ESTIMATION OF THE ACCURACY OF A SHAPE-FROM-SILHOUETTE
MARKERLESS MOTION CAPTURE SYSTEM

Lars Mündermann1, Ajit Chaudhari1, Gene Alexander1, Thomas P. Andriacchi1,2,3

1Division of Biomechanical Engineering, Stanford University, Stanford, CA
2Bone and Joint Center, Palo Alto VA, Palo Alto, CA
3Department of Orthopedic Surgery, Stanford University Medical Center, Stanford, CA
lmuender@stanford.edu

INTRODUCTION

The most common methods for accurate capture of three-dimensional human motion require a laboratory environment and the attachment of markers or fixtures to the body segments. These laboratory conditions can cause adaptive changes to normal patterns of locomotion that introduce experimental artifact. For example, it has been shown that attaching straps to the thigh or shank alters joint kinematics and kinetics (Fisher 2003). Thus our understanding of normal and pathological human movement would be enhanced by a method that allows capture of human movement without the constraint of fixtures or markers placed on the body. Markerless methods are not widely available because the accurate capture of human motion without markers is technically challenging. A recently reported method in the computer vision literature using shape-from-silhouette (SFS) reconstruction offers an attractive approach (Cheung 2003). However, the factors influencing the accuracy of this method have not been addressed. The purpose of this study was to evaluate the accuracy of SFS reconstruction of a body for different camera placements, number of cameras, camera resolution and object contours.

METHODS

The analysis was conducted in a virtual environment using a realistic human form generated with a 3D laser scanner (CyberScanner, Monterey, CA) (92461 vertices; average 3D inter-point distance 4±1mm). Camera placement for SFS was evaluated using 16 virtual snapshots simulating 640×480 cameras, each with a field of view of 80°. The uniform arrangement of the 16 cameras approximated a hemisphere with each camera fully capturing a 3×2×1.5m viewing volume and directed towards the center of this volume.

Figure 1: SFS reconstruction using four cameras.

SFS 3D point clouds were constructed from points along rays that extend from the camera positions through points along the corresponding silhouettes (Figure 1) choosing 8 out of the 16 total cameras, resulting in 12870 combinations. Accuracy was evaluated by calculating the percentage of vertices of the original mesh with a shortest distance to the point cloud smaller than a threshold value. Subsequently, the accuracy of SFS based on the number of cameras (4, 8, 16) and resolution of the camera (320×240, 640×480, 1280×960) was evaluated using the optimal camera placement. In each case, one camera was placed above the object while the remaining cameras were placed uniformly around the object in a circular plane at height 1m with one camera directly in front of the object. All cameras
were directed towards the center of the rectangular viewing volume. The SFS point cloud density was determined by the distance between points along silhouettes and the spacing along the rays. The distance between neighboring points along silhouettes corresponded to pixel distance, and the distance of neighboring points along rays was set to 4mm (approximately the inter-point distance of the original mesh). SFS 3D point clouds were evaluated by calculating the shortest distances to the original mesh using Geomagic Studio (Raindrop Geomagic, NC).

RESULTS AND DISCUSSION
SFS reconstruction was sensitive to camera placement. The most accurate reconstruction based on shortest distance between the original mesh and SFS point clouds was obtained by placing one camera directly above and all others at a height of 1m, uniformly distributed in a circle around the object. Cameras with higher resolutions decreased the standard deviation (Figure 2a,b,c; Table 1). The number of points in the point clouds varied approximately proportional to the resolution. Similar results were obtained when all silhouettes were resampled with a fixed point distance of 4mm, resulting in similar numbers of points in the point clouds. While significant improvements (P < .001) were obtained, the relative changes were minor, suggesting that a 320×240 resolution may be sufficient to capture the body surface. Greater reduction in error occurred by increasing the number of cameras (Figure 2d,e,f; Table 1; P < .001). These results suggest that increasing the number of cameras is more effective than improving camera resolution. In general, SFS reconstructions were not able to capture surface depressions such as eye sockets and lack accuracy in narrow spaces (Figure 2f). For instance, the arm pits and groin region yield greater deviations than other regions of the body. Nevertheless, most of lower and upper limb surfaces were reconstructed very accurately with 8 or more cameras. This study demonstrates the feasibility of developing an accurate markerless motion capture system on the basis of SFS.

Figure 2: Absolute shortest distance between original mesh and SFS point clouds dependent on camera resolution (a) 320×240, (b) 640×480, (c) 1280×960 for 8 cameras and number of cameras (d) 4, (e) 8, (f) 16 for 640×480 cameras.

Table 1: Standard deviation and maximum distance between the original mesh and SFS point clouds and number of points in SFS point clouds dependent on camera resolution (320×240, 640×480, 1280×960) for 8 cameras and number of cameras (4, 8, 16) for 640×480 cameras.

<table>
<thead>
<tr>
<th>Camera Resolution</th>
<th>SD [mm]</th>
<th>Max [mm]</th>
<th>Points</th>
</tr>
</thead>
<tbody>
<tr>
<td>320×240</td>
<td></td>
<td>11.4</td>
<td>54.6</td>
</tr>
<tr>
<td>640×480</td>
<td>9.3</td>
<td>58.6</td>
<td>76381</td>
</tr>
<tr>
<td>1280×960</td>
<td>8.8</td>
<td>56.9</td>
<td>161276</td>
</tr>
<tr>
<td>4</td>
<td>19.8</td>
<td>91.0</td>
<td>92735</td>
</tr>
<tr>
<td>8</td>
<td>9.3</td>
<td>58.6</td>
<td>76381</td>
</tr>
<tr>
<td>16</td>
<td>6.3</td>
<td>40.1</td>
<td>71706</td>
</tr>
</tbody>
</table>

REFERENCES

ACKNOWLEDGEMENTS
Funding provided by NSF #03225715 and VA #ADR0001129. The authors gratefully acknowledge valuable inputs by Stefano Corraza.