MVA2007 IAPR Conference on Machine Vision Applications, May 16-18, 2007, Tokyo, JAPAN

Assessment of Post-Stroke Functioning using Machine Vision

Sonya Allin
Carnegie Mellon University
Pittsburgh, PA

Deva Ramanan
Toyota Technical Institute
Chicago, IL

Abstract

We present a system to automatically assess the functional performance of stroke survivors along axes defined by the Arm Motor Ability Test (AMAT). The upper body motion of seven stroke survivors was measured in a laboratory environment using a commercial motion capture device and a novel kinematic tracker of our design. Statistics generated by each individual were related to expert-determined assessments of functional health. Results indicate several kinematic targets that correlate with and predict opinions of health. These include recorded motion of the torso during the performance of tasks and flexion about the elbow on the hemiparetic (impaired) side. We show that both kinematic statistics can be cheaply and robustly measured with video cameras while still preserving their diagnostic value. Such cheap and robust measurements will ultimately facilitate assessment outside of clinics and in places where functioning is valuable to individuals, such as homes and workplaces.

1 Introduction

Stroke is one of the leading causes of chronic disability in the United States. Evidence increasingly suggests that individuals who have been disabled by stroke can achieve functional motor improvements for many months, if not years, after their injury [14]. Occupational and physical therapy, however, is not commonly covered by U.S. insurance for the length of time that post stroke recovery is possible. In this paper, we suggest the use of cheap and robust monitoring technologies to help fill gaps in therapeutic care. More specifically, we explore the capacity for robust kinematic tracking tools to extract statistics of upper body motion that correlate with and predict expert opinions of functional health after a stroke. Our motivating vision is one in which these clinically meaningful statistics can be cheaply and robustly recorded, thereby enabling retrospective reporting about motions that take place in the real world.

Perhaps the main contribution of this paper to the field of computer vision is its use of clinical metrics to evaluate the quality of kinematic tracking. Recent years have seen many instances in which kinematic reconstructions, be they based on single or multi-view camera input, are numerically judged relative to motion capture data [13]. Evaluation criteria include squared error in 3D point reconstruction, joint angle error, or false-positive/missed-detection rates based on the overlap between estimated body configurations and configurations provided by motion capture. In the many cases where motion capture is not available, numeric evaluations have been based on comparisons with kinematic estimates or re-projection error [7]. Even given ground-truth motion capture data, however, the right evaluation metric is not clear [5]. We circumvent these issues altogether. In our case, a good kinematic reconstruction is one that can relate well to and predict a clinician's functional score.

To develop our system we used motion capture to identify kinematic features that contain diagnostic information relating to the performance of different functional tasks. We are now reproducing the diagnostic content of salient kinematics using more cost effective and portable devices.

2 Methods

The clinical assessment we chose to be our 'ground truth is called the Arm Motor Ability Test (AMAT). This test was chosen from a range of instruments employed by therapists for post-stroke evaluation. The test has been shown to have high inter-rater reliability, sensitivity to change, and concurrent validity with other leading assessments, including the Wolf Motor Function Test [6]. To perform the assessment, therapists watch as clients simulate the performance of several upper body tasks, like raising a comb to the hair or dialing a telephone (see Table 1). Each subtask is scored on a 5 point scale (0-4) according to the quality of the underlying motion (i.e. its smoothness, fluidity) as well as its functional efficacy. An individual who receives a 0 may be completely flaccid on his or her hemiparetic side and unable to complete any given task on the assessment. An individual who receives a 4, by contrast, may have motion that is indistinguishable from an individual who has never had a stroke. Various post-stroke symptoms may be manifest among individuals with intermediate scores, such as muscle synergy (i.e. increased flexion in characteristic patterns [2]), spasticity, or jerky, uncoordinated motions. In practice, therapists tend to conflate evaluations of quality and functional efficacy on the AMAT [6]. In work presented here, then, we asked therapists to use a single scale for their performance evaluations.

To test our ability to extract mobility statistics from visual data that correlate with scores on the AMAT, seven stroke survivors were asked to perform tasks on the AMAT assessment while seated at a desktop. The desktop was instrumented with six commercial camcorders, as illustrated in Figure 1. Each camera recorded at a rate of 30 frames per second and was synchronized with the others using red LEDs placed close to the camera lenses.

