A Region-of-Influence Measure for Automatic Skinning

J.P. Lewis[1*] and Ken Anjyo[2]

[1]Computer Science, Stanford University, California, USA
[2]OLM Digital, Tokyo, Japan.
Email: zilla@computer.org,anjyo@olm.co.jp

Abstract

Skinning algorithms are widely used in film animation and video games to produce the deformation of a character's skin surface as a function of the motion of a skeleton or other control objects. Popular traditional skinning algorithms require the definition of a region of influence for each control object. These regions of influence have most often been manually defined by painting influence weights. Recent techniques offer hope of automating this task, but require an underlying measure of co-movement. Simple linear correlation is inadequate for characterizing co-movement on commonly occurring influence areas (such as some on the human face) where the movement is time correlated but in different directions. In this paper we introduce a measure that reflects time-correlated movement regardless of the direction of the movement. The resulting influence regions are illustrated using facial motion capture.

Keywords: Computer graphics, character animation, skinning, correlation

1 Introduction

Skinning algorithms describe the motion of a character's skin under the influence of underlying controls such as skeleton joint angles, motion capture markers, or a lower-resolution "animator friendly" mesh. These algorithms are employed in essentially all films with animated characters, as well as in video game characters. Despite its importance and wide use, skinning has only recently become a popular subject for graphics research, while skinning practice has largely been based on industry implementations of several simple techniques.

Many widely used skinning techniques are variations of linear blend skinning, in which the motion of a skin vertex is a weighted sum of the motion of that vertex as rigidly transformed by several surrounding skeletal frames. Commercial animation packages such as Maya and Softimage have generalizations of this technique in which the skin vertex position is a weighted sum of the motion of various underlying control objects.

The weights used in the linear blend skinning approaches are usually manually defined. This process is both labor-intensive, in that it involves painting one weight map over the character skin for each influence object (e.g. each bone in the underlying skeleton), and difficult, in that the weights only affect the geometry indirectly and in combination. In fact, these weight maps are notoriously difficult to generate, and often result in artifacts that can only be hidden through iterative experimentation [1]. Although existing commercial packages have automatic weight-assignment features, these are based on simple heuristics and are only suitable for low-quality results, or as a first approximation for later manual refinement.

Recently several algorithms for automatically defining weights have been introduced [1, 2, 3]. While the weight maps produced by these techniques reduce artifacts by definition (since the weights minimize the least squares error to a set of example poses), they rely on definitions of influence that are most appropriate for articulated characters. In particular, it is assumed that a vertex moves in the same general direction as the controller.

While this is most often the case, there are many important exceptions, particularly on the face, and if volume-conserving skin motion is desired. For example, in puckering the mouth to produce an "oo" sound, the lips move forward, while the corners of the mouth move inward at the same time. This motion is highly correlated (in a general sense), and influence regions for a sophisticated deformer should consider this effect. Unfortunately, a standard linear correlation measure would not detect this situation, since the motion of the mouth corner is essentially perpendicular to the lip motion, despite being clearly correlated in time.

In this paper, we introduce a measure suitable for automatically defining regions of influence for a so-

*current address: Weta Digital, Wellington NZ

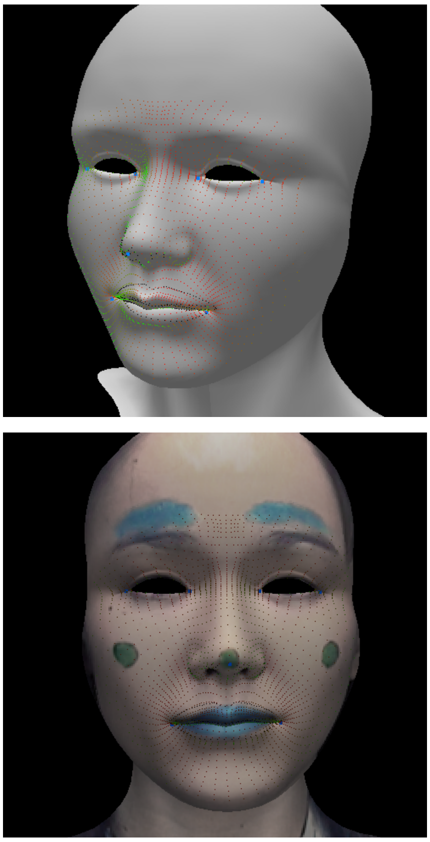

Figure 1: Model-based face tracking. Face mesh (top) and with projected texture (bottom).

phisticated deformer based on time-correlated motion. The relative motion may be in any direction, including perpendicular (as in the example just mentioned), or in even opposite directions (as may be the case with some creatures for the skin above and below the eyes in a “squint”). While this influence measure may be used with various skinning and deformation algorithms, the focus of the paper is on the measure itself.

