Multidim Syst Sign Process (2013) 24:317–330
DOI 10.1007/s11045-012-0181-8

Tele-immersive environments for rehabilitation activities: an empirical study on proprioception

Kenton McHenry · Peter Bajcsy · Mike Frogley · Rob Kooper

Received: 31 March 2011 / Revised: 5 January 2012 / Accepted: 14 February 2012 /
Published online: 1 March 2012

Abstract Those with proprioceptive impairments can use their other senses as the proprioceptive feedback. We hypothesize that tele-immersion systems can aid in this supplemented proprioception by providing novel visual perspectives of one's own body. In particular we are interested as to whether the real-time 3D reconstructions used in tele-immersive systems are better or worse than conventional 2D views (i.e. video) with regards to aiding a task requiring proprioception. The objective of this study is to investigate the stated hypothesis by quantifying and ranking the various visual and auditory cues available in a tele-immersive system as they are used during an assigned task. The paper briefly describes a portable immersive VR system, our methodology for quantifying task performance, and results from our experiments with wheelchair basketball athletes.

Keywords Human factors in stereoscopy · Tele-immersion · Real time 3D reconstruction · Proprioception · Real world uses for 3D technology

1 Introduction

In the area of assistive technologies for the aged and disabled there have been several research efforts to show the use of virtual reality (VR) for educational and rehabilitation purposes (Hinton and Connolly 1992; Wilson and Foreman et al. 1996; Kumar and Rahman et al. 1997;

K. McHenry (✉) · P. Bajcsy · M. Frogley · R. Kooper
National Center for Super computing Applications, University of Illinois at Urbana-Champaign,
Urbana, IL, USA
e-mail: mchenry@illinois.edu

P. Bajcsy
e-mail: peter.bajcsy@nist.gov

M. Frogley
e-mail: frogley@illinois.edu

R. Kooper
e-mail: kooper@illinois.edu

Sveistrup 2004). VR spaces can be viewed as environments where humans are immersed into synthetic three-dimensional worlds which they can then move around and interact in 3D (Burdea and Coiffet 2003). The objectives of past research experiments have typically been to quantify the benefits of VR for various types of disabilities by doing tasks in such environments (Viau and Levin et al. 2004). These tasks would cover a wide range of activities including balance and posture, locomotion, upper and lower extremity functions, and exercise and pain tolerance (Sveistrup 2004). The assessment of the task performance aims at providing qualitative and quantitative measures about the performance improvements of daily living skills (meal preparation, spatial memory, and cognitive function) and about the transfer of training from a lab into a real life. The motivation behind our work is to design a methodology for quantitative evaluations of the VR spaces for rehabilitation activities, and to gather any evidence that VR spaces with extended immersive capabilities are a useful technology for supporting rehabilitation applications such as those focusing on regaining proprioception.

Proprioception refers to one's unconscious awareness of their body configuration, an important sense we often ignore until it is lost due to disabilities such as lower extremity paralysis (Jerosch and Prymka 1996). It has been shown that proprioceptive capability is critical for executing motor tasks (Fuchs and Thorwesten et al. 1999; Danion and Boyadjian et al. 2000). Past studies using VR assumed that presenting visual cues in VR improves spatial knowledge (Wilson and Foreman et al. 1996; Jung and Bajcsy 2006). We hypothesize that presenting this 3D+time visual information (i.e. dynamically changing 3D reconstructions) to human subjects with impaired proprioception should lead to further improvements in proprioceptive capability (Bajcsy and Frogley et al. 2009). With regards to the evaluation of VR spaces for such rehabilitation purposes one must consider those methodologies suitable for quantitative evaluations of the rehabilitation efficiency as well as the level of automation needed for evaluating large scale experiments. The methodology for evaluating VR spaces also maps closely to what human sensory inputs are being evaluated. Visual and/or auditory inputs are the only feedback signals in an immersive environment unless additional devices are introduced (e.g. haptic). According to (Hinton and Connolly 1992), "visual and auditory feedback signals are slower, less automated, and less programmed than the normal proprioceptive feedback." In light of this we expect some improvement in comparison with presenting lower dimensional cues such as 2D video or audio cues. The evaluation methodology we use involves simple floor markers that can be easily monitored to represent virtual walls. Visual and auditory cues are presented to human subjects who are asked to perform some task. While the task is performed accuracy and speed measurements are acquired automatically for later interpretation. The ability to automatically track collisions with the virtual wall markers is important in that it allows us to quantify results from a decent number of human trials in a completely objective manner. We describe the immersive system used as well as the details of this evaluation methodology in the sections that follow.

