolution detail of the target subject. It also sharpens the creases on the mouth and eyes, without introducing much noise.

3.4 Aging and De-aging

It may be desirable to change the perceived age of a face mesh. For example, we may want to make an actor look older or younger. The goal is to create a plausible older version of a young face, and vice-versa. Because facial detail plays such a key role in our perception of age, and because scans for the same individual taken at different ages are not available, hallucinating aging is a challenging task.

Blanz and Vetter [1999] would perform aging by linear regression on the age of the meshes in the set, finding an “aging” vector. However, this approach suffers from the same problem as interpolation: wrinkles do not line up, and detail is blurred. It also does not solve the problem of ghosting and disregards existing detail. This approach has another problem that is highlighted by de-aging. The “aging” vector captures large-scale features of aging, such as sagging cheeks. However, a particular older face has more of these “aged” features in some areas and less in others; therefore, subtracting the aging vector to de-age often creates a strange-looking face (e.g., “negative” sagging of cheeks).

Aging falls neatly into our framework. We select a young face and an old face. To age, we start with the image of the young face, and coerce it to match statistics of the old face. The resulting image contains the detail of the young face, with wrinkles and pores sharpened and elongated to adjust to the statistics of the old face.

To make the adjustment convincing, we also need to change the underlying coarse facial structure. Our hierarchical decomposition of face meshes suggests a way to make such deformations. Prior

to the displacement map, our remeshing scheme decomposes each face into a marker mesh and four levels of detail. In this case, we can take the marker mesh and lower levels of details from the young mesh (since these coarse characteristics are individual and do not change with age), and the higher levels of details from the old mesh.

An example of this process is shown in Figure 9 for the two male subjects in Figures 9a and d. Adjustment of details is shown in Figures 9b and e, and adjustment of details along with coarse changes in Figures 9c and f. Near corners of the eyes and the forehead, the young face is adjusted to have the highly directional wrinkles of the old face. The young face also acquires the creases below the sides of the mouth. The de-aged face has its wrinkles smoothed (for example, on the cheek), but retains sharpness in the creases of the mouth and eyelids.

4 Comparison to Alternative Methods

In this section, we compare our approach to simple alternatives, focusing only on aging because comparisons for other applications are analogous.

Our comparison focuses on adding detail to a young face after it has been warped to fit the coarse geometry of an old face, as described in the previous section. A warped young face is shown in Figure 10a, and the results of aging it to match the details of an old face (Figure 10b) are shown for our method in Figure 10c and for three alternative approaches in Figures 10d-f:

1. **Noise:** facial details could be simulated by adding noise. For example, in Figure 10d, we show the result of adding Perlin noise to the displacement image of Figure 10a. Clearly, the face does not look realistic.
2. **Sharpen:** aging could be simulated by sharpening the warped mesh of a young face, thereby deepening creases and enlarging pores (Figure 10e). While sharpening approximates aging in some areas, it does not respect the orientations of details in the target old face (e.g., wrinkles near the corner of the eyes), and it sharpens creases that should not get sharper with age (e.g., the crease of the mouth).