All video was captured at a resolution of 740x240 and compressed to MPEG2 in real time using an AXIS 250S Video Servers. Compressed data was stored on external hard drives and subsequently spatially down-

Figure 1: At left, the experimental setup. At right, a colored jersey used to facilitate tracking.

Table 1: Examples of elements on the Arm Motor Ability Test (AMAT). Performance of each subtask is evaluated on a 5 point scale (0-4). Evaluations consider both quality of motion and functional efficacy. In this paper, all reported scores are the sum of scores across performed subtasks.

Task	Subtask
Cut Meat	1. Pick up knife and fork 2. Cut meat 3. Fork to mouth
Sandwich	4. Pick up sandwich 5. Sandwich to mouth
Use Spoon	6. Pick up spoon 7. Collect dried kidney bean 8. Spoon to mouth

sampled by a factor of 4. Images that were processed, then, were 185x120 pixels in dimension. While recordings were made, each subject wore a colored jersey to facilitate the localization of individual limb segments of the upper body. They also wore a selection of IR reflective markers on key anatomical locations of the upper body, so that we had access to 3D reconstructions provided by commercial motion capture.

As subjects performed the AMAT, an expert in occupational therapy generated a functional score based on his or her observations. The assessments that were generated for each subject are given in Table 2.

Table 2: Subject demographics. "Side" denotes the side of the brain that was damaged by stroke. "Dom" is the subject's dominant hand. "AMAT" is the subject's AMAT score. All tasks were performed with the side of the body that was most affected by stroke. 'FMA' is a score on the upper body portion of an assessment that tests specifically for muscle synergy called the Fugl-Meyer [2].

ID	Sex	Age	Side	Dom	AMAT	FMA
1	M	75	Right	Right	56	64
2	M	60	Left	Right	32	37
3	M	47	Right	Right	45	51
4	M	82	Left	Right	52	60
5	M	58	Left	Right	63	73
6	F	78	Right	Right	36	40
7	F	63	Right	Right	52	58

2.1 Data Processing

A. 'Potential limb segments were detected based on their color. For each color of the jersey that was worn, we trained a quadratic logistic regression classifier, as in [10]. A single image from each camera view was used for training purposes; pixels inside selected limb segments were labeled 'positives and remaining pixels labeled 'negatives'. At every frame, detected image regions corresponding to each color were grouped into ellipsoidal 'blobs'. Each blob was parameterized according to its location, size and eccentricity.

B. 'Potential limb segments were filtered with simple kinematic trees. To remove blobs that did not relate to limb segments, we employed two simple kinematic trees in a fashion inspired by [12, 10]. Each tree was made up of three segments spanning a single arm, and defined by the following equation:

$$P(p_1, p_2, p_3|I) = \prod_{(i,j)\epsilon E} P(p_i|p_j) \prod_{i=1}^{3} P(p_i|I) \qquad (1)$$

In this equation each p_i is a limb segment (hand, lower arm or upper arm) and E denotes the links between segments in a given three link model. $P(p_i|p_j)$ was determined based on a single training image from each view. The relative angle between the major axes of any two limb segments was computed in training images as well as the difference between any two adjacent segments centroids. $P(p_i|p_j)$ was defined to be a normal distribution centered at these idealized angles and relative locations. $P(p_i|I)$, by contrast, was determined based on the color and eccentricity of a given blob relative to given ideals.

C. Filtered 'parts were combined across multiple views. Next, the filtered limb segments were combined across multiple views to yield 3D estimates of each limb segment's location. To do this, each 2D ellipsoid was first re-parameterized as a set of four points; two points were located at extreme positions on the ellipsoid's major axis and two points were located on the minor axis. Sets of 2D points were combined across multiple views by solving the following linear system, as in [4]:

$$\sum_{i}^{N}(w_i(Id - u_i u_i'))P = \sum_{i}^{N}(w_i(Id - u_i u_i'))c_i \qquad (2)$$

Here, i is an index referring to an individual camera view, P is a reconstructed 3D point, c_i is the ith camera's center, and u_i is the direction of the ray extending from the camera's center to the point. Id is the identity matrix. Finally, w_i is an independent weighting factor for each view, indicating the level of "trust we have in that view. This, in our application, was based on the value of filtered kinematic trees.

