Babak Taati*, Jennifer Campos, Jeremy Griffiths, Mona Gridseth and Alex Mihailidis

Vision-based categorization of upper body motion impairments and post-stroke motion synergies

Abstract: Upper body motion impairment is a common after-effect of a stroke. We are developing a novel multisensory therapy system that makes use of augmented feedback and is expected to enhance rehabilitation outcomes. The first phase of the project involved developing algorithms to automatically differentiate between normal and impaired upper body motions. Seven healthy subjects performed two types scripted actions: (a) elbow flexion and extension and (b) reaching over via elbow flexion and shoulder flexion and adduction. Each action was repeated 10 times as each participant felt most comfortable and also 10 times simulating a common post-stroke impaired motion according to the literature. Principal component analysis was applied to the upper body trajectories during each action, as observed with a Kinect sensor, to extract the dominant modes of motion. Three classification algorithms and up to three motion modes were examined in order to distinguish between normal and impaired motions. A statistical analysis of the Kinect skeletal tracking data vs. manual annotation confirmed a significant bias in the tracking of an elevated shoulder joint. Despite this bias, leave-one-subject-out cross validation experiments confirmed the effectiveness of the proposed methods in detecting impaired motions (accuracy >95%), even when the impairment involved elevating a shoulder. A single, most dominant motion mode provided sufficient information for the binary classification task. The high accuracy rates validate the use of vision-based pose tracking technologies in identifying motion deficiencies.

Keywords: Kinect; mobility assessment; motion impairment; motion synergy; stroke rehabilitation; tracking bias; virtual reality.

DOI 10.1515/ijdh-2014-0326

Received May 3, 2013; accepted June 22, 2013; previously published online July 18, 2014

Introduction

Hemiparesis, or weakness on one side of the body, is a common after-effect of a stroke. A well-known body illusion, known as the mirror box illusion, is used in rehabilitation training exercises that have been shown to be effective in restoring mobility [1]. During the treatment, the patient enters both arms into a box with a mirror in the middle so that the view of the affected arm is replaced by the mirror reflection of the healthy arm. Simultaneous motion, or attempted motion, of both arms results in artificial visual feedback indicating that it is, in fact, the affected arm that is moving, thus promoting neurorehabilitation.

A closely related body illusion, the rubber hand illusion [2], offers promising potential in strengthening the visual feedback from the mirror box. In this illusion, a plastic arm is placed on a table next to the real arm, which is hidden by a screen, and both the rubber hand and the real hand are simultaneously touched by a brush for a few minutes using a rhythmic motion. The temporal synchrony of the seen movements on the rubber hand and the felt movements on the real hand causes the observer to embody the rubber hand so the perceived location of the real hand is shifted toward the seen location of the rubber hand.

Objectives

The main objective of this work is to reproduce and combine the aforementioned illusions in a virtual environment. It is expected that augmenting mirror box therapy with the rubber hand illusion will strengthen the sense of ownership over the artificial visual feedback and, thus, will enhance the rehabilitative potential. The virtual environment will also enable a wider range of motion than a physical mirror box.

The project takes advantage of state-of-the-art computer vision and markerless skeleton tracking systems to

***Corresponding author: Babak Taati**, PhD, Scientist, Intelligent Assistive Technology and Systems Lab (IATSL), University of Toronto, Toronto, Ontario, Canada; and Toronto Rehabilitation Institute, University Health Network, 500 University Ave., Toronto, Ontario, M5G 2A2, Canada, E-mail: Babak.Taati@uhn.ca

Alex Mihailidis: Intelligent Assistive Technology and Systems Lab (IATSL), University of Toronto, Toronto, Ontario, Canada; and Toronto Rehabilitation Institute, Toronto, Ontario, Canada

Jennifer Campos and Jeremy Griffiths: Toronto Rehabilitation Institute, Toronto, Ontario, Canada

Mona Gridseth: Intelligent Assistive Technology and Systems Lab (IATSL), University of Toronto, Toronto, Ontario, Canada

tackle the engineering challenges and to counter some of the weaknesses identified among many of the virtual rehabilitation technologies [3], for example, in the user interface or in the interaction methods with the virtual world. The final system is being targeted at an affordable price ($\leq \$200$) to enable wide deployment and home usage. The system is also aimed at being easily configurable so it can be adapted to target a range of mobility impairments.