2 Skinning Algorithms

The widely used linear blend skinning algorithm was developed in commercial animation software packages by the early 1990s (and has several names), although [4] introduced related ideas. In recent years skinning has become a focus of academic research. Examples include a body of research that largely addresses several well known artifacts in linear blend skinning (see [5, 6, 7] and references therein for a review) and example-based skinning approaches that allow artists to directly sculpt or digitize the desired skin deformation at various poses [3, 8, 9, 10, 11]), as well as approaches for automatically defining the weights in linear blend skinning, to be described below.

Linear blend skinning defines the final position of a surface vertex as a weighted combination of that vertex as rigidly transformed by several relevant surrounding skeletal frames. This may be notated

$$\hat{\mathbf{p}} = \sum_k w_k \mathbf{T}_k \, \mathbf{p} \tag{1}$$

where $\hat{\mathbf{p}}$ is the final position of the vertex and $\mathbf{T}_k$ are the transforms associated with several skeletal frames. The transform $\mathbf{T}_k$ can be expanded as

$$\mathbf{T}_k \equiv \mathbf{T}_k^{\delta} \mathbf{T}_k^{0^{-1}} \mathbf{T}_{\mathbf{p}}^{0}$$

where $\mathbf{T}_p^0$ is the transform from the surface containing $\mathbf{p}$ to the world coordinate system, $\mathbf{T}_k^0$ is the transform from the stationary skeletal frame k to the world system ($\mathbf{T}_k^{0^{-1}} \mathbf{T}_p^0$ together represent $\mathbf{p}$ in the coordinate system of skeletal frame k), and $\mathbf{T}_k^{\delta}$ expresses the moving skeletal frame k in the world system. As mentioned earlier, defining the weights w_k is a tedious process because of the large number of weights (NK weights for a model with N vertices and K bones) and the fact that the effect of an individual weight on the geometry is indirect. Further, the weights that produce the best shape in a particular pose are different than those for other poses, so the artist must find a compromise set of weights that produces the best result under the expected views and movement of the character.

Recently several authors have introduced automatic algorithms for obtaining the influence weights required in equation (1) [1, 2]. The idea in these approaches is to solve for the weights that cause the character skin to best match a set of given character skins in various poses, under the correct assumption that sculpting or digitizing the desired skin shape in a set of poses is an easier and more verifiable task than manually authoring influence weights.

These automatic weight algorithms require the definition of an *influence set* – the subset of bones that are relevant to the movement of each vertex and thus appear in the sum (1). In [1] the spread of transformed vertex positions relative to a particular bone is used as a measure of the influence of that bone, with a small spread indicating that the vertex is somewhat rigidly related to that bone. In [2] the best bones are those that when transforming a vertex result in the smallest sum-squared error over the animation relative to the desired vertex positions.

While linear blend skinning was developed for skeleton driven animation, the general approach has immediate extensions wherein the skeleton transforms are replaced by transforms defined by other influence objects. For example, the in-out breathing motion of a character’s torso can be effected by using at each vertex a defined fraction of a small animated scaling transform centered in the torso. In the case of marker-driven facial animation, the influence of each surrounding marker on a vertex can similarly be defined as a weighted sum of the marker movement (considered as a translation transform), and it has been argued that this leads to more natural and less rigid skin motion than is obtained with the competing blendshape approach [12]. Blend skinning is also often employed in conjunction with other approaches. In particular example-based skinning is often layered as a correction over an underlying blend skinning solution, to produce the required deformation in regions where the underlying blend skinning produces poor results (e.g. elbows) or the desired deformation is too complex to be reproduced using only blend skinning (e.g. muscle bulges, major wrinkles).

