ABSTRACT

Establishing inter-mesh dense correspondence is a key step in the process of constructing Morphable Models. The most successful approaches to date reduce this 3D correspondence problem to a 2D image morphing one by applying an interpolant in UV space - a space in which the manifold of the face is flattened into a contiguous 2D atlas. Contiguous UV spaces are natural products of the laser scanning devices popular for use in Morphable Model construction, but a wide gamut of devices can now be used to capture 3D data that do not yield a UV space representation conducive to warping. In this paper we explore how to optimally construct contiguous UV spaces from both annotated and non-annotated facial meshes with efficient cylindrical and spherical projection operations.

Index Terms— dense correspondence, cross parametrisation, Morphable Model, UV space

1. INTRODUCTION AND PREVIOUS WORK

Arguably the most important stage of Morphable Model construction is establishing dense correspondence. Regardless of what statistical treatment is performed on the aligned data, poor alignment of the facial meshes will yield poor results.

This general problem of cross parameterization is well studied for generic meshes, or meshes with broad restrictions (i.e. the family of all genus 0 meshes) [1, 2, 3]. Whilst powerful, these methods are computationally expensive and over engineered for the somewhat more constrained case of facial meshes. As a result, work to date in the Morphable Model community takes a simpler approach to cross parametrization.

Recognising that the UV or texture space produced from laser scanning is typically a contiguous embedding of the mesh in a Cartesian 2D space, any 2D non-rigid image alignment technique applied between two UV spaces infers an alignment of the corresponding 3D data (figure 2).

In the original Morphable Model formulation, Blanz and Vetter [4] utilised optical flow to find pixel-wise correspondence in UV space. This worked well on the dataset used to build the model - all 200 subjects were of a similar age and all displayed a neutral expression - so an intensity-based optical flow was able to achieve good results, although a flow regularisation step was required to deal with drift artefacts. Patel and Smith [5] addressed these shortcomings by utilising a Thin Plate Spline [6] (TPS) interpolation driven by manual annotations to find dense correspondences in UV space, at the cost of having to landmark meshes by hand.

It is notable that Patel and Smith showed that annotation-driven techniques like TPS produce superior dense correspondence when compared to unguided methods like optical flow, emphasising the importance of having good annotations for deformable model construction. Good annotations are equally important for images, and in that domain it is com-
The geometry of a mesh is defined as

2.1. Data representation

Topology is encoded in a triangle list $T = \{t^T_1, t^T_2, \ldots, t^T_m\}$, $t_i = [t^R_i, t^G_i, t^B_i, t^U_i, t^V_i, t^W_i, t^T_i] \in \mathbb{R}^7$ where $t_i$ is the $i$th triangle index. Note that triangle definitions are strictly ordered, providing a notion of a front and back face to each triangle.

Texture is given in a separate 2D space $I$, and the relationship between the texture space and the geometry encoded in per-vertex texture coordinates $C = [c^R_t, c^G_t, c^B_t, c^U_t, c^V_t, c^W_t, c^T_t] \in \mathbb{R}^7$, $c_i = [c^R_i, c^G_i] \in \mathbb{R}^2$. The $i$th vertex colour value is selected by $I(c_i)$, with inter-vertex $c$ values being assigned by linear interpolation of the texture coordinate across the triangle face.

Mesh $\mathbf{M} = \{\mathbf{X}, \mathbf{T}, \mathbf{C}, I\}$ is thus comprised of $n$ vertices and $m$ triangles, and a textured mesh is given by $\mathbf{M}_t = \{\mathbf{X}, \mathbf{T}, \mathbf{C}, I\}$.

2.2. Dense Correspondence

In 3D statistical deformable models, we are interested in generating novel mesh instances. Consider the case of a simple linear model of shape and texture which has a fixed topology $T_m$ and texture mapping $C_m$. A new geometry and texture instance is generated from a linear combination of orthogonal bases (which might have been found using Principal Component Analysis)

$$\mathbf{X}^* = \mathbf{X}_\mu + \sum_{\alpha_i} \alpha_i \mathbf{X}_i \quad \mathbf{I}^* = \mathbf{I}_\mu + \sum_{\beta_i} \beta_i \mathbf{I}_i \quad (1)$$

where $\alpha_i \in \mathbb{R}$, $\beta_i \in \mathbb{R}$ are weightings on basis of shape and texture respectively. Such a model combined with the topology information generates a mesh $\mathbf{M}^* = \{\mathbf{X}^*, \mathbf{I}^*, \mathbf{T}_m, \mathbf{C}_m\}$.

Concentrating on the shape (all the following can be applied to texture without loss of generality) we note that this formulation enforces that all shape bases have the same number of vertices. Furthermore, for this mathematical model to yield face-like instances, we need require that the $j$th vertex across all shape bases has a particular unique semantic meaning. The conjecture is that a linear combination of points that share semantic meaning (e.g. nose tip) will have the same meaning in the generated mesh (that is, a mixture of nose tips wont produce a left eyeball). The property that between two meshes the $j$th vertex has a shared semantic meaning is dense correspondence.