2 Tele-immersive system

One of the drawbacks of many VR technologies is the limited capability of real-time fusion of a physical world with a virtual world as well as the photo-realism of the real world rendering. Many VR spaces require participants to be equipped with tracking devices in order to acquire the location and orientation of a particular part of the body and fuse it with the virtual world (Wartell and Hodges 2002). Another common limitation is that VR systems tend to not be portable making it non-trivial to deploy at arbitrary locations. In

Fig. 1 *Left:* A TYZX G2 stereo camera with onboard CPU and networking. *Right:* An example of the color and depth images produced by the G2 camera

other words, the access to rehabilitation using VR technology has been limited due to limited availability, high cost, and complexity of deployment at places where the disabled or aged citizens spend significant amounts of time. Our VR environment was constructed in a manner to allow for photo-realistic rendering of a physical world in real time with a goal to deliver real-time fusion of the physical world with a virtual world without the need to equip humans with tracking devices. In addition we wished to do this in a way that is cost efficient and in a way that is easily portable to enable access to more communities.

Our efforts to develop a cheaper more portable tele-immersive system utilizes commercial off the shelf hardware solutions and tying them together in custom software. As described below this involves purchasing a number of stereo cameras and displays while focusing our own efforts on the needed software development. In order to transport our current system these cameras and displays will have to be moved along with a single machine. This machine is responsible for processing the data the cameras provide, creating a single 3D video stream, and transmitting this stream to the various displays which can be located locally and/or remotely.

We have purchased and custom configured an immersive system consisting of multiple stereo cameras, a server receiving depth and color streams from stereo cameras, and several large LCD displays for displaying the 3D+time information. The stereo cameras are manufactured by TYZX Inc. and consist of two CMOS imagers, a custom stereo application-specific integrated circuit (ASIC), a Field Programmable Gate Array (FPGA), a Digital Signal Processing (DSP) co-processor running the Linux operating system (OS), and onboard Ethernet (Fig. 1). The two imagers are constructed with a 6 cm baseline and have lenses that cover a 44° field of view. These stereo cameras produce depth maps and corresponding color images at a resolution of 500×312 at a frame rate of 33 frames per second.

Given the above hardware components we designed code for the optimal placement of stereo cameras given space and cost constraints, acquisition of the data from multiple stereo cameras, colorimetric calibration of color images, transformation of all obtained depth maps into a common 3D coordinate system, and rendering of 3D clouds of points in real time. The code for the geometric calibration of the stereo cameras has been provided by the ven-

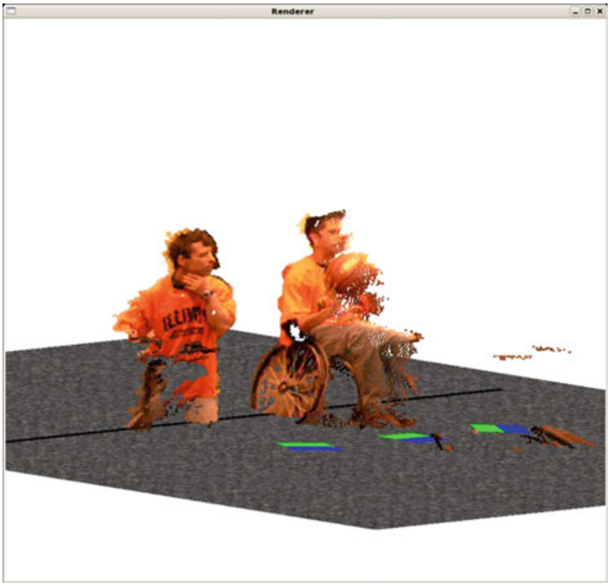

Fig. 2 An example of the 3D reconstruction produced by the TYZX stereo cameras overlaid on top of a virtual floor

dor but only provides a very coarse estimation of the camera locations and orientations. We limited our use of the registration code in our experiments. The resulting system represents the virtual world within a windowed application running on ones desktop (Fig. 2). Elements reconstructed from the 3D cameras are rendered within this window and manipulated with the mouse (e.g. rotated, scaled). This ability to move around the 3D space is something not available in conventional 2D video feeds. We expect the additional information to aid users in tasks within the virtual environment. Further information about our tele-immersive systems can be found in our other published works (Bajcsy and Johnson et al. 2008; Johnson and Bajcsy 2008; Malik and Bajcsy 2008, 2009; Lee and McHenry et al. 2009).

3 Methodology

For this study we collaborate with the wheelchair basketball team on the campus of the University of Illinois at Urbana-Champaign (UIUC). Those forced to be in a wheelchair lack the normal proprioception one has to know where their lower extremities are in space. This proprioception however is still important for maneuvering on a basketball court and for avoiding collisions that often occur between the wheelchairs. It is important for players to have some perception of where their lower body (and wheelchair) is in space. It is also important for them to do this without necessarily looking down at their lower body the whole time. Training and practice (i.e. repeated physical activity) are usual means of building a form of perception into wheelchair basketball players. Here we investigate a tele-immersive system as a means to provide cues for lower body proprioception. In terms of using supplemental perspectives of oneself to know how they are oriented in space, the 3D information provided by the tele-immersive system provides more than would be obtained from a 2D video (i.e. depth and other information derived from depth). We examine the usefulness of multiple sensory cues,

