IMPROVING THE ANALYSIS OF HUMAN MOVEMENT USING MARKERLESS MOTION CAPTURE

Eugene J. Alexander[1] and Thomas P. Andriacchi[1,2,3]

Department of Mechanical Engineering, Stanford University[1]   Department of Orthopedics, Stanford University[2]

VA Palo Alto RR&D Center[3]

INTRODUCTION

Motion capture is an integral part of musculoskeletal biomechanics, generating critical kinematic information and providing a reality check on forward dynamic simulations. There is a need for improved methods to quantify human movement to advance our understanding of musculoskeletal function. In the past the techniques of motion capture have included goniometers, magnetic devices, active and passive markers. Each of these techniques has drawbacks, either in experimental complexity, limitations on the types and locations of activities, or in encumbering a subject. Recent research has focused on the acquisition of motion capture data strictly from video observations of the subject. This purpose of this paper is to describe the development of a new technique for markerless motion capture that uses a richly textured spandex body suit, a hierarchy of models, and a modification of the soft assign/deterministic annealing point matching algorithm.

METHODOLOGY

Hardware.

Video motion was captured using commercial, color digital video cameras with commercially available camera calibration software (1) is used to calculate the distortion, internal, and external camera calibration parameters. The subjects wear a spandex body suit (Figure 1). The area of interest here is the left leg, which is of a distinct blue color and densely covered in stars.

Algorithms

The overall philosophy of this approach was to build a 3D template that may require a lot of user intervention, then specialize that 3D template to each individual subject in a way that may require some user intervention, in order to have little or no user interaction during the bulk of the data processing. The 3D template is used to facilitate building the subject specific model. The subject specific model is used to track the subject during the various activities. In all cases, after the video data has been dewarped, the stars must be extracted from the video data.

Star Extraction Extracting the stars from the video data is done using the color information . The distinctive blue color of the left leg is use to segment the limb from the background. A watershed

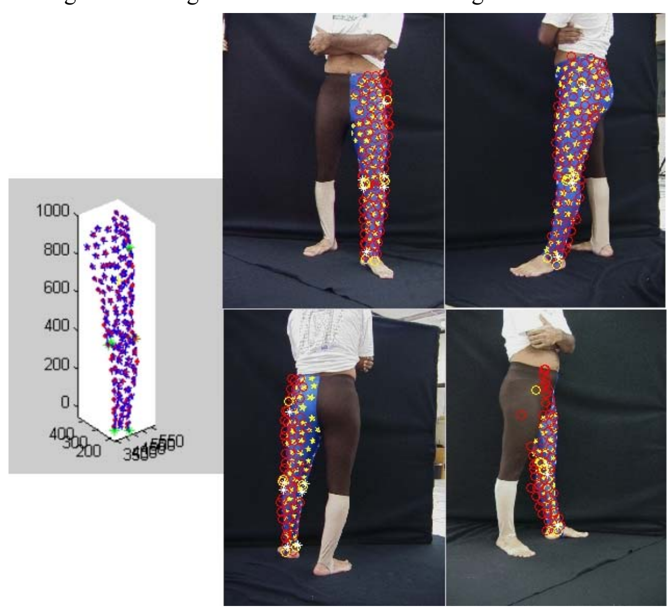

Figure 1. Subject wearing the spandex suit, viewed from four camera locations. Circles are the camera plane projection of the 3D model, shown on the left.

style segmentation algorithm is then used within the perimeter of the leg to find the individual stars. Then an edge detection algorithm with a star shaped structuring element is used to find the edges and then vertices of the stars.

3D template. The 3D template is built once for each suit. Data is acquired with a mannequin wearing the spandex suit. With multiple camera observations, the stars are extracted as above. Working in camera pairs, the user is prompted to select a number (>3) of points from each camera. These points are used to calculate a projective transform from one camera to the next. A simple nearest neighbor algorithm is then used to correspond all of the associated points. The points that are seen by more than one camera are then used to calculate 3D locations of stars. This 3D data is stored along with a spline fit through the data.

Subject Specific Model. The subject specific model building software is then run. Data is acquired with the subject standing in the prescribed reference pose. The star extraction software is applied. The user is then prompted to initialize the location of the 3D template in 3D in such a pose as to bring the back projection of the 3D template roughly into alignment with the subject data (Figure 1). The general tracking algorithm (described below) is then applied to affinely deform the 3D template to fit the subject data. After the user confirms that the model is in fairly good alignment, a spline deformation algorithm is applied that allows the 3D template to deform to match the subject data. Once the user confirms that this deformation is correct (implying a subject specific model of 3D locations of stars, vertices, and spline) the model is stored.

