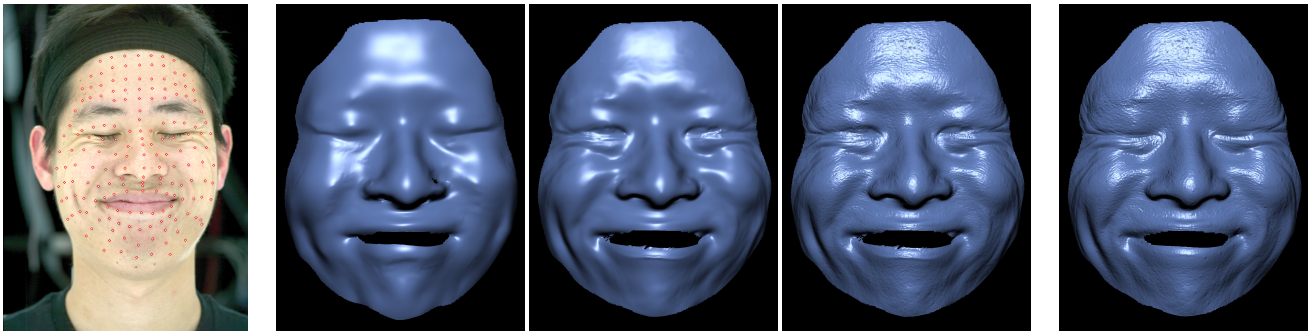
Facial Performance Synthesis using Deformation-Driven Polynomial Displacement Maps

Wan-Chun Ma[⋆†] Andrew Jones[⋆] Jen-Yuan Chiang[⋆] Tim Hawkins[⋆] Sune Frederiksen[⋆] Pieter Peers[⋆]
Marko Vukovic[‡] Ming Ouhyoung[†] Paul Debevec[⋆]

University of Southern California[⋆]
Institute for Creative Technologies

National Taiwan University[†]

Sony Pictures Imageworks[‡]



Motion capture markers | Deformed neutral mesh | Adding medium frequency displacements | Adding high frequency displacements | Ground truth geometry

Figure 1: *We synthesize new high-resolution geometry and surface detail from sparse motion capture markers using deformation-driven polynomial displacement maps; our results agree well with high-resolution ground truth geometry of dynamic facial performances.*

Abstract

We present a novel method for acquisition, modeling, compression, and synthesis of realistic facial deformations using polynomial displacement maps. Our method consists of an analysis phase where the relationship between motion capture markers and detailed facial geometry is inferred, and a synthesis phase where novel detailed animated facial geometry is driven solely by a sparse set of motion capture markers. For analysis, we record the actor wearing facial markers while performing a set of training expression clips. We capture real-time high-resolution facial deformations, including dynamic wrinkle and pore detail, using interleaved structured light 3D scanning and photometric stereo. Next, we compute displacements between a neutral mesh driven by the motion capture markers and the high-resolution captured expressions. These geometric displacements are stored in a *polynomial displacement map* which is parameterized according to the local deformations of the motion capture dots. For synthesis, we drive the polynomial displacement map with new motion capture data. This allows the recreation of large-scale muscle deformation, medium and fine wrinkles, and dynamic skin pore detail. Applications include the compression of existing performance data and the synthesis of new performances. Our technique is independent of the underlying geometry capture system and can be used to automatically generate high-frequency wrinkle and pore details on top of many existing facial animation systems.

Keywords: Polynomial displacement maps, Facial performance synthesis

1 Introduction

The appearance and expressiveness of facial performances are greatly influenced by complex deformations of the face at several scales. Large-scale deformations are driven by muscles and determine the overall shape of the face. Medium-scale deformations are mainly caused by skin wrinkling, and produce many of the expressive qualities in facial expressions. Finally, at the skin mesostructure there is fine-scale stretching and compression which produces subtle but perceptually significant cues. This complex behavior is challenging to reproduce in virtual characters with any combination of artistry and simulation.

Currently, creating realistic virtual faces often involves capturing textures, geometry, and facial motion of real people. However, it is difficult to capture and represent facial dynamics accurately at all scales. Face scanning systems can acquire high-resolution facial textures and geometry, but typically only for static poses. Motion capture techniques record continuous facial motion, but only at a coarse level of detail. Straightforward techniques of driving high-resolution character models by relatively coarse motion capture data often fails to produce realistic motion at medium and fine scales. This limitation has motivated techniques such as wrinkle maps [Oat 2007], blend shapes [Pighin et al. 1998; Joshi et al. 2003; Hawkins et al. 2004; Zhang et al. 2004], and real-time 3D scanning [Zhang et al. 2004; Zhang and Huang 2006]. However, these methods either fail to reproduce the non-linear nature of skin deformation, are labor-intensive, or do not capture and represent all scales of skin deformation faithfully.

In this work, we introduce a novel automated method for modeling and synthesizing facial performances with realistic dynamic wrinkles and fine scale facial details. Our approach is to leverage a real-time 3D scanning system to record training data of the high-resolution geometry and appearance of an actor performing a small set of predetermined facial expressions. Additionally, a set of motion capture markers is placed on the face to track large scale deformations. Next, we relate these large scale deformations to the deformations at finer scales. We represent this relation compactly in

the form of two deformation-driven *polynomial displacement maps* (PDMs), encoding variations in medium-scale and fine-scale displacements for a face undergoing motion (Figure 1).

Similar to polynomial texture maps (PTMs) [Malzbender et al. 2001], deformation-driven PDMs use biquadratic polynomials stored as textures to model the data. However, our deformation-driven PDMs differ from PTMs in three significant aspects. First, PDMs model geometric deformations instead of changes in scene radiance. Second, PTMs have never been driven by changes in geometry. Finally, unlike PTMs used to date, our driving parameters (not just the coefficients) can vary over the image space to better model complex facial expressions. The PDM representation yields a relatively compact model which allows synthesis of realistic medium-scale and fine-scale facial motion using coarse motion capture data.

The principal contributions of our work are:

1. Deformation-driven polynomial displacement maps, a compact representation for facial deformations.
2. A novel real-time acquisition system for acquiring highly detailed geometry based on structured light and photometric stereo.
3. A novel method that is able to generate highly detailed facial geometry from motion capture marker locations making use of PDMs describing the subject's appearance.

2 Related Work

In this section an overview of related work is given. We limit this discussion to three categories: real-time 3D scanning, skin meso-structure acquisition and modeling, and facial performance synthesis.

Real-time 3D Scanning Several real-time 3D scanning systems exist that are able to capture dynamic facial performances. These methods either rely on structured light [Rusinkiewicz et al. 2002; Zhang et al. 2004; Davis et al. 2005; Zhang and Huang 2006], use photometric stereo [Wenger et al. 2005; Malzbender et al. 2006], or a combination of both [Nehab et al. 2005; Jones et al. 2006]. These methods are not suited for our purpose, either because they do not attain the necessary acquisition rate to capture the temporal deformations faithfully, or they are too data-intensive, or they do not provide sufficient resolution to model facial details. The presented acquisition system is capable of capturing high-resolution geometry of dynamic facial performances at 30 fps.