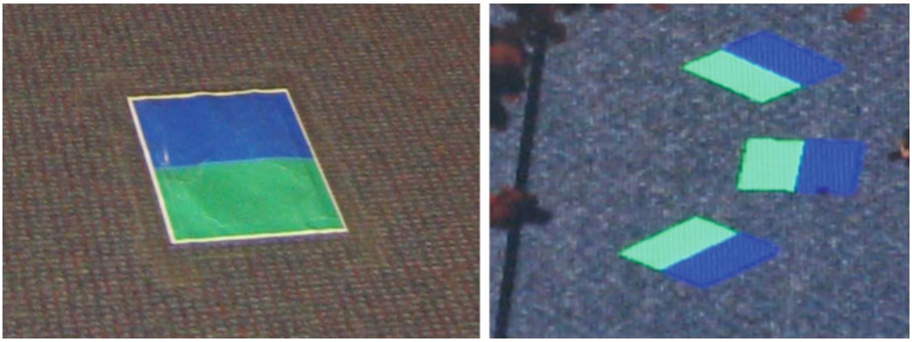

Fig. 3 The two colored floor markers used to represent virtual walls (*left:* physical marker taped to the floor, *right:* virtual markers shown on the floor of the tele-immersive environment). These fairly simple lambertian targets made up of pure *blue* (0, 0, 255) and pure *green* (0, 255, 0) can be found automatically. The intersection of the two colors represent a virtual wall that should not be crossed (Color figure online)

some obtained from the 3D depth information of the tele-immersive system and some from more conventional 2D feeds. To compare the cues we present each individually and quantify the users overall accuracy and speed at completing an assigned exercise.

The technical problems of quantitative evaluations of large numbers of rehabilitation exercises are addressed by using easily tracked colored markers to represent virtual walls and automatically assigning scores based on how close subjects approach these walls without colliding into them. In our experiments we use virtual walls represented by markers consisting of blue and green halves placed on the physical floor (Fig. 3). These lambertian markers made up of pure blue (0, 0, 255) and pure green (0, 255, 0), are fairly easy to find automatically by simply thresholding the color values of pixels within captured frames. Groups of blue and green can be found using simple clustering techniques such as K-Means and matched according to proximity in order to find the entire marker. The intersection between the two colors will represent a virtual wall that should not be crossed. An estimate of ones distance from the virtual wall can be determined automatically by observing the number of occluded green pixels on a particular marker from a suitably placed conventional 2D camera (e.g. the ceiling, Fig. 4). The system is initialized without anyone present in the view so the rough number of green and blue pixels on each marker is known beforehand. As a user approaches a marker they occlude the cameras view of some of the pixels making up that marker allowing us to estimate which virtual wall is being approached and roughly how close the user is to it. We detect collisions with the wall, as well as how far a user has passed through the wall by counting the number of occluded blue pixels. Since all we measure is the ratio of occluded colors the colors can be used interchangeably as far as the software is concerned. In our experiments however green is always placed closer to the user and blue further away, thus occluded blue will indicate collisions.

These virtual wall markers are registered across physical and virtual spaces in that they exist both on the physical floor and the virtual floor on the screen containing the users 3D reconstruction. Thus a user can see how close they are to the marker by either looking down in the real world or at the screen showing the reconstructed virtual world. The task subjects are asked to perform will involve moving a wheelchair to reach the virtual marker in such a way that one half (one color) of the marker is occluded by the wheelchair front as viewed by the ceiling camera. Again the goal here is to evaluate the usefulness of such 3D information as opposed to conventional 2D video feeds. Though in this scenario the user has the ability

Fig. 4 The floor markers observed from a ceiling mounted camera. Each pixel in the image is thresholded to separately find the bluest and greenest pixels in order to identify the markers. These pixels are then clustered in pairs to find the two halfs of the marker. In the image above the marker pixels are replaced with *pure blue* (0, 0, 255) and *pure green* (0, 255, 0) (Color figure online)

to look down if they wanted to one can envision scenarios were this is not a possibility due to the setup and only the 2D or 3D feed is available.

We will evaluate the effectiveness of the 3D information, or cue, by measuring the accuracy and speed by which users approach the virtual walls using it, ideally without colliding into the virtual walls. In order to draw comparisons we also measure the accuracy and speed by which users approach the virtual walls using a number of more conventional cues. The types of cues explored include:

- **3D Virtual Environment**—A 3D reconstruction of the user rendered on top a virtual floor containing virtual wall markers. The view can be adjusted so as to allow one to select a view best suited to gauge ones position in space to the markers. We set this view beforehand to a nearly overhead view so as to allow one to easily judge their distance from the markers.
- **2D Color**—A conventional 2D video feed from a camera in front of the user. We acquire this through a G2 camera using only the color video component. The camera is placed under the display the subjects look at during the experiments. Users must estimate their position based on their appearance in the video.
- **2D Depth Map** —A depth map feed from a camera in front of the user. We again acquire this through a G2 camera under the display using only the depth image produced by the camera. Depth can be derived based on the brightness of the pixels in the image, with brighter pixels being closer to the camera.
- **View changing (Color/Depth)**—Three cameras are used, one in front of each marker. As before either a color or depth map is shown to the user. This time however there are multiple cameras. A single view is still shown to the user where the view is selected based on the users proximity to the camera. Depth can be derived from this view by knowing

ones appearance in “smart” view, a view that is always in front of the target you are trying to approach.

