OPTIMAL UV SPACES FOR FACIAL MORPHABLE MODEL CONSTRUCTION

James Booth, Stefanos Zafeiriou*†

Imperial College London
Department of Computing
180 Queen's Gate, SW7 2AZ, London, UK

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-

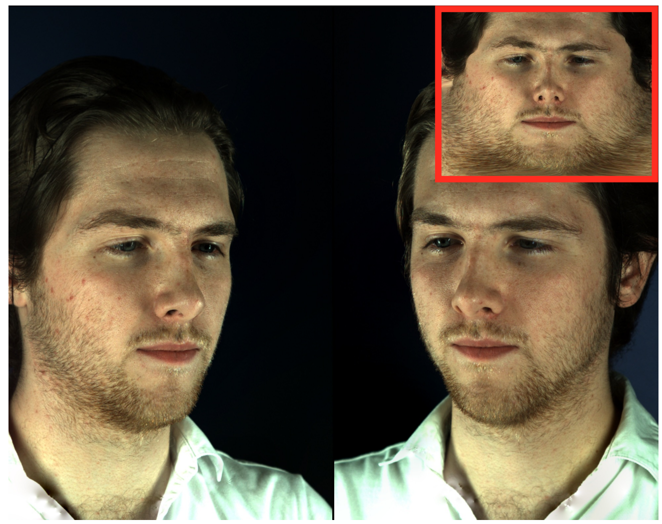

Fig. 1. *Main:* Typical UV map produced from a photometric stereo capture system. The mapping onto the 3D mesh is non-contiguous and thus not suitable for image-based cross parameterisation techniques. *Inset:* An ideal UV space for the same mesh, where the texture has a contiguous embedding. In this space we can reduce the 3D mesh correspondence problem into a 2D image morphing one. In this paper, we explore how to go from the former texture representation to the latter.

*James Booth acknowledges an EPSRC DTA from Imperial College London

†The work of Stefanos Zafeiriou was partially funded by the EPSRC project EP/J017787/1 (4D-FAB)

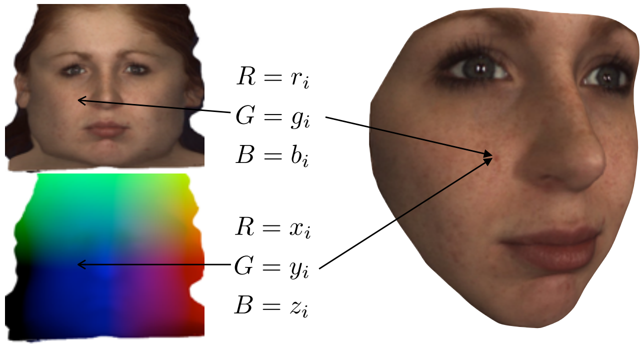

Fig. 2. Texture information (top left) is naturally expressed in UV space but spatial data (bottom left) can be embedded too. An interpolant such as TPS can be applied between a group of mesh's UV spaces to 'line up' each UV space. Sampling colour from aligned textures and spatial information from aligned spatial images allows for the constuction of meshes in dense correspondence (right).

mon practice to utilize an Active Appearance Model [7, 8] (AAM) to automatically find landmarks (which may be refined by iteration of the AAM or manually checked by hand to ensure quality). Cosker [9] demonstrated that this automatic annotation technique can be successfully applied to 3D meshes by fitting an AAM in the UV space. This application places an important constriction on the UV map - it must be possible to produce a UV space from a mesh without any annotations at all, so that an initial attempt at fitting an AAM can be done.

The ability to construct a suitable contiguous UV space for facial meshes is clearly of central importance to 3D deformable model construction, so it is somewhat surprising that an analysis of subject has not been reported previously. This is most likely due to the fact that the output of Cyberware laser scanners, the devices used for acquiring data in the above three examples, is naturally a contiguous UV space [4]. However the advent of highly accurate structured light and photometric stereo systems means there has been an explosion of devices in the last decade that can produce data suitable for Morphable Model construction that don't provide a suitable UV space for establishing dense correspondence in (figure 1). This paper formalizes the desirable qualities of a UV space, and explores two simple but effective UV space mappings that, when applied to any 3D facial data, produce contiguous UV spaces suitable for Morphable Model construction.