Acquisition and Modeling of Skin Meso-structure Modeling and capturing fine wrinkle details is a challenging problem for which a number of specialized acquisition and modeling techniques have been developed. Haro et al. [2001] and Golovinskiy et al. [2006] model static pore detail using texture synthesis. While these techniques are suitable to enhance static geometry, they do not model wrinkle or pore deformations over time. Recently, Oat [2007] demonstrated how linear interpolation of artist-modeled wrinkle maps can be used for real-time rendering. These methods model wrinkle and pore detail either statistically or artistically, making the creation of an exact replica of a subject's skin detail difficult.

A different approach is to model skin detail by measuring it from live subjects. Sánchez [2006] relies on normal maps to model skin meso-structure, captured using photometric stereo from few static expressions. Dynamic normal variation in skin meso-structure for intermediate facial poses is obtained using trilinear interpolation. Bickel et al. [2007] record dynamic facial wrinkle behavior from motion capture and video of an actor. A pattern of colored makeup is employed to improve shape-from-shading to detect wrinkle indentations in these regions. A non-linear thin shell model is used to recreate the buckling of skin surrounding each wrinkle. Recently, Bickel et al. [2008] extended [Bickel et al. 2007] by using radial basis functions to interpolate medium-scale wrinkles and generate new facial performances. While these systems [Sánchez 2006; Bickel et al. 2007; Bickel et al. 2008] estimate realistic facial geometry, they are mostly limited to larger scale wrinkles, and rely on (a form of) linear data interpolation to generate intermediate expressions. Our system captures not only wrinkles but also dynamic fine-scale pore detail. We represent training data as a biquadratic polynomial function driven by a sparse set of motion capture markers positions. Our representation is both compact and maintains the non-linear dynamics of the human face.

Facial Performance Synthesis Performance capture techniques [Williams 1990] use the recorded motion of an actor to drive a performance of a virtual character, most often from a set of tracked motion capture markers attached to the actor's face. Mapping the set of tracked markers to character animation controls is a complex but well-studied problem. We focus on methods for deriving accurate high-resolution animation from relatively sparse motion capture data.

Parke [1972] introduced linear expression blending models. Blend shapes have become an established method for animating geometric deformation, and can be either defined by an artist or estimated automatically [Joshi et al. 2003]. Several techniques [Pighin et al. 1998; Joshi et al. 2003; Hawkins et al. 2004; Zhang et al. 2004] have used blend shapes to simulate detailed facial performances by linearly interpolating between a set of images or geometric exemplars with different facial expressions. A drawback of this approach is that it can be difficult to use linear blend shapes to reproduce the highly non-linear nature of skin deformation. Skin tends to stretch smoothly up to a point and then buckle nonlinearly into wrinkles [Cerda and Mahadevan 2003]. Furthermore, relating blend shapes to motion capture data is a non-trivial task.

Physically based simulation models [Terzopoulos and Waters 1990; Lee et al. 1995] use underlying bio-mechanical behavior of the human face to create realistic facial animations. Sifakis et al. [2005] determined individual muscle activations from sparse motion capture data using an anatomical model of the actor. Synthesizing detailed animations from such performance capture data would require very detailed models of facial structure and musculature, which are difficult to accurately reconstruct for a specific performer.

Our method is more closely related to blend shapes than to physical based simulations. As in blend shapes, we capture set of training expressions. However, unlike blend shapes, we do not directly use the captured expressions during synthesis, but use a compact representation that encodes the non-linear behavior of the deformations as a function of motion capture marker positions.

3 Training data acquisition

Before introducing deformation-driven polynomial displacement maps in Section 4, we overview our real-time geometry acquisition system. This system is used during the training phase of our technique to record highly detailed facial geometry for a set of short expression sequences.

Setup Our real-time 3D capture system uses a combination of structured light and photometric stereo to obtain high-resolution face scans, and consists of a stereo pair of high-resolution high-speed cameras synchronized to a high-speed DLP video projector and a spherical gradient illumination device similar to that of Ma et al. [2007]. Six grayscale sinusoidal structured light patterns at varying scales and a full-on pattern are output by the high-speed video projector running at 360 frames per second. From the stereo camera pair and the structured illumination, a base geometry is triangulated. After each structured light sequence we generate four gradient illumination patterns and an additional diffuse tracking pattern with the spherical lighting apparatus for computing photometric normals. Additionally, we place 178 tracking dots on the actor's face so that each frame of motion can be registered in a common texture space; the marker motion also serves as the basis for the parameter space for facial detail synthesis. Two lower-resolution cameras are placed to the sides to assist with motion capture marker tracking.

Geometry Reconstruction We triangulate geometry based on camera-to-camera correspondences computed from the ratios of the sinusoidal structured light patterns to the full-on pattern. Photometric surface normals are computed from the spherical gradient patterns as in [Ma et al. 2007]. We then use the photometric normals to add fine-scale detail to our base geometry as in [Nehab et al. 2005]. This allows details such as dynamic wrinkles and fine-scale stretching and compression of skin pores to be captured in real-time.

Because our gradient illumination patterns are captured at different points in time, we correct for subject motion as in [Wenger et al. 2005] using the optical flow algorithm of Brox et al. [2004]. We compute this flow between the first gradient pattern and the tracking pattern, and then use this flow to warp the four gradient-lit images to the same point in time. This allows for accurate calculation of surface normals using ratios of the gradient-lit images. We do not compensate for motion in the structured light patterns because the optical flow would lose stereo correspondences. However, slight errors due to motion in the structured light geometry are acceptable, since it is refined by the photometric normals afterwards, which corrects for these errors.

We register the complete set of 3D training models to a common texture space determined by the motion capture tracking dots to achieve the initial alignment. We then re-use the optical flow algorithm of Brox et al. [2004] to achieve alignment at the level of fine-scale features. Facial skin is often lacking in high-frequency diffuse texture features needed for accurate traditional optical flow. We instead leverage the fact that skin is rich in high-frequency geometric details such as pores, cracks, and wrinkles to achieve accurate optical flow. To do this, we integrate the computed normal maps to derive fine-scale displacement maps per frame. As can be seen in the two leftmost columns of Figure 6, these maps contain much more texture information than the original diffusely lit images. After this final warp, surface details become well aligned in a consistent texture space.

Training To capture the range of facial deformation, we capture several short sequences as the subject transitions from the neutral expression to various strong expressions such as those seen in the top row of Figure 3. From each transition, we select between 10 and 30 frames to use as input to the PDM fitting process (Section 4), including the neutral start point, the extreme expression end points, as well as intermediate deformations. This allows for the non-linear character of wrinkle formation and other fine-scale deformations to be modeled by the PDM.