2.3. Ideal UV spaces

An ideal UV space is a two dimensional space $\mathbf{U} \subset \mathbb{R}^2$ that, for all $\mathbf{u} \in \mathbf{U}$, satisfies the following two properties:

- There exists a bijective mapping between $\mathbf{X}$ and $\mathbf{U}$
  $f(\mathbf{x}) \mapsto \mathbf{u} \quad f^{-1}(\mathbf{u}) \mapsto \mathbf{x} \quad (2)$
- For infinitesimal changes $d\mathbf{x}$ along the manifold and $d\mathbf{u}$ in UV space
  $f(\mathbf{x} + d\mathbf{x}) \mapsto \mathbf{u} + d\mathbf{u} \quad (3)$
3. IDEAL UV PROJECTIONS

We now analyse two simple projections that can be performed on 3D mesh data that yield ideal UV space representations - cylindrical unwrapping, and spherical unwrapping. For both techniques we will for now assume that the facial mesh has been in some way optimally placed at the origin pointing down the $z$ axis, and with the top of the head pointing up the $y$ axis. We will later consider what the optimal placement is for each technique.

Note that these projections are applied to $X$ only. Building the contiguous UV space given the adjusted vertex locations is trivial - it simply requires an orthographic rendering of the projected mesh, which can be efficiently computed with OpenGL.

**Cylindrical Unwrap:** Let $u = (\theta, z')$, where $f_c(x) \mapsto u$ is decomposed into two mappings:

$$\theta \leftarrow \arctan \left( \frac{x}{z} \right) \quad z' \leftarrow y \quad (4)$$

where we enforce $\theta \in [-\pi, \pi)$, and note that the discontinuity in $\theta$ occurs at the back of the head.

**Spherical Unwrap:** Let $u = (\phi, \theta)$, where $f_s(x) \mapsto u$ is decomposed into two mappings:

$$\phi \leftarrow \arctan \left( \frac{-x}{z} \right) \quad \theta \leftarrow \arccos \left( \frac{y}{r} \right) \quad (5)$$

where $r = \sqrt{x^2 + y^2 + z^2}$, and we enforce $\phi \in [-\pi, \pi)$, $\theta \in [-\pi/2, \pi/2]$ so that both discontinuities occur at the back of the head.

3.1. Unwrapping optimally

In all cases, we are seeking an optimal translation $y$ to position the mesh before applying equation 4 or 5

$$x_i \leftarrow x_i - y \quad \forall i \in [1, 2, ..., n] \quad (6)$$

**With annotations:** In this case we have access to a set of $t$ annotations $X' = [x'_1, x'_2, ..., x'_t]$ for a face. Assuming that all landmarks should be placed as close to the surface of the sphere or cylinder as possible, we can find the optimal translation $y$ by solving the nonlinear least squares problem 1

$$\min_{y, r} \sum_{j=1}^{t} |r - ||y - x'_j||z|^2 \quad (7)$$

In the cylindrical case, we discount $y$ values from each of the landmarks, solving for a 2D translation in the $z - x$ plane.

We use Coope’s linear least squares reformulation of the problem [10] to achieve stable accurate solutions.

1The optimum radius $r$ is calculated as a necessary by-product, but is not of great interest as all it provides is a scale factor in the case of both mappings.
Fig. 5. *Shaded, left:* The mean, first second and third Principal Components of a Morphable Model constructed using cylindrical unwrap and TPS. Each component is shown added (+) and subtracted (-) from the mean. Both texture (left) and shape (right) components are shown. Texture components are visualized rendered on the mean face. *Right:* The result of fitting the model to a subject smiling.

**Without annotations:** A good heuristic that leverages the symmetry present in the human face is simply to place the centre of mass of the mesh at the origin

\[ y = \frac{1}{n} \sum_{j=1}^{n} x_j \]  \hspace{1cm} (8)

4. RESULTS

Figures 3 and 4 show contiguous UV maps computed using both the spherical and cylindrical unwrapping methods for two subjects captured using a DI4D capture system. Each unwrapping is shown for the cases where annotations are present and absent.

Cylindrical unwrapping does a good job of flattening the majority of the facial surface under both the optimal and centre of mass approaches. The spherical map performs well, but does introduce issues around the eye sockets, where aggressive bending leads to small folds occurring in the UV space.

5. CONCLUSION

The ability to robustly generate contiguous UV spaces for Morphable Model construction is critical. Cylindrical unwrapping has been shown to be a simple but powerful technique for generating such spaces in an optimal way from arbitrary data, expanding the types of data that Morphable Models can be constructed from. Further work could explore how this technique coupled with a bespoke AAM could be used to automatically annotate arbitrary 3D facial data.
6. REFERENCES


