Repeatability and Steadiness of Fingertip Force using Depth Feedback

Keming Cao, Gokul Swamy, Sayan Ray Chaudhuri, and Pamela Cosman
Dept. of Electrical and Computer Engineering, University of California, San Diego
La Jolla, California 92093
Email: k5cao@ucsd.edu, pcosman@ucsd.edu

***Abstract*—An inexpensive Kinect color/depth camera can be used to give feedback to physical therapy patients. We compare the steadiness and repeatability of finger force on a deformable object for physical therapists given no feedback, as well as for untrained subjects who are provided with feedback from force sensors, or with feedback from Kinect depth, or with no feedback. For force repeatability, it is found that untrained subjects and physical therapists perform comparably, and that both force feedback and depth feedback allow the untrained subjects to achieve a given force more repeatably than the physical therapists with no feedback. In contrast, for steadiness, physical therapists outperform untrained subjects when neither group has feedback. Force feedback, but not depth feedback, makes the untrained subjects steadier than the therapists. Objects of different deformability are also compared.**

I. Introduction

Patients undergoing physical therapy typically receive instructions from physical therapists (PTs) in a clinic, with printed pictures showing the correct movements. However, once the patient is at home, maintaining motivation and achieving the correct motion can be difficult, and compliance rates can be low [1]–[4]. Some systems use a Microsoft Kinect depth/color camera or other sensors to evaluate and give feedback on whether the exercise motion is correct, which can also help with motivation (e.g., [5]–[10]).

In addition to motion, therapy exercises often involve force production. A patient might squeeze a ball to strengthen hand grip, or a patient's home caregiver might push down on the patient's arm while the patient raises the arm in a resisted shoulder flexion exercise. In this work, we examine whether Kinect depth-based feedback can help people achieve a target force production. There is some previous work on estimating force on an elastic object using a grayscale camera [11] or using a depth camera [12]. We examine how well feedback based on depth information compares to no feedback and to force feedback in assisting untrained people to achieve a target force or maintain a steady force. Our metrics for assessing repeatability and steadiness come from our prior work [13], in which we examined force feedback but not depth feedback.

II. Experiment Design

A. Hand and finger segmentation

We consider the task of pushing on a sponge with the index finger, with the hand resting on a table. A Kinect camera is fixed about 30 inches above the table. Obtaining fingertip depth involves segmenting the hand region and localizing the fingertip. In [14]–[16], hand segmentation is performed by skin color matching with YCbCr color channels. Skin color segmentation is achieved with adaptive color matching in [15]. In [17]–[19], hand segmentation is achieved by simply thresholding the depth image, because in their application the hand is up in the air and does not contact any object.

A variety of methods have been used for fingertip exact position localization. An Adaboost classifier is used in [14], as there is only one finger pointing up in the image. In [15], [18], [20], k-curvature is used to find sharp outside corners in the mask, which correspond to fingertips when the hand is outstretched. In [17], k-means is used to cluster fingertip and arm points. Assuming the fingertip is semicircular, [16] applies a Circle Hough Transform after Canny edge detection, while [21] uses template matching. In [19], a 3D geodesic shortest path is used to locate the fingertip.

In our application, the hand and sponge are within a predetermined area on the table. The finger has a force sensor which is worn like a glove tip (see Fig. 1). Within the region-of-interest (ROI), we use k-means clustering with $k = 4$ for skin, background, sponge and sensor colors, where the known gray color of the sensor is used as one of the initial values. Around the fingertip point in the last frame, we pick out the largest connected component. After a binary opening operation with a 5×5 disk template to de-noise, a convex hull operation then finds the fingertip point as the convex hull point closest to the fingertip point in the previous frame. The operations are shown in Fig. 1.

B. User Interfaces

Our experiment aims to compare force repeatability and steadiness under three conditions: no feedback, force sensor feedback, and depth feedback. Our force sensor (fingerTPS from Pressure Profile Systems, Inc.) is shown in Fig. 2 (left). The GUI main window, shown in Fig. 2 (right), has a control panel on the left, a force bar with target line in the middle, and the real-time force curve plot on the right. The subject watches the center bar, whose height changes with the force feedback from the sensors.