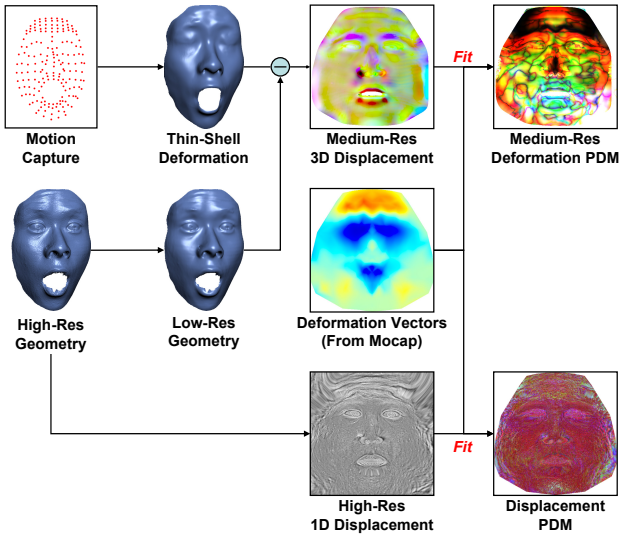


Figure 2: Training pipeline *Our geometric displacements are defined as the difference between high-resolution training data and a neutral mesh deformed to match the motion capture markers. We fit two polynomial displacement maps to the displacement data: a medium-frequency facial deformation PDM and a high-frequency skin detail PDM. Both maps are parameterized based on the motion of the tracking markers.*

4 Deformation-driven Polynomial Displacement Maps

Deformation-driven PDMs are based on the observation that medium-scale and fine-scale changes in surface shape correlate with larger-scale deformations in the corresponding facial region. For example, the formation of horizontal forehead wrinkles correlates with the larger-scale compression of the surface in a direction transverse to the wrinkles. Similarly, skin pores and fine wrinkles can become stretched or flattened according to the local stretching of the skin at coarser scales. In this section, we describe how we construct PDMs to represent these deformations based on the high-resolution training data and tracked motion capture markers.

The mathematical form of deformation-driven PDMs is the same as for traditional PTMs:

$$D_{u,v}(d_1, d_2) = a_0(u,v)d_1^2 + a_1(u,v)d_2^2 + a_2(u,v)d_1d_2 + a_3(u,v)d_1 + a_4(u,v)d_2 + a_5(u,v). \quad (1)$$

Here $D_{u,v}$ is the local displacement at point (u, v), and d_1 and d_2 are measures of low-frequency deformation evaluated at point (u, v). We limit our measurement of large-scale deformation to the two dimensions d_1 and d_2 in order to keep the number of PDM coefficients as small as possible. We describe our method of computing the best 2D parameterization of large-scale deformation in Section 4.1.

Our process for computing PDMs based on the captured training set of motion sequences is illustrated in Figure 2. We express our displacements D relative to the motion of a low-resolution base mesh. We derive this base mesh using a linear thin shell interpolation technique to deform a neutral mesh to the basic shape of the current expression, as in [Bickel et al. 2007]. However, rather than

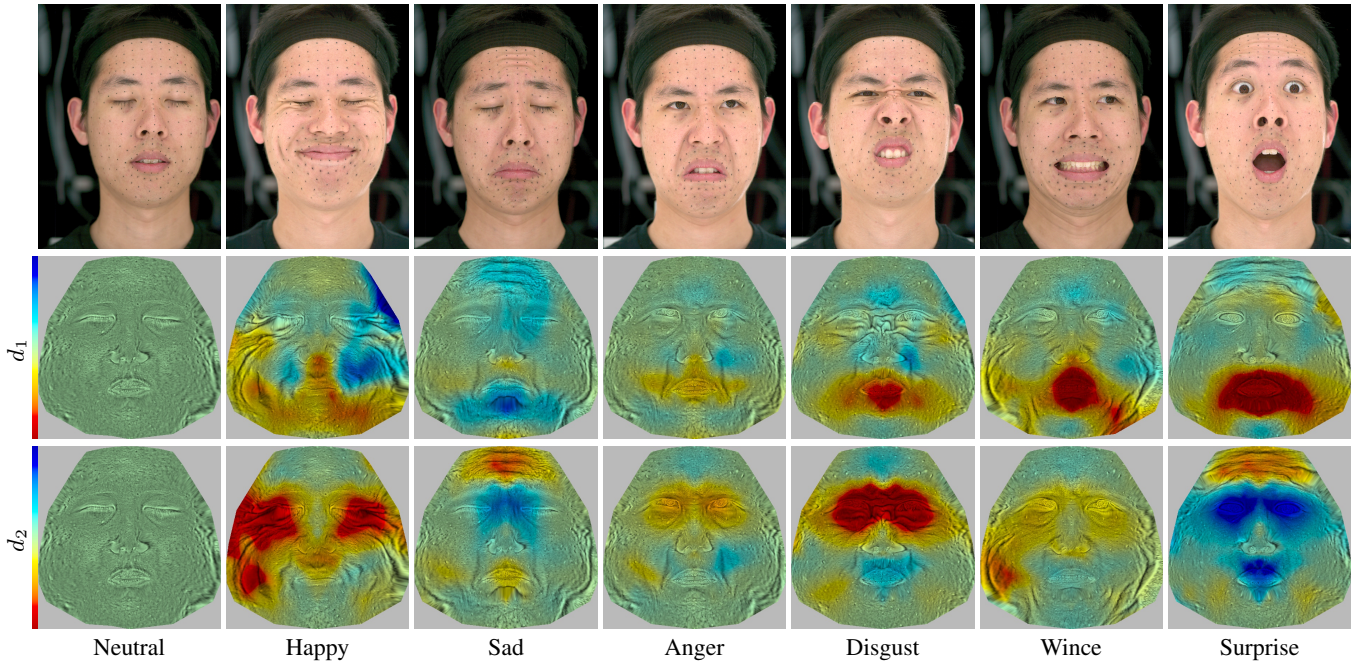


Figure 3: Top row *The maximum deformation seen in each of the training expressions.* **Bottom rows** *For each expression, the largest two PCA deformation parameters as defined in Equation 3. We display the parameters as false-color images where red = -2, blue = 2. The parameters are overlaid on a high-frequency displacement map to show the correlation between high-frequency displacement and low-frequency deformation.*

applying the thin shell deformation to a detailed neutral mesh, we instead deform a smooth neutral mesh. We then encode differences between the deformed neutral mesh and the sequences of high-resolution scans using the deformation-driven PDMs. This significantly reduces the data compared to the original scans as the model becomes a single smooth neutral mesh and a set of deformation-driven PDMs.

We found that trying to fit both the medium-scale and fine-scale facial dynamics to a single deformation-driven PDM tended to attenuate fine-scale detail to better fit the medium-scale displacements. For this reason, we fit a combination of two PDMs, one for medium-scale deformations of several millimeters and a second one for fine-scale deformations on the order of one millimeter. We first fit a medium-scale deformation-driven PDM to the 3D scans, and then fit a fine-scale deformation-driven PDM to a high-pass filtered version of the residual. This process can also reduce storage, since the medium-scale displacements can be computed at a lower resolution than the fine-scale displacements, and since the fine-scale displacements need only be computed in the 1D direction normal to the mesh. We discuss the details of the fitting process in Section 4.2.

Scaled to the same resolution, these separately fit deformation-driven PDMs can be combined back into a single deformation-driven PDM by simply adding their respective coefficients. In practice, however, we apply the medium-scale PDM to the geometric vertices and use the evaluated fine-scale PDM for GPU displacement mapping.