The influence set measures in [1, 2] are suited to most situations arising in skeletal animation, but a more flexible measure of influence is sometimes needed for layered and non-skeletal skinning applications. For example, in layering a sculpted example-based correction over a local region, it may be desirable to define the interpolation of the examples with respect to a small set of surface vertices or motion capture markers (since the surface is what is being corrected). The relevant set of vertices is those that have correlated movement. However, "correlated" should be understood in a general sense, in that there are cases where the x movement of one vertex may be strongly correlated with the y (rather than x) movement of another vertex. Facial motion exhibits several such examples, including the puckered-mouth "oo" motion mentioned earlier. In the next section we define an influence measure suitable for both these cases and the more conventional cases of parallel movement.

3 Improved Influence Measure

In defining an influence measure, correlation is the first approach that probably comes to mind. The correlation of two one dimensional signals is proportional to the inner product of the signals with their mean removed. Unfortunately it is not immediately obvious how to extend this to the three-dimensional spatial motion of two vertices over time. In addition, we hope to define points as being correlated even if the direction of their movement is different, provided the motion is strongly synchronized over time. Simply taking the magnitude of the spatial movement is also inadequate for our purposes, since it would allow movement in effectively random directions. We seek to identify movement that has a strong functional relationship, to serve as a reliable basis for example-based interpolation and other deformation approaches.

Consider maximizing the expected inner product of the movement of two points with respect to a rotation that aligns this movement,

$$\max_{\mathbf{O}} E[\mathbf{p}^T\mathbf{O}\mathbf{q}] = \frac{1}{N}\sum_t \mathbf{p}_t^T\mathbf{O}\mathbf{q}_t$$

where $\mathbf{p}_t, \mathbf{q}_t$ are mean-subtracted time series of two points under consideration and $\mathbf{O}$ is an orthogonal matrix that will rotate $\mathbf{q}$ to have maximum expected inner product with $\mathbf{p}$. (This problem has some resemblance to canonical correlation, although it differs in that the desired transform is restricted to be orthogonal).

The objective is then

$$\frac{1}{N}\sum_t \mathbf{p}_t^T\mathbf{O}\mathbf{q}_t + \operatorname{tr}\mathbf{L}(\mathbf{O}^T\mathbf{O} - \mathbf{I})$$

with a symmetric Lagrange multiplier $\mathbf{L}$ on the orthogonality constraint $\mathbf{O}^T\mathbf{O} = \mathbf{I}$.

The derivative is

$$\frac{1}{N}\sum \mathbf{p}_t\mathbf{q}_t^T + 2\mathbf{O}\mathbf{L} = 0$$

which we rewrite as

$$\mathbf{O}\mathbf{L} = \mathbf{M} = -\frac{1}{N}\sum \mathbf{p}_t\mathbf{q}_t^T$$

Here the right hand side matrix $\mathbf{M}$ is given, while the left hand side is the product of an orthogonal and a symmetric matrix. Thus we seek a decomposition of $\mathbf{M}$ into orthogonal and symmetric matrices.

A decomposition of this form is given by the polar or QS decomposition. Polar decomposition can be accomplished by singular value decomposition,

$$\mathbf{M} = \mathbf{U}\mathbf{D}\mathbf{V}^T = (\mathbf{U}\mathbf{V}^T)(\mathbf{V}\mathbf{D}\mathbf{V}^T)$$

where $\mathbf{U}\mathbf{V}^T$ is the desired rotation $\mathbf{O}$, and $\mathbf{V}\mathbf{D}\mathbf{V}^T$ is a symmetric positive semi-definite matrix.

As a simple example and verification of this technique, take $\mathbf{p} = [1,0,0]^T, \mathbf{q} = [0,1,0]^T$. With the singular value decomposition of $\mathbf{p}\mathbf{q}^T$, form the rotation matrix $\mathbf{O} = \mathbf{U}\mathbf{V}^T$. The product $\mathbf{O}\mathbf{q}$ is $[1,0,0]^T$, i.e., $\mathbf{q}$ is rotated to the direction of $\mathbf{p}$.