2. THEORY

2.1. Data representation

The geometry of a mesh is defined as $\mathbf{X} = [\mathbf{x}_1^T, \mathbf{x}_2^T, ..., \mathbf{x}_n^T]^T$, $\mathbf{x}_i = [x_x^i, x_y^i, x_z^i]^T \in \mathbb{R}^3$ where $\mathbf{x}_i$ is termed the ith vertex. Topology is encoded in a triangle list $\mathbf{T} = [\mathbf{t}_1^T, \mathbf{t}_2^T, ..., \mathbf{t}_m^T]$, $\mathbf{t}_i = [t_1^i, t_2^i, t_3^i]$, $t_j^i \in \{\mathbb{Z}^+ \mid t_j^i \leq n\}$ where $\mathbf{t}_i$ is the ith 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 $\mathbf{I}$, and the relationship between the texture space and the geometry encoded in per-vertex texture coordinates $\mathbf{C} = [\mathbf{c}_1^T, \mathbf{c}_2^T, ..., \mathbf{c}_n^T]^T$, $\mathbf{c}_i = [c_u^i, c_v^i] \in \mathbb{R}^2$. The ith vertex colour value is selected by $\mathbf{I}(\mathbf{c}_i)$, with inter-vertex $\mathbf{c}$ values being assigned by linear interpolation of the texture coordinate across the triangle face.

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

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 $\mathbf{T}_m$ and texture mapping $\mathbf{C}_m$. A new geometry and texture instance is generated from a linear combination of orthogonal bases (which might have been found using Principal Component Analysis)

$$\mathbf{X}^* = \mathbf{X}_\mu + \sum_i^{k_\alpha} \alpha_i \mathbf{X}_i \qquad \mathbf{I}^* = \mathbf{I}_\mu + \sum_i^{k_\beta} \beta_i \mathbf{I}_i \tag{1}$$

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 jth 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 jth 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 $\mathbf{U}$

$$f(\mathbf{x}) \mapsto \mathbf{u} \qquad f^{-1}(\mathbf{u}) \mapsto \mathbf{x} \tag{2}$$

- For infinitesimal changes $d\mathbf{x}$ along the manifold and $d\mathbf{u}$ in UV space

$$f(\mathbf{x} + d\mathbf{x}) \mapsto \mathbf{u} + d\mathbf{u} \tag{3}$$

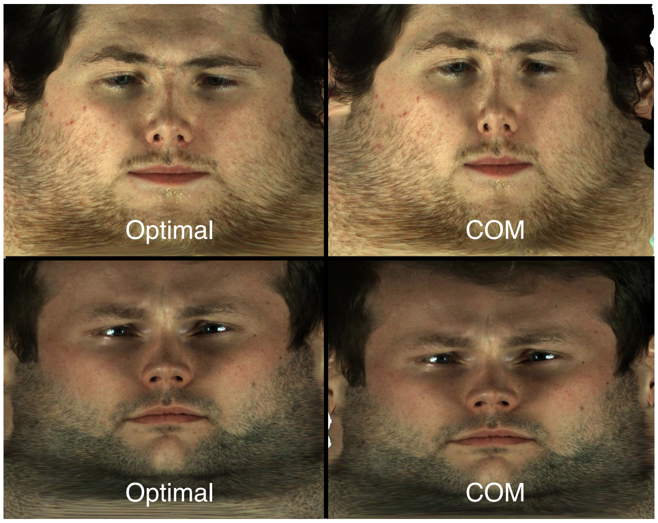

Fig. 3. Cylindrical unwrap projection for two subjects both optimally performed based on annotations (left) and crudely aligned around the centre of mass (right)

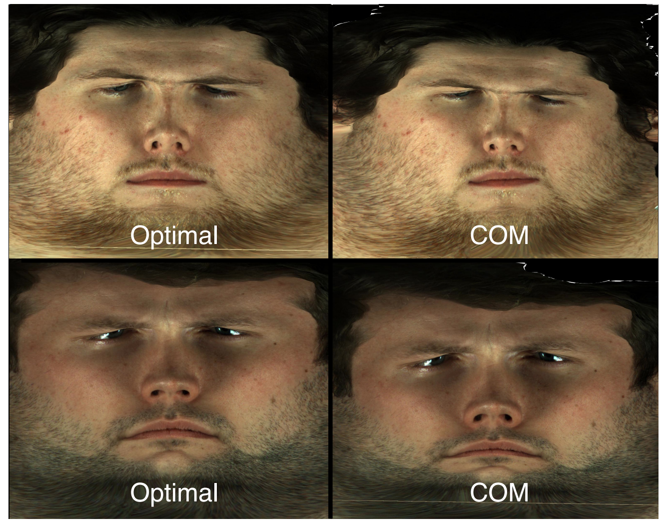

Fig. 4. Spherical unwrap projection for two subjects both optimally performed based on annotations (left) and crudely aligned around the centre of mass (right)

Note that this enforces contiguous UV space for a contiguous mesh.

