2011 IEEE International Conference on Rehabilitation Robotics
Rehab Week Zurich, ETH Zurich Science City, Switzerland, June 29 - July 1, 2011

Computer Vision-Based Classification of Hand Grip Variations in Neurorehabilitation

José Zariffa, *Member, IEEE*, and John D. Steeves
International Collaboration On Repair Discoveries (ICORD)
University of British Columbia and Vancouver Coastal Health Research Institute
Vancouver, BC, V5Z 1M9, Canada
zariffa@icord.org, steeves@icord.org

Abstract— **The complexity of hand function is such that most existing upper limb rehabilitation robotic devices use only simplified hand interfaces. This is in contrast to the importance of the hand in regaining function after neurological injury. Computer vision technology has been used to identify hand posture in the field of Human Computer Interaction, but this approach has not been translated to the rehabilitation context. We describe a computer vision-based classifier that can be used to discriminate rehabilitation-relevant hand postures, and could be integrated into a virtual reality-based upper limb rehabilitation system. The proposed system was tested on a set of video recordings from able-bodied individuals performing cylindrical grasps, lateral key grips, and tip-to-tip pinches. The overall classification success rate was 91.2%, and was above 98% for 6 out of the 10 subjects.**

Keywords:robotic rehabilitation, hand posture, computer vision, cylindrical grasp, lateral key grip, tip-to-tip pinch.

I. Introduction

A variety of interactive assistive devices have been recently developed as tools for rehabilitation after neurological damage, such as spinal cord injury (SCI) and stroke [e.g. 4, 7, 10, 11, 14]. In particular, robotic rehabilitation devices with virtual reality environments are seen as a way to decrease the amount of repetitive manual labor by therapists, as well as provide patients with more practice trials and increased enjoyment via the variety of virtual reality tasks. In short, virtual reality rehabilitation devices can complement the efforts of therapists, improve long-term compliance of patients with their rehabilitation programs and provide quantitative feedback on a patient's performance. In addition, in-home devices can extend and/or maintain the rehabilitation gains achieved as an in-patient.

Upper limb interactive rehabilitation devices have focused mostly on improving upper-arm and forearm functions. The hand, however, is significantly more complex, and as a result most devices have employed only simplified hand interfaces (e.g. a cylindrical grasp module). Given the crucial importance of improving hand function to regaining greater independence in activities of daily living (ADL) [1], there is a need for more sophisticated hand rehabilitation technology, as well as improved hand function assessment tools [13].

One possible avenue to achieve more reliable and quantitative hand function monitoring and to provide immediate feedback to the user and/or therapist is to use computer vision approaches for evaluating hand posture. Since several robotic rehabilitation devices include virtual reality environments [e.g. 7], hand posture recognition could be integrated into these interfaces by asking the user to assume the correct type of grip in order to accomplish a virtual task.

Computer vision approaches to hand posture recognition have been extensively investigated in the context of Human Computer Interaction (HCI, see [5] for a thorough review), but the application of these results to a rehabilitation setting has been very limited. In addition, most systems that have been introduced are designed to recognize hand postures and movements that often bear little relation to the types of grips that would be desirable for most ADL. In particular, rehabilitation will involve more "closed" postures, which may be more difficult to discriminate than postures with outstretched fingers.

The goal of the present study is two-fold. First, we present an image processing and classification methodology that is appropriate for the discrimination of rehabilitation-relevant hand postures. Second, we demonstrate that this goal can be met with low-cost technology that could easily be deployed in a clinical setting.

II. Methods

A. Data Collection

Video was recorded from two webcams (LifeCam Show, Microsoft Corporation, Redmond, WA, USA), one placed 30 cm above the hand and the other 30 cm to the side. Each camera was positioned such that only the hand was visible, against a uniform blue background. No markers were placed on the hand.

10 able-bodied subjects were recruited for the study (6 male, 4 female). Each subject was asked to perform three hand postures that are of pragmatic interest during rehabilitation after a neurological injury and fundamental to ADLs: a cylindrical grasp, a lateral key grip, and a tip-to-tip pinch (Fig. 1). This choice of postures was based on the grips used in several functional assessments, including the Action Research Arm Test (ARAT [3]), the Fugl-Meyer assessment

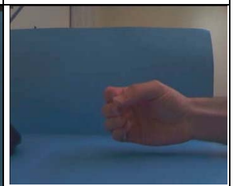
Grip Type	Top View	Side View
Cylindrical Grasp		
Lateral Key Grip		
Tip-to-Tip Pinch		

Figure 1: Examples of cylindrical grasp, lateral key grip and tip-to-tip pinch, from the two view-points used by the classifier.