- **Audio**—A tone is generated according to how close the user is to a particular target. The tone is generated using the ceiling camera tracking the targets. As users approach a particular target a tone is generated such that the frequency of the tone is higher if the user is closer to the target.
- **Top down**—We show the user the video feed from the ceiling camera used to track the interactions with the virtual wall markers. This particular 2D color view is ideally placed for this particular task allowing users to directly see how close they are two a particular target.
- **None** —No artificial cue is used. Instead the user is allowed to look directly at the floor to see how close they are to the virtual wall markers. We will use this cue as our baseline as we expect it to the one allowing for the highest accuracy since it is the most natural.

In our experiments we evaluate 7 cues: 3D, 2D color, 2D color with view changing, 2D depth map with view changing, audio, top down, and none. In addition we add a second depth video cue with view changing. We chose one cue, this cue being particularly foreign, to do twice to acquire a sense of the amount of learning taking place between trials. While presenting each cue, a human subject is asked to move from the base location past a black line on the floor to one of the virtual walls, stop at the boundary of green and blue as accurately as possible, come back to the black line and then proceed to the next target. There are three targets on the floor and three repetitions of the movements to the three targets. This task is based on the clover exercise used by wheelchair basketball players.

With the marker regions identified in the image marker scores can be tallied by counting the number of observed colored pixels with respect to those observed during an initialization phase (where the markers are guaranteed to not be occluded):

$$score_i\,(t) = \left|\frac{green_t}{green_{total}} - \frac{blue_t}{blue_{total}}\right| \times 100$$

In the above equation $green_{total}$ and $blue_{total}$ are the number of the colored pixels visible during initialization and $green_t$ and $blue_t$ are the number of colored pixels visible at time t during the experiment. The score during any time t is taken to be the maximum marker score (i.e. the one resulting from the virtual wall closest to the subject):

$$score\,(t) = \max_i(score_i\,(t))$$

where i is the index of the marker used (i = 1, 2, or 3 in our experiments with 3 markers). Scores are between 0 and 100 with high values being better indicating the user got right up to the wall without colliding into it. Because of the simple nature of the markers this process can be performed fairly reliably producing score versus time plots that we can then analyze (see Fig. 5).

The actual assessment of the accuracy of reaching each target is computed by considering the cases that can occur in the score plots as illustrated in Fig. 6. Notice it is not as simple as looking at the peaks. Single peaks are produced when the user gets near a virtual wall without colliding into it. On the other hand if the user does collide with the wall double peaks will be produced as the score drops based on how far the user passes through the wall (occluding more and more of the blue portion of the marker). If the user completely passes of over the marker two separate peaks will be produced, one as the user first approached the marker and one as the user re-approached the marker from the wrong side. If the user does not completely cross over the marker the two peaks will be connected. In this case the pit between the two

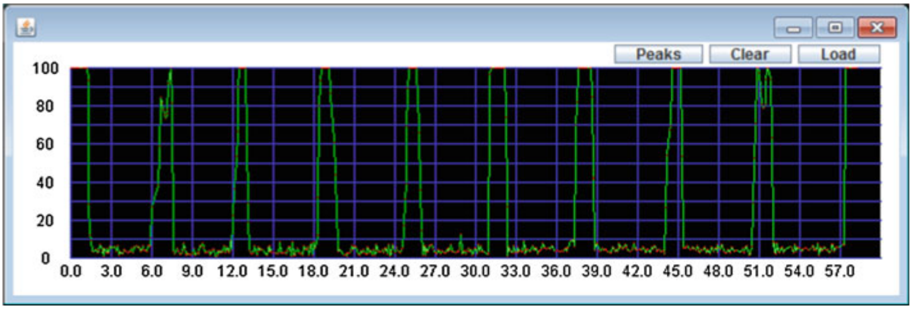

Fig. 5 An example of the score vs. time plots obtained from the subject experiments. Shown here is an ∼60 s run where a subject approached markers 10 times creating 10 peaks in the plot

Fig. 6 The four degenerate peak possibilities that can occur within the score plots. *Left* to *right:* a short peak occurring when a subject approaches the wall but never actually reaches it, a double peak occurring when a subject collides with the wall, no peak occurring when a subject fails to get close to the wall, and a split double peak occurring when a user not only collides with the wall but completely passes through it

peaks is the score we are interested in, indicating a penalized score proportional to how far they passed through the wall. The final case occurs when the user fails to get anywhere near the marker, producing no peaks. The cases of no peaks and completely separated peaks rarely, if ever, occur in our experiments thus we do now worry about identifying them.

