The Power of Texture:

A New Approach for

Surface Capture of the Human Hand

M. Ian Graham

Senior Honors Thesis Version 0.1
April 30, 2004

Thesis Advisor:
James Kuffner
(Robotics Institute, Carnegie Mellon University)

Computer Science Department
School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213

Abstract

Capturing the details of a moving surface with a system of cameras is a complex problem which has been receiving a great deal of attention in recent literature. Current methods for surface capture of human actors ("skin capture") do not yield results which are sufficiently realistic for use in the depiction of characters for computer animation. A skin capture system with the ability to produce an animated 3D model whose deformations are indistinguishable from the original actor's skin would enable animators to produce characters with realistic skin using less time and effort.

I propose a new system for surface capture of the human hand, a subject which has proven difficult to depict in a realistic manner for character animation in films and computer games. My approach will use structural knowledge of the human hand to very quickly estimate the position and shape of the surface of the skin and rapidly converge to an accurate representation of the current state of the actor's hand. Also, the method used is generalizable for use in capture of other articulated subjects whose structural knowledge is available.

Keywords: Surface Capture, Motion Capture, Model Reconstruction, Surface Reconstruction, Computer Animation, Computer Vision.

Contents

1 Introduction

Computer-based motion capture is a technology with a history of over twenty years stretching back to the early 80s and the first successful attempts to record and analyze human motion with computers. [Sturman] Since its creation it has grown into a versatile and well-used set of algorithms and tools utilized in medicine, sports, and visualization in general, but especially in entertainment. Motion capture systems observe a moving real-life character and record joint angles for the subject, allowing for reproduction of the motion given a kinematic model of the character. The utility of this technology in the context of entertainment is immediately obvious: When a character model for a movie or game must be animated in a believable way, it can be moved using data recorded from a real-life subject, eliminating the need for artists to painstakingly configure animation keyframes by hand. Optical motion capture, the most prevalent method in use today, uses a network of cameras to observe specially-designed markers placed on the subject. These markers convey information about joint locations from which it is relatively easy to extract joint angle information.

A problem with current motion capture systems is that even perfect reproduction of the changes in joint angles over time does not always lead to a well-animated model. The reason for this is that current capabilities to determine a character's overall appearance (including surface shape) from joint angles are severely limited. For example, knowing that a human's elbow is bent at 90 degrees will let one position the upper and lower arm relative to each other, but this does not provide rich information about surface deformation caused by muscles tensing or relaxing and general movement of the skin. "Soft-skinning" techniques attempt to increase the realism of deformations by applying vertex blending to animation of the character's skin. This allows surface points to be influenced by multiple joint vertices, and typically tools are provided so that artists can easy specify the vertex weights and decide how a given point will move as a function of any or all joint positions. This produces results which are more pleasing to the eye, but can add a large amount of overhead and repeated guesswork for the artist. [Stokdyk *et al.*]

Surface capture of moving subjects is a much newer technology which represents a possible solution to the problems outlined above. The objective of

surface capture is similar to that of motion capture—a moving subject is observed and data recorded over time. However, in the case of surface capture, the goal is to record not merely joint angles, but the shape and change in shape over time of the surface of the subject. The ability to determine these deformations as a function of time would allow for direct reproduction of the surface itself, rather than having to attempt to reproduce it algorithmically or through manual configuration by artists. Current surface capture systems for moving characters, or "skin capture" systems, are entirely optical (camera-based), due to the simple fact that attaching a significant apparatus to the subject tends to change the shape of the skin.

Until very recently, the use of a surface capture system to produce an animated model was unfeasible due to the complexity of such systems, and the use of one for animation in entertainment was unheard of. In [Sand *et al.*], a deformable model of an individual's body is created by using silhouettes to observe deformations for various joint configurations and using interpolation to produce deformations for new sets of joint angles. This allows for the animation of a character's overall body motion through new poses in a manner which generates

rough but fairly realistic-looking movement of the skin of the trunk and limbs.

The human hand is one of the most difficult subjects for realistic computer animation due to elaborate articulation and large amounts of occlusion when attempting to capture real-world data optically. It is quite possible (and fairly common practice) to perform traditional motion capture of the hand using cameras placed densely around a smaller capture area. The joint angle information thus retrieved may be used to position a rigid or soft-skinned model with accurate joint configuration, but there are disadvantages to animating a hand in this manner. Firstly, even reduced-size motion capture markers made specifically for hand capture restrict the motion of the hand, and there are several natural configurations which cannot be performed for fear of disrupting markers—a tight fist is one of these. Secondly, the surface deformations of the hand are extremely elaborate and involve flesh which moves and wrinkles as well as bone and ligament structure which can be plainly visible through the skin. Therefore, even soft-skinning will produce deformations which are very obviously unrealistic. The creation of a model from silhouettes is also prohibitive due to the extreme amounts of occlusion and overlap for many

configurations of the hand.

