Facial Cartography: Interactive Scan Correspondence (Supplement)

Cyrus A. Wilson[1]   Oleg Alexander[1]   Borom Tunwattanapong[1]   Pieter Peers[1,2]
Abhijeet Ghosh[1]   Jay Busch[1]   Arno Hartholt[1]   Paul Debevec[1]

[1]USC Institute for Creative Technologies   [2]The College of William & Mary

The following sequence of screenshots illustrates a sample user interaction session: applying the Facial Cartography approach to find correspondences between neutral and an extreme expression for a desired animation mesh. The text below details these steps. The accompanying video illustrates another (different) example of user interaction.

1. Selection of data

The entire dataset consists of approximately 30 expression scans, including high-resolution geometry and photometric textures and normal maps. The user has specified the location of this dataset in Figure 1 (top left). One of the expressions in the dataset, (expression 2), is a neutral pose; it was used as the basis of our low-resolution animation mesh (Figure 1 bottom left).

The user has selected non-neutral expression 8 (Figure 1 bottom right). Here the user is inspecting the ground-truth scan data, with the expression 8 detail maps applied to the full-resolution geometry. Note that the expression data is in fact shown here in *wireframe* mode, though it appears solid due to the density of the data. There is a substantial difference in resolution between the original scanned mesh and the animation mesh we will be using. In subsequent steps the user will compute a correspondence solution which gives a desirable result for this low resolution animation mesh, such that:

1. blending between neutral and expression 8 vertex positions and detail maps will give a visually consistent result (for downstream blendshape animation applications);
2. increasing the blend weight for the expression 8 vertex positions and detail maps gives a result which is visually consistent with the ground truth data for the expression; and
3. the solution meets other user-desired constraints such as a regular distribution of mesh deformation in the corners of the eyes and mouth.

2. Initial state of the optimization

The user is inspecting the initial state of the optimization, by way of the proxy called the Active Visage (Figure 2 top right). This specific visualization of the active visage can be manipulated freely by the user: viewpoint, lighting, selected channels of neutral detail maps, wireframe (Figure 2 bottom right), etc. It serves as a means for the user to assess the current correspondence solution, as well as a means for interaction: selection of vertices and application of Directable Forces.

This optimization estimate is also shown in the optimization *parameter* domain—before any iteration has begun—in Figure 2 bottom center. The user is visualizing a highpass-filtered texture map from the expression scan, with a superimposed wireframe showing the optimization estimate. Arrows depict the value of the forces on the active visage, at iteration zero, after remapping into the optimization domain.

The main difference between the neutral and extreme expression is that the scan subject has opened his mouth wide. The correspondence process will involve sliding the active visage along the target expression manifold until corresponding features from the neutral expression occupy the correct location on the target expression manifold. The animation mesh topology includes holes for eyes and mouth; therefore we will be able to open the mouth of the active visage while still applying Internal forces.

3. Bootstrapping the optimization

The opening of the mouth is a significant change: the target expression manifold now includes the geometry of the mouth interior, which does not exist in the neutral expression scan. As a general hint to what the correspondence should do, the user has prepared an initial target for the Shape force. This was quickly created by roughly deforming the animation mesh in 3D in external software (Autodesk Maya) and then loading it here. This hint does not specify explicit correspondence information, nor does it lie along the target expression manifold. However, since the hint mesh vertices have a 1:1 relation to those in the active visage, this Shape force target provides a general cue for where the vertices of the active visage should go in 3D, by sliding along the manifold.