The first property is a mathematical formulation of correspondence which has to hold in order for interpolations in UV space to be interpreted along the mesh manifold. The second property formalizes the intuition that the local topology in UV space has to map to local topology on the mesh. This property means that neighbouring vertices on the mesh will be neighbours in UV space, and allows us to have a shared topology $\mathbf{T}$ between UV vertices and 3D vertices.

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 $\mathbf{X}$ 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 $\mathbf{u} = (\theta, z')$, where $f_c(\mathbf{x}) \mapsto \mathbf{u}$ is decomposed into two mappings:

$$\theta \hookleftarrow \arctan\left(\frac{x}{z}\right) \qquad z' \hookleftarrow y \tag{4}$$

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

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

$$\phi \hookleftarrow \arctan\left(\frac{-x}{z}\right) \qquad \theta \hookleftarrow \arccos\left(\frac{y}{r}\right) \tag{5}$$

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 $\mathbf{y}$ to position the mesh before applying equation 4 or 5

$$\mathbf{x}_i \hookleftarrow \mathbf{x}_i - \mathbf{y} \qquad \forall i \in [1, 2, ..., n] \tag{6}$$

With annotations: In this case we have access to a set of t annotations $\mathbf{X}' = [\mathbf{x}'_1, \mathbf{x}'_2, ..., \mathbf{x}'_t]$ 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 $\mathbf{y}$ by solving the nonlinear least squares problem [1]

$$\min_{\mathbf{y},r} \sum_{j=1}^{t} |r - ||\mathbf{y} - \mathbf{x}'_j||_2|^2 \tag{7}$$

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.

[1]The 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.