Any four 3 dimensional points that were reconstructed from corresponding ellipses defined a cylinder in space. We used these cylinders to estimate the location and orientation of target limb segments. Figure 2 illusrates resulting reconstructions. A visual comparison of kinematic reconstructions achieved with our system and a commercial motion capture system (VICON) is illustrated in Figure 3.

D. Resulting 3D data was trimmed. Once the motion corresponding to the performance of a given task was reconstructed, data was cropped. All the raw

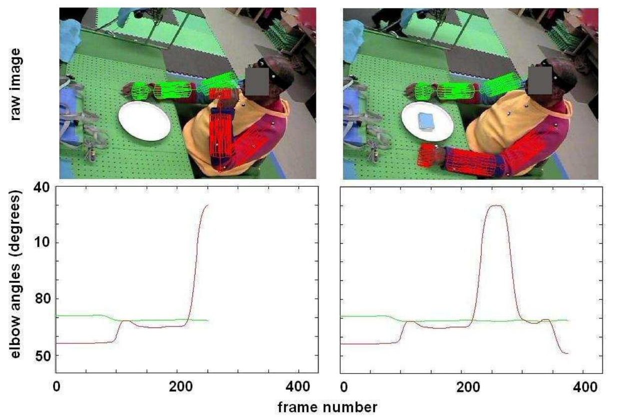

Figure 2: 3D reconstructions are shown in the top row and overlaid on the image data. Estimates of elbow angles over time, for both the right (green) and left (red) arm, are shown at the bottom.

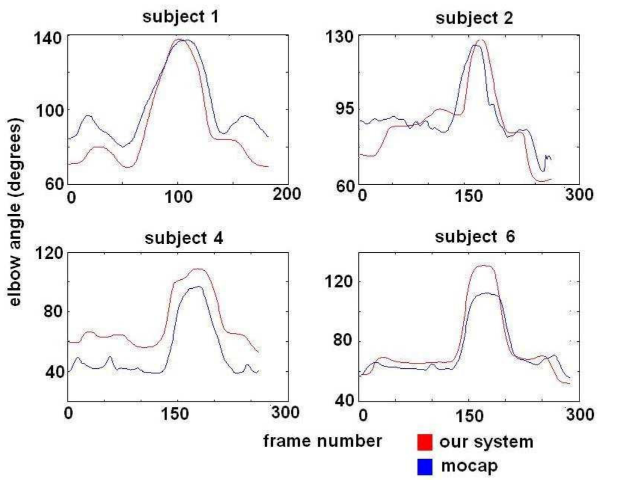

Figure 3: Elbow angles on the hemiparetic side, over time, for four subjects. Estimates are shown both based on motion capture (in blue) and on our system (in red).

movement data was first passed through a median filter of width 15 and a simple segmentation scheme was applied. The beginning of a subject's movement was said to correspond to the point when the velocity of his or her hemiparetic hand surpassed 5 percent of its maximum. The end of movement was located at the final valley in the hand's velocity profile. This scheme was used to segment both motion capture data as well as the motion data that was generated by video cameras.

E. Kinematic statistics were computed. Finally, various kinematic statistics were computed in an effort to locate information that correlated with expert opinions. Here, we present results that relating to two statistics: motion of the torso and flexion about the elbow. We focus on torso motion because stroke survivors have shown a tendency to compensate for distal impairments by moving this part of the body [11]. We focus on elbow angles because stroke survivors frequently suffer from spasticity and muscle synergy at this location; this is defined in part by excess flexion [2].

Torso motion was defined in the video data as the derivative of the mean trajectory of the two reconstructed shoulders. In the motion capture data, torso motion was defined based on the mean trajectory of IR markers located on either acromion process. Elbow flexion in the video data involved first locating the three dimensional points at the intersections of reconstructed upper and lower arm segments. Elbow flexion was defined by this point and points at the shoulder and wrist. In motion capture data, elbow flexion was determined by markers placed on the elbow, the acromion process, and wrist.

