Towards a Low Cost Realistic Human Face Modelling and Animation Framework

A. M. Woodward, P. J. Delmas
University of Auckland, Dept. Computer Science.
awoo016@ec.auckland.ac.nz

Abstract

Realistic 3D human faces have found application in a myriad of different computer situations. The easily recognisable human face provides an emotive bridge for the personalization of many human-computer interactions in today's society. Face models based on real people are thus widely sought after. This paper presents computer vision techniques for low cost acquisition of raw data, a computer graphics model for facial animation, and an interface between the two which binds the process into a system for facial synthesis and animation. This paper addresses a low cost approach in terms of acquisition technique, making the process viable for a wide range of environments and audiences. The presented solution is capable of running on standard hardware. Results for a variety of test subjects are presented and commented on.

Keywords: Binocular stereo, facial animation, abstract muscle model, layered tissue model

1 Introduction

Facial animation has a wide range of computer applications such as the gaming and cinema industries, medicine, social agents and avatars. Thus a great deal of work has been employed to advance these various techniques and applications [2]. As commodity hardware increases in performance, a renaissance in more advanced and physically accurate techniques for facial animation will come to the fore.

Although facial animation is the domain of computer graphics, there is a need for the acquisition of 3D data from real human subjects. In the absence of sophisticated scanning equipment, low cost techniques must be explored. Binocular stereo is a strong choice since it has readily available texture data, is passive to the scene, and 3D positions can be reconstructed from the acquired depth map.

In this paper, the process of acquiring 3D data is firstly presented. Stereo vision techniques are utilized to acquire a dense disparity map of the face. The mapping between a generic model and the acquired 3D data allows us to obtain an animated 3D reconstruction of the test subject. Next, the facial animation system is described, comprising a physics based model of human tissue. Finally, a variety of expressions are generated and commented on with regards to muscle formulation and effectiveness.

2 Acquisition of 3D Data using Binocular Stereo

The human face provides a unique challenge for stereo vision. Within a single image, we are presented with many areas of differing texture, surface reflectance properties, and colouration. As a broad collective, these properties can differ quite markedly between individuals, when considering aspects such as ethnicity and age. For human faces we require a dense disparity map for 3D reconstruction purposes. For a comparison specifically regarding algorithm operation on human faces, see [6]. The KZ1 algorithm was chosen for its relatively robust inference scheme on a wide variety of signals that may compose a facial image.

2.1 Lab Setup

A stereo rig was setup consisting of two Canon A80 cameras with standard epipolar geometry. The cameras were placed in a vertical configuration as depicted in Fig. 1 to take into account the horizontal symmetry of human faces, which can influence the result of stereo correspondence between pixels. Thus, knowing both the distance between cameras (e.g. baseline) and their focal length, the reconstruction formulae are straightforward:

\[
X = \frac{b \cdot x_{\text{left}}}{x_{\text{left}} - x_{\text{right}}} \quad Y = \frac{b \cdot y}{x_{\text{left}} - x_{\text{right}}} \quad Z = \frac{b \cdot f}{x_{\text{left}} - x_{\text{right}}}
\]
Where $b$ is the baseline, and $f$ is the focal distance. Using the standard epipolar geometry, both cameras will possess the same $f$ and the disparity measure can be restricted to one dimension. The disparity is the positional difference of a point’s projection into the two image frames.

2.2 The KZ1 Stereo Algorithm

Surface reconstruction when utilizing pairs of stereo images requires the matching of corresponding pixels between both images. The resulting surface is known as the disparity space image (DSI) [1], existing within the disparity space volume $(x,y,d)$. The KZ1 algorithm was chosen for its strong regularization properties and inference across homogeneous signal regions, making it applicable to human faces [5].

2.2.1 Formulation

The KZ1 algorithm uses an energy minimisation approach where the energy function to be minimized is

$$E(f) = E_{\text{data}}(f) + E_{\text{smoothness}}(f) + E_{\text{visibility}}(f)$$  \hspace{1cm} (1)

The data term imposes photo-consistency. Under the lambertian assumption corresponding pixels will have the same intensity. The smoothness term penalizes the assignment of different disparities to neighbouring pixel positions in the 4-neighbourhood system. This helps in providing local coherence and the regularization avoids spurious noise in the result. Discontinuities can be preserved through an appropriate choice of metric for the smoothness penalty function. Lastly, the visibility term handles the visibility constraint and considers only the case where 3D-points at different depths can interact. The energy $E$ of the system is minimized using the $\alpha$-expansion technique to find a strong local minimum for the DSI solution. See [5] for a full description of the algorithm and its implementation.