[6], the Sollerman Hand Function Test [12] and the Graded and Redefined Assessment of Strength, Sensibility and Prehension (GRASSP [9]).

Each grip was modeled for the subjects by the investigator before the videos were recorded, and one practice run was performed to familiarize subjects with the protocol. All subjects held their hand in approximately the same orientation. Following onscreen prompts, the subjects held each posture for 5 seconds while video was recorded at 15 frames per second with a resolution of 352 x 288 pixels. 10 trials were recorded for each posture over the course of a single session. The total data set therefore consisted of 7500 frames for each posture (5 seconds x 15 frames/s x 10 subjects x 10 trials). The protocol of the study was approved by the University of British Columbia human ethics committee.

B. Video Processing

The video data was processed using the OpenCV library in C++. The image processing was fully automated.

First, a background subtraction was performed on each recorded frame to extract the image of the hand. In order to avoid corruption of the hand contour by shadows projected by the hand onto the background, a difference metric was devised. For each pixel in the image, the three values shown in (1)-(3) were computed:

$$RG = (R - G)/2^D \qquad (1)$$
$$RB = (R - B)/2^D \qquad (2)$$
$$GB = (G - B)/2^D \qquad (3)$$

Where R, G, B are the red, green, and blue values of the pixel and D is the number of bits used to represent each value (in our case D = 8). The values in (1)-(3) represent a pixel by the relationships between the colors rather than by their absolute values, and are therefore more sensitive to qualitative changes in color than to general darkening due to shadows. The difference metric for each pixel was then defined as shown in (4):

$$M=2^D\sqrt{(RG_F\text{-}RG_B)^2+(RB_F\text{-}RB_B)^2+(GB_F\text{-}GB_B)^2} \qquad (4)$$

Where RG_F, RB_F and GB_F correspond to pixels in the frame of interest, whereas RG_B, RB_B and GB_B correspond to pixels in a reference frame obtained at the beginning of each trial. The reference frame contains the same background, but no hand. Pixels for which M was greater than one standard deviation above the mean of M over all pixels in the image were considered to be part of the hand and were retained. The remaining pixels were set to 0.

The resulting background-subtracted image was then smoothed slightly using morphological filtering (one round of erosion followed by dilation, both using a rectangular kernel whose size was empirically set at 3x3). The largest contour present in the image was identified and taken to represent the hand (images in which no contour was larger than 1% of the total image area were deemed to be empty). In order to enable meaningful comparison between recordings, the image was then shifted so as to be centered on the center of gravity of the identified contour, and rotated so that the long axis of the contour was always in the same direction. An example of a processed frame is given in Fig. 2.

Three types of features were then extracted from the processed frame:

- The seven Hu invariant moments of the image were computed, using the greyscale version of the image [8].
- A contour signature was computed, following the methodology described by Bader et al. [2]. The description of the contour shape that this signature provides is invariant to translation and scaling. The resulting signature was then sampled at 40 equally spaced points around the contour, where each point is described by two values (an arc width function and an arc height function, see [2] for details).
- An inner contour feature was designed to describe the presence of any "holes" in the image, as would occur

Figure 2: Example of a processed frame (cylindrical grasp, top view), showing the results of the background subtraction, centering, and detection of outer and inner contours.

for example in a cylindrical grasp viewed from the top. The three components of this feature were the area of the largest inner contour detected (if any), and the two coordinates of its center of gravity.

In total, the resulting feature set contains 90 elements per camera (7 Hu moments, 80 contour signature values, and 3 inner contour descriptors). Combining the information from both cameras, each frame is described by a 180-element feature vector. The combination of features used is a combination of previously described methods (the Hu invariants and the contour signature), as well as an inner contour feature chosen based on the characteristics of the postures of interest in our application.

The Hu invariants and inner contour feature are useful in that they help to characterize the region within the outer contour of the hand. This type of information is important because the hand postures do not have outstretched fingers that result in distinctive contours, as is often the case in HCI applications. The elements of the feature vector were weighted so as to give each type of feature equal weight: each of the Hu invariants for a given camera had a weight of 80/7, each of the contour signature values had a weight of 1, and each of the inner contour features had a weight of 80/3. Upon testing the system, however, it was empirically determined that including the Hu invariants from the top-view camera was detrimental rather than beneficial to performance, possibly due to the similarity between the lateral key grip and tip-to-tip pinch from this angle (Fig. 1). These elements were therefore not included in the final feature vector, which contained 173 elements.