There is currently no versatile method to quickly and easily capture and reproduce both the motion and deformation of the human hand.

2 Conception

2.1 Objective

The goal of my work is the design of a surface capture system specialized to handle the unique challenges offered by the human hand as a subject, yet still generalizable for use with other subjects. It should provide a method to observe the hand of a real-life actor, record data from its motion, and then reproduce an animated surface which matches the original's structure as closely as possible.

Challenges specific to surface capture of the hand include:

- Heavy articulation
- Large amounts of occlusion
- Elaborate skin deformation which is not easily reproduced algorithmically,

due to:

- Bunching and wrinkling of the flesh of the palm
- Greater concentration of flesh on the front of the hand than on the back, resulting in different deformation properties for each
- Skeletal structure which often directly deforms the skin, including knuckles and tendons

My goal, then, is to design a system which overcomes these problems and records accurate surface deformations of the human hand over time, while simultaneously being easy to use and having little required calibration time.

2.2 Initial Concerns

Using an optical capture system is an immediately obvious choice given the context; any substantial apparatus will change the way the hand deforms, and therefore it is ideal to observe the subject with cameras.

However, rather than attempting to solve a pure computer vision problem and

trying to extract as much data as possible from raw images of an unaltered human hand in an arbitrary scene, a surface capture system has the freedom to customize its capturing conditions just as motion capture does, to enable easy extraction of as much information as possible from the scene. Though it is generally straightforward to segment human skin from the rest of an image using red and green chromaticity [Fritsch *et al.*], controlling the lighting of the scene and providing a darkened background, for example, will make segmentation much easier.

An immediate observation is that attaching motion capture-style markers to the hand is prohibitive for surface capture. Though markers in a full-body context are relatively unobtrusive, due to the size of even hand-specific markers and the increased number of placements required by the heavy articulation of the hand, significant modification of surface structure would result. Even if the new surface introduced by the markers themselves might be intelligently removed, they still obstruct the skin beneath and reduce the amount of surface area which may be analyzed. Therefore, it becomes necessary to devise another method of communicating extra information about the structure of the hand.

2.3 "Real-life Texture Customization"

Motion capture markers provide information about the positions of the joints of the hand. The ideal way to duplicate this marker functionality without changing the shape of the hand is to somehow customize only the hand's texture (color). If the real-life texture of the hand may be modified, then marks which function identically to motion capture markers could be placed easily at relevant locations. These could be used to approximate joint angles as in a motion capture system and retrieve information about the configuration of the hand's skeletal structure as well as the size of all parts of the hand if this information is needed for the subject.

Additionally, the ability to modify the hand's texture in any way would enable the transmission of even more additional information through images captured by cameras. One way in which this could help the process of surface capture a great deal is by transmitting per-pixel locality information, or communicating via color information about what part of the hand a given pixel represents. There are many ways to go about this, but perhaps the simplest and most

straightforward is to assign a solid color to each finger of the hand. Using this method, one could, for example, specify that the thumb and only the thumb will be colored solid red. On seeing a red pixel, an algorithm could immediately know that it was part of the thumb. Specific methods for utilizing this information will be described below.

How, then, can the texture of the hand be changed? Two methods, each with particular advantages and disadvantages, are the use of (washable) paint to color the skin and the use of a thin glove. A glove offers the advantage of quick application and reusability. However, obtaining a proper fit for a specific hand is a concern, and avoiding significant modification of surface characteristics is the biggest problem. Even tight-fitting latex gloves can change the deformations of skin by generating new and different wrinkles at bent joints. Therefore, a glove should be made of an even less-obtrusive material such as that of a sheer stocking. The use of such a material, however introduces new problems including limited ability to change surface color and minimal durability.

If a water-based paint is used, painting the hand avoids the problem of changing the deformations of the skin, and is completely unobtrusive if the paint used will not smear after drying. However, with paint, application becomes a problem—how can a specific texture be applied quickly and consistently to a subject? If only making marks in a pattern similar to motion capture markers application is easy, but if an arbitrary texture must be applied then time consumption and precision of application are both large issues. A specialized appliance for automated paint application might help with this, but would likely be expensive to produce. Because of the absence of serious issues apart from application time, a paint-based approach was decided on for early development.