3 Interface Between the Raw Data and the Facial Animation System

The interface between the raw data and its integration into a facial animation framework involves the selection of a few important landmarks on the raw data and matching them to their equivalent positions on a generic model possessing our pre-configured animation system. Radial Basis Functions (RBF) [3] are used to interpolate the generic model to the raw data, and finally, vertices of the generic model are moved closer towards the raw data surface by considering their closest points.

The rationale behind using a generic model is that it firstly allows a great deal of prior knowledge and system to be infused along with its geometric data. The raw data obtained may not be suited for direct integration into an animation framework due to noise and vertex distribution. Facial regions can be defined once on the model, and can then become applicable to a range of different facial data. Jaw definitions, eye regions, and the scalp are important examples of this. The muscle system used in this application was specifically matched to the generic model and warped accordingly via RBF interpolation.
4 Facial Animation Model

For the purpose of realistic facial animation, a physically based, layered tissue model was created. Since the facial tissue is not just a single layer of skin, a mass-spring system has been used to create this layered tissue model which represents the epidermal, fascial and skull layers. This is an attempt to more faithfully represent the anatomical nature of human facial tissue. The benefit of such a model is that the interaction between the nodes and springs provides a more naturalistic behaviour for the skin tissue, stretching and pulling as forces are applied. The automatic creation of the tissue structure is firstly described, followed by the abstract muscle definitions, and forces present within the system. The basis of this implementation is formed primarily from [9], and [10].

4.1 Spring and Node Creation

Given a mesh of facial data, it is first necessary to create the layered tissue model. Fascia and bone nodes need to be generated. This proceeds as follows:

1. Calculate the surface normals at each defined vertex in the facial mesh. This vertex position becomes the position for an epidermal node.
2. Go an increment downward below the face surface, in a direction collinear to the normal, this defines the position of a fascia node.
3. Go a further increment downward in the same direction, this defines the bone node which is assumed attached to the skull surface.
4. Interconnect all nodes in the configuration as presented in Fig. 3.

Each facet of the input mesh represents a triangular prism element with a certain rest volume and location (as depicted in the above diag). Together these form a skin patch over the face. Muscle forces are applied to fascia nodes to generate muscle contractions. The dermal-fatty layer could easily be extended by adding another surface to define a dermal layer and a subcutaneous fatty tissue layer.

4.2 Forces

The various forces presented in the following sections, when taken holistically, form a model for facial tissue. Once the layered tissue model is in place, abstract muscles apply a contraction force to the fascia nodes within the system. By doing so the tissue reacts to this application until it reaches a new equilibrium, resulting in a reconfigured facial state.

4.2.1 Spring Forces

Spring forces are the effect of the various springs interacting between the nodes, to either push or pull. This is the most important element of our model in order produce a skin tissue continuum. The force $g_j$ is the force on node $i$ due to node $j$.

$$g_j = c_j (l_j - l_\tau) s_j, \quad g_i = -g_j$$

- $c_j$ is the biphasic spring constant that models the stress-strain relationship of the tissue
- $l_j$ is the rest length for spring $j$
- $l_\tau$ is the rest length of the tissue
- $s_j = (x_j - x_i) / l_j$ is the spring direction vector for spring $j$

4.2.2 Volume preservation

The volume preservation force allows for tissue form restitution. It helps preserve the facial tissue and is therefore proportional to the observed change in volume and displacement of the volume element’s nodes.

$$q_i^c = k_1 (V^c - \bar{V}^c) n_i^c + k_2 (p_i^c - \bar{p}_i^c)$$

Figure 4: An element of the tissue model, the spring configuration over the face, and a close up view in-program.
\[ \mathbf{f}_{\text{linear}} = ak \mathbf{r} \frac{\mathbf{xv}_1}{|\mathbf{xxv}|} \]  

- \( k \) is the muscle activation co-efficient
- \( a = 1 - \frac{\cos(\theta)}{\cos(\omega)} \) is the angular weighting with respect to the muscle vector

\[ r = \cos\left( \frac{|\mathbf{xxv}| - |\mathbf{v}_2\mathbf{v}_1|}{|\mathbf{v}_2\mathbf{v}_1| (\text{fallfinish} - \text{fallstart})} \right) \frac{\pi}{2} \]

Where \( \text{fallfinish} \) specifies the radial extreme of the muscles influence, and \( \text{fallstart} \) specifies the radial distance where the muscle force influence begins to taper to zero at \( \text{fallfinish} \). \( \mathbf{x} \) is the current node, \( \mathbf{v}_1 \) is the muscle origin, \( \mathbf{v}_2 \) is the muscle insertion point in the tissue, thus \( \mathbf{xxv} \) is a vector from \( \mathbf{x} \) to \( \mathbf{v}_1 \).

