Simulation versus Embodied Agents: Does either induce better human adherence to physical therapy exercise?

Douglas Brooks, Yu-ping Chen, and Ayanna M. Howard

***Abstract*— This research investigates proper movement correlation as well as the overall perception of human subjects' interaction with a simulated agent and an embodied agent in a physical therapeutic scenario. Using computer vision techniques coupled with the Microsoft Kinect to quantify reaching kinematics, correlation was assessed by aliging movements with a Vicon Motion Capture System as well as determining how well the specific exercises were mimicked. The results indicate that this approach is a viable alternative to Motion Capturing Systems for assessing certain movements during therapy. The results also indicate that there is some dependence on the use of an embodied agent as opposed to a simulated agent when assessing adherence.**

I. INTRODUCTION

Physical robots versus virtual agents? Which is better for improving back-and-forth interaction with humans, whether for education [1], rehabilitation [2], or just gaming [3]? With the advances in visualization and realism of virtual agents in 3D environments, some believe that much of what robots achieve with regards to human-agent interaction can also be achieved with an intelligent, less expensive, simulated version. Although robot enthusiasts would argue otherwise, by defending the practical and theoretical importance of physical embodiment in human-agent interaction, there are very limited studies that formalize the benefits in using physical versus virtual agents.

A. Brief Literature Review

The majority of the studies that address this issue typically focus on comparing differences based on elements of human perception and engagement. To date, researchers have solely used a survey analysis approach to derive an answer to the simulated versus physical agent question. For example, Takeuchi et al. [4] wanted to determine whether or not there were any clear advantages to using a simulated agent versus an embodied agent when presenting information to a human audience. The authors used an on-screen (simulated) agent in comparison to ASIMO (the embodied agent) to present a weather forecast using a multimodal presentation markup language. Using a post-session questionnaire, the authors found that ASIMO was rated higher in areas such as human likeness, goodness of the presentation, and the user's overall interest. However, the on-screen agent received higher scores for participant comprehension, focus on the presentation, and the agent's ability to accurately point to the objects.

Powers et al. [5] researched differences between simulated and embodied agents as they pertain to the disclosure of health or sensitive information. Using four different scenarios – Computer agent on a computer monitor, computer agent projected life-size on a screen, remote robot projected life-size on a screen, and collocated robot – in a between-subjects test, participants answered certain questions regarding general health habits. The researchers deduced that choosing between an embodied or simulated agent was very task specific. For tasks that involve a significant amount of information transmission but relatively little social rapport (e.g., information kiosks), disembodied agents should suffice. Likewise, for tasks that require users to reveal personal information, disembodied agents may be preferable. However, for tasks that are more relationship-oriented (e.g., a home companion), a collocated robot would seem to be best.

Finally, Lee et al. [6] questioned the significance of embodiment as it pertains to social agents and their tactile interaction. Using the touch sensors on the Sony Aibo [7] in comparison to mouse clicks when interacting with a virtual version of the same platform, subjects were allowed to interact with their respective agent for 10 minutes. Again using a post test survey, the researchers determined that physical embodiment has an added value for people's social interaction with agents by increasing the social presence.

B. The Role in Physical Therapy

While each of the aforementioned bodies of work give insight on this topic with respect to social interaction, none incorporate a physical metric based approach that would provide a more concrete analysis. More specifically, for therapeutic robotics, a metric based approach would provide a more concrete analysis regarding the physical attributes and effectiveness of the overall treatment. Therefore, the recent literature induces a question in the field of therapeutic robotics. Namely, "Does an embodied agent induce better human adherence to physical therapy exercise than a simulated agent regardless of user perception?" The intellectual merit of this question stems from the progression of using robotics to administer aid during physical rehabilitation [8]. As such, the research presented in this paper takes a novel approach of applying computer vision algorithms coupled with human-agent interaction to determine the significance of using a simulated agent versus an embodied agent during the mimicking of a physical exercise.

D. Brooks & A. Howard are with the Department of Electrical & Computer Engineering, Georgia Institute of Technology, Atlanta, GA 30308, USA douglas.brooks@gatech.edu & ayanna.howard@gatech.edu

Y. Chen is with the College of Health & Human Sciences, Division of Physical Therapy, Georgia State University, Atlanta, GA 30302, USA ypchen@gsu.edu

Section II details the methodology for applying the computer vision algorithms to quantify each subject's physical attributes based upon literature in rehabilitation. Section III describes the experimental protocol for administering the tests, and Section IV is a presentation of the results. Finally, Section V is a discussion of the implications from the resulting data, and Section VI concludes the paper with a direction for future work.

II. VISION-BASED ALGORITHMS TO QUANTIFY THERAPEUTIC EXERCISE

There are several methods that can be applied in order to determine proper adherence to physical exercise. Often, researchers utilize a motion capture system to accurately track a subject's limbs during the exercise and analyze the complete movement after the data has been stored [9]. However, when considering in-home treatment, it is impractical to administer this method of tracking. As an alternative, robust computer vision algorithms can enable segmentation and tracking of the patient's movements while simultaneously ensuring the feasibility of incorporating a cost-effective and space-efficient device (e.g. a camera). With the introduction of the Microsoft Kinect, obtaining depth information for a more detailed segmentation is relatively straightforward. Many researchers in the field of computer vision have migrated to utilizing this groundbreaking technology for several applications [10]. As such, this research uses the Microsoft Kinect's Infrared (IR) Projector in order to extract subjects' movements in 3-Dimensions (3-D). The next sections describes the method that was implemented for segmenting, normalizing, and analyzing subject upper-limb motions. This approach is a progression from previous work in [11].