In order to automate as much of the analysis as possible we begin by applying non-maximal suppression to the plots in order to obtain an initial set of peaks. Non-maximal suppression eliminates small peaks near larger peaks an acts a s a pre-processing step to remove noise from the score plot. Each remaining peak is then examined to see if it lies near another found peak and adjusted to sit at the bottom of the double peak valleys show in Fig. 6. As double peaks indicate collisions with the virtual wall we choose this valley, indicating how far through the wall the subject penetrated, as the score for the peak. The equation below describes the score assignment per approached virtual wall according to the detected peaks:

$$score_{approach} = \begin{cases} \max\left(score\left(t\right)\right) : t_{start} < t < t_{end} & \text{if one peak} \\ \min\left(score\left(t\right)\right) : t_{\max_1} < t < t_{\max_2}, t_{start} < t_{\max_1}, t_{\max_2} < t_{end} & \text{if two peaks} \end{cases}$$

Peaks are manually checked and adjusted if need be to ensure scoring is accurate. The average accuracy score per subject and per cue is computed from all scores assigned to each

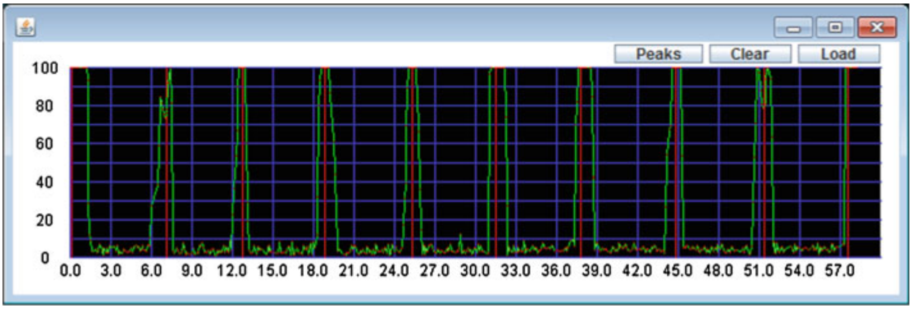

Fig. 7 An example of the score versus time plots obtained from the subject experiments with the peaks overlaid

target since each experiment consists of three targets and three repetitions (ten target scores overall when the starting location is included). In our results we use the average and standard deviation of these scores as a measure of overall accuracy. Figure 7 shows the peaks used to score one of the obtained score versus time plots.

4 Experiments

To quantify the usefulness of immersive environment cues towards proprioception in a real world environment we utilized the above methodology to evaluate the performance of 39 human suggests in wheelchairs as they carried out a clover exercise. In this exercise participants move from a starting point to one of 3 points in front of them. As they move to each point they attempt to stop as close as possible to the point, move back to the starting point, and move on to the next point. This is repeated 3 times for each of the points. In our experiment the starting point is a line on the floor and the 3 target points are virtual wall markers that will be used to automatically measure their accuracy at reaching the point. The participants attempt to do this task using only one of the cues described above, using the information it provides to judge where they are in the space and to guide them to each point (i.e. virtual wall marker). As mentioned previously we conduct the experiment with the view change depth cue twice in order to observe any effect learning may have between the trials. We also attempt to capture the effect of long term learning has on our obtained results by having a number of our subjects repeat all experiments at a later date.

In order to limit the amount of extra information available to the subjects performing the tasks, apart from the utilized cue, subjects are asked to wear sound dampening ear protectors during all but the audio experiments (preventing them from hearing the crunching of the paper targets as they ran over them). Subjects are asked to keep their eyes on the screen in all experiments other than the one using no artificial cues. Subjects are given the ability to practice the exercises prior to each cue so that they can become familiar with it. This is a necessary step for several cues such as the 2D color cue where the subject needs this practice in order to know how they appear in the 2D video feed when over the targets. We conducted 376 experiments (39 subjects times 8 cues plus 8 re-tests), acquiring 376 temporal accuracy recordings and extracting ∼3,760 accuracy scores for the times when subjects reached the targets.

The average accuracy score per cue and average speed to complete each task using each cue are shown in Fig. 8. To evaluate the effect of experience on performance subjects were

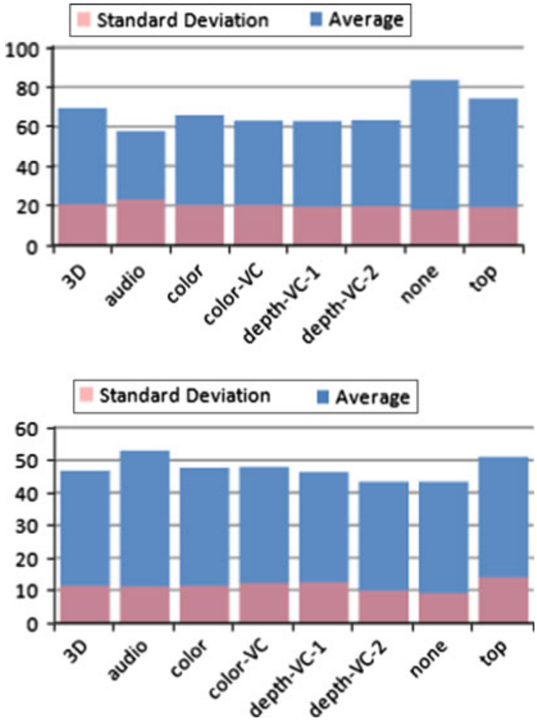

Fig. 8 *Top:* Average and standard deviation of accuracy score per cue. *Bottom:* Average and standard deviation of time per cue

asked to return two months after completing the initial experiment. The scores and speeds for these re-tests were then compared to their original scores and speeds. The average accuracy score per cue and average speed to complete each task using each cue are shown in Fig. 9. In these plots cues are labeled with a "1" if from the first experiment and "2" if from the second experiment. Note, we did a re-test of the depth with view change in each experiment so those are numbered "1, 2" for the first experiment and "3, 4" for the second experiment.