The GUI for depth feedback is similar, with a depth bar with target line. The height of the bar changes based on the fingertip depth measurement, with some processing to minimize jumping. Specifically, let d_t represent the Kinect

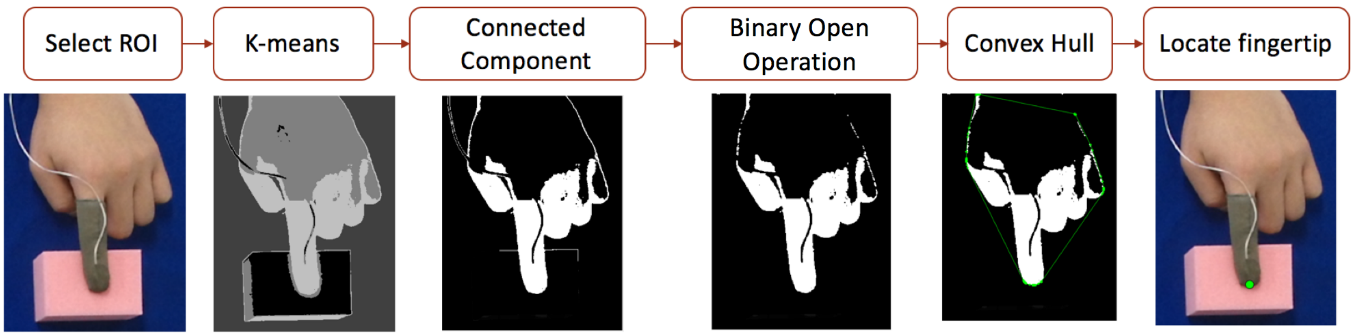

Fig. 1. Flowchart for fingertip detection

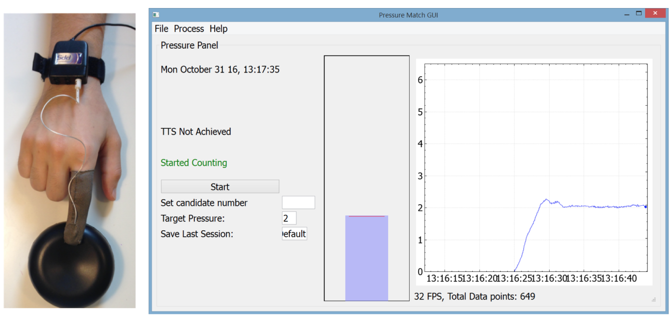

Fig. 2. FingerTip force sensor (left) and GUI for fingertip force sensor (right)

depth measurement temporally averaged over the 5 frames up to and including frame t and spatially averaged over the 5×5 pixel square around the fingertip point. To minimize fluctuations, the displayed bar height v_{visual}^{t+1} at time $t+1$ is kept unchanged unless the depth difference equals or exceeds a threshold T:

$$v_{visual}^{t+1} = \begin{cases} v_{visual}^{t} & \text{if } d_{t+1} - d_t < T \\ d_{t+1} & \text{otherwise} \end{cases} \quad (1)$$

where T was set to 3. The system operates at about 16 frames per second.

C. Depth-force relation for test objects

We used two therapy sponges (North Coast Medical Inc., product number NC59550, pink and yellow) with different softness. To obtain the force-depth relation curve, we collected 4 pairs of depth and force sequences for each sponge. Because the sampling rate is different for the force sensor (40/s) and depth camera (around 16/s), we interpolated the data pairs to the same length of 160 points. We fit the points with a 4th order polynomial:

$$force = a*\Delta_{depth}^4 + b*\Delta_{depth}^3 + c*\Delta_{depth}^2 + d*\Delta_{depth} + e \quad (2)$$

The data and fitted curves are shown in Fig. 3. For a given depth, the force for the pink sponge is higher than that of the yellow sponge, indicating that the pink sponge is stiffer. The parameters of the fit are in Table I.