The skeleton tracking information is processed to identify impaired modes of motions by preprocessing short segments of scripted actions. This information will then be used to modify the profile of subsequent motions in real-time in order to exaggerate impaired motions, for instance by amplifying a reduced range of motion or exaggerating impaired synergies. Finally, the modified motion is displayed in a virtual environment in real-time and in synchronization with tactile feedback.

Contributions

This paper presents results from the first phase of the project, which focuses on the development of posture categorization preprocessing algorithms and preliminary validation experiments on healthy subjects. Because measuring range of motion deficits is relatively easy via skeleton tracking, the experiments presented here focus on the more challenging task of identifying motion synergies. Motion synergies are a common after-effect of a stroke and refer to the undesirable couplings of human degrees of freedom.

We have also performed preliminary experiments to validate the skeleton tracking algorithm in a specific scenario related to this research. The results identify an inherent bias in tracking the shoulder joint. We provide a short discussion on the consequences of such limitations and how they might affect the assessment of posture and motion.

Methods

The system employs a consumer depth sensor, a Microsoft Kinect (Microsoft, Redmond, WA, USA) to capture live video and depth images and to track major skeletal joints of the body in real-time. The PrimeSense™ open source Open NI and NITE libraries were used to interface with the device. In the final system, touch feedback will be provided via a wearable glove, equipped with small vibrating motors, and synchronized with the visual display through a digital I/O board and a digital solid-state relay board. The touch feedback interface was not utilized during the experiments reported here as it did not relate to identifying impaired motions.

Experimental dataset

Two of the most common upper limb flexion synergies present in post-stroke patients were selected according to the Chedoke-McMaster stroke assessment 2005 and were simulated by healthy adults. The selected actions were (a) reaching over to the opposite shoulder via elbow flexion and shoulder flexion and adduction (Figure 1) and (b) elbow flexion (Figure 2). The stereotypical post-stroke movement synergy associated with flexing the elbow is hiking the shoulder while reaching across the body and is typically accompanied by rotating the trunk [4].

Figure 1 Sample color, depth, and skeleton tracking images in six representative frames over two sequences of reaching across. The top row illustrate a “normal” movement pattern, whereas the bottom row illustrates a simulated “impaired” motion synergy of rotating the trunk when reaching across, common among the post-stroke population.

Figure 2 Sample color, depth, and skeleton tracking images in six representative frames over two sequences of elbow flexion. The top three rows illustrate a “normal” movement pattern, whereas the bottom three rows illustrate a simulated “impaired” motion synergy of hiking the shoulder when flexing the elbow, common among the post-stroke population.

Seven healthy adults (4 males, 3 females) were recruited and asked to use their right arms to perform each of the two actions 20 times in front of the sensor, 10 times normally as they felt most comfortable and 10 times simulating the impaired motion. The 3-D

coordinates of the upper body skeletal positions were normalized into a subject-centered canonical coordinate frame independent of the body size and the viewing angle [5]. The 3-D displacement of all upper body skeletal joints in consecutive frames was computed.

Considerations of tracking accuracy

As the real-time skeleton tracking system used in the experiments (i.e., the Kinect [6]) is a relatively new technology (introduced in late 2010), its accuracy levels and limitations are not fully known, and the existing literature on the topic is sparse. Most notably, Dutta [7] evaluated the depth sensing accuracy of the device and, more recently, Clark et al. [8] validated its skeleton tracking capabilities. When compared to a marker-based Vicon motion capture system as the ground truth, these studies concluded that the Kinect provided subcentimeter accuracy in depth sensing in a sizable workspace (covering a range from 1 to 3 m from the sensor and a field of view of 54°) and its 3-D skeleton tracking had comparable intertrial reliability. However, Clark et al. also observed tracking biases in estimating the position of the torso and the pelvis.

During our experimentations, it was observed that the skeleton tracking library sometimes estimated joint positions that visually appeared to be systematically biased in certain ways. These biases were mostly observed in the tracking of the elevated shoulder joint and were different from the earlier published results reporting biases in the pelvis and the torso. Therefore, a series of measurements were made to confirm and quantify this visual observation. With the growing interest in applying the Kinect and similar commercially available systems, for example, Asus WAPI Xtion (Asus, Fremont, CA, USA) in gait assessment and other applications in rehabilitation, monitoring, and assistive technologies, it is hoped that the results from these experiments will contribute to obtaining a better understanding of the capabilities and limitations of these technologies.

Automatic categorization of motion impairments

The system is expected to observe a human subject performing a short scripted motion, such as flexing his or her elbow and to automatically analyze the motion characteristics of the upper body and detect possible impaired motions.