Based on the average accuracy values, the top three cues are "none", "top-down" fixed color video, and "3D". Based on the average time values, the top three cues are "none", a side view proximity-based depth video after learning, and a side view proximity-based depth video. The "none" cue is used as a baseline, being the most natural with the subject simply looking down at the floor, and thus was expected to be the best. It is also of little surprise that the "top down" view was 2nd best since the camera is perfectly placed for the objectives of this task (motivating our use of it for automated scoring). Having the "3D" cue with the next highest accuracy gives some weight to the usefulness of such information over more traditional 2D views. A Welch's t-test between the "none" and "top down" cues with each of the other cues shows all but two p-values less than the commonly accepted 5% threshold indicating statistical significance. The p-value between the "top down" results and the "color" results has a value of 6.4%. The p-value between the "top-down" results and the "3D" results has a value of 30.3% indicating almost a 1 in 3 chance that null hypothesis is true and that there is no difference between the two samples. A t-test between the "3D" cue against each of the other cues shows p-values greater than 5% for 4 of the other cues, with the "audio" and "none" cues having p-values much less than 5%. The three view changing cues have p-values between 14.2 and 17%. The "color" cue however has a p-value

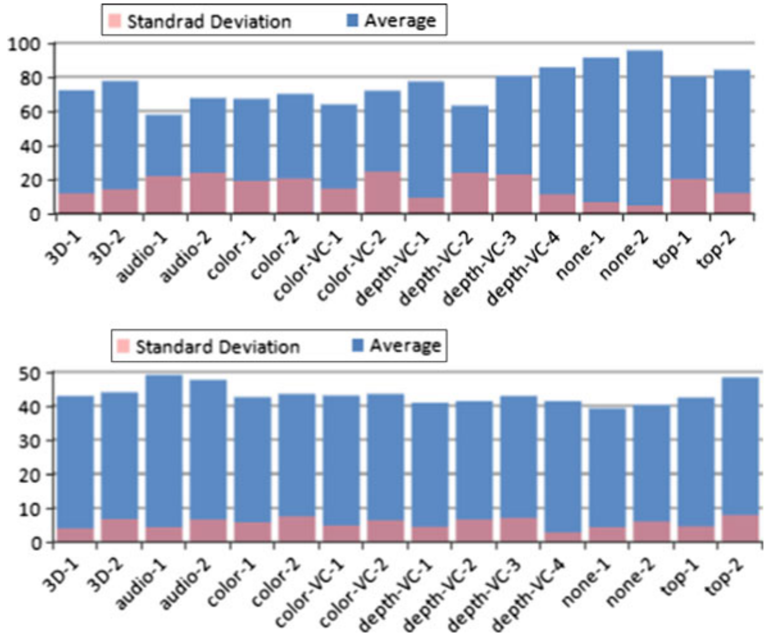

Fig. 9 Results for subjects who repeated the experiments 2 months later. *Top:* Average and standard deviation of accuracy score per cue. *Bottom:* Average and standard deviation of time per cue

of 42.2% indicating that the "3D" cue may not be better than the conventional 2D "color" view.

Learning does appear to have an effect on our results as well. In all but one case the average score for the subsequent tests were higher. The ordering of cues based on accuracy remains mostly the same as well. It is interesting to note, however, that the accuracy from the depth maps with view changes do appear to improve steadily after the first attempt, as if the users are training themselves to use this type information (which we imagine is somewhat foreign to most people).

Qualitative data, gathered in the form of questionnaires filled out by subjects after they completed the experiment, showed responses that were aligned with the quantitative results. Users on average rank the top down view as the best for the task with 3D being their next choice. The depth view, side view, and audio cues followed in that order. We note that the photo-realistic rendering capabilities of the VR space used for this experiment is far from ideal as the stereo cameras have a relatively low resolution. On average users felt that the responsiveness of the system was too low. Though the cameras can produce depth maps at $\sim$30 fps due to network limitations and rendering the resulting frame rate was $\sim$10 fps. This lag was found to be a hindrance to users as they tried to accurately approach targets solely from this input. Nonetheless, users still performed better with the 3D based information than other 2D based perspectives which had higher frame rates. Users rated the automated view switching tested with some of the cues lowest however. Subjects stated that they found it easier to learn what an accurate approach looked like in a single 2D view rather than having the view change. Subjects felt that the view change in fact added confusion rather than helping. When asked what feature they would most like to see improved within the VR system subjects on average ranked faster response time as number one, followed by spatial resolution. With regards to the intuitive nature of the cues, subjects found the top-down view by far the most intuitive, followed by the 3D view. With regards to achieving the best accuracy and speed however, subjects felt that “none” was best (our baseline) followed by the 3D view.