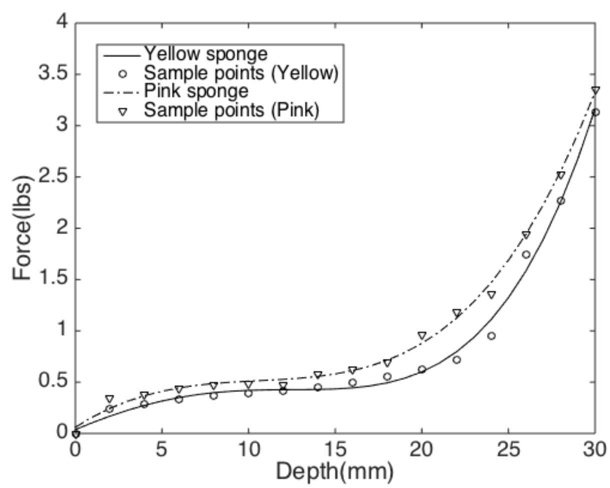

Fig. 3. Depth-Force mapping curves

TABLE I
PARAMETERS OF THE FIT FOR FORCE VS. DEPTH

	a	b	c	d	e
Pink	1.9754e-06	2.4166e-04	-0.0090	0.1091	0.0606
Yellow	1.0593e-05	-2.0505e-04	-0.0023	0.0708	0.0380

D. Data Collection

Our subjects are of two types: untrained people, and physical therapists (PTs). Untrained subjects pressed the sponge under three conditions: no feedback, force feedback, and depth feedback. Under the no-feedback condition, the subject would be asked first to press the sponge with moderate force to establish a reference value. They would then be asked to remember the force they used that first time and repeat it 5 times with no feedback. With force (or depth) feedback, we set a target of the same moderate force (or corresponding depth) that the person used for their no-feedback condition. The subject watches the feedback bar in the GUI and tries to make it reach and remain at the target line (corresponding to the target force or target depth). We record 650 samples (at 40 samples per second) for each press, which is about 16s.

We used 30 untrained subjects. Subjects signed an informed consent document and were paid \$10 for their time. The test used both the dominant and non-dominant hand (in randomized order). When the sensor was first put on each hand, a sensor calibration was performed so the output force values would be calibrated in pounds. The subject had a test run to familiarize themselves with the GUI. For each hand, the case of no feedback was done first to establish moderate force for that subject, and then depth feedback and force feedback were conducted in a randomized order.

Subjects (N=8) in the PT group were asked to press the sponge with no feedback for each hand and each sponge. They were not tested with feedback, since PT force steadiness and repeatability were considered to be, for physical therapy applications, a gold standard against which untrained subjects, both with and without feedback, would be compared.

III. Data Analysis

The distribution of force target values used by the subjects in the no-feedback case is shown in Fig. 4. Most of the target values are above 1 lb, where the depth-force curves of Fig. 3 are sensitive for depth. Example plots of measured force vs. time are shown in Fig. 5 for the cases of no feedback (left), force feedback (middle) and depth feedback (right). The horizontal line is the target line that subjects try to reach. Since it takes a few seconds to reach a stable state, we take the last 300 sample points (7.5s) in our analysis. For the case of no feedback, we take the time average of the first press as the target, and analyze the remaining 5 presses. So for all three conditions, we use 5 data trails for each hand/sponge case for each subject. We use μ_i $(i = 1, 2, 3, 4, 5)$ to denote the mean over time for the 5 trails.

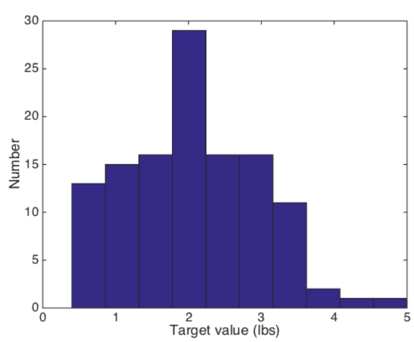

Fig. 4. Distribution of force target values

A. Force Repeatability

The degree to which the mean value of one trial is close to the mean value of another trial shows how well the subject can remember and repeat a force value. So we use the standard deviation from the target value of the 5 mean values as our measurement r of repeatability:

$$r = \sqrt{\frac{\sum_{i=1}^{5}(\mu_i - target)^2}{5}} \tag{3}$$

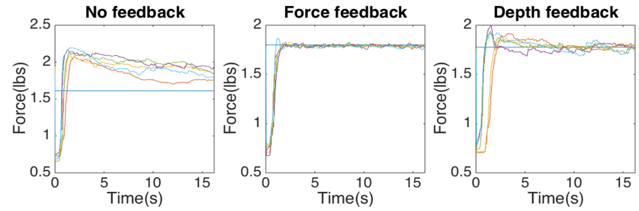

Fig. 5. Example curves of force vs. time

Perfect repeatability would correspond to $r = 0$. For the untrained subjects, we have 120 values of r for each feedback condition (30 subjects $\times$ 2 hands $\times$ 2 sponges), and we have 32 values for the PTs (no feedback). Fig. 6 plots these values.

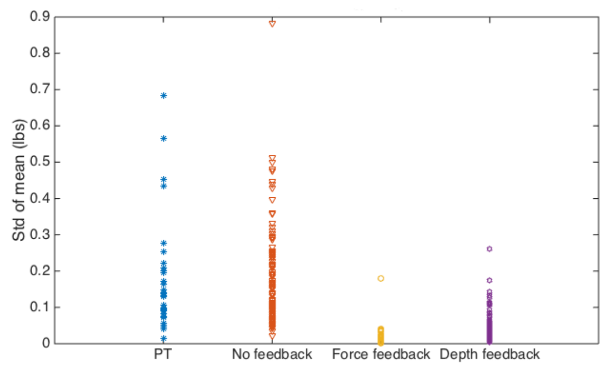

Fig. 6. Repeatability data for physical therapists and untrained people

Our data is paired (data is collected from the same person under different conditions), so we define the difference between no feedback (NF) and force feedback (FF) as $D1 = r_{NF} - r_{FF}$ between NF and depth feedback (DF) as $D2 = r_{NF} - r_{DF}$, and between DF and FF as $D3 = r_{DF} - r_{FF}$. To test for normality, the QQ-plots for D1, D2 and D3 are in Fig. 7, and a Jarque-Bera test rejects the hypothesis that they are normally distributed.

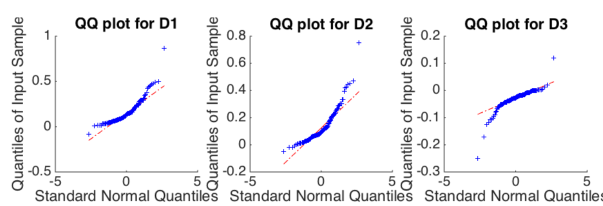

Fig. 7. QQ plot for $D1$, $D2$ and $D3$

We can apply a t-test because the sample size $n = 120$ is large, and we can also use a non-parametric test of significance. For all tests in this paper, we consider a significance level of 5%. We expect that feedback is better than no feedback, and that FF is better than DF, so we use a one-tailed t-test with null hypothesis that the mean difference is zero, and alternative hypothesis that the differences are positive. For $D1$, $D2$ and $D3$, the mean differences are 0.164, 0.132, and 0.032 (lbs), respectively, and the null hypothesis is rejected, with p $\ll$ .001. This means that untrained subjects, with the help of either force feedback or depth feedback, achieve better repeatability than in the no-feedback case, and also force feedback leads to better repeatability than depth feedback. The non-parametric Wilcoxon signed rank test considers the null hypothesis that median difference is zero, against the alternative hypothesis that the median difference is positive. For $D1$, $D2$ and $D3$, the median differences (lbs) are 0.136, 0.096, and 0.0268, respectively, and the null hypothesis is rejected, with p $\ll .001$, confirming the result of the t-test.