3 Results

We report results as they relate to the recorded performance of the 'sandwich' task in the battery of AMAT tasks. A selection of kinematic statistics that were shown to positively correlate with the functional opinion of experts follows:

(1) Mean torso displacement. As in the studies of [8], individuals who were more functionally impaired tended to exhibit more average motion of the torso during the performance of tasks. At the top of Figure 4 we compare computations of torso motion based on motion capture data and our kinematic data. Diagnostic information is carried by all computed statistics. The p-value for linear regressions relating aggregate AMAT scores to the motion capture statistic is less than .0001, and the r squared value for this fit alone is .897. Likewise, the p-value for the regression of our kinematic tracking statistics onto AMAT scores is less than .001, and the r-squared value is over .9. For regressions of the same statistics onto the Fugl-Meyer (FMA) scores, there are slightly higher p-values. The FMA measures muscle synergy in stroke survivors more specifically that the AMAT.

(2) Standard deviation in elbow flexion. A slightly more subtle statistic found to correlate with functional scores is the range of motion recorded about the elbow during functional motion; this is illustrated at the bottom of Figure 4. A linear regression relating the standard deviation in elbow flexion and AMAT score carries a p-value of .06 (for motion capture) and .06 (for our kinematic reconstructions). Limits in elbow range of motion reflect a fairly common post-stroke muscle synergy, which is characterized in the upper body by excess flexion [2]. When movement kinematics are regressed on assessment scores that specifically test for synergies, similarly significant p-values are achieved. Regressions relating standard deviation of elbow flexion and scores on the FMA yield p-values of .05 both for motion capture and our kinematic tracking statistics.

4 Discussion

This paper presents initial work directed at the creation of a system which will automatically assess individuals who are recovering from a stroke based on perceived kinematics. Results indicate that it possible to use kinematic statistics that are generated by a robust, region based kinematic tracker to assess the functional behavior of stroke survivors. The accuracy of vision based kinematic trackers as they currently stand, then, may be enough to produce clinically meaningful applications.

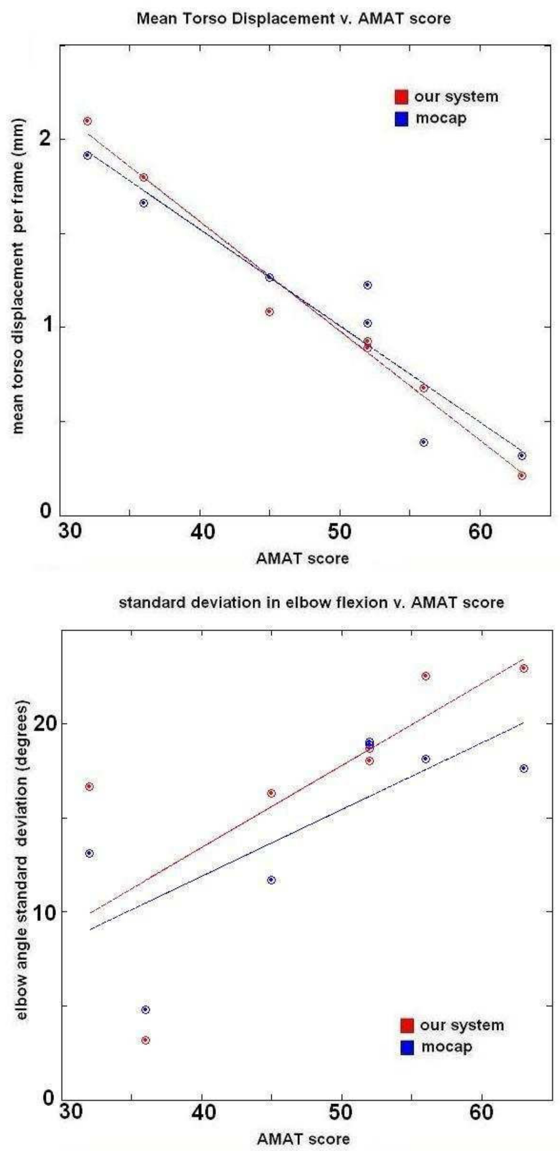

Figure 4: At top, estimates of elbow range of motion v. AMAT score. At bottom, estimates of mean torso displacement per frame v. AMAT score. Red is the statistic from our system, blue is from motion capture.