5 Discussion

The conducted experiments evaluated the effectiveness of a 3D virtual environment in gaining a notion of proprioception with regards to carrying out a particular task. Towards this goal we proposed a setup by which to automatically evaluate the accuracy of a carried out task and made the software available online[1] as a tool for future experiments. We used the proposed methodology to compare the 3D cues from the virtual environment to more conventional tele-communication cues such as 2D video used in video conferencing. In our experiments the “3D” cues were among the top 3 in terms of obtained accuracy, along with the “none” cue used as a baseline (and expected to be the best) and the fixed “top down” cue which was ideal for this given task. With regards to the “none” cue we would point out that the main drivers of tele-immersive systems have been long distance collaboration, education, and training. The value of merging users from distant locations into a common virtual environment is very high for citizens with disabilities since tele-immersive systems could minimize the travel needed to receive the best rehabilitation and/or training. In such remote training sessions, the “none” cue of looking down is simply not available. More interestingly, however, was the rank of the 3D cue with regards to the 2D video and audio cues representing the more conventional means for tele-communication. Among these the “3D” cue had an on average higher accuracy. However a statistical significance test did show that the “3D” results and the 2D “color” video results may not be significantly different. As noted by the users the resolution and responsiveness of the tele-immersive system was felt to be low.

A second round of experiments did show that learning did seem to take place between the experiments with average accuracy scores increasing for nearly all cues, suggesting some value in such information for a possible regaining of proprioception over time. As commercial off the shelf technology for 3D tele-immersive systems becomes more prevalent and cheaper we hope to build higher quality systems that may more clearly show a value in 3D information over conventional 2D video information towards proprioception.

Acknowledgments The research is funded by grant 490630 from the National Science Foundation IIS 07-03756. We would like to thank Dr. Brad Hedrick, the Director of the Division of Disability Resources and Education Services (DRES) at UIUC and Dr. Kenneth Lloyd Watkin from the Speech and Hearing Department at UIUC for their contributions to the overall project, and the following students for assisting during the experiments and with some data analyses: Rahul Malik, Hyu-Jung Na, Dong-Wook Kim and Kelly Veit.

References

Bajcsy, P., Frogley, M., et al. (2009). Design and use of immersive environments for regaining proprioceptive abilities. In *Workshop on multimedia aspects in pervasive healthcare in conjunction with 2009 IEEE international conference on multimedia and Expo* (*ICME*).

Bajcsy, P., Johnson, M., et al. (2008). Real time localization and 3D depth estimation across disparate sensing systems: Toward hazard aware and tele-immersive spaces. In *The 15th International conference on systems, signals and image processing* (*IWSSIP*).

Burdea, G., & Coiffet, P. (2003). *Virtual reality technology*. New York: Wiley-IEEE.

[1] https://isda.ncsa.uiuc.edu/svn/teeve/trunk/WallMarkerApplet.

Danion, F., Boyadjian, A., et al. (2000). Control of locomotion in expert gymnasts in the absence of vision. *Journal of Sports Sciences, 18*(10), 809–814.
Fuchs, S., & Thorwesten, L., et al. (1999). Proprioceptive functions in knees with and without total knee artroplasty. *American Journal of Physical Medicine and Rehabilitation*, *78*(1), 39–45.
Hinton, D., & Connolly, C. (1992). *Examining advanced technologies for benefits to persons with sensory impairments*. Chantilly: Scientific Applications International Corporation (SAIC).
Jerosch, J., & Prymka, M. (1996). Proprioception and joint stability. *Knee Surgery, Sports Traumatolgy, Arthroscopy, 4*(3), 171–179.
Johnson, M. J., & Bajcsy, P. (2008). Integration of thermal and visible imagery for robust foreground detection in tele-immersive spaces. In *The 11th International Conference on Information Fusion*, pp. 1–8.
Jung, S., & Bajcsy, R. (2006). A framework for constructing real-time immersive environments for training physical activies. *Journal of Multimedia, 1*(7), 9–17.
Kumar, V., Rahman, T., et al. (1997). Assistive devices for people with motor disabilities. In J. Webseter (Ed.), *Wiley Encyclopedia for Electrical and Electronic Engineers*, John Wiley & Sons.
Lee, K., McHenry, K., et al. (2009). "Characterizing human subjects in real-time and three-dimensional spaces by integrating thermal-infrared and visible spectrum cameras. In *Workshop on multimedia aspects in pervasive healthcare in conjunction with 2009 IEEE international conference on multimedia and Expo* (*ICME*).
Malik, R., & Bajcsy, P. (2008). Automated placement of multiple stereo cameras. In *OMNIVIS'2008, the 8th workshop on omnidirectional vision, camera networks, and non-classical cameras held in conjunction with ECCV 2008.*
Malik, R., & Bajcsy, P. (2009). Optimal stereo camera placement under spatially varying resolution requirements. In *2nd International conference on immersive telecommunications*.
Sveistrup, H. (2004). Motor rehabilitation using virtual reality. *Journal of Neuro Engineering and Rehabilitation*, *1*(1), 10.
Viau, A., Levin, M., et al. (2004). Reaching in reality and in virtual reality: A comparison of movement kinematics. *Journal of Neuro Engineering and Rehabilition*.
Wartell, Z., & Hodges, L. F. (2002). A geometric comparison of algorithms for fusion control in steroscopic HTDs. *IEEE Transactions on Visualization and Computer Graphics*, *8*(2), 129–143.
Wilson, R., Foreman, N., et al. (1996). Transfer of spatial information from a virtual to a real environment in physically disabled children. *Disability Rehabilitation*, *18*(12), 633–637.