C. Posture Classification

Classification of new frames was performed using a K-Nearest Neighbors classifier with three categories, corresponding to the three grip types. The number of neighbors used was 20 and the result was determined by majority voting. The performance of the classifier was evaluated using 10-fold cross-validation, using the 10 trials from each subject in turn as the testing set and the data from the remaining 9 subjects (90 trials total) as the training set.

III. Results

The classification success rate was 94.1% for cylindrical grasps, 93.1% for lateral key grips, and 86.4% for tip-to-tip pinch, resulting in an overall classification success rate of 91.2%. The details of the mistaken cases are provided in the confusion matrix shown in Table 1. The tip-to-tip pinch (thumb to index finger) was the most difficult posture to classify accurately and was confused with a key grip in 13.6% of cases. The cylindrical grasp was the most distinctive of the three postures: it was accurately classified most often, and other postures were almost never mistaken for cylindrical grasps.

The variations in classification accuracy (all grip types combined) between different subjects are shown in Fig. 3, and ranged from 58.6% to 99.7%. 8 out of 10 subjects had accuracies above 90%, and 6 out of 10 were above 98%.

Table 1: Classification performance for individual grip types

Selected grip / True grip	Cylindrical	Key	Tip-to-tip
Cylindrical	*94.1%*	5.6%	0.3%
Key	0.9%	*93.1%*	6.0%
Tip-to-tip	0%	13.6%	*86.4%*

Once the classifier was trained, the system could interactively classify postures from a live video feed at a rate of approximately 3 frames per second per camera. The image processing steps (background subtraction, image smoothing and centering, feature vector extraction) could be performed at approximately 7 frames per second per camera, with the classification step accounting for the difference in frame rate. This performance was measured on a mid-range laptop (Intel® Core™ 2 Duo CPU, 2.1GHz). As our focus here was on classification accuracy, no attempt was made to optimize performance for speed.

IV. Discussion

We have presented a hand posture recognition system and evaluated its performance on a set of postures relevant to neurorehabilitation and recovery of function in ADLs. The system is low-cost (approximately \$200 for hardware components) and easy to use. It does not use any markers and requires very little setup, such that it could easily be used in a clinical setting with any personal computer. The frame rate during interactive classification was low, but could be increased with modest improvements to hardware and code optimization. Note that for applications in interactive rehabilitation devices, very rapid finger movements are not expected, and there is no need to classify the posture with high frequency, making modest frame rates acceptable.

The system's classification accuracy was satisfactory and very high for the majority of subjects, except for two subjects. This can be attributed to each of those subjects positioning their hand somewhat differently than the nine other subjects whose hand postures constitute the training set, though still not so differently as to constitute an improper grasp. With a larger training set containing a wider range of posture profiles, the number of subjects for whom classification accuracy is low would be expected to decrease, as the system become more flexible with regards to postural variations such as the angle of pronation/supination.

The fact that the cylindrical grasp showed the highest classification accuracy is most likely due to the presence of a distinctive inner contour in the top view for this grip type, as illustrated in Fig. 2. This feature is dependent on the hand being empty but, if an object was being held, alternative algorithms could be used to identify the inner contour of the hand (for example, based on the color difference between the hand and the object) and we expect that the assessment performance would not be significantly impacted.

Computer vision-based classification of grip type can be used to monitor the hand function of rehabilitation subjects, and therefore has direct applications in improving interactive upper limb robotic rehabilitation devices with virtual reality components. Having now developed an image processing and classification approach suited to this task, the next step should be the development of a hand posture recognition system whose training set includes examples of impaired grips, because users of robotic rehabilitation devices will in many cases not be able to perform a "normal" grip. The process of building a data set of hand posture video recordings from individuals with cervical SCI is ongoing in our group.

A related goal is to create a computer vision-based system capable not only of recognizing an impaired grip, but of assessing the level of impairment. This objective could be achieved by associating each example of an impaired grip in the training set with hand-related clinical scores measuring the extent of disability. A new observation of impaired grip could then be assessed using the clinical scores of the most similar examples in the training set, for a given type of intended grip (in the context of an assessment, the subject is provided with instructions and the intended grip is therefore known). Such a system would have important applications in clinical trials for neurorehabilitation interventions, as it would provide an automated and quantitative outcome measure for hand impairment.