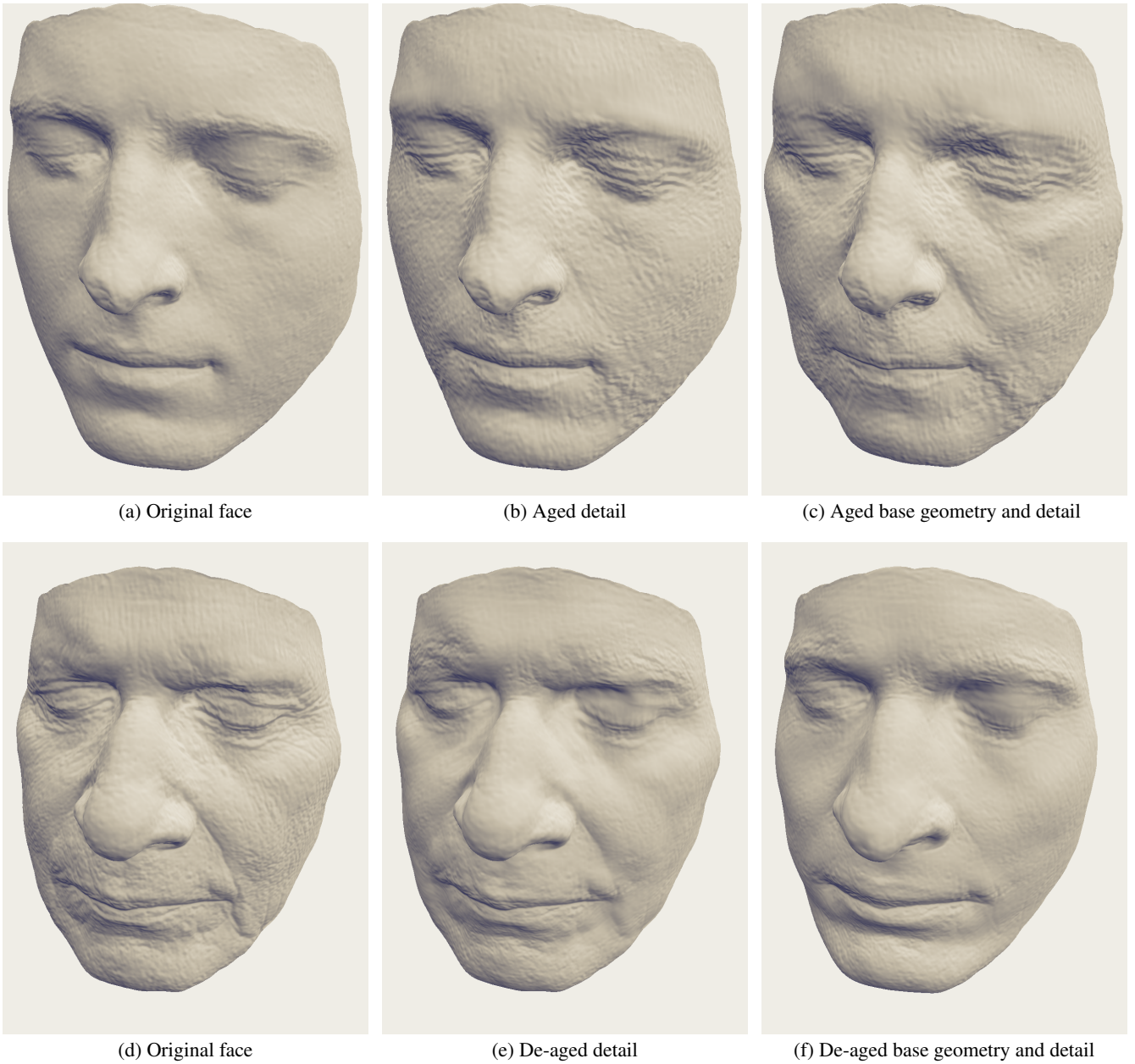


Figure 9: Adjusting detail. Top row contains a young face getting older, and bottom row contains an old face getting younger. The young face is adjusted using the statistics and base mesh of the old face, and vice-versa. First column has original faces, second column has synthesized details, and third column has an additional coarse adjustment.

3. **Detail transfer:** a face could be aged by replacing its displacement image with one extracted from an old face (Figure 10f). This approach amounts to simply warping the old face to coarsely fit the young face. It completely overwrites existing high frequency detail of the young face, which removes important individual features and creates new wrinkles without respecting existing ones. For example, notice that eyes and lips in Figure 10f do not retain the individual characteristics of those in Figure 10a.

Our method combines the strengths of these three approaches under one framework: 1) added noise is scaled and oriented to match the statistics of the aged face; 2) features are sharpened only where the aged face contains sharp features; and 3) existing details (e.g., wrinkles) are modified rather than replaced.

5 Dissemination

As a final result, we make available to the public the statistics extracted from the high-resolution face geometry of 149 subjects of various age, gender, and race. The ages of our subjects range between 15 and 83 years, and there are 114 male and 35 female subjects. Most of the subjects are Caucasians (81), with Asians making up the second largest group (63), and African-Americans the smallest (5). For all faces, we make available both the Heeger-Bergen statistics (full histograms at 4 orientations, 4 scales, and 16×16 tiles) and the reduced statistics extracted from the high resolution scan. These statistics should make it possible for other researchers to add details to low-resolution scans of new faces and investigate new applications beyond the ones considered in this paper.