4 Results

Figs. 3-5 illustrate the influence measure applied to data from model-based face-tracking sessions (Figures 1,2). Fig. 3 shows the discovered influence region for a selected point near the mouth. Here the influence region is defined as those points whose motion has a normalized correlation of more

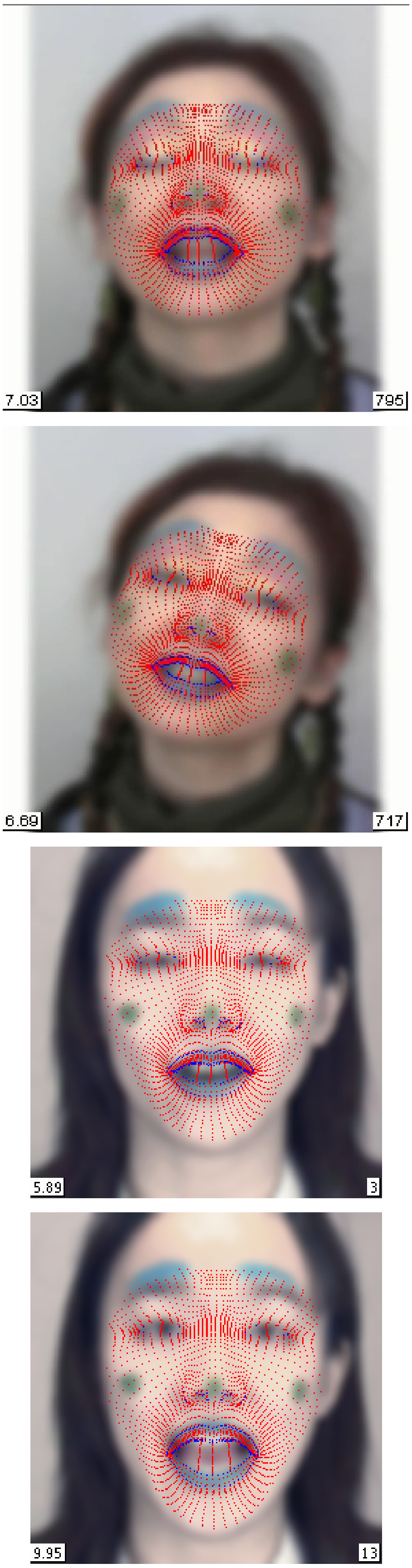

Figure 2: Face tracking performances. Overlaid points (red in the .pdf) are vertices of the aligned 3D model from Fig. 1.

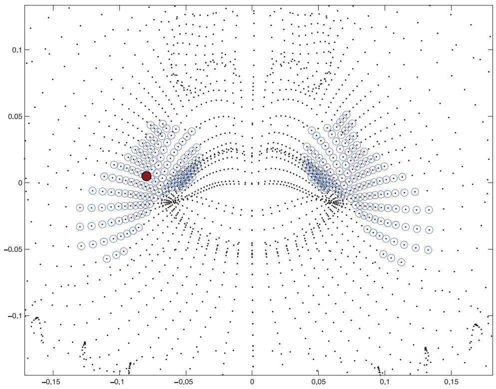

Figure 3: Vertices having more than .75 correlation coefficient with respect to the indicated point are circled. The largely symmetric mouth movement in this performance has been detected.

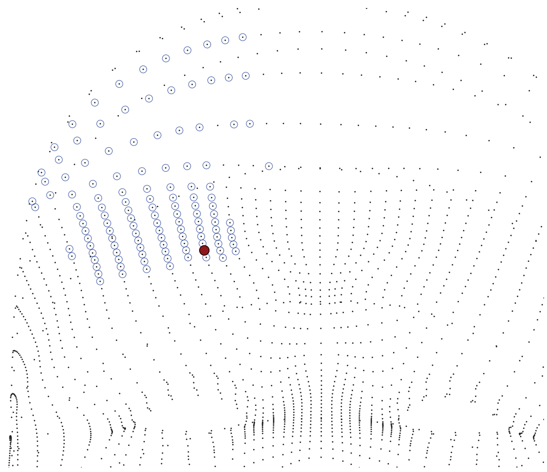

Figure 4: Vertices having more than .9 correlation coefficient with respect to the indicated point are circled. Enlarge to see details in the .pdf.

than .75 with the probe point after applying the aligning rotation **O**. Fig. 4 shows the discovered influence region for a selected point near the eyebrow. On this performance the influence region is not bilaterally symmetric. Fig. 5 shows the time history of a vertex near the mouth for which the aligning rotation significantly increases the discovered correlation.