This mesh is shown in yellow (Figure 3 top) and compared with the neutral animation mesh (in purple). The 3D displacement between this hint mesh and the active visage 3D vertex locations is used to compute a 3D Shape force, which is then remapped to the (2D) optimization parameter domain by remapping onto the target expression manifold. This remapping of the shape force is shown on the active
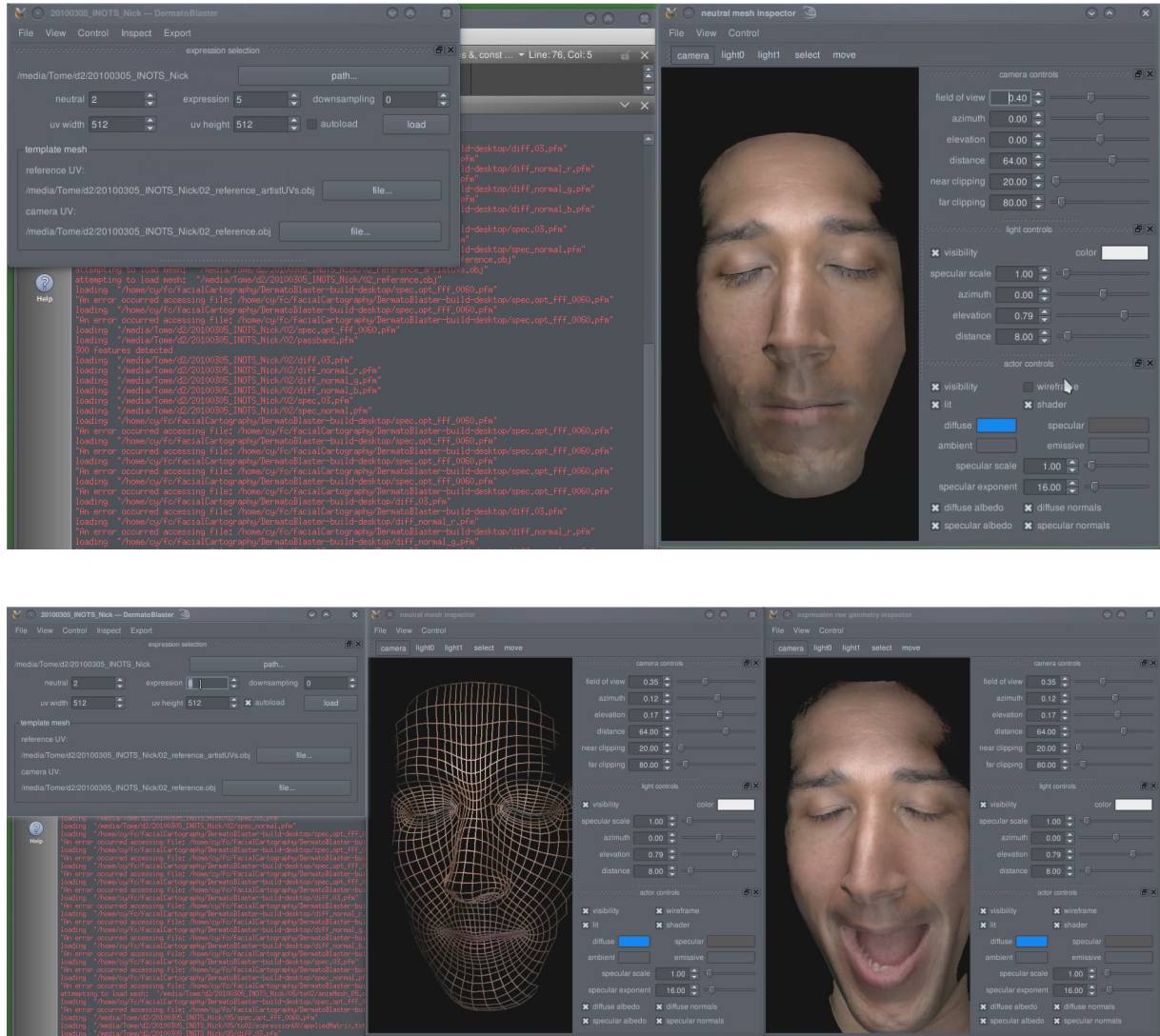
Figure 1: *Selection of dataset for correspondence. Top row: Neutral expression scan. Bottom row: The neutral animation mesh (left) and the high resolution expression scan (right).*

visage (Figure 3 bottom right). The sum of this and the Internal forces is shown in the optimization domain (Figure 3 bottom center).

When bootstrapping the optimization using this hint, we do not wish to rely on the rough deformation of the animation mesh having been done with extreme care (e.g., preventing polygons from collapsing, etc.) Therefore the user has enabled both the shape force (labeled "targetForceActive" in the UI), and internal forces ("internalForcesActive"). The user activates the simulation (via the ">" play button, in the iteration controls at the bottom of the optimization domain window). The active visage then slides along the target expression manifold to more closely resemble the 3D shape of

the hint mesh, while preserving vertex spacing and ordering via the Internal forces.

4. Directable force for recovering from a local optimum

At any point during the process, the estimate may become trapped in a local optimum. This is one type of situation in which the user may want to apply Directable forces. In this case some vertices of the lip have become trapped within the mouth of the target expression surface (Figure 4 top). The user selects these vertices and drags them to exert Directable forces. This interaction can be performed in the active visage visualization domain, in which case the Directable forces are

remapped into the optimization domain, or the manipulation can be performed directly in the optimization domain.

If the simulation is paused, then Directable forces simply displace the vertices according to the distances they have been dragged. This can be useful if other forces are resisting movement out of the local optimum.

However in this example (Figure 4) the user leaves the simulation running. In this mode the Directable forces work in concert with the other forces; to the optimization they appear no different from the other types of forces. Here the Shape and Internal forces remain active: the user gives the vertices a nudge; then the Shape and Internal forces snap the vertices into place (Figure 4 bottom).