2.4 A Two-Phase Algorithm

Considering the two pieces of information which may be delivered by a customized hand texture as outlined above, a simple two-phase algorithm presents itself which operates on image data from multiple calibrated cameras:

- Phase 1: Quick Approximation with "Markers"
  - Use the painted "markers" to perform motion capture-style determination of joint angles and digit lengths.

- Use this information to configure a generic CG hand model. First digits should be resized to appropriate lengths, then joints bent to get a rough synthetic approximation of the real hand's configuration.

- Phase 2: Detailed Fitting with Locality Information
  - Now that the digit lengths and pose of the synthetic model are accurate, modify the shape of the surface of the model and "fine-tune" it into correspondence with reality.
  - Use locality information to perform error reduction methods on isolated parts of the model.

2.5 Details of Phase 1

The first phase of the algorithm is relatively straightforward as it largely follows the same process as standard motion capture, merely using painted marks rather than markers. This process can be done through standard images captured by camera and completed by entirely new software, or could even be separated from the rest of the algorithm and performed by a commercial motion capture system facilitated by the use infrared-reflective paint.

The generic hand model prescribed in the description of this phase should ideally be similar to CG hand models which are used today to render hands based on motion capture data. It should include both a kinematic model (to aid in the determination of digit lengths and joint configuration) and a vertex-blended surface which deforms given a joint configuration. The resultant surface will have all the problems associated with common soft-skinned hand models as described in the Introduction, but will serve as an early approximation of the hand's surface. From this model, a direct representation of the surface itself, such as a sampling of points for a NURBS surface, may be acquired for use in Phase 2.