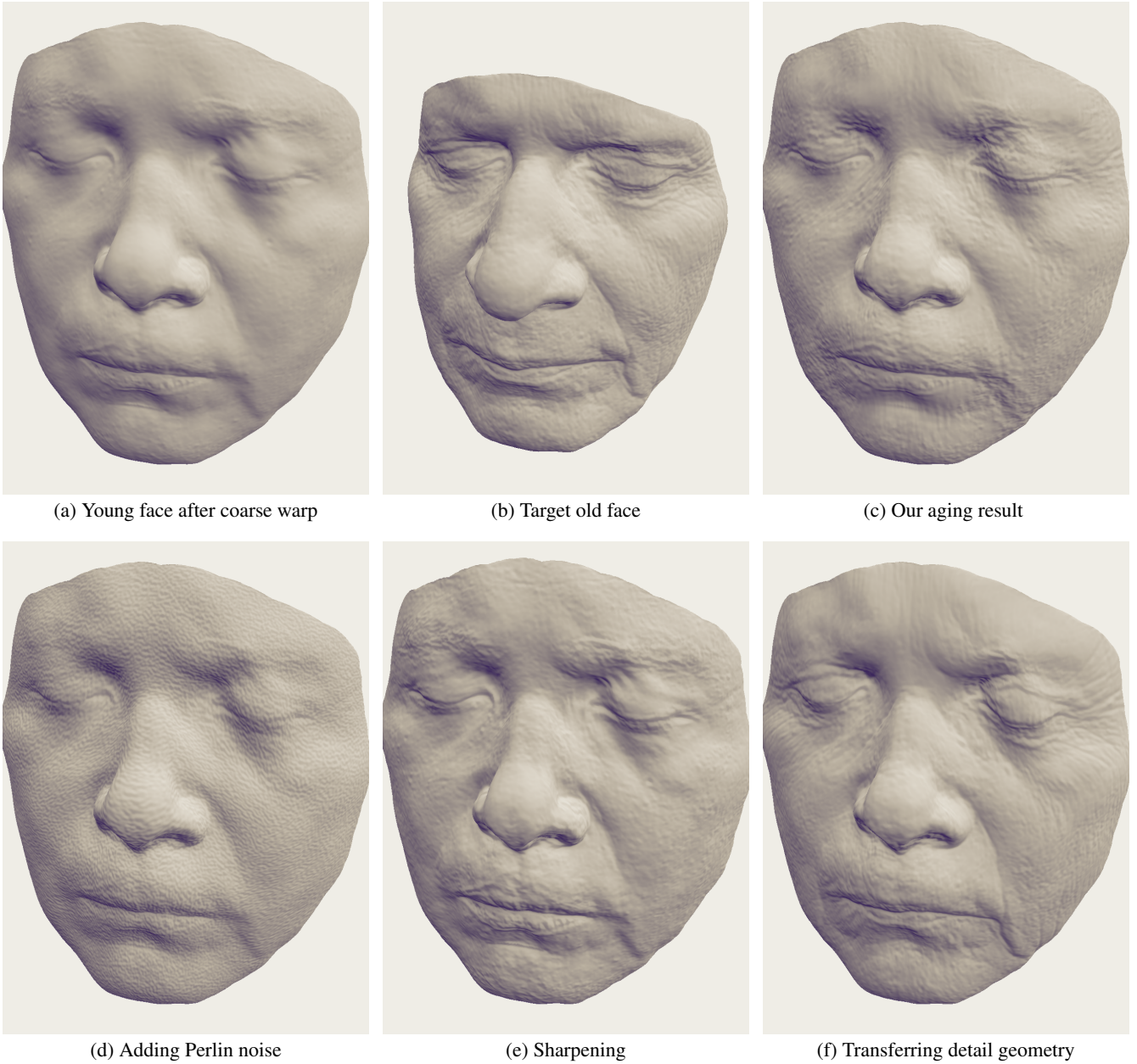


Figure 10: Comparison of our aging method with simple alternative approaches: (a) shows a young face after low-resolution deformation, but without any detail adjustment, and (b) has the target old face. Our synthesized result is shown in (c), and the bottom row contains alternative approaches: adding noise (d), sharpening (e), and direct geometry transfer (f). Each of these has deficiencies: (d) contains detail that is too uniform, (e) does not have oriented wrinkles (e.g., eye corner) and overly sharpened features (e.g., mouth crease), and (f) loses some individual characteristics of (a) (e.g., eyes).

6 Conclusion and Future Work

We have presented a method for analyzing and synthesizing facial geometry by separating face meshes into smooth base meshes and displacement images, extracting statistics from the images, and synthesizing new faces with fine details based on extracted statistics. This work takes a small step towards more realistic facial models and suggests several directions for further research.