Principal component analysis (PCA) of the displacement vectors was used to compute the dominant modes of motion over the course of each action and the top n PCA dimensions, along with their corresponding singular values, were used as the feature set in a binary classification to distinguish between normal and simulated impaired motions. Seven-fold leave-one-subject-out cross validation was used to train and validate three binary classification algorithms, each separately trained for identifying impaired motions while flexing the elbow or while reaching across. The classification algorithms included logistic regression (LR) [9], binary support vector machines (SVM) with radial basis function kernels [10], and Random Forests (RF) [11]. The SVM and the RF are state-of-the-art machine learning algorithms [12] whereas the LR is a simple linear classifier used here as a baseline.

Assessing the skeleton tracking in impaired motions

With an elevated shoulder, it was visually observed that the 3-D skeleton tracking repeatedly underestimated the degree by which the shoulder was elevated by a large margin (Figure 2). To confirm

this observation, in seven videos simulating a motion with a hiked shoulder (one of each subject), 10 frames were randomly selected for manual annotation. Two human annotators were separately asked to mark the location of the left and the right shoulder of the subject at each frame via a mouse click and the elevation angle of the connecting lines (i.e., the angle with the horizontal axis in the 2-D image plane) was computed.

For comparison, the same angle with the horizontal axis was also computed based on the 3-D tracking data. To calculate the angle, the 3-D line segment connecting the estimated location of the left and right shoulder joints, tracked in the depth sensor coordinate frame, and was first transformed into the color camera coordinates via the default color-depth registration of the Kinect (i.e., a rigid 3-D transformation). The 3-D line was then projected onto the image plane based on the default extrinsic and intrinsic calibration parameters of color camera.

Ideally (i.e., in the absence of any inaccuracies or biases in the tracking of the shoulder joints, angles computed from the skeleton) tracking data would be very close to those from the manual annotations. If the error in 3-D tracking is larger than that of the 2-D annotations, it is still expected that the angles be without a bias. That is, whereas the variance of the angles computed from the 3-D tracking could be larger, their mean will ideally be the same as that of the manual annotations.

Admittedly, identifying the shoulder joints visually and via user mouse clicks is not always very accurate. Nevertheless, it is reasonable to assume that the human annotators will mark the left and the right shoulder similarly and their annotations will lead to an unbiased estimate of the elevation angle over the 70 frames. Furthermore, the agreement between the two manual annotators was used as a baseline to confirm the validity of the annotations. Various statistical comparisons between the angles from human annotations and the 3-D tracking were used to assess the bias in the 3-D tracking of elevated shoulder joints.

Results

The SVM correctly identified simulated impaired motions with high accuracy (~96% and 97%); see Table 1. The RF also provided good, albeit slightly less consistent results (98% and 90%). A single motion mode (i.e., $n=1$, the most dominant motion over the course of an action) provided

Table 1 Classification accuracy in distinguishing between normal and simulated impaired motion.

Action	n	Classifier		
		Logistic regression	Support vector machine	Random forest
Reaching Across	1	54.3	97.1	97.9
	3	61.4	85.7	97.1
Elbow Flexion	1	64.3	95.7	90.0
	3	59.3	85.0	92.9

sufficient information to discriminate between normal and impaired motions.

The high rates of accuracy validate the use of commercially available and affordable vision-based pose tracking technologies in identifying motion deficiencies. This encourages further work into the development of the fully integrated tactile and visual virtual reality system for inducing sensory conflicts and targeted rehabilitation therapy. Future work includes validation experiments with real (i.e., not simulated) impaired motions through testing with actual stroke survivors, real-time exaggeration or attenuation of undesirable synergies, and human subject rehabilitation efficacy trials.

Verifying a tracking bias

The correlation coefficient (ρ) between the sequence of 70 numbers further illustrates the agreement between the two human annotators ($\rho_{h1h2}=0.98$) and the disagreement between the human annotators and the 3-D data ($\rho_{3D-h1}=0.20$ and $\rho_{3D-h2}=0.15$). The smaller coefficients indicate that the angles computed from the 3-D skeletal joints not only underestimated the shoulder elevation angle but also were less correlated with the true elevation angle.

Finally, the Student's t-test and its nonparametric equivalent, the Wilcoxon signed rank (WSR) test, also rejected the hypothesis that the 70 angles computed from the 3-D joints were taken from distributions with the same mean (or median in the case of the nonparametric test) as that of the 2-D angles while strongly confirming that the sequence of the two human annotators indeed were from the same distribution.