5. Image forces for alignment

The major work for the correspondence computation is performed by the Image forces. These act on the correspondence estimate in such a way to increase the similarity between a) visualizations of the active visage, mapped with neutral detail map data; and b) ground-truth visualizations of the target expression geometry, mapped with target expression detail map data. This analysis-by-synthesis approach is aimed at our desire to faithfully reproduce the appearance of the target expression in renderings; and to find the solution which best achieves such fidelity for our specific combination of low-resolution animation mesh and high-resolution detail maps.

In our current implementation we compute image forces on visualizations from five virtual viewpoints, mapped with highpass-filtered versions of the diffuse color texture from the expression scans. These views are shown in Figure 5 (top). The image forces for four of these viewpoints are modulated using the "imageForcesActive" control in the UI; and the fifth frontal view using the "omegaForceActive" control in the UI.

Image forces are computed at a number of sparse feature locations in these viewpoint images. These feature locations are detected once in the detail map (Figure 5 top) and then projected into the active visage visualization images at each iteration. The user has specified that at this time an initial set of 300 features should be used.

The local cross-correlation calculation is performed using a window centered at each feature location. If a peak has been detected, then the feature is shown in the target expression viewpoint visualization, displaced according to the centroid of the peak position. This displacement, together with a confidence measure, are used to define a sparse image force vector. These sparse image forces are then remapped into the optimization domain, resampled onto mesh vertices, and combined with the image forces from other viewpoints as well as Internal forces, etc.

The net force is shown in the optimization domain (Figure 5 bottom left). The green force arrows are superimposed on an overlay; the overlay consists of the expression scan texture, and the neutral scan texture as it would map into the optimization domain given the current estimate. At this stage in the process, registration is quite good on the nose (note the sharpness of features due to their alignment) but needs to be improved on the cheek (note the blurriness due to the offset of pronounced features).

The user has scaled the relative weight of the image forces to 4.00. After some iteration (Figure 5 bottom right) the registration on the cheek is improving.

6. Directable force for guiding the optimization

The user noticed that the correspondence process required additional direction on the chin. In Figure 6 the simulation has been paused, and the user is in the process of moving certain vertices closer to proper registration. The user then increased the number of sparse feature locations used for the Image force computations. This improves coverage in this area, in which this scan subject's skin is a bit smoother than on his nose and cheeks. After improving registration in the area, the user opted to leave a soft constraint in place by updating the Shape force target (via the "use current" button).

7. Refining the solution

The user is now toggling between neutral and expression textures in the optimization view to check the registration, which is satisfactory (Figure 7 top). The 3D layout of the mesh vertices also meets expectations overall, though it could be improved slightly at the corner of the scan subject's eye. The scan subject's eyelids are very smooth, leaving some uncertainty in an objective correspondence measurement. Therefore it is up to the creative direction of the user to obtain a result which will perform as desired in animation applications.

To tweak only certain vertices, the user pins all other vertices to constrain them against further updates (Figure 7 bottom). All forces are disabled except for Internal forces. The user activates the simulation, and through Directable and Internal forces tweaks the correspondences to relax the mesh in this area.

The user then saves out the current optimization estimate for testing. If further refinement is desired, it can be performed now, or later—after loading the intermediate optimization estimate back into the software.