The kinematics presented here are by no means the only statistics, however, that carry diagnostic information relating to stroke. Velocity and smoothness of perceived motion, for example, are movement targets that are also known to be influenced by stroke. Future work, then, must explore a more complete range of motor statistics.

In addition, future work must explore the degree to which we can degrade our visual signals while still yielding clinically meaningful movement information. Our interest is in tracking that takes place quickly, with fewer cameras, in much less constrained environments (like homes or workplaces). The degree to which our region based approach is robust to signal degradations has yet to be explored. Other approaches, like those infused with physics [15], or based on voxel carving or visual hull constructions [3], may be similarly capable of yielding clinically significant statistics of motion, but perhaps less able to perform in a wide variety of environments.

References

[1] A. Agarwal and B. Triggs. Learning to track 3d human motion from silhouettes. 2004.

[2] S. Brunnstrom. Movement therapy in hemiplegia: a neurophysiological approach. Harper and Row, New York, New York, 1st edition, 1970.

[3] German K.M. Cheung, S. Baker and T. Kanade. Shape-From-Silhouette of Articulated Objects and its Use for Human Body Kinematics Estimation and Motion Capture, in Proceedings of IEEE Conference on Computer Vision and Pattern Recognition 2003 (CVPR'03), Vol. 1, pages 77-84, Madison WI, June 2003.

[4] R. Collins, O. Amidi and T. Kanade. An active camera system for acquiring multi-view video. pages 517-520, September 2002.

[5] David A. Forsyth, Okan Arikan, Leslie Ikemoto, James O' Brien, Deva Ramanan, Computational Studies of Human Motion: Part 1, Tracking and Motion Synthesis, Foundations and Trends in Computer Graphics and Vision Volume 1 Issue 2/3 (255pp), 2006

[6] B. Kopp, A. Kunkel, H. Flor, T. Platz, U. Rose, K. Mauritz, K. Gresser, K. McCulloch, and E. Taub. The arm motor ability test (amat): Reliability, validity and sensitivity to change. Arch. Phys. Med. Rehab., 78:615-620, 1997.

[7] M. Lee and R. Ne Nevatia. atia. Dynamic human pose estimation using mark markov chain monte carlo approach. In IEEE Workshop orkshop on Motion and Video ideo Computing Computing, pages 168175, 2005.

[8] M. Levin, J. Desrosiers, D. Beauchemin, N. Bergeron, and A. Ro- chette. Development and validation of a scale for rating motor compensations used for reaching in patients with hemiparesis: the reaching performance scale. Phys Ther., 84(1):8-22, Jan 2004.

[9] Medicare benefit policy manual, ch. 7. Centers for Medicare and Medicaid Services.

[10] Ramanan, D., Forsyth, D. A., Zisserman, A. "Strike a Pose: Tracking People by Finding Stylized Poses." Computer Vision and Pattern Recognition (CVPR), San Diego, CA, June 2005

[11] A Roby-Brami, A Feydy, MCombeaud, E Biryukova, B Bussel and M Levin. Motor compensation and recovery for reaching in stroke patients. Acta Neurol Scand., 107(5):369-81, May 2003.

[12] H. Sidenbladh, M. Black, and D. Fleet. Stochastic tracking of 3d human figures using 2d image motion. ECCV, 2:702-718, 2000.

[13] L. Sigal, S. Bhatia, S. Roth, M. Black, and M. Isard. Tracking loose- limbed people. Computer Vision and Pattern Recognition (CVPR), 2004.

[14] E. Taub, G. Uswatte, and D. Morris. Improved motor recovery after stroke and massive cortical reorganization following constraint-induced movement therapy. Phys Med Rehabil Clin N Am., 14(1):77-91, 2003.

[15] C. Wren and A. Pentland, Dynamic models of human motion, In Proceedings of IEEE FG'98, Nara, Japan, April 1998.