Comparing data from untrained people and PTs, we do not have paired data. The QQ plots are shown in Fig. 8. The Jarque-Bera test rejects normality, so we use the non-parametric Wilcoxon rank sum test for two independent samples. There are thee comparisons, NF and PT, FF and PT, DF and PT. For all three, the null hypothesis for the rank sum test is that the data come from distributions with the same median. For NF and PT, it is a two-tailed test. Since we expect that feedback will make untrained subjects perform better than PTs, for FF and PT data (likewise for DF and PT), this is a one-tailed test, with alternative hypothesis that the median for FF (likewise for DF) is less than that for the PT data. The test shows that with either depth feedback or force feedback, untrained subjects are more repeatable in their finger force production than PTs (p $\ll 0.001$). The test fails to reject the null hypothesis when comparing PT and NF data (p ≈ 0.5) so untrained people with no feedback are performing comparably to PTs with no feedback. Table II lists the mean and median values for NF, FF, DF and PT repeatability scores.

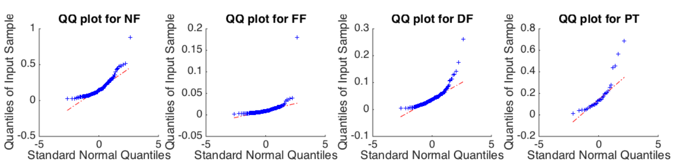

Fig. 8. QQ plot for NF, FF, DF and PT

TABLE II
MEAN AND MEDIAN VALUES OF REPEATABILITY

	NF	FF	DF	PT
mean (lbs)	0.1768	0.0124	0.0444	0.1769
median (lbs)	0.1455	0.0089	0.0368	0.1332

In comparing the two different sponges (paired data) we use a two-tailed t-test and signed rank test, for all the data combined (PT data, as well as data from untrained subjects with and without feedback). We fail to reject the hypothesis that the mean (or median) difference equals zero (p-values of 0.3239 and 0.1333). Apparently, the two different degrees of softness are not affecting the repeatability.

B. Force Steadiness

We use σ_i $(i = 1, 2, 3, 4, 5)$ to denote the standard deviation over time for the 5 trails. Since σ_i gives an indication of how steadily a person is holding the force, we use the average of the 5 σ_i values, denoted s, as our measure of steadiness. Perfect steadiness would correspond to $\sigma_i = 0$ for all i, and so the steadiness score s would be zero. Fig. 9 plots the steadiness scores for the PT, NF, FF, and DF data.

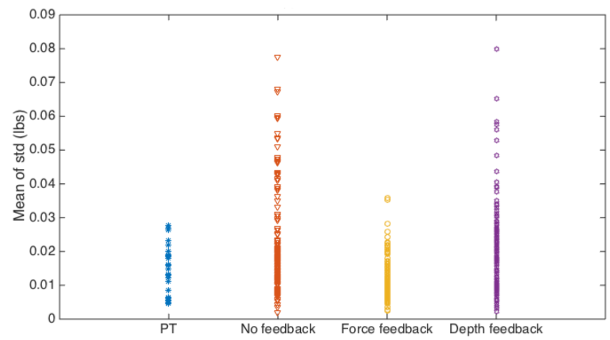

Fig. 9. Steadiness data for physical therapists and untrained people

We examine the paired differences as before, defining $D1 = s_{NF} - s_{FF}$, $D2 = s_{NF} - s_{DF}$, and $D3 = s_{DF} - s_{FF}$. The Jarque-Bera test rejects our null hypothesis that they are normally distributed.

When we use the t-test, because we expect that no feedback will have worse (higher values of) steadiness than feedback, and DF will be worse than FF, it is a one-tailed test for which the null hypothesis is that the mean difference is zero, against the alternative hypothesis that the mean difference is positive. The mean differences for $D1$, $D2$, and $D3$ are 0.0123, 0.0023, and 0.01. The null hypothesis is rejected for $D1$ and $D3$ with p-values $\ll 0.001$, which means that subjects produce a more constant force when they are given force feedback, compared to being given no feedback or depth feedback. We fail to reject the null hypothesis for $D2$, with p-value of 0.06. This means that depth feedback does not help the subject produce a more steady force, although since the p-value is 0.06, this is marginal in significance. The reason why DF has no improvement in steadiness over NF may be because the Kinect depth camera, when pointed at a motionless sponge on the table, produces raw depth readings that fluctuate by $\pm$3mm. With the spatial and temporal averaging prior to feedback display, those fluctuations were almost always within $\pm$1mm, but occasionally may cause the subject to adjust force. With the non-parametric Wilcoxon signed rank test, the results were similar to those from the t-test.