The NITE skeleton tracking library provides a 3-value (0, 0.5, 1) confidence level for each estimated joint position, where 1 is the highest confidence and 0 is the lowest. The confidence values of the estimated position of both shoulder joints, provided by the skeleton tracking library, were at the maximum level (i.e., 1) in all 70 frames. In other words, any discrepancy between the 3-D tracking data and the 2-D annotations could not be attributed to low confidence tracking.

The results in assessing the tracking of an elevated shoulder highlight the need for a better understanding of the algorithms applied in the skeleton tracking libraries and their underlying assumptions if their usage is to become more widespread in gait and posture assessment. It is interesting, however, to note that despite the inability of the system to track an elevated shoulder joint properly, the developed posture categorization system worked well with a high accuracy of 95.7%. This was only slightly lower

(1.4 percentage points when using SVM) than the accuracy rates obtained in identifying trunk rotation. Perhaps even the slight elevation detected in the placement of the shoulder, combined with other cues from the orientation of the shoulder-elbow and other segments, was sufficient in detecting this compensation mode. It is, however, reasonable to assume that better skeleton tracking would improve the overall identification of compensations.

Discussion

When observing short scripted motions, the most dominant mode of motion, as identified by PCA, provided sufficient information to discriminate between normal and simulated impaired motions with high accuracy. Inherent biases were observed in the tracking of an elevated shoulder via the Kinect sensor. The automatic detection of an impaired motion involving an elevated shoulder was only marginally affected by this bias. However, further research in identifying and quantifying such biases is required so that they can be taken into account when designing rehabilitation and health monitoring systems that use the Kinect.

Current work in completing the first phase of the research focuses on experiments with post-stroke survivors to validate the results obtained in classifying simulated impaired motions. Future work toward the full implementation of the sensory-conflict rehabilitation system involves the real-time augmentation of impaired motions, duplicating the rubber hand illusion in a virtual environment via real-time tracking and simultaneous touch feedback, combining the rubber hand illusion with the mirror box illusion via real-time motion augmentation and clinical trials.

Acknowledgments: This work was partially supported through a MITACS NCE Strategic Project post-doctoral fellowship. The authors would also like to thank the Toronto Rehabilitation Institute.

References

1. Altschuler EL, Wisdom SB, Stone L, Foster C, Galasko D, Ramachandran VS. Rehabilitation of hemiparesis after stroke with a mirror. *Lancet* 1999;353:2035–36.
2. Botvinick M, Cohen J. Rubber hands “feel” touch that eyes see. *Nature* 1998;391:765.
3. Rizzo A, Kim GJ. A SWOT analysis of the field of virtual reality rehabilitation and therapy. *Presence: Teleoperators Virtual Environ* 2005;14:119–146.

4. Gowland C, Stratford P, Ward M, Moreland J, Torresin W, Van Hulle-naar S, et al. Measuring physical impairment and disability with the Chedoke–McMaster stroke assessment. *Stroke* 1993;24:58–63.
5. Taati B, Wang R, Huq R, Snoek J, Mihailidis A. Vision-based posture assessment to detect and categorize compensation during robotic rehabilitation therapy. *Proc 4th IEEE/RAS–EMBS Intl Conf Biomedical Robotics and Biomechatronics*, 2012.
6. Shotton J, Fitzgibbon A, Cook M, Sharp T, Finocchio M, Moore R, et al. Real-time human pose recognition in parts from a single depth image. *Proc 24th IEEE Intl Conf Computer Vision and Pattern Recognition*, 2011.
7. Dutta T. Evaluation of the Kinect™ sensor for 3-D kinematic measurement in the workplace. *Appl Ergonomics* 2011;43:645–9.
8. Clark RA, Pua YH, Fortin K, Ritchie C, Webster KE, Denehy L, et al. Validity of the Microsoft Kinect for assessment of postural control. *Gait Posture* 2012;6:372–7.
9. Agresti A. *Categorical data analysis*. New York: Wiley-Interscience, 2002.
10. Joachims T. Making large-scale SVM learning practical. In: Schölkopf B, Burges C, Smola A. *Advances in kernel methods: support vector learning*. Boston, MA: MIT Press, 1999:41–56.
11. Breiman L. Random forests. *Mac Learn* 2001;45:15–32.
12. Caruana R, Niculescu-Mizil A. An Empirical comparison of supervised learning algorithms. *Proc 23rd Intl Conf Machine Learning*, 2006.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.