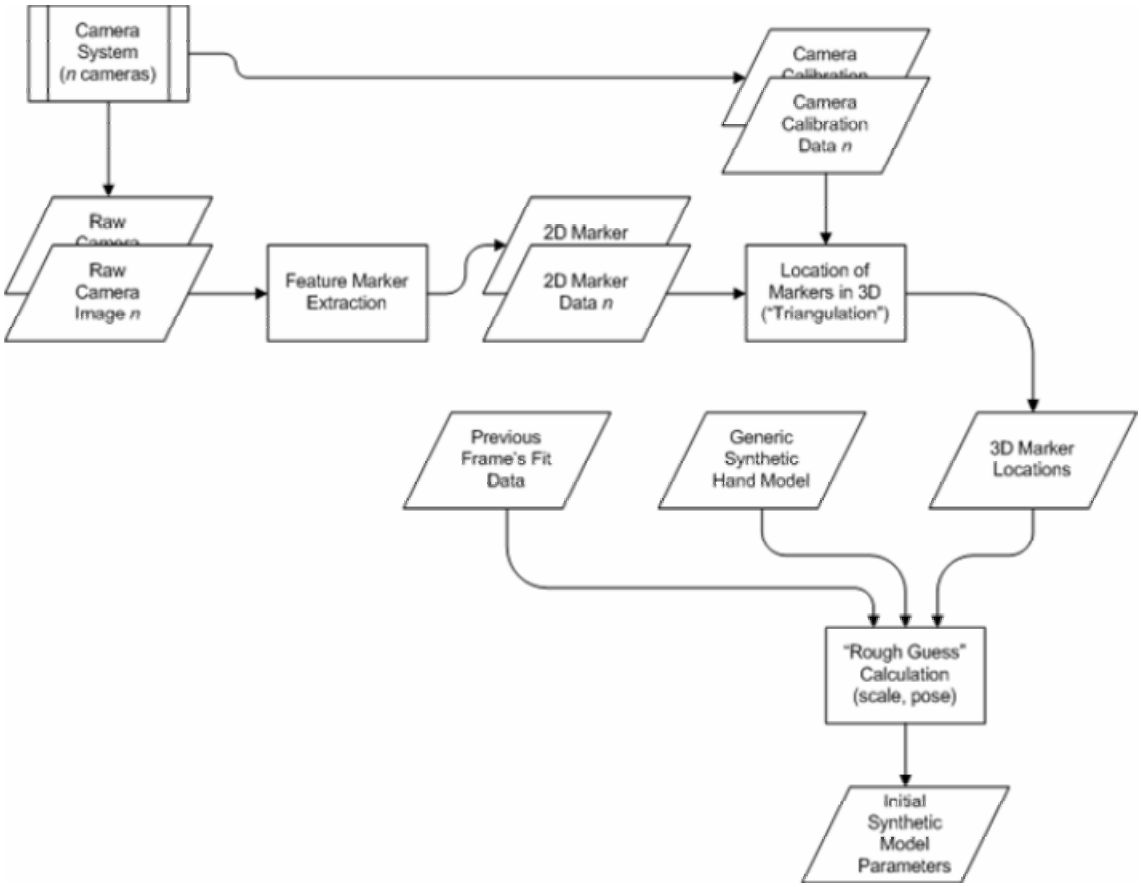

Figure 1: Flowchart of Phase 1

2.6 Details of Phase 2

Phase 2 represents the bulk of the work performed in the algorithm. Phase 1 has given a surface which is an approximation of the real hand, and though the general shape of the hand is correct, the details of the surface itself are not. How can these be changed to match the real-world hand? The most straightforward idea is to establish an error metric and reduce this error as much as possible. There are a number of ways to go about this, including

several possible error metrics and several possible error reduction methods. First, some error metrics:

- **Marker locations:** Markers are accounted for in Phase 1, but to make sure that joint configuration stays consistent during error reduction, the use of difference in marker location as part of the error measure would be helpful.
- **Surface contours(silhouettes):** As shown in [Sand *et al.*] and many other works, this can be an excellent indicator of surface detail, particularly when multiple camera views are involved. However, the comparison of the contour from a 2D image with a 3D synthetic model is not straightforward, as it becomes necessary to either generate a contour from the synthetic model or to project the real hand's contour into 3D for direct comparison with the synthetic model.
- **3D surface points:** If the cameras observing the subject are set up properly, stereo vision might be used to generate 3D surface points of the real hand which could be straightforwardly compared with those of the synthetic model.

Numerous error methods could be used, and given the approaches described in the next section even a straightforward gradient descent method has a good chance of converging quickly because of the relative proximity of the real surface data to the rough synthetic approximation generated in Phase 1.

2.7 Using Locality Information

The true power of the texture-provided locality information which was proposed in section 2.3 is the potential to dramatically reduce the amount of work necessary during error reduction. Stated more accurately, being able to localize areas of both the synthetic model and the real model and determine correspondence between the two allows for reduction on isolated parts of the model rather than over the entire set of surface points. This makes error reduction on surface contours particularly powerful—if specific digits are identified by distinct colors, then minimum errors of single fingers may be determined without worrying about the rest of the hand.