8. Testing the solution

The user tests the current optimization estimate (Figure 8). This is performed by morphing between neutral and expression vertex positions for the animation mesh, while simultaneously blending detail maps from the expressions, as would occur in a downstream high-quality blendshape animation rig.
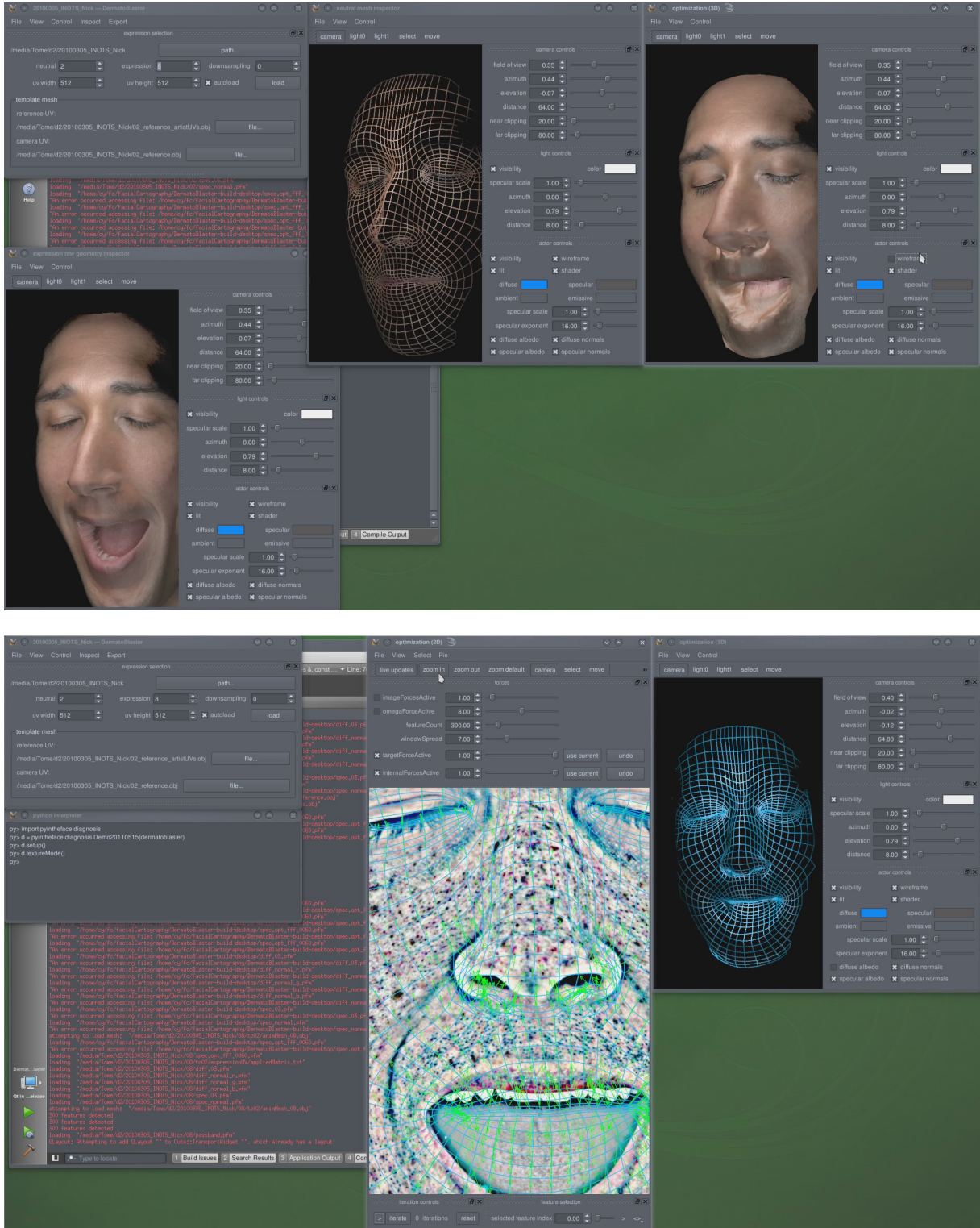
Figure 2: *Initializing the optimization. Top row: The active visage visualized in its initial state (right). Bottom row: The correspondence estimate—prior to any iterations—in the optimization* parameter *domain (center).*
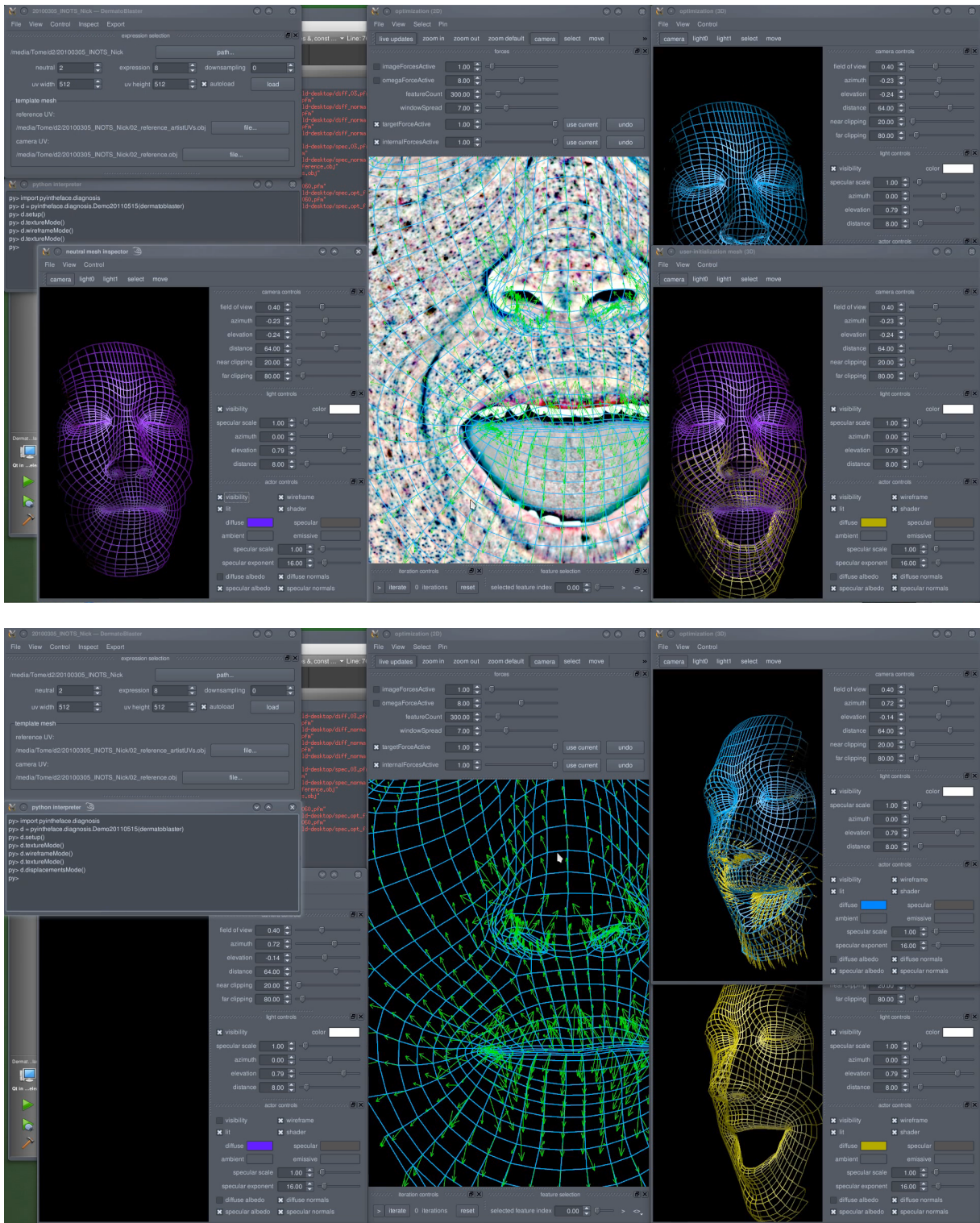
Figure 3: *Shape force for initial alignment. Given the extreme change around the mouth, the user has roughly deformed the animation mesh (in 3D; no correspondence information) to use to bootstrap the optimization. Top row: The roughly deformed animation mesh (right), compared to the neutral animation mesh (left). Bottom row: Shape force exerted on the active visage to make it more closely resemble this roughly deformed mesh.*
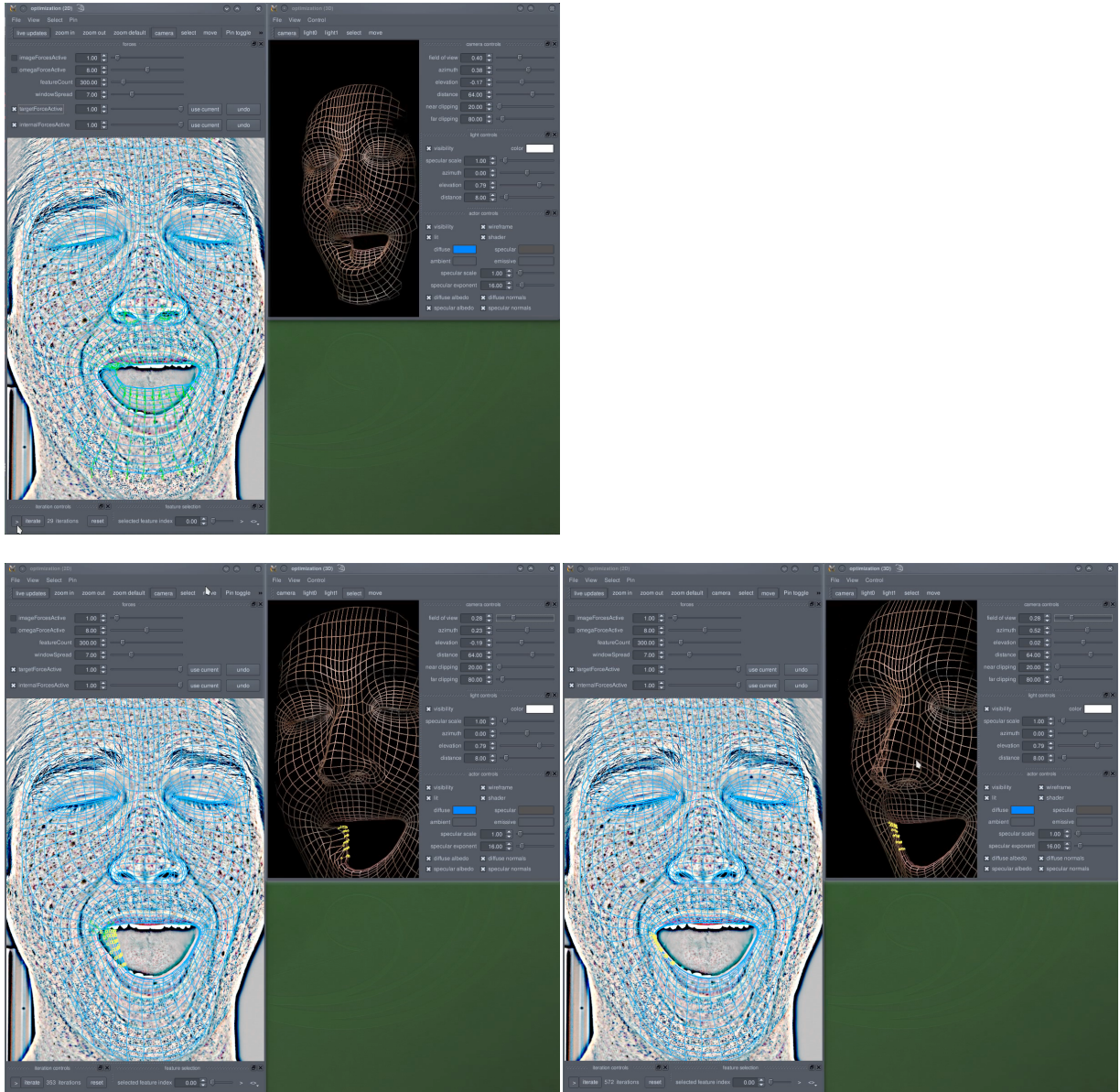
Figure 4: *Directable force for recovering from a local optimum. Top row: Initial bootstrapping with shape force to resemble the initial hint, with internal force active to maintain smoothness and ordering. Bottom row: Applying directable force to move some vertices out of a local optimum.*