Author Biographies

Kenton McHenry received a Ph.D. degree in computer science from the University of Illinois at Urbana-Champaign in 2008 after completing a B.S. degree from California State University of San Bernardino. He is currently a Research Scientist at the National Center for Supercomputing Applications (NCSA) and lead of the Image, Spatial, and Data Analysis (ISDA) group. His background is in computer vision with interests in the areas of image segmentation, object/material recognition, and 3D reconstruction. At NCSA he has applied this experience towards the task of building cyber-infrastructure. In collaboration with National Archives Applied Resreach Kenton has worked on a series of tools focused around the need for file format conversions in digital archives. One such tool, NCSA Polyglot, is a file format conversion service that was designed as an extensible, scalable, and practical solution to accessing data among the many contemporary (and legacy) file formats available. The Polyglot service is built on top of another tool developed by the ISDA group, the Software Server, a background processes that runs on desktop machines to provide API-like access to installed applications. These servers turn traditional desktop software into web based services and are what allow Polyglot to carry out a potentially large number of format conversions. More recently Kenton and the ISDA group are investigating means of providing immediate free searchable access to the upcoming 1940 census data release. Consisting of over 3 million un-transcribed digitized census forms the goal of this work is to explore methods of allowing users to find information within the image data without months of effort by thousands of human transcribers. Continuing with the theme of applied computer vision Kenton has recently become engaged in the GroupScope project. With the goal of studying large group behaviour, GroupScope brings together overlapping collections of video and audio and attempts to automate and/or aid in the identification of interactions between the people observed.

Peter Bajcsy received his Ph.D. in Electrical and Computer Engineering in 1997 from the University of Illinois at Urbana-Champaign (UIUC) and a M.S. in Electrical and Computer Engineering in 1994 from the University of Pennsylvania (UPENN). He is currently with the National Institute of Standards and Technology (NIST). Peter's area of research is large-scale image-based analyses and syntheses using mathematical, statistical and computational models while leveraging computer science fields such as image processing, machine learning, computer vision, and pattern recognition.

Mike Frogley has a Bachelor's degree in Secondary Education with an emphasis on Broad Field Social Studies and a Master's degree in Special Education with an emphasis in Learning Disabilities. Having played basketball prior to his accident, Mike began coaching at his former high school following the car accident that resulted in his disability in 1986. In the spring of 1988, he started playing wheelchair basketball for his local club team. In 1989, he began studying and playing wheelchair basketball at the University of Wisconsin-Whitewater. After completing his college athletic career, Mike spent one year as the interim head coach before being hired full-time. After coaching at UW-Whitewater for four years, Mike was hired to coach the men's and women's wheelchair basketball teams at the University of Illinois where he continues to coach the men's team today. His teams have combined for eleven national championships. Mike served as an assistant coach with the Canadian Women's National Wheelchair Basketball Team that won gold in the Atlanta Paralympics in 1996 and was the head coach of the Canadian Men's National Wheelchair Basketball Team that won gold in Sydney in 2000, Athens in 2004 and silver in Beijing in 2008 before stepping down to spend more time with his family. He has been involved in numerous studies examining physical activity and disability. He has conducted wheelchair basketball camps around the world; has written on wheelchair basketball; and has produced an instructional DVD in wheelchair basketball. He is actively involved in the development of wheelchair basketball at all levels.

Rob Kooper has earned his M.S. degree from the College of Computer Science, Georgia Institute of Technology, GA, 2001. During his masters degree he worked on context aware systems resulting in a seminal paper about context aware systems called CyberGuide, as well as systems for Augmented Reality and Virtual Reality which resulted in publications in the NY Times and showcased on PBS and CNN. Rob Kooper joined the National Center for Supercomputing Applications (NCSA) at the University of Illinois at Urbana-Champaign (UIUC) in 2001. Rob has worked on spectral image analyses and on management and retrieval of information from large databases using fuzzy statements. Currently he is working on scalability options using high performance computing (HPC) and cloud computing environments as well as on extraction of information from complex documents, analyses of the extracted information and visualization of large datasets.