4.1 Parameterizing Low Frequency Deformation

To generate and make use of PDMs, we must create an input parameter space which characterizes local coarse-scale facial motion in a well-conditioned manner. To generate these parameters, we first create a coarse triangle mesh over the set of motion markers. At each vertex V_i of this coarse mesh we define the low-frequency deformation at time t as $\mathbf{S_i(t)}$ by conjoining the 3D positional offset $\mathbf{O_i}(t)$ of the vertex with two additional values $E_i^u(t)$ and $E_i^v(t)$ representing the large-scale in-plane strain of the surface. This forms a 5D deformation space:

$$\mathbf{S_i(t)} = \{\mathbf{O_i}(t), E_i^u(t), E_i^v(t)\}.$$

The vertex position offsets $\mathbf{O_i}(t)$ are computed by first applying a rigid transformation R to the coarse mesh to best match the neutral pose, thereby correcting for overall head pose. $\mathbf{O_i}(t)$ is then simply the difference between the transformed vertex position $R(\mathbf{P_i}(t))$ and the neutral vertex position $\mathbf{P_i^{ref}}$.

The large-scale strains E^u and E^v are estimated from the coarse mesh vertex positions of all vertices $N_2(V_i)$ connected to V_i by a path of two or fewer edges. Similar to the treatment of light source direction in [Malzbender et al. 2001], we project the positions of $N_2(V_i)$ into the local texture coordinate system $(\mathbf{u}, \mathbf{v})$. We approximate 2D strain as the difference between the standard deviation of the projected positions $\mathbf{P_j}$ of $V_j \in N_2(V_i)$ in the current deformation and the standard deviation in the reference neutral expression:

$$\begin{aligned} E_i^u(t) &= \sigma\{\hat{\mathbf{u}} \cdot \mathbf{P_j}(t)\} - \sigma\{\hat{\mathbf{u}} \cdot \mathbf{P_j^{ref}}\}, \\ E_i^v(t) &= \sigma\{\hat{\mathbf{v}} \cdot \mathbf{P_j}(t)\} - \sigma\{\hat{\mathbf{v}} \cdot \mathbf{P_j^{ref}}\}. \end{aligned} \tag{2}$$

To find a suitable 2D parameterization for the PDM domain, we perform principal components analysis (PCA) on the 5D deformation vectors $\mathbf{S_i}$ over all captured deformations. This determines the most important axes of large scale shape variation in the neighborhood of V_i. Prior to PCA, we scale the strain values (E_i^u, E_i^v) by

Eigenvectors	1	2	3	4	5
Energy	76.84%	92.52%	98.46%	99.76%	100%

Table 1: *Energy averaged across the face represented by a subset of eigenvectors of the 5D deformation vectors. Two eigenvectors can model over* 90% *of the training data.*

$\sqrt{|N_2(V_i)|}$ to account for the lower noise of this aggregate measure relative to the noise in the single measurement $\mathbf{O_i}$. We select the first two principal components $\hat{\mathbf{Q}}_\mathbf{i}$ and $\hat{\mathbf{R}}_\mathbf{i}$. We found that the eigenvalues decreased very quickly after the first two, indicating most of the variation in $\mathbf{S}$ could be well captured by our choice to use only a two-dimensional PDM parameterization. Examples of eigenvalues, averaged across the face are shown in table 1. This analysis shows that most of the eigenvalues at each motion capture marker decay quickly. By choosing the best two dimensions for each motion capture marker, we can model over 90% of the training data.

Finally, we derive the final PDM domain axes over the coarse mesh by a smoothing process which assures the deformation bases of adjacent vertices do not differ excessively. We accomplish this by comparing each basis vector to the average of the corresponding basis vectors at adjacent vertices. We successively replace the worst case outlier vector over the entire mesh with the average of the adjacent basis vectors. We reorthogonalize and renormalize these vectors at each step. This process is repeated until the worst case outlier lies within a threshold angle of the neighborhood-averaged vector. We denote the result of smoothing $\hat{\mathbf{Q}}_\mathbf{i}$ and $\hat{\mathbf{R}}_\mathbf{i}$ by $\hat{\mathbf{q}}_\mathbf{i}$ and $\hat{\mathbf{r}}_\mathbf{i}$. The input parameters to the PDM at the coarse mesh vertex V_i at time t are then simply:

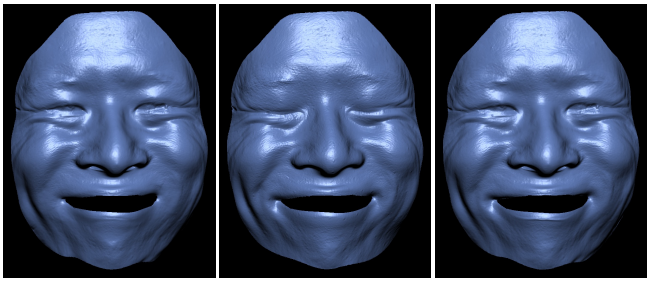
$$d_1(P_i,t) = \mathbf{E_i(t)} \cdot \hat{\mathbf{q}}_\mathbf{i}, \qquad (3)$$
$$d_2(P_i,t) = \mathbf{E_i(t)} \cdot \hat{\mathbf{r}}_\mathbf{i}.$$

To extend these deformation values over the entire mesh, we interpolate $\mathbf{E}$, $\hat{\mathbf{q}}$, and $\hat{\mathbf{r}}$ from their values at the vertices V using barycentric interpolation over the coarse triangle mesh.

Our choice to include the vertex offsets $\mathbf{O_i}$ themselves in the deformation vector is perhaps counterintuitive, as mechanical properties are typically invariant with respect to simple translation. However, we found that local shape deformation correlates significantly with these vertex offsets. We believe this is due to the strong influence of the underlying bone structure on the skin deformation. For example, we expect a skin patch under a fixed strain will nonetheless change shape as it slides over different bony facial features. The thin shell model does not account for such effects, and they therefore must be accounted for by the PDM.

Figure 4 shows a visual comparison between a mesh synthesized using the results of a 5D PCA and a mesh using just the two dimensions of principle strain. While the differences can be subtle, they are perceptually important and most notable in dynamic sequences. The inclusion of vertex offsets generates particular improvement in the shape of the mouth and lower jaw. Compared to only using 2D strain (without the positional offsets $\mathbf{O_i}$), the 5D PCA does not require any extra storage, adds only a limited amount of preprocessing, and results in lower errors.

We note that we use an absolute strain formulation for E rather than the more traditional relative strain because the units of absolute strain are distance, which facilitates common analysis with the positional offsets $\mathbf{O_i}$. In addition, we have chosen to neglect bending and shear strains in our current implementation. We hypothesize that over the restricted domain of facial motion, the five dimensions we do analyze function as effective proxies for the omitted dimensions. Examples of d_1 and d_2 evaluated over the face for different facial expressions can be seen in Figure 3.



2D strain only | Ground truth | Incl. positional offset

Figure 4: *Comparison of PDM results with and without positional offsets included in the deformation metric.* **Left:** *PDM results using 2D strain only.* **Center:** *ground truth.* **Right:** *PDM results including offsets in the initial deformation vector. Including offsets produces a better match to ground truth, as can be seen in the shapes of the corners of the lips, the dimple on the subject's right cheek, and on the chin.*

In concurrent work, Bickel et al. [2008] propose an alternative strain based on the relative length of each edge that connects motion capture markers. By interpolating edge strains, this approximation could be adapted to work with our technique. However, we opted for the per-vertex variance-based strain configuration because of its simplicity, effectiveness and compactness.