Within the existing framework, several areas of improvement may lead to more plausible facial synthesis. The statistical model we use can be augmented to include correlations within and between steerable pyramid filter outputs [Portilla and Simoncelli 2000]. Moreover, the noise introduced at the first stage of synthesis could

be made more realistic by incorporating information about the existing overall geometry, the results of an explicit large-scale wrinkle model, or even image-based information from photographs. These refinements would improve the ability of our synthesis to capture long, continuous features (e.g., long wrinkles across the forehead), at the expense of additional processing and complexity in our statistical representation.

It would be interesting to expand this model into the larger context of face synthesis, rendering, and animation. This involves extending our model to capture not only statistics of geometric detail, but also the correlations between shape, appearance, and movement. For example, we might capture the fact that pores appear in both geometry and reflectance or that the shape of wrinkles depends

on expression, and use this information to produce compelling syntheses that are consistent with all available data.

Moving beyond the applications presented here, we would like to explore the use of the statistical model for identifying or classifying faces. The results presented here suggest that the statistics we use offer cues about the age and gender of an individual, and represent a heretofore underutilized source of information that may complement overall shape and appearance for the purposes of identification. More generally, the availability of high-resolution scans of faces from a broad population may enable new classes of applications in comparative analysis and realistic rendering.

Acknowledgements

We would like to thank Fredo Durand and Jason Lawrence for initial discussions about this work, Diego Nehab for help with mesh processing, and Bernd Bickel, Chien Tu, Agata Lukasinska, and Natalie Charton for processing the face data. We would also like to thank the Tigglyph reviewers and the anonymous SIGGRAPH reviewers for their valuable comments. This work is partially supported by the National Science Foundation under grants CCF-0347427, CCR-0093343, and IIS-0121446, by the Air Force Research Laboratory under grant FA8650-04-1-1718, and by the Sloan Foundation.