In comparing with the PTs, we have unpaired data, and we use the Wilcoxon rank sum test. In comparing NF and PT data, from the results of [13], we expect that PTs can be steadier than untrained subjects when nobody has any feedback, so the null hypothesis is that the medians are equal, against the alternative hypothesis that median(PT) $<$ median(NF). The null hypothesis is rejected with p=0.0026, so PTs are steadier in finger force production.

When the untrained subjects are given force feedback, we expect they can hold a force more steadily than PTs who have no feedback. So the null hypothesis is that the medians are equal against the alternative hypothesis that median(PT) $>$ median(FF). The null hypothesis is rejected with p=0.0088.

Finally, when untrained subjects are given depth feedback, it is not obvious whether they will be better or worse than PTs who have no feedback. So we use a two-tailed rank sum test. The null hypothesis is rejected with p=0.0084. So untrained subjects, given depth feedback, are not as steady in force production as the PTs who have no feedback. Table III lists the mean and median s scores for NF, FF, DF and PT.

TABLE III
MEAN AND MEDIAN STEADINESS SCORES

	NF	FF	DF	PT
mean (lbs)	0.0237	0.0114	0.0214	0.0145
median (lbs)	0.0187	0.0101	0.0185	0.0132

Lastly, in a comparison of the two sponges (paired data, two-tailed t-test), the null hypothesis of no difference is rejected (p $\ll$ .001). For the pink sponge, the mean s value is 0.0201, whereas for the yellow sponge it is 0.0168. It is easier to maintain steady force for the softer object.

IV. DISCUSSION AND CONCLUSIONS

Repeatability and steadiness are important in different situations. Steadiness is important for diagnosing certain illnesses, and plays an important role in many functional activites, such as holding a spoon with liquid, or grasping and writing with a pen. With no feedback, physical therapists were able to hold forces more steadily than untrained people. Our experiment found that the depth feedback in our system did not improve steadiness over no feedback, whereas force feedback did lead to significant improvement.

In other applications, repeating a given force level may be important when a patient is trying to strengthen certain muscles or regain range of motion. For example, a patient squeezes on a ball or presses on a bar on successive days at home, and is trying to make improvements, but does not know whether they are actually improving. After knee surgery, a patient may sit with legs extended and a physical therapist pushes down on the knee to help with regaining range of motion. In the clinic, the therapist could show the patient's spouse how to do this. But at home, the spouse is concerned about causing pain and is unsure whether the force being used is the same as in the clinic. For situations like these, a system would be useful which can assist a patient or home caregiver to achieve the same average level of force. Our results suggest that an inexpensive Kinect camera may help patients achieve average force targets. Both FF and DF allowed untrained people to be more repeatable on force than physical therapists. In this experiment, because the person was repeating their force 5 times over immediately after setting their own reference level, they would be most likely to remember it. In real situations, where a day may pass between the clinic visit and the home exercise, a person is less likely to remember the force level, so we expect that the difference between the no-feedback case and the feedback system could be even larger than was found in this study.

Acknowledgement: This research was supported by NSF under Award IIS-1522125.

REFERENCES

[1] K. Jack, S. M. McLean, J. K. Moffett, and E. Gardiner, "Barriers to treatment adherence in physiotherapy outpatient clinics: a systematic review." *Manual Therapy* 15.3 (2010): 220-228.

[2] S. M. McLean, M. Burton, L. Bradley, and C. Littlewood, "Interventions for enhancing adherence with physiotherapy: A systematic review." *Manual Therapy* 15.6 (2010): 514-521.

[3] L. Yardley, M. Donovan-Hall, K. Francis, and C. Todd, "Attitudes and beliefs that predict older peoples intention to undertake strength and balance training." *J. of Gerontology Series B: Psychological Sciences and Social Sciences* 62.2 (2007): 119-125.

[4] J. Miller, A. Litva, and M. Gabbay, "Motivating patients with shoulder and back pain to self-care: can a videotape of exercise support physiotherapy?" *Physiotherapy* 95.1 (2009): 29-35.