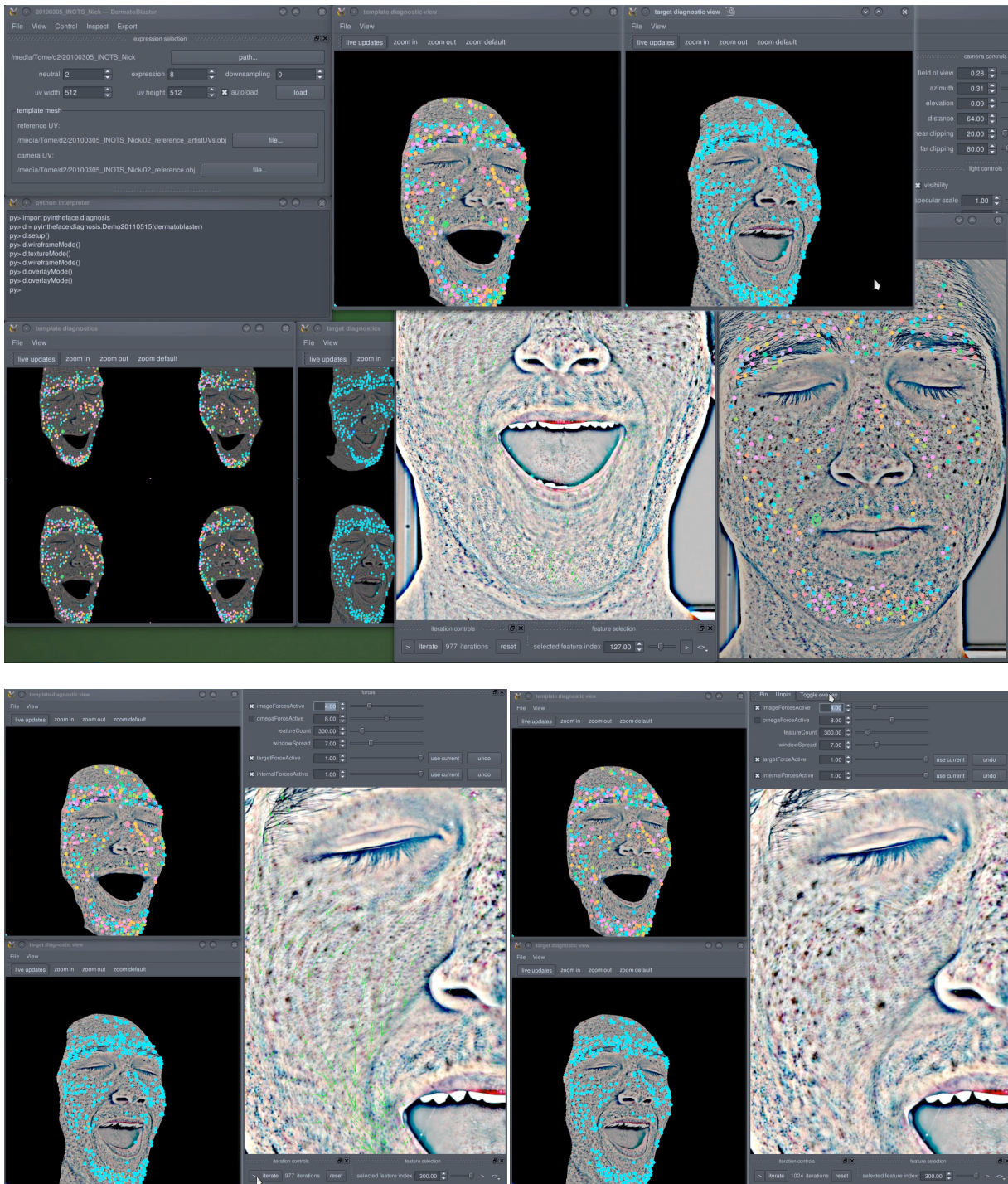
Figure 5: *Image forces for alignment. Top row: The viewpoints for active visage visualization (left), and the initial set of feature locations at which image forces are computed. Bottom row: Applying image forces with the small initial set of features and initial relative scale of 4 (left); assessing registration thus far (right).*
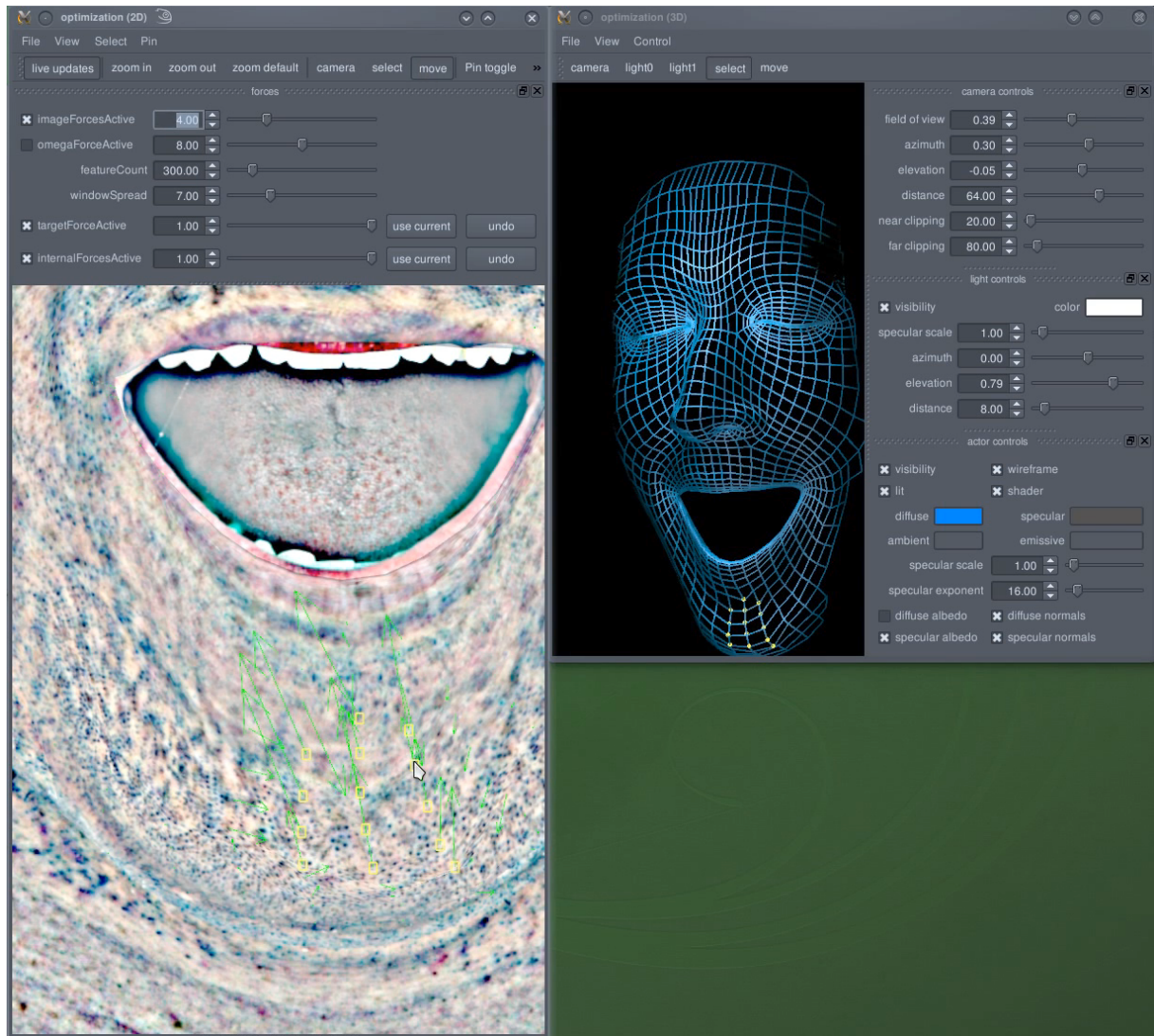
Figure 6: *Directable forces with iteration paused to improve the registration in some areas.*