4.2 Optimal Fitting of Deformation-driven PDMs

We compute optimal polynomial coefficients for Equation 1 at each texture point using the measured displacement values and the derived deformation input parameters. Given the sequence of measured displacement coordinate values f_t at a point, we compute the PDM coefficients as the least-squares solution to the equations:

$$\begin{bmatrix} d_{11}^2 & d_{21}^2 & d_{11}d_{21} & d_{11} & d_{21} & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ d_{1t}^2 & d_{2t}^2 & d_{1t}d_{2t} & d_{1t} & d_{2t} & 1 \\ \frac{\gamma}{\sigma_{\hat{\mathbf{q}}}^2} & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{\gamma}{\sigma_{\hat{\mathbf{r}}}^2} & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{\gamma}{\sigma_{\hat{\mathbf{q}}}\sigma_{\hat{\mathbf{r}}}} & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{\gamma}{\sigma_{\hat{\mathbf{q}}}} & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{\gamma}{\sigma_{\hat{\mathbf{r}}}} & 0 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \\ a_4 \\ a_5 \end{bmatrix} = \begin{bmatrix} f_1 \\ \vdots \\ f_t \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

We include regularization terms to account for the possibility that one or both of the input parameters may not have exhibited sufficient variation in the training set, which could make recovery of the non-constant coefficients of the PDM unstable. We found that the regularization was effective for low values of the regularization constant γ, such that no degradation in the fidelity of the fitting was noticeable.

We recover two deformation-driven PDMs for each subject: one for medium-scale 3D displacement at 512×512 pixel resolution and one for fine-scale 1D displacement normal to the mesh at 1024×1024 resolution. For the medium-scale displacement, we fit each coordinate of displacement independently, yielding 18 total PDM coefficients. We also fit a deformation-driven PTM to the time-varying diffuse albedo measurements, yielding an additional 18 PTM coefficients.

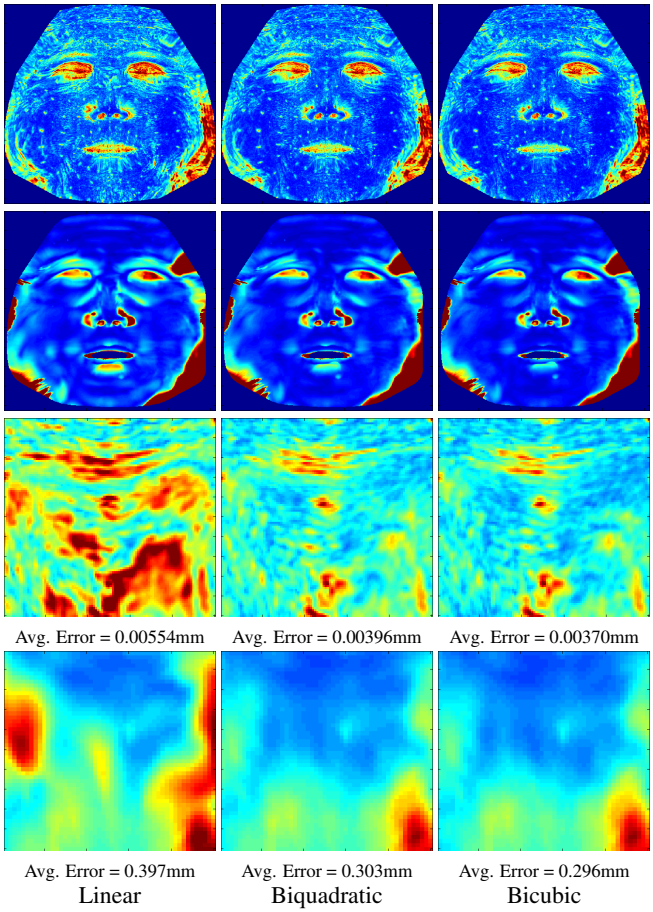


Figure 5: *Fitting errors for displacement maps for linear, biquadratic, and bicubic PDM modeling. (Row 1) Errors for high-frequency displacement map. (Row 2) Errors for medium-frequency displacement map. The large errors near the nostrils and on the edges of the face are due to errors in geometry acquisition near glancing angles. The large errors on the eyes are due to the lack of geometric registration. (Row 3) Detail area (bridge of nose) showing error for high-frequency displacement map. (Note: false-color mapping has been rescaled to the range of error in the detail area.) (Row 4) Same detail area showing error for medium-frequency displacement map. Significant improvements can be seen between the linear and biquadratic models, while minimal improvement is seen between biquadratic and bicubic.*

4.3 Effect of PDM order on accuracy

We compared our results using biquadratic PDMs with those obtainable from linear PDMs and bicubic PDMs. Error comparisons for these different cases are shown in Figure 5. Although the linear PDMs perform well for much of the face, there are a number of areas where the benefit of biquadratic PDMs can be easily seen. While biquadratic polynomials cannot fit the training data precisely (it represents a vast reduction in data), most of the perceptually important aspects of skin deformation are modeled effectively. Bicubic PDMs and higher-degree polynomials do not capture much more information while requiring substantially more storage (10 coefficients per data channel rather than 6). Furthermore, higher order approximations also carry the danger of over-fitting the data.

5 Motion Synthesis

Once training data has been captured and deformation-driven PDMs have been derived, we can synthesize highly detailed models according to a performance recorded with standard facial motion capture. In this work, we use the same markers used in the training sequences to record novel facial performances not in the training set. This makes synthesis and rendering of detailed facial geometry for each frame a straightforward process:

1. Deform the low-resolution neutral mesh to the motion capture points using linear thin shell interpolation.
2. Evaluate the deformation vector $\mathbf{S}$ and the deformation axes $\hat{\mathbf{d}}_\mathbf{1}$ and $\hat{\mathbf{d}}_\mathbf{2}$ over the motion capture points.
3. Interpolate the deformation axes and deformation vectors over the mesh texture space, forming the dot products d_1 and d_2 at each surface point.
4. Evaluate the medium-scale deformation-driven PDM and deform the mesh vertices according to the computed 3D offsets.
5. Evaluate the fine-scale deformation-driven PDM to form a 1D displacement map.
6. Render the deformed geometry and displacement map on the GPU.

We have implemented off-line and real-time on-line rendering systems. For off-line rendering steps 1-5 are performed on the CPU. For real-time rendering we generate the linear thin shell mesh on the CPU with a reduced vertex count of 10k (versus 200k for the off-line rendering system) to maintain a frame rate of 20fps on an Intel Pentium 4 Xeon and an nVidia 8800GTS. Next, we evaluate the two PDMs using the GPU and add the displacement to the thin shell mesh. Examples of synthesized displacement maps, as well as a synthesized diffuse albedo map, can be seen in Figure 6. Rendered results, discussed in Section 6, can be seen in Figure 8.

Novel Marker Placement In practice, the performance data would likely be captured at a different time than the training data and thus would involve a new application of the motion capture markers. One industry-standard technique for obtaining maximally consistent marker placements uses a thin plastic mold of the actor's face with small holes drilled into it to guide the marker placement, usually to within a millimeter of repeatability. Accommodating marker placements with greater deviation would require a remapping step to evaluate the PDM as follows: First, the new motion capture markers (observed in a neutral position) must be mapped onto the neutral mesh acquired during training. Second, if the density of motion capture markers is different, it will be important to scale the values $E_i^u(t)$ and $E_i^v(t)$ accordingly. Because such mappings and corrections involve some error, we expect the best results will result from using approximately the same marker locations during performance capture and training. However, we believe this is not excessively restrictive since markers are often placed relatively repeatably in practice.