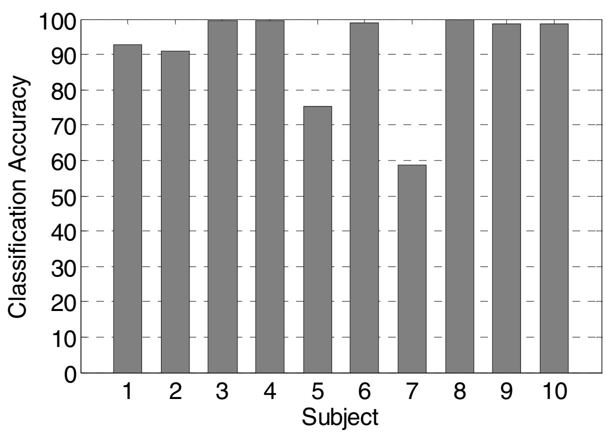


Figure 3: Classification accuracy per subject

ACKNOWLEDGMENTS

The authors would like to express their gratitude to all of the subjects who took part in this study. This work was supported by the Rick Hansen Institute Spinal Cord Injury Translational Research Network and the Sensorimotor Computation Major Thematic grant from the Peter Wall Institute for Advanced Studies.

REFERENCES

[1] K. D. Anderson, "Targeting recovery: Priorities of the spinal cord-injured population," J. Neurotrauma, 21 (10) (Oct): 1371-83, 2004.

[2] T. Bader, R. Rapple, and J. Beyerer, "Fast invariant contour-based classification of hand symbols for HCI," In *CAIP* 2009, volume 5702 of *LNCS*, p. 689-696. Springer, 2009.

[3] D. Carroll. "A quantitative test of upper extremity function," J Chronic Diseases 18:479-491, 1965.

[4] R. Colombo, F. Pisano, S. Micera, A. Mazzone, C. Delconte, M. C. Carrozza, P. Dario, and G. Minuco, "Robotic techniques for upper limb evaluation and rehabilitation of stroke patients," IEEE Trans. Neural Sys. and Rehab. Eng., 13 (3) (Sep): 311-24, 2005.

[5] A. Erol, G. Bebis, M. Nicolescu, R. D. Boyle, and X. Twombly. "Vision-based hand pose estimation: A review," Comp. Vis. Im. Und., 108: 52-73, 2007.

[6] A. R. Fugl-Meyer, L. Jaasko, I. Leyman, S. Olsson, and S. Steglind. "The post-stroke hemiplegic patient. 1. a method for evaluation of physical performance," Scand. J. Rehab. Med. 7 (1): 13-31, 1975.

[7] S. J. Housman, K. M. Scott, and D. J. Reinkensmeyer. "A randomized controlled trial of gravity-supported, computer-enhanced arm exercise for individuals with severe hemiparesis," Neurorehab. Neural Re., 23, (5) (Jun): 505-14, 2009.

[8] M.K. Hu, "Visual pattern recognition by moment invariants," IRE Trans. Info. Theory, 8 (2): 179-87, 1962.

[9] S. Kalsi-Ryan, A. Curt, M. G. Fehlings, and M. C. Verrier, "Assessment of the hand in tetraplegia using the graded redefined assessment of strength, sensibility and prehension (GRASSP): Impairment versus function," Top. Spinal Cord Inj. Rehabil., 14 (4): 34-46, 2009.

[10] H. I. Krebs, N. Hogan, M. L. Aisen, and B. T. Volpe. "Robot-aided neurorehabilitation," IEEE Trans. Rehab. Eng., 6 (1) (Mar): 75-87, 1998.

[11] T. Nef, M. Mihelj, and R. Riener. "ARMin: A robot for patient-cooperative arm therapy," Med. Biol. Eng. Comp., 45 (9) (Sep): 887-900, 2007.

[12] C. Sollerman, and A. Ejeskar. "Sollerman hand function test. A standardised method and its use in tetraplegic patients," Scand. J. Plastic Reconstruct. Hand Surg. 29 (2) (Jun): 167-76, 1995.

[13] J. D. Steeves, D. Lammertse, A. Curt, J. W. Fawcett, M. H. Tuszynski, J. F. Ditunno, P. H. Ellaway, et al. "Guidelines for the conduct of clinical trials for spinal cord injury (SCI) as developed by the ICCP panel: Clinical trial outcome measures," Spinal Cord, 45(3) (Mar): 206-21, 2007.

[14] B. T. Volpe, P. T. Huerta, J. L. Zipse, A. Rykman, D. Edwards, L. Dipietro, N. Hogan, and H. I. Krebs. "Robotic devices as therapeutic and diagnostic tools for stroke recovery," Arch. Neurol., 66, (9) (Sep): 1086-90, 2009.