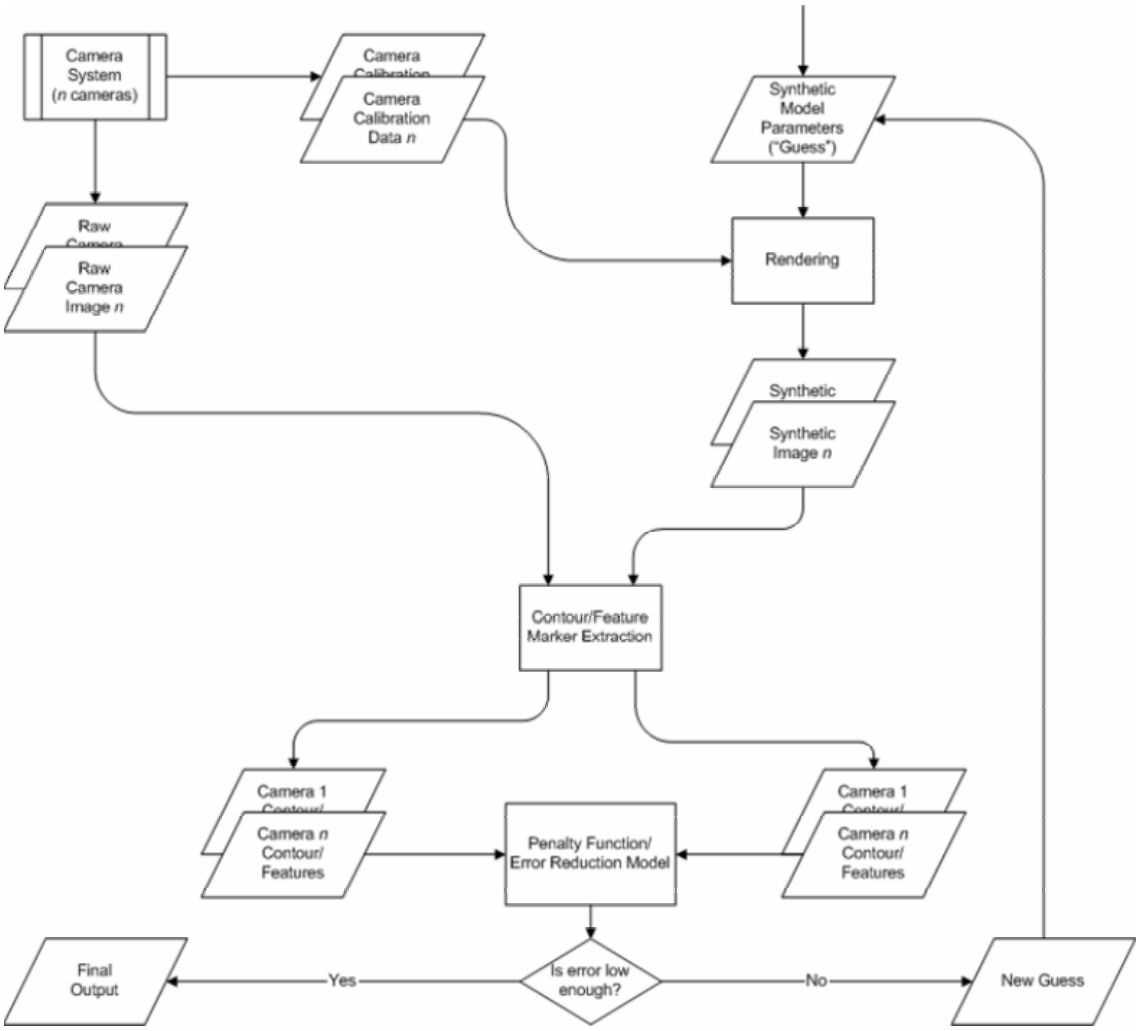

Figure 2: Flowchart of Phase 2

3 Discussion

Recently there has been a good deal of work in reconstructing different human body types as functions of reduced parameter sets such as height and weight. [Allen, *et al.*] If similar work were applied to the hand, the accuracy of the approximation generated by Phase 1 here could be greatly increased, which has

the potential to speed up convergence in Phase 2 by a great deal.

Development of a prototype system is currently in progress. During development, Phase 2 itself has been greatly aided by the locality information provided by a custom texture, as this allows for the error reduction process to be greatly accelerated. However, the error reduction loop as seen in Figure 2 still requires a good deal of work to calculate the error metrics used to compare the synthetic model with the real-world data. Improvements in this process, as well as further analysis of the effectiveness of different error reduction methods, are necessary to streamline the performance of this algorithm.

Another possibility for improvement may come from machine learning techniques: Intelligent feature selection [Kira, *et al.*] might eliminate the need to consider many of the surface parameters and lead to quick convergence towards a visually-pleasing surface. Exploration of this possibility will be an objective of future research.

4 Conclusion

The surface capture system proposed here allows for versatile capture of the human hand and use of the models produced in computer animation. Its compensation for the large amount of articulation of the hand using the aid of a customized hand texture lets it straightforwardly resolve the ambiguity present in everyday images of the human hand. At the same time, this algorithm may be generalized for use with any subject for which a synthetic model (including skeletal and kinetic information, plus a rough surface) is available. It should function well for subjects which have high articulation or high levels of occlusion and may have a custom texture applied via one of the methods outlined here. There remain definite performance issues in the execution of the algorithm, and further research into quick error reduction methods and other ways to reduce computation in the detailed fitting process should yield an approach with the potential to become a standard similar to the motion capture systems of the present day.

References

Sturman, David. *A Brief History of Motion Capture for Computer Character Animation*. SIGGRAPH 1994, Course 9.

Stokdyk, S., Hahn, K., Nofz, P., Anderson, G. *Spiderman: Behind the mask*. Special Session of SIGGRAPH 2002.

Sand, P., McMillan L., Popovic, J. *Continuous Capture of Skin Deformation*. ACM Transactions on Graphics, Vol. 22, No. 3, (Proceedings of SIGGRAPH 2003, San Diego, CA, July 27-31), pages 578-586.

Fritsch, J., Lang, S., Kleinehagenbrock, M., Fink, G.A., Sagerer, G. *Improving Adaptive Skin Color Segmentation by Incorporating Results from Face Detection*. Proceedings of IEEE Int. Workshop on Robot and Human Interactive Communication (ROMAN), Berlin, Germany, September 2002.

Allen, B., Curless, B., Popovic, Z. *The space of all body shapes: reconstruction and parameterization from range scans*. ACM Transactions on Graphics (Proceedings of SIGGRAPH 2003, San Diego, CA, July 27-31).

Kira, K., Renedell, L. *A practical approach to feature selection*. Proceedings of the Ninth International Conference on Machine Learning. Aberdeen Scotland, 1992.