References

- BANDO, Y., KURATATE, T., AND NISHITA, T. 2002. A simple method for modeling wrinkles on human skin. In *Pacific Conference on Computer Graphics and Applications*, 166–175.
- BARSKY, S., AND PETROU, M. 2001. Colour photometric stereo: simultaneous reconstruction of local gradient and colour of rough textured surfaces. In *Eighth IEEE International Conference on Computer Vision*, vol. 2, 600–605.
- BLANZ, V., AND VETTER, T. 1999. A morphable model for the synthesis of 3d faces. In *Proceedings of ACM SIGGRAPH 99*, ACM Press/Addison-Wesley Publishing Co., New York, NY, USA, 187–194.
- BOISSIEUX, L., KISS, G., MAGNENAT-THALMANN, N., AND KALRA, P. 2000. Simulation of skin aging and wrinkles with cosmetics insight. In *Computer Animation and Simulation 2000*, 15–27.
- BROX, T., AND WEICKERT, J. 2006. Level set segmentation with multiple regions. *IEEE Transactions on Image Processing*.
- DEBEVEC, P., HAWKINS, T., TCHOU, C., DUKER, H.-P., SAROKIN, W., AND SAGAR, M. 2000. Acquiring the reflectance field of a human face. In *Proceedings of ACM SIGGRAPH 2001*, 145–156.
- DEBEVEC, P., WENGER, A., TCHOU, C., GARDNER, A., WAESE, J., AND HAWKINS, T. 2002. A lighting reproduction approach to live-action compositing. *ACM Transactions on Graphics (SIGGRAPH 2002)* 21, 3 (July), 547–556.
- DECARLO, D., METAXAS, D., AND STONE, M. 1998. An anthropometric face model using variational techniques. In *Proceedings of ACM SIGGRAPH 98*, ACM Press, New York, NY, USA, 67–74.
- DYN, N., LEVINE, D., AND GREGORY, J. A. 1990. A butterfly subdivision scheme for surface interpolation with tension control. *ACM Trans. Graph.* 9, 2, 160–169.
- EFROS, A. A., AND LEUNG, T. K. 1999. Texture synthesis by non-parametric sampling. In *ICCV '99: Proceedings of the International Conference on Computer Vision-Volume 2*, IEEE Computer Society, Washington, DC, USA, 1033.
- FUCHS, M., , BLANZ, V., LENSCH, H., AND SEIDEL, H.-P. 2005. Reflectance from images: A model-based approach for human faces. Research Report MPI-I-2005-4-001, Max-Planck-Institut für Informatik, Stuhlsatzenhausweg 85, 66123 Saarbrücken, Germany. Accepted for publication in IEEE TVCG.
- GUSKOV, I., VIDIMČE, K., SWELDENS, W., AND SCHRÖER, P. 2000. Normal meshes. In *Proceedings of ACM SIGGRAPH 2000*, ACM Press/Addison-Wesley Publishing Co., New York, NY, USA, 95–102.
- HARO, A., ESSA, I., AND GUENTER, B. 2001. Real-time, Photo-Realistic, physically based rendering of fine scale human skin structure. In *Rendering Techniques*, 53–62.
- HEEGER, D. J., AND BERGEN, J. R. 1995. Pyramid-based texture analysis/synthesis. In *Proceedings of ACM SIGGRAPH 95*, ACM Press, New York, NY, USA, 229–238.
- HERTZMANN, A., JACOBS, C. E., OLIVER, N., CURLESS, B., AND SALESIN, D. H. 2001. Image analogies. In *Proceedings of ACM SIGGRAPH 2001*, ACM Press, New York, NY, USA, 327–340.
- IGARASHI, T., NISHINO, K., AND NAYAR, S. 2005. The appearance of human skin. Tech. Rep. CUCS-024-05, Department of Computer Science, Columbia University, June.
- LARBOULETTE, C., AND CANI, M.-P. 2004. Real-time dynamic wrinkles. In *Computer Graphics International*, IEEE Computer Society Press, Greece.
- LEE, A., MORETON, H., AND HOPPE, H. 2000. Displaced subdivision surfaces. In *Proceedings of ACM SIGGRAPH 2000*, ACM Press/Addison-Wesley Publishing Co., New York, NY, USA, 85–94.
- LIU, C., SHUM, H.-Y., AND ZHANG, C. 2001. A two-step approach to hallucinating faces: Global parametric model and local nonparametric model. In *CVPR (1)*, 192–198.
- LOOP, C. 1987. *Smooth Subdivision Surfaces Based on Triangles*. Master's thesis, University of Utah.
- MATUSIK, W., ZWICKER, M., AND DURAND, F. 2005. Texture design using a simplicial complex of morphable textures. *ACM Trans. Graph.* 24, 3, 787–794.
- NEHAB, D., RUSINKIEWICZ, S., DAVIS, J., AND RAMAMOORTHY, R. 2005. Efficiently combining positions and normals for precise 3d geometry. *ACM Transactions on Graphics* 24, 3, 536–543.
- PERLIN, K. 1985. An image synthesizer. In *Computer Graphics (Proceedings of ACM SIGGRAPH 85)*, ACM Press, New York, NY, USA, 287–296.
- PIGHIN, F., AND LEWIS, J. 2005. Digital face cloning. In *SIGGRAPH 2005 Course Notes*.
- PORTILLA, J., AND SIMONCELLI, E. P. 2000. A parametric texture model based on joint statistics of complex wavelet coefficients. *International Journal of Computer Vision* 40, 1, 49–70.
- RUSINKIEWICZ, S., BURNS, M., AND DECARLO, D. 2006. Exaggerated shading for depicting shape and detail. *ACM Trans. Graphics (Proc. SIGGRAPH)* 25, 3.
- SIMONCELLI, E. P., AND FREEMAN, W. T. 1995. The steerable pyramid: a flexible architecture for multi-scale derivative computation. In *ICIP '95: Proceedings of the 1995 International Conference on Image Processing (Vol. 3)-Volume 3*, IEEE Computer Society, Washington, DC, USA, 3444.
- VLASIC, D., BRAND, M., PFISTER, H., AND POPOVIĆ, J. 2005. Face transfer with multilinear models. *ACM Trans. Graph.* 24, 3, 426–433.
- WEYRICH, T., MATUSIK, W., PFISTER, H., NGAN, A., AND GROSS, M. 2005. Measuring skin reflectance and subsurface scattering. Tech. Rep. TR2005-046, Mitsubishi Electric Research Laboratories (MERL).
- WU, Y., THALMANN, N. M., AND THALMANN, D. 1995. A dynamic wrinkle model in facial animation and skin ageing. *The Journal of Visualization and Computer Animation* 6, 4, 195–206.
- WU, Y., KALRA, P., AND THALMANN, N. M. 1997. Physically-based wrinkle simulation & skin rendering. In *Computer Animation and Simulation '97*, Eurographics, 69–79.