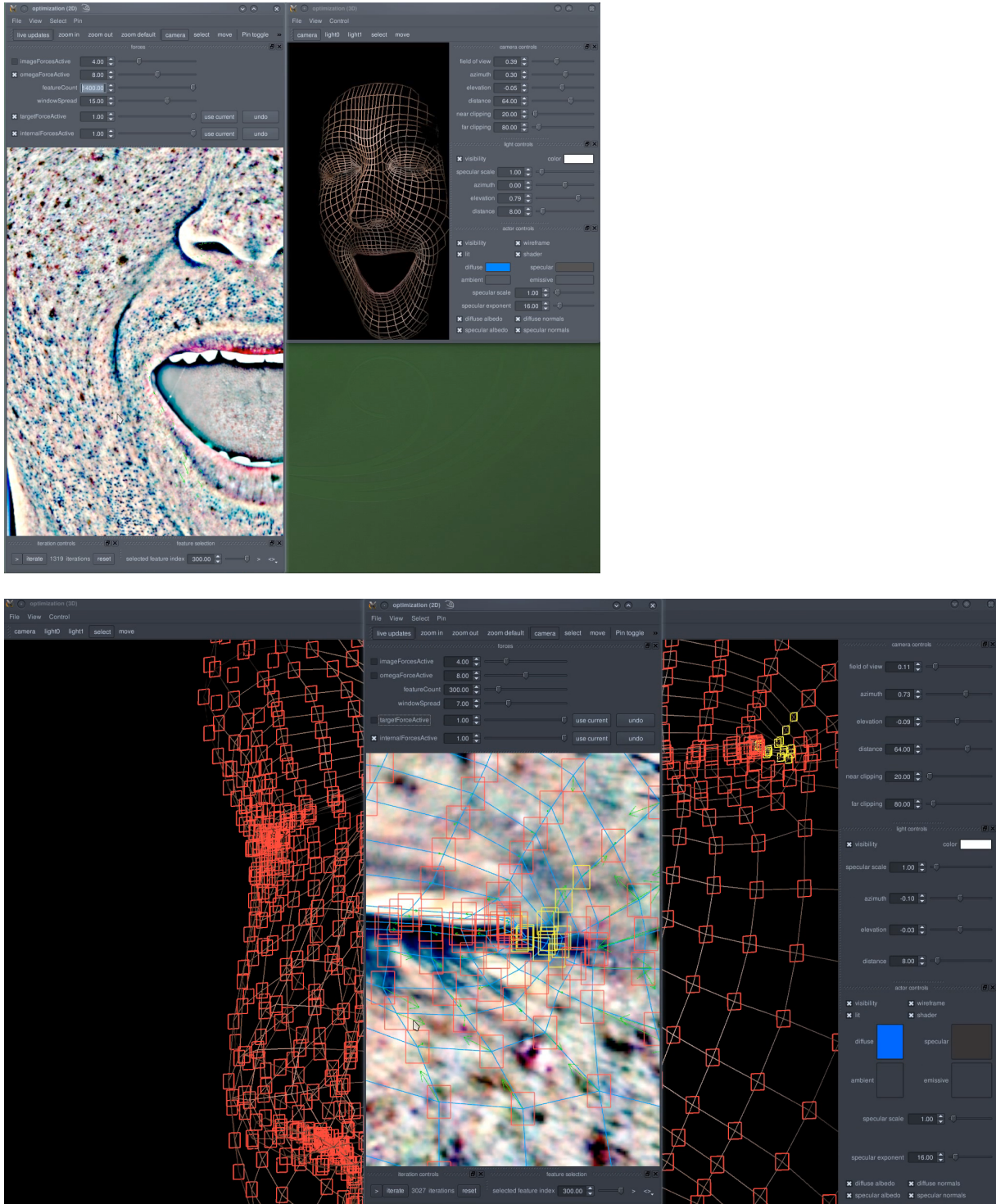
Figure 7: *Refining the solution. Top row: Arriving at our solution; checking registration and 3D layout of the vertices. Bottom row: refinement of selected vertices around the corner of the eye, with other vertices pinned in place (red).*

Figure 8: *Blend shape animation to test the final correspondences between neutral and expression scan. The renderings blend maps from both expressions.*