[5] V. Fung, A. Ho, J. Shaffer, E. Chung, and M. Gomez, "Use of Nintendo Wii Fit in the rehabilitation of outpatients following total knee replacement: a preliminary randomised controlled trial." *Physiotherapy* 98.3 (2012): 183-188.

[6] J. E. Deutsch, J. A. Lewis, and G. Burdea, "Technical and patient performance using a virtual reality-integrated telerehabilitation system: preliminary finding." *IEEE Trans. Neural Systems and Rehabilitation Engineering* 15.1 (2007): 30-35.

[7] E. J. Lyons, and C. Hatkevich, "Prevalence of behavior changing strategies in fitness video games: Theory-based content analysis." *J. of Medical Internet Research* 15.5 (2013): e81.

[8] S. Bamberg, A. Y. Benbasat, D. M. Scarborough, D. E. Krebs, and J. A. Paradiso, "Gait analysis using a shoe-integrated wireless sensor system." *IEEE Trans. on Info. Tech. in Biomedicine* 12.4 (2008): 413-423.

[9] R. Secoli, M. H. Milot, G. Rosati, D. J. Reinkensmeyer et al., "Effect of visual distraction and auditory feedback on patient effort during robot-assisted movement training after stroke." *Journal of Neuroengineering and Rehabilitation* 8.1 (2011): 1.

[10] Y. J. Chang, S. F. Chen, and J.D. Huang, "A Kinect-based system for physical rehabilitation: A pilot study for young adults with motor disabilities." *Research in Developmental Disabilities* 32.6 (2011): 2566-2570.

[11] X. Wang, G. K. Ananthasuresh, and J. P. Ostrowski, "Vision-based sensing of forces in elastic objects." *Sensors and Actuators A: Physical* 94.3 (2001): 142-156.

[12] B. Frank, R. Schmedding, C. Stachniss, M. Teschner, and W. Burgard, "Learning deformable object models for mobile robot navigation using depth cameras and a manipulation robot." *Proc. of the IEEE Intl. Conf. on Robotics & Automation (ICRA)*. 2010.

[13] K. Khan, D. Qin, D. Glaser, A. Nalci, C. Printz, C. McElroy, and P. Cosman, "Study of finger force variability for physical therapists and untrained subjects." *Signal Processing in Medicine and Biology Symposium (SPMB)* IEEE, 2014.

[14] B. H. Oh, J. H. An, and K. S. Hong. "Mobile User Interface using a Robust Fingertip Detection Algorithm for Complex Lighting and Background Conditions." *Int. Proc. of Computer Science and Information Technology* 27 (2012): 143-147.

[15] C. Wang, and B. Yuan. "Robust Fingertip Tracking with Improved Kalman Filter." *Intelligent Computing Theory*. Springer International Publishing, 2014: 619-629.

[16] M. Do, T. Asfour, and R. Dillmann. "Particle filter-based fingertip tracking with circular hough transform features." *Proc. of the 12th IAPR Conf. on Machine Vision Applications*, 2011

[17] Z. Feng, S. Xu, X. Zhang, L. Jin, Z. Ye, and W. Yang. "Real-time fingertip tracking and detection using Kinect depth sensor for a new writing-in-the air system." *Proc. of the 4th Intl. Conf. on Internet Multimedia Computing and Service*. ACM, 2012.

[18] F. T. Cerezo. "3D hand and finger recognition using Kinect." *The 19th ACM Conference on Computer and Communications Security*, 2012.

[19] H. Liang, J. Yuan, and D. Thalmann. "3D fingertip and palm tracking in depth image sequences." *Proc. of the 20th ACM International Conference on Multimedia*. ACM, 2012.

[20] T. Lee, and T. Hollerer. "Handy AR: Markerless inspection of augmented reality objects using fingertip tracking." *The 11th IEEE International Symposium on Wearable Computers*, 2007.

[21] K. Oka, Y. Sato, and H. Koike. "Real-time fingertip tracking and gesture recognition." *IEEE Computer Graphics and Applications* 22.6 (2002): 64-71.