6 Results

To demonstrate our technique, we captured facial performances of two subjects. For each subject, we captured the six training expressions shown in Figure 3. In choosing our captured expressions, we focused on the inner and outer motion of the brow, motion of the mouth corners, nose wrinkling, cheek dimpling, brow furrowing, and basic jaw movement.

We then captured several facial performances for each subject. Although the idea is that these performances need only consist of motion capture marker motion, we continued to acquire real-time

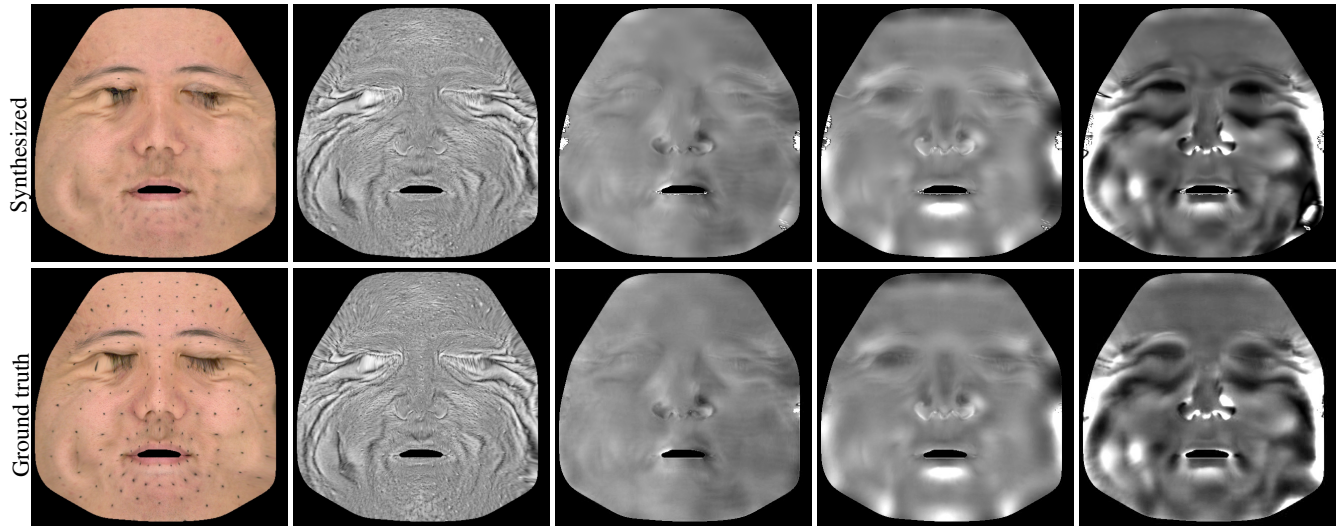


Figure 6: *Comparison of synthesized (top row) and ground truth (bottom row) albedo map and medium-frequency and high-frequency displacement map components for the "happy" expression.*

high resolution face scans to serve as "ground truth" validation data for the synthesized sequences. We used the derived PDMs to reconstruct sequences that were part of the training set (Figure and Figure 8, rows 1 and 3) as well as novel performances not of the training set (Figure 8, rows 2 and 4). We found that in both cases our synthesis algorithm produces results which are largely indistinguishable from the ground truth geometry sequences. Several of the performances contain significant global head motion which did not appear to pose problems for model fitting.

The top two rows of Figure 8 show that while the thin-shell deformation model provides general face shape of the subject, it fails to reproduce medium-scale details such as brow ridges and facial musculature. The medium-scale PDMs add large wrinkles and definition to the brow. The fine-scale 1D PDM adds the remaining fine wrinkles and pore detail, making the synthesized model a close approximation of the ground truth.

Figure 7 shows the effect of skin mesostructure deformation near the cheek for a "happy" expression. This deformation, seen in the ground truth geometry (center) and reproduced by the PDM in (right), can be seen as pore stretching across the cheek and the formation of a few fine-scale wrinkles below the eye. In contrast, mapping a static displacement map from the neutral expression to the "happy" geometry (left) does not reproduce these effects. We believe these nonlinear changes in mesostructure may be important to synthesizing realistic expressions since they affect aspects of skin appearance.

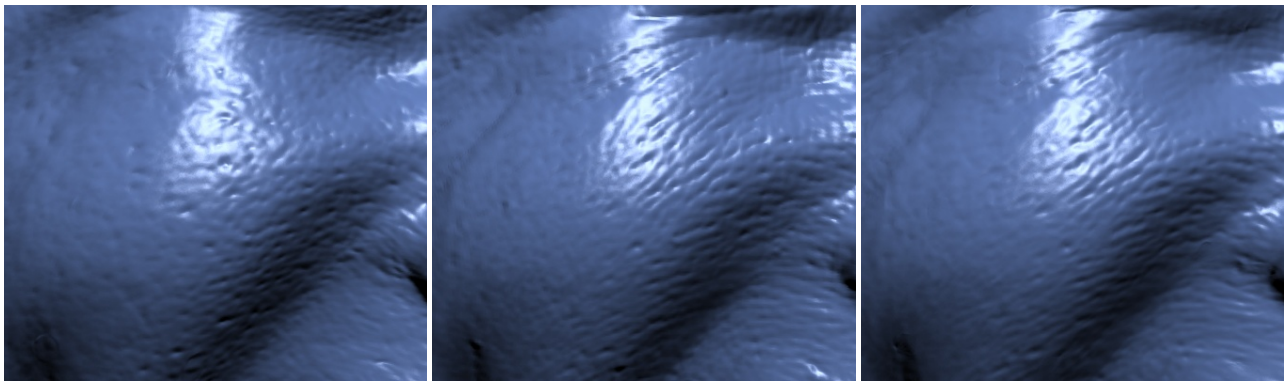
In designing our expression set, we deliberately did not break down expressions into individual facial action units. Our fitting process inherently segments the captured data into usable subexpressions by choosing different PCA parameters for different facial regions (Figure 3). To test this part of the algorithm, we captured a new motion capture performance where the actor produced an asymmetric smile and raised a single eyebrow (Figure 8, row 2). The synthesized geometry effectively combines elements from multiple training expressions to closely approximate the ground truth.

The low-frequency deformation parameters can also be used to model and synthesize other attributes such as facial reflectance. In addition to displacement, we fit a three-channel PTM to the dynamic surface reflectance recorded by the video cameras parameterized by the same facial deformation space. We performed semi-automatic dot removal as in [Guenter et al. 1998] to create a clean texture, though some black smudged remained in the images. The bottom two rows of Figure 8 and the final examples in the video show results texture-mapped with synthesized PTMs. In these renderings, the PTMs successfully encode changes in surface shading due to self-occlusion. However, the eyes are not realistic. This is because during training, we did not require that the actor's eyes remain open or closed and as a result, the eyes can be assigned distorted texture containing both eyelid and eye color. In general, while this technique is successful at generating the shape and appearance for most of the face, we believe that realistically modeling the eyes and the skin immediately around them (as well as the inner mouth and lips) will require different and more sophisticated techniques. Additional polynomial texture maps could be used to model other skin properties such as changes in specularity and subsurface scattering caused by facial deformation.