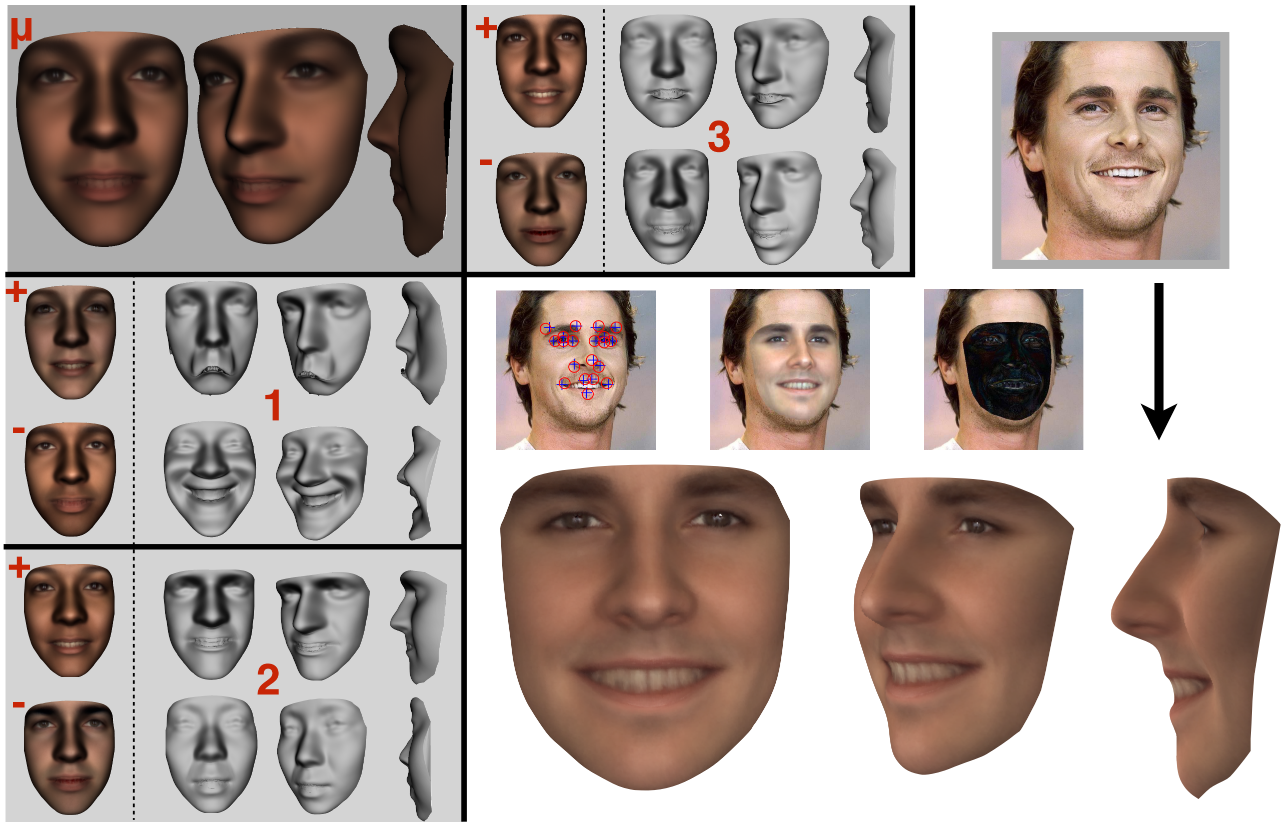

Fig. 5. *Shaded, left:* The mean, first second and third Principal Components of a Morphable Model constructed using cylindrical unwrap and TPS. Each component is shown added (+) and subtracted (-) from the mean. Both texture (left) and shape (right) components are shown. Texture components are visualized 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

$$\mathbf{y} = \frac{1}{n}\sum_{j=1}^{n}\mathbf{x}_j \tag{8}$$

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.[2] 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.

[2]Results generated with Menpo, http://www.menpo.io

To demonstrate the effectiveness of the cylindrical unwrapping approach, a Morphable Model was constructed using this mapping coupled with TPS as an interpolant. Approximately 300 facial meshes from the *BU4D-FE* database [11] were manually annotated with 68 fiducial points. The subjects are displaying a mix of facial emotions, which yields emotion-specific principal components, which are visualized in figure 5. The model demonstrates impressive detail retention under various facial expressions, evidence that the proposed techniques are performing well in establishing dense correspondence.

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

[1] Marc Alexa, "Merging polyhedral shapes with scattered features," in *Shape Modeling and Applications, 1999. Proceedings. Shape Modeling International'99. International Conference on*. IEEE, 1999, pp. 202–210.

[2] Emil Praun, Wim Sweldens, and Peter Schröder, "Consistent mesh parameterizations," in *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*. ACM, 2001, pp. 179–184.

[3] Vladislav Kraevoy and Alla Sheffer, "Cross-parameterization and compatible remeshing of 3d models," in *ACM Transactions on Graphics (TOG)*. ACM, 2004, vol. 23, pp. 861–869.

[4] Volker Blanz and Thomas Vetter, "A morphable model for the synthesis of 3d faces," in *Proceedings of the 26th annual conference on Computer graphics and interactive techniques*. ACM Press/Addison-Wesley Publishing Co., 1999, pp. 187–194.

[5] Ankur Patel and William AP Smith, "3d morphable face models revisited," in *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*. IEEE, 2009, pp. 1327–1334.

[6] Fred L. Bookstein, "Principal warps: Thin-plate splines and the decomposition of deformations," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 11, no. 6, pp. 567–585, 1989.

[7] Timothy F Cootes, Gareth J Edwards, Christopher J Taylor, et al., "Active appearance models," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 23, no. 6, pp. 681–685, 2001.

[8] J. Alabort i medina and S. Zafeiriou, "Bayesian active appearance models," in *Proceedings of IEEE Intl Conf. on Computer Vision & Pattern Recognition*, June 2014.

[9] Darren Cosker, Eva Krumhuber, and Adrian Hilton, "A facs valid 3d dynamic action unit database with applications to 3d dynamic morphable facial modeling," in *Computer Vision (ICCV), 2011 IEEE International Conference on*. IEEE, 2011, pp. 2296–2303.

[10] ID Coope, "Circle fitting by linear and nonlinear least squares," *Journal of Optimization Theory and Applications*, vol. 76, no. 2, pp. 381–388, 1993.

[11] Lijun Yin, Xiaochen Chen, Yi Sun, Tony Worm, and Michael Reale, "A high-resolution 3d dynamic facial expression database," in *Automatic Face & Gesture Recognition, 2008. FG'08. 8th IEEE International Conference on*. IEEE, 2008, pp. 1–6.