Data tracking. For each of the data sets, the operator initializes the location of the subject specific model at one time step and the software performs all other needed operations. The initialization is similar to that described in the subject specific model generation. For the first few frames about the initialization frame, the general tracking algorithm is applied. This algorithm is an adaptation of the deterministic annealing/softassign approach for point matching of Gold et al. (2). This approach simultaneously estimates the correspondence between point sets and the transformation between the point sets. As developed the algorithm works on matching 3D to 3D or 2D to 2D point sets; in this work the algorithm has been modified to match 3D point sets to their 2D projections, and to include multiple projections at once. This was done by including the camera projection matrix in the algorithm and by extending the correspondence matrix to multiple cameras. The objective funtion to be minimized is then

$$E_{3D-2D} = \sum_{c=1}^{C} \sum_{j=1}^{J_c} \sum_{k=1}^{K} m_{c,j,k} * \left\| X_{c,j} - P_c(A * Y_k) \right\|^2 \quad (1)$$

where m is the correspondence matirx weighting factor, X is the camera plane data, Y is the previous 3D location of the model, A is the affine transformation, P is the world to camera plane projection function, C is the number of cameras, Jc is the number of features in that camera plane, and K is the number of features on the 3D model.

For all of the other frames, after an initial track has been established, the inter-frame 2D to 2D camera data is used to provide an initial estimate of camera correspondences. This is done by applying the 2D to 2D point matching algorithm. As the star labels are known in the current frame, if they correspond to a 2D point in the next frame that star will then be labeled. For the 2D points that are unambiguously labeled, some 3D points can be calculated, providing an initial estimate of the pose in the new frame. This estimate is then used in the general tracking algorithm, greatly speeding convergence.

After the model has been matched with affine transforms, the spline fit of the data is deformed to exactly match the data points. In this manner the shape of the limb segments are allowed to freely deform. Once the model has been driven through the entire video sequence, those points that have been seen by less than two cameras are discarded. The point clouds can then be used to estimate limb segment pose, using any of a number of algorithms (3-5).

RESULTS

A sample result is shown in Figure 2. The subject data is shown as outlines of the stars. The white boxes are the back projection of the 3D subject specific model onto the camera plane. In almost all cases the data back projects correctly; in those cases where a star is only observed by one camera, it back projects correctly in that camera plane but may show up out of place in another camera. This data is discarded prior to limb segment pose estimation.

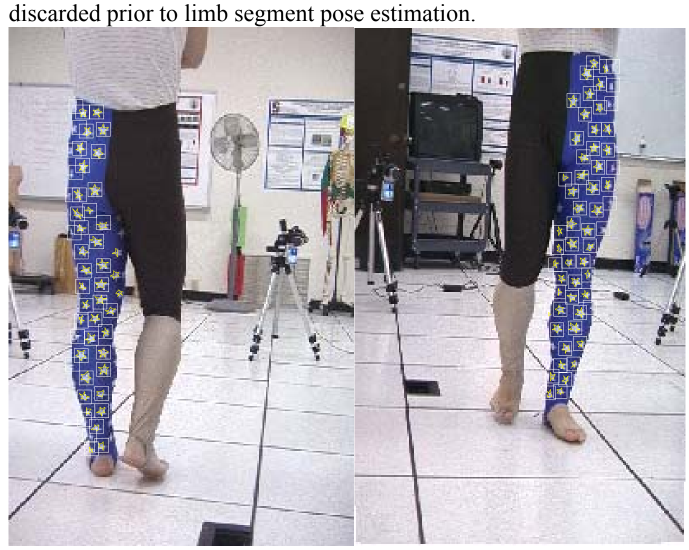

Figure 2. Results from two of six camera views. The subject data is shown in the star outlines. The back projection onto the camera planes are shown as squares.

DISCUSSION

This research has demonstrated that by modifying the softassign algorithm to work on 2D projections of 3D point sets and using the rich feature sets of the spandex body suits, markerless motion capture is practical. One significant advantage of this technique as compared to standard markered motion capture can be seen immediately: there are roughly 150 stars on one limb, and with 10 vertices per star that leads to 1500 points that are tracked for limb segment pose estimation. This is two orders of magnitude more data than is available for the most sophisticated markered motion capture (3,4).

REFERENCES

1. Bouguet, J.Y.; "Camera Calibration Toolbox for Matlab", http://www.vision.caltech.edu/bouguetj/calib_doc/.
2. Gold S., Lu C.P., et al. 1995, "New Algorithms for 2D and 3D Point Matching: Pose Estimation and Correspondence", Advances in Neural Information Processing Systems , 957 – 964.
3. Andriacchi, T.P., Alexander, E.J., et. al ., 1998, "A Point Cluster Method for In Vivo Motion Analysis: Applied to a Study of Knee Kinematics", J. Biomech Eng, 120(12) 743-749,.
4. Cappello, A., Cappozzo, A., et al. 1997, "Multiple anatomical landmark calibration for optimal bone pose estimation", Hum Mov Sci 16(2-3) 259-274.
5. Spoor, C.W., Veldpaus, F.E. 1980, "Rigid Body Motion Calculated from Spatial Coordinates of Markers", J Biomech 13, 391-393.

ACKNOWLEDGEMENTS

This work was supported in part by grants from the Defense Advanced Research Projects Agency, the National Science Foundation, and the National Institutes of Health.