In the accompanying video we show a synthesized performance containing a short speech sample. Even though the training dataset contained no explicit phonemes, the synthesized performance exhibits realistic lower-face wrinkles. A failure case can be seen in the subtle lip pout in Figure 8 (g,h) which produces an inaccurate crease below the lower lip. We believe this problem is caused by the presence of asymmetric lip deformation which was not represented in the training set. This problem could be eliminated by capturing a larger set of training expressions that more completely spans the space of desired performance motions.

7 Conclusions and Future Work

We have presented deformation-driven polynomial displacement maps and demonstrated their application in modeling and synthesizing dynamic high-resolution geometry for facial animation. A high-resolution real-time 3D geometry acquisition system was built that is capable of capturing facial performances at the level of wrinkle and pore details. Furthermore, performance-driven polynomial displacement maps, a novel compact representation for facial deformation, was presented. We demonstrate that this compact representation provides a high level of visual fidelity, comparable to that achievable with hardware-intensive real-time scanning techniques.



Static displacement map | Ground truth | Polynomial displacement map

Figure 7: *Skin pores change shape as the face deforms. This non-linear stretch can not be represented as a single displacement map.* **Left:** *facial detail obtained by applying a static neutral expression displacement map to a smiling face.* **Center:** *ground truth using real-time acquired displacement map.* **Right:** *high frequency displacement map generated using fitted polynomial displacement map, exhibiting accurate pore stretching and shearing effects comparable with ground truth.*

Finally, we showed that the performance-driven PDMs are suited to synthesize new expressions that are not part of the original training dataset using only motion capture marker positions of the new facial expression.

Our technique yields accurate reconstructions of medium-scale and fine-scale geometry over most of the face, however, we have not yet optimized the technique to reproduce the perceptually important detailed motions near the boundaries of the lips and eyelids. For this, it may be necessary to provide better tracking of the contours of the mouth and eyes in order to register and interpolate facial detail up to these boundaries. We believe the performance of the technique for dialog performances could be improved by adding a short speech exemplar to the training data.

We have so far only explored the use of motion capture to drive our recovered PDMs. While motion capture with markers is an industry-standard technique, alternative markerless methods have been developed [Zhang et al. 2004]. Such systems use template geometry or global optimization to smooth the tracked data and lose some high-frequency information. Our PDM technique could be driven by marker-less data as long as the quality of the motion data is reasonable.

It would be interesting to evaluate the performance of the detail synthesis using other facial animation techniques such as key-framing or blend shapes to provide the smooth base motion geometry. Likewise, our detail synthesis could be used to add fine details to performances captured with systems such as that of Zhang et al. [2004].

At this moment, the current method only allows geometry/texture and motion capture data of the same subject. One significant area of future work is motion capture retargeting in order to drive PDMs from arbitrary facial motion capture data. Different subjects have significantly different facial detail based on age, gender, ethnicity, and bone structure. While some existing work has mapped expressions between different subjects [Vlasic et al. 2005; Bickel et al. 2008], facial performance retargeting between actors may require reinterpreting the many different stylistic motions that are unique to a specific actor.

Acknowledgements

The authors wish to thank Michael Kennedy, Jay Busch, Abhijeet Ghosh, Ian McDowall, Tom Pereira, Monica Nichelson, Jeff Fisher, Bill Swartout, Randy Hill, and Randolph Hall for their support and assistance with this work. This work was sponsored by the U.S. Army Research, Development, and Engineering Command (RDECOM) and the University of Southern California Office of the Provost. The high-speed projector was originally developed by a grant from the Office of Naval Research under the guidance of Ralph Wachter and Larry Rosenblum. The content of the information does not necessarily reflect the position or the policy of the US Government, and no official endorsement should be inferred.

References

BICKEL, B., BOTSCH, M., ANGST, R., MATUSIK, W., OTADUY, M., PFISTER, H., AND GROSS, M. 2007. Multi-scale capture of facial geometry and motion. *ACM Transactions on Graphics 26*, 3 (July), 33:1–33:10.

BICKEL, B., LANG, M., BOTSCH, M., OTADUY, M., AND GROSS, M. 2008. Pose-space animation and transfer of facial details. In *2008 ACM SIGGRAPH / Eurographics Symposium on Computer Animation*, ACM, New York, NY, USA, 57–66.

BROX, T., BRUHN, A., PAPENBERG, N., AND WEICKERT, J. 2004. High accuracy optical flow estimation based on a theory for warping. In *Proc. of the 8th European Conference on Computer Vision*, Springer-Verlag, Prague, Czech Republic.

CERDA, E., AND MAHADEVAN, L. 2003. Geometry and physics of wrinkling. *Physical Review Letters 90*, 7 (February), 074302+.

DAVIS, J., NEHAB, D., RAMAMOORTHI, R., AND RUSINKIEWICZ, S. 2005. Spacetime stereo: A unifying framework for depth from triangulation. *IEEE Transactions on Pattern Analysis and Machine Intelligence 27*, 2, 296–302.

GOLOVINSKIY, A., MATUSIK, W., PFISTER, H., RUSINKIEWICZ, S., AND FUNKHOUSER, T. 2006. A statistical model for synthesis of detailed facial geometry. *ACM Transactions on Graphics 25*, 3 (July), 1025–1034.

GUENTER, B., GRIMM, C., WOOD, D., MALVAR, H., AND PIGHIN, F. 1998. Making faces. In *Proceedings of SIGGRAPH 98*, Computer Graphics Proceedings, Annual Conference Series, 55–66.

HARO, A., GUENTER, B., AND ESSA, I. 2001. Real-time, photorealistic, physically based rendering of fine scale human skin structure. In *Rendering Techniques 2001: 12th Eurographics Workshop on Rendering*, 53–62.

HAWKINS, T., WENGER, A., TCHOU, C., GARDNER, A., GÖRANSSON, F., AND DEBEVEC, P. 2004. Animatable facial reflectance fields. In *Rendering Techniques 2004: 15th Eurographics Workshop on Rendering*, 309–320.

JONES, A., GARDNER, A., BOLAS, M., MCDOWALL, I., AND DEBEVEC, P. 2006. Performance geometry capture for spatially varying relighting. In *CVMP 2006*.

JOSHI, P., TIEN, W. C., DESBRUN, M., AND PIGHIN, F. 2003. Learning controls for blend shape based realistic facial animation. In *2003 ACM SIGGRAPH / Eurographics Symposium on Computer Animation*, 187–192.

LEE, Y., TERZOPOULOS, D., AND WATERS, K. 1995. Realistic modeling for facial animation. *Proceedings of SIGGRAPH 95*, 55–62.

MA, W.-C., HAWKINS, T., PEERS, P., CHABERT, C.-F., WEISS, M., AND DEBEVEC, P. 2007. Rapid acquisition of specular and diffuse normal maps from polarized spherical gradient illumination. In *Rendering Techniques*, 183–194.

MALZBENDER, T., GELB, D., AND WOLTERS, H. 2001. Polynomial texture maps. In *Proceedings of ACM SIGGRAPH 2001*, Computer Graphics Proceedings, Annual Conference Series, 519–528.

MALZBENDER, T., WILBURN, B., GELB, D., AND AMBRISCO, B. 2006. Surface enhancement using real-time photometric stereo and reflectance transformation. In *Rendering Techniques 2006: 17th Eurographics Workshop on Rendering*, 245–250.

