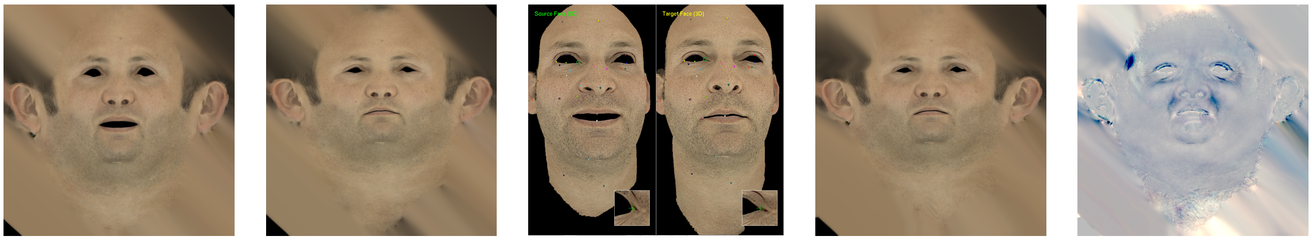
Vuvuzela: A Facial Scan Correspondence Tool

Ryosuke Ichikari   Oleg Alexander   Paul Debevec

USC Institute for Creative Technologies   oalexander@ict.usc.edu



(a) (b) (c) (d) (e)

Figure 1: *(a) Source. (b) Target. (c) Vuvuzela workflow. (d) Warped source. (e) Difference between warped source and target.*

1 Introduction

When scanning an actor's face in multiple static facial expressions, it is often desirable for the resulting scans to all have the same topology and for the textures to all be in the same UV space. Such "corresponded" scans would enable the straightforward creation of blendshape-based facial rigs. We present Vuvuzela, a semi-automated facial scan correspondence tool. Vuvuzela is currently being used in our facial rigging pipeline, and was one of the key tools in the Digital Ira project.

2 Our Approach

Our rig building process begins by scanning an actor's face in our Light Stage X device [Ghosh et al. 2011]. We capture a set of about 30 static facial expressions, roughly corresponding to Action Units from the Facial Action Coding System [Ekman and Friesen 1978]. We also capture a master "neutral" expression, which becomes the target scan in our correspondence pipeline.

Rather than storing our scans as geometry and textures, we choose instead to store our scans as images. Each one of our scans is stored as a set of 4K, 32 bit float EXR images, including diffuse, specular, specular normals, and a high resolution point cloud. The maps are in a cylindrically unwrapped UV space, representing our ear to ear data. However, the UV space differs slightly for each expression scan.

Vuvuzela exploits this image-based scan represantation by doing the scan correspondence in 2D rather than 3D. Vuvuzela takes as input two scans: one of the expressions as the source and the neutral expression as the target. Vuvuzela provides an OpenGL UI, allowing the user to interact with the scans in 3D. The scans are rendered with the diffuse textures only, and all of the correspondence processing uses only the diffuse textures.

The user clicks corresponding points in the source and target scans, such as corners of the eyes and lips, and other facial landmarks. We found that we don't need to put dots or markers on the face during scanning, because there is plenty of naturally occuring texture in the face, especially when over-sharpened. The placement of the correspondence points doesn't have to be exact—the points are used only as an initialization by our algorithm.

Once enough points have been placed, the user presses the Update button, which triggers our correspondence algorithm. The result is displayed to the user and the UI offers several modes to preview the quality of the correspondence, including a "blendshape" slider blending both geometry and/or texture. The user can then add, delete, or edit points, and repeat the process until a high quality correspondence is achieved.

Our algorithm has three steps and runs in 2D. First, we construct a Delaunay triangulation between the user supplied points and apply affine triangles to roughly pre-warp the source diffuse texture to the target. Second, we use GPU-accelerated optical flow to compute a dense warp field from the pre-warped source diffuse texture to the target. Finally, we apply the dense warp to each one of our source texture maps, including diffuse, specular, specular normals, and point cloud. The result is the source scan warped to the target UV space. The submillimeter correspondence is able to align individual pores across the majority of the face.

Some expressions are more challenging to correspond than others. Especially expressions with lots of occlusions, like mouth open to mouth closed. In such cases, optical flow will fail to get a good result. We assist optical flow in two ways. First, we paint black masks around occlusion regions in both source and target diffuse textures. Second, we mark some points as "pinned" and those points are rasterized into small black dots at runtime. Using both of these techniques in combination usually produces good results even in the toughest cases.

A useful byproduct of Vuvuzela is the ability to generate blendshapes directly from the corresponded scans. First, we remesh the neutral scan, creating an artist mesh with artist UVs. Then we load the artist mesh into Vuvuzela and export the blendshapes for all the scans by looking up vertex positions in the warped point clouds. All the texture maps are also warped into the artist UV space, which is simply an additional affine triangles 2D warp. The result is a set of blendshapes and texture maps ready to hand off to the facial rigger.

References

EKMAN, P., AND FRIESEN, W. 1978. *Facial Action Coding System: A Technique for the Measurement of Facial Movement.* Consulting Psychologists Press, Palo Alto.

GHOSH, A., FYFFE, G., TUNWATTANAPONG, B., BUSCH, J., YU, X., AND DEBEVEC, P. 2011. Multiview face capture using polarized spherical gradient illumination. *ACM Trans. Graph. 30*, 6 (Dec.), 129:1–129:10.

SIGGRAPH 2013, July 21 – 25, 2013, Anaheim, California.

ACM 978-1-4503-2261-4/13/07