5 Discussion

We have defined a linear measure of model point relatedness or influence. This measure reflects co-movement of the points but does not require that the movement directions be parallel. While this is only one (small) component of a full skinning system, automatic identification of potential influence regions has until recently been an open prob-

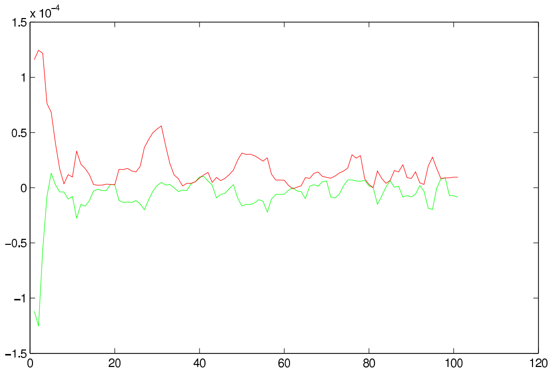

Figure 5: Graph of the inner product of the movement of two vertices over 100 frames. Light line (green in the electronic copy): before application of the aligning rotation from section 3. Darker line (red in the electronic copy): after the aligning rotation.

lem, requiring error-prone and sometimes extensive manual work. A further improvement would be to use mutual information rather than a correlation-based approach, as mutual information can capture any functional relationship rather than purely linear correlations.

6 Acknowledgments

This work was partly supported by the Japan Science and Technology Agency, CREST project.

References

[1] A. Mohr and M. Gleicher, "Building efficient, accurate character skins from examples," *ACM Transactions on Graphics*, vol. 22, no. 3, pp. 562–568, Jul. 2003.

[2] D. L. James and C. D. Twigg, "Skinning mesh animations," *ACM Trans. Graph.*, vol. 24, no. 3, pp. 399–407, 2005.

[3] T. Rhee, J. Lewis, and U. Neumann, "Real-time weighted pose-space deformation on the GPUs," *Computer Graphics Forum, Proc. of Eurographics 2006*, vol. 25, no. 3, pp. 439–448, 2006.

[4] N. Magnenat-Thalmann, R. Laperriere, and D. Thalmann, "Joint-dependent local deformations for hand animation and object grasping," in *Proc. Graphics Interface*, 1988, pp. 26–33.

[5] B. Merry, P. Marais, and J. Gain, "Animation space: A truly linear framework for character animation," *ACM Trans. Graph.*, vol. 25, no. 4, pp. 1400–1423, 2006.

[6] L. Kavan and J. Zara, "Spherical blend skinning: A real-time deformation of articulated models," in *2005 ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games*. ACM Press, April 2005, pp. 9–16.

[7] X. Yang and J. Zhang, "Realistic skeleton driven skin deformation," in *Technical Session on Computer Graphics and Geometric Modeling (TSCG)*. Springer, 2005.

[8] T. Kurihara and N. Miyata, "Modeling deformable human hands from medical images," in *SCA '04: Proceedings of the 2004 ACM SIGGRAPH/Eurographics Symposium on Computer Animation*. ACM, 2004, pp. 355–363.

[9] P. G. Kry, D. L. James, and D. K. Pai, "EigenSkin: Real time large deformation character skinning in hardware," in *Proceedings of the 2002 ACM SIGGRAPH Symposium on Computer Animation (SCA-02)*, S. N. Spencer, Ed. New York: ACM Press, Jul. 2002, pp. 153–160.

[10] P.-P. J. Sloan, C. F. Rose, and M. F. Cohen, "Shape by example," in *SI3D '01: Proceedings of the 2001 symposium on Interactive 3D graphics*. New York, NY, USA: ACM Press, 2001, pp. 135–143.

[11] J. P. Lewis, M. Cordner, and N. Fong, "Pose space deformation: A unified approach to shape interpolation a nd skeleton-driven deformation," in *Siggraph 2000, Computer Graphics Proceedings*, ser. Annual Conference Series, K. Akeley, Ed. ACM Press / ACM SIGGRAPH / Addison Wesley Longman, 2000.

[12] G. Borshukov, "personal communication," Electronic Arts, 2005.