NEHAB, D., RUSINKIEWICZ, S., DAVIS, J., AND RAMAMOORTHI, R. 2005. Efficiently combining positions and normals for precise 3d geometry. *ACM Transactions on Graphics 24*, 3 (Aug.), 536–543.

OAT, C. 2007. Animated wrinkle maps. In *SIGGRAPH 2007: ACM SIGGRAPH 2007 courses*, ACM, New York, NY, USA, 33–37.

PARKE, F. I. 1972. Computer generated animation of faces. In *ACM'72: Proceedings of the ACM annual conference*, ACM, New York, NY, USA, 451–457.

PIGHIN, F., HECKER, J., LISCHINSKI, D., SZELISKI, R., AND SALESIN, D. H. 1998. Synthesizing realistic facial expressions from photographs. In *Proceedings of SIGGRAPH 98*, Computer Graphics Proceedings, Annual Conference Series, 75–84.

RUSINKIEWICZ, S., HALL-HOLT, O., AND LEVOY, M. 2002. Real-time 3d model acquisition. *ACM Transactions on Graphics 21*, 3 (July), 438–446.

SÁNCHEZ, M. A. 2006. *Techniques for performance-based, real-time facial animation*. PhD thesis, University of Sheffield.

SIFAKIS, E., NEVEROV, I., AND FEDKIW, R. 2005. Automatic determination of facial muscle activations from sparse motion capture marker data. *ACM Transactions on Graphics 24*, 3 (Aug.), 417–425.

TERZOPOULOS, D., AND WATERS, K. 1990. Physically-based facial modelling, analysis, and animation. *Journal of Visualization and Computer Animation 1*, 2, 73–80.

VLASIC, D., BRAND, M., PFISTER, H., AND POPOVIĆ, J. 2005. Face transfer with multilinear models. *ACM Transactions on Graphics 24*, 3 (Aug.), 426–433.

WENGER, A., GARDNER, A., TCHOU, C., UNGER, J., HAWKINS, T., AND DEBEVEC, P. 2005. Performance relighting and reflectance transformation with time-multiplexed illumination. *ACM Transactions on Graphics 24*, 3 (Aug.), 756–764.

WILLIAMS, L. 1990. Performance-driven facial animation. In *Computer Graphics (Proceedings of SIGGRAPH 90)*, 235–242.

ZHANG, S., AND HUANG, P. 2006. High-resolution, real-time three-dimensional shape measurement. *Optical Engineering 45*, 12.

ZHANG, L., SNAVELY, N., CURLESS, B., AND SEITZ, S. M. 2004. Spacetime faces: high resolution capture for modeling and animation. *ACM Transactions on Graphics 23*, 3 (Aug.), 548–558.

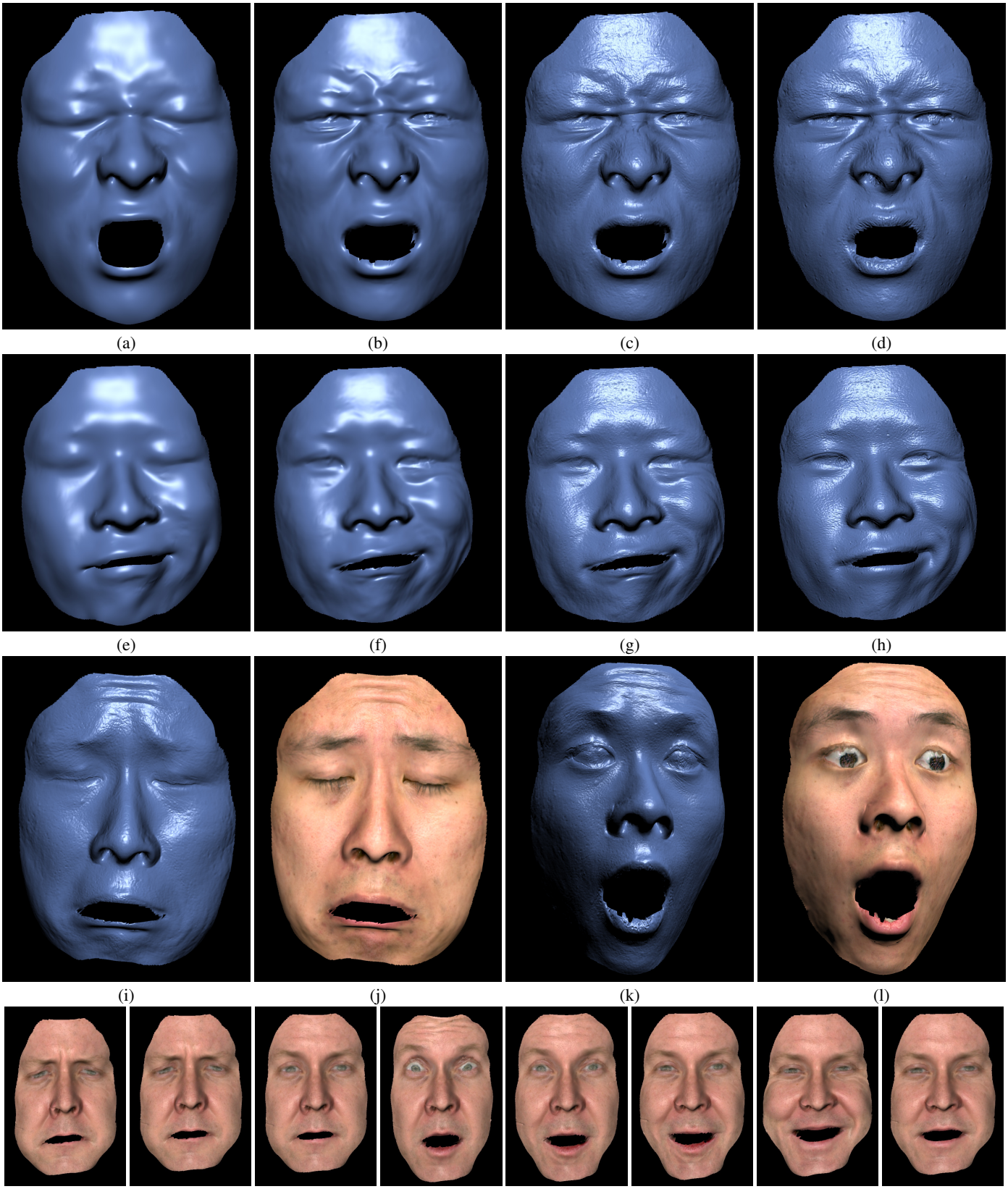


Figure 8: Results generated using deformation-driven PDMs and sparse motion capture points. *(a-h) show the the three layers of our model: (a,e) a deformed neutral mesh, (b,f) synthesized medium-frequency displacement, and (c,g) synthesized medium and high-frequency displacement. (d,h) show ground truth geometry for comparison. (a-d) shows a re-synthesized expression from the training set while (e-h) shows a synthesized performance frame with asymmetric deformation not in the training set. (i,k) show synthesized high resolution geometry of two extreme expressions, alongside the same expressions mapped with reflectance from a PTM in (j,l). The bottom row shows selected frames from from a longer performance with synthesized geometry and albedo.*