**Ellipsoid Muscle** The ellipsoid muscle acts like a string bag. Points are squeezed towards the muscle origin. It possesses only a radial weighting to nodal force application, and has no angular weighting. Because of its function this muscle is also known as a sphincter muscle, its definition being that of an ellipsoid, having a major and two minor axes. The general formula is similar to the linear muscle except that there is no angular weighting to it.

\[ \mathbf{f}_{\text{ellipsoid}} = kr \frac{\mathbf{xv}_1}{|\mathbf{xxv}|} \]  

### 4.2.4 Muscle Forces

Two muscle models are used to apply forces to the fascia nodes; namely the linear muscle and the ellipsoid muscle. These abstract muscles are defined and placed in physiologically correct regions of the face. The abstract muscle allows for an important decoupling between the facial mesh and muscle position. It provides a simple and more intuitive way of manipulating a face to create an expression, has the advantage of not being tied to any specific facial mesh, and henceforth, can be easily applied to many different data sets. 20 muscles were placed within the face according to [9], and as depicted in Fig. 4.

**Linear Muscle** The linear muscle is also known as a vector muscle; having an origin where the muscle is attached onto the bone, and another point which is inserted into the skin tissue. Nodes that lie within a certain angular zone to the muscle vector, and within a certain radial distance from it’s origin, are directly affected by this muscle.
4.2.6 Expression Generation

Expressions at the basic level are the combination of various muscle contractions to generate a re-configuration of the facial state. Since we already have defined our muscle types and their function, expression generation is trivial. An expression is therefore just a set of contraction values for each muscle in the face. To animate our face from a neutral state, one needs only to progressively scale these contraction values from zero to unity over time. Once an expression is designed, muscle contraction coefficients are saved for later recall.

4.3 Results

It can be found that there are six universal categories of facial expression that all cultures can recognise. These categories are sadness, anger, joy, fear, disgust, and surprise [9]. Falling within these categories are a range of possible intensities and variations in expression detail. The following diagrams demonstrate the systems ability to reproduce these categories. What is not possible to convey through pictures alone is the smoothness of animation produced by the tissue model. An evaluation of the system for one timestep took 0.016 seconds with 8160 springs comprising the model. The following table presents various terms including the spring constants used for the system.

Table 1: Table of Settings for the System

<table>
<thead>
<tr>
<th>Nodal Mass (m)</th>
<th>?</th>
<th>Time Step Δt</th>
<th>Epidermal Spring</th>
<th>Fascia Spring</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>30.0</td>
<td>0.02</td>
<td>60.0</td>
<td>30.0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Dermal Spring 1</th>
<th>Dermal Spring 2</th>
<th>Muscle Spring</th>
</tr>
</thead>
<tbody>
<tr>
<td>70.0</td>
<td>80.0</td>
<td>10.0</td>
</tr>
</tbody>
</table>

5 Conclusion

This paper has presented a low cost solution for the acquisition and animation of a human model based on real human data. The framework employs current research into stereo algorithms and a physically based model for animation, providing superior results over simpler geometric deformation alternatives. Although the processing cost is higher than in some other models, the complex interplay between the nodes, springs, and the forces that affect them validate the choice of model needed to attain realism. Future work can be placed on analysing the affect of different resolution models, the local configuration of springs for each tissue element, and also how forces are applied to the system. Bone nodes remain in a fixed position throughout the simulation; this assumption leads to over-attachment for some regions of the face and deserves future inspection for improvement. Fully automatic methods for the mapping between the generic model and raw data are being explored. Along with a tissue model, important features such as modelling hair, lips, eyes, and ear geometry can be investigated. A more advanced reflectance model should be incorporated into the final result, and adding fine wrinkles as a further extension to accommodate for low facial tessellation could be pursued. This solution could be used as a foundation in various areas such as interactive avatars, game and movie characters, and teleconferencing over low bandwidth environments. As computers becoming increasingly more powerful we will no doubt see the realization of more advanced techniques and interactive tools in these areas.

References


