Model Flowing: Capturing and tracking of deformable geometry

Kjell Reuterswärd* Linköping University
John Flynn† Digital Domain
Doug Roble‡ Digital Domain
Ken Museth§ Linköping University

1 Introduction

In this sketch we present a new markerless deformable model capture system which is more accurate and controllable than previous methods. In recent years there has been considerable interest in applying computer vision methods to capture the geometry of deformable objects [Zhang et al. 2004] such as human faces and cloth. For many applications, markerless methods are desirable since they allow the possibility of recovery of geometry, texture and lighting at the same time. Recent image-based methods employing optical flow have successfully been applied to facial animations in actual production[Borshukov et al. 2003].

We assume that the initial pose of the deforming object is known. Artists typically build the geometry for an object from a range scan of the object, computing the deformation for this geometry is ideal. After optical flow analysis of the image sequences from each camera, the 2D motion of each vertex can be used to triangulate the approximate movement of the vertex in 3D[Borshukov et al. 2003]. Unfortunately, since optical flow is an ill-posed problem and is computed independently for each image, this triangulation will likely have significant errors due to matching errors and smoothing in the original optical flow. We have modified the basic optical flow algorithm so that the movement of the 3D vertex is solved for directly and in all views simultaneously. This technique incorporates the epipolar constraint across the cameras, reducing the search space and resulting in higher accuracy and less drift.

2 Capture Setup

We use an array of synchronized and calibrated cameras that are placed so that there is considerable overlap in their views. Unlike some systems though, the cameras do not have to be color calibrated, static, or of the same type. This allows us to use a hero camera potentially on set.

3 Our Approach: Model Flowing

We directly solve for the global 3D deformation using all available image data. We call this ”model flowing”. The general strategy is to combine image information from all the camera-feeds into a single cost function which is then minimized to find the best model deformation. We assume the model for the first frame is known, and we are trying to solve for the vertex positions at the second frame. At each minimization step, for each camera, we synthesize the second frame image using the texture from the first frame, the model at the first frame and the current estimate of the model vertex positions at the second frame. The difference between the synthesized second frame and the actual second frame is used as the cost function. This cost is a function of the vertex positions at the second frame so reducing the cost function will deform the model.

Since our method directly incorporates the epipolar constraint, it reduces the ambiguity of optical flow based methods. This improves performance compared to these methods where the 3D solution essentially is a combination of many separate solutions in 2D.

*kjere757@student.liu.se
†jflynn@d2.com
‡doug@d2.com
§museth@acm.org

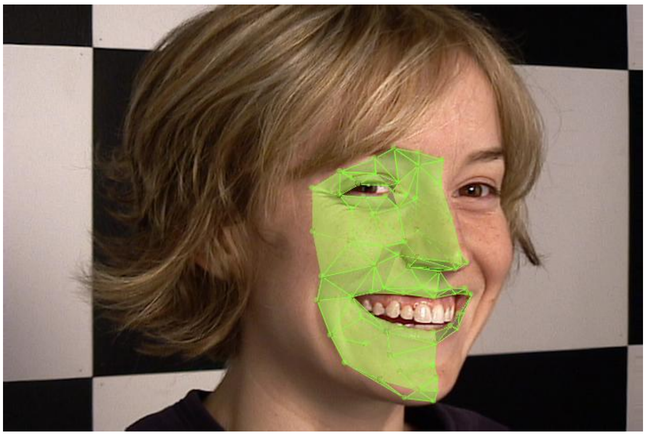

Figure 1: Deforming polygonal mask accurately tracking the subject’s face using Model Flowing from 3 camera views.

Care must be taken when performing the minimization, since there are a large number of parameters and a complex cost function, a naive approach would quickly become intractable. The Jacobian is sparse so we can use a sparse Levenburg-Marquardt method. Since analytic derivatives with respect to each parameter are not available, and numeric differencing would be prohibitively expensive, we approximate them as a function of the image derivatives in a method similar to that of [Lucas and Kanade 1981]. We have found that the addition of spring forces to the cost function greatly improves the stability. While we do not attempt to accurately model any physical system, these spring forces are a simple approximation to the elasticity of many deformable surfaces such as cloth and human skin. Our system also allows for locking of individual vertices, this way an artist can lock down certain areas of a face, and for example only track the movements of the mouth.

4 Extensions

Currently our method makes no attempt to model occlusion. The method could be extended to use the visibility from the current frame as an estimate of the visibility for the next frame. This would be accurate everywhere except at occluding boundaries. The robustness of the cost function could also be improved to handle incorrect matches in these areas. Our present implementation has constant spring forces everywhere in the model - it would be a very simple extension to allow an artist to specify different spring constants for different areas of the model.

References

BORSHUKOV, G., PIPONI, D., LARSEN, O., LEWIS, J. P., AND TEMPELAAR-LIETZ, C. 2003. Universal capture: image-based facial animation for ”the matrix reloaded”. In *Proceedings of the SIGGRAPH 2003 Conference on Sketches & Applications*, 1–1.

LUCAS, B., AND KANADE, T. 1981. An iterative image registration technique with an application to stereo vision. In *International Joint Conference on Artificial Intelligence ’81*, 674–679.

ZHANG, L., SNAVELY, N., CURLESS, B., AND SEITZ, S. 2004. Spacetime faces: high resolution capture for modelling and animation. In *In Proceedings of the 2004 SIGGRAPH Conference*, 548–558.