A. Image Segmentation

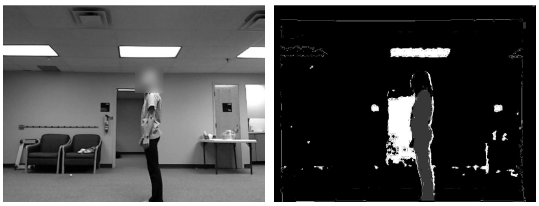
Many approaches to segmenting human motion within a sequence of images use pose estimation of the human at sequential times via a model [12], which generally requires a strong segmentation of the foreground and background as well as individual body parts, or a large dataset of previously stored poses [13], which require devices for sufficient data storage. In order to nullify the need for the previously mentioned processes, an algorithm known as the Multimodal Mean (MM) is used. A detailed explanation of the MM algorithm can be found in [14]; a summary is given here.

1) Multimodal Mean Overview: In the MM, each background pixel is modeled as a set of average possible pixel values. When subtracting the background, each pixel I_t in the current frame is compared to each of the background pixel means (averages) to determine whether it is within a predefined threshold. The background model for a given pixel is a set of K mean pixel representations called cells. An image pixel I_t is a background pixel if each of its color components is within a predefined threshold for that color component E_x of one of the background means [14].

Each background cell B_i is represented as three running sums for each color component $S_{i,t.x}$ and a count $C_{i,t}$ of how many times a matching pixel has been observed in t frames. It is a background pixel if a cell B_i can be found whose mean for each color component x matches within E_x of the corresponding color component of I_t:

$$\left(\bigwedge_x |I_{t.x} - \mu_{i,t-1.x}| \leq E_x \right) \wedge (C_{i,t-1} > T_{FG}), \qquad (1)$$

where T_{FG} is a small threshold number of times a pixel value can be seen and still considered to be foreground. When a pixel I_t matches a cell B_i , the background model is updated by adding each color component to the corresponding running sum $S_{i,t.x}$ and incrementing the count $C_{i,t}$. When a pixel I_t does not match cells at that pixel position, it is declared to be foreground [14].



(a) (b)

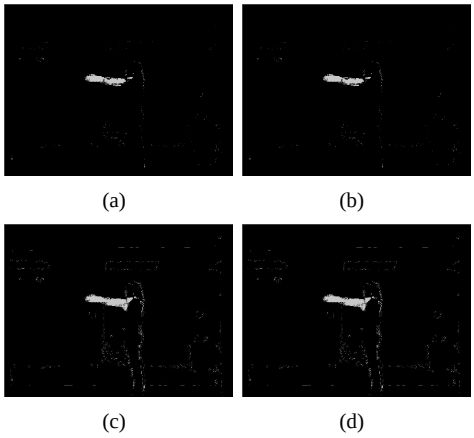
Fig. 1. (a)Original Image. (b)Result of MM algorithm as a standalone method for segmenting the IR projected image.

2) Analysis of the Multimodal Mean: The MM works well for typical red, green, and blue (RGB) and grayscale images. However, when applying this approach as a standalone process to the Microsoft Kinect's IR projected image, a noisy segmentation is produced. An IR projector is essentially an IR laser that passes through a diffraction grating and converts into many IR pixels. Each pixel is unique to the IR camera, which determines the pixel's coordinates, thus allowing real-time 3-D information. The downfall is that the IR projector produces quite a bit of noise due to lighting conditions, low resolution, and other factors, which undoubtedly causes the MM image segmentation process to produce a noisy segmentation as shown in Figure 1.

One way of alleviating this issue would be to segment the Kinect's RGB images using the MM algorithm and map the segmented pixels to corresponding pixels in the depth images. However, after testing this procedure, it was realized that often times the images produced by the Kinect's RGB camera were not properly aligned with those produced by the IR projector. Therefore, one would need to manually determine the proper alignment between the two cameras with each collection of data, nullifying the desire for autonomy. Hence, it was determined that it was best to use *a priori* information. If each person that interacts with the robot is located at a known distance, segmenting the person from the background is straightforward. By removing all data that is outside of the distance values produced by the IR projector, representing the location of the subject, a less noisy silhouette is produced, see Figure 2.

However, as seen in the figure, there are still a few artifacts produced by the IR projector. The noisy silhouettes

Fig. 2. Silhouette of subject given a known distance from the IR projector.



(a) (b) (c) (d)

Fig. 3. MM results for different values of MCDTH and CTH: (a) MCDTH = 0.001, CTH = 4, (b) MCDTH = 1.000, CTH = 4, (c) MCDTH = 0.001, CTH = 10, (d) MCDTH = 1, CTH = 10

(coupled with subject movement) were tested with the MM algorithm using different values for the maximum component difference threshold (MCDTH), which is the pixel difference threshold between frames, as well as the cell threshold (CTH), which is the number of cells in a given window used to calculate the recency value $R_{i,t}$ (how often a pixel matching cell B_i was observed within a recent window) [14]. Figure 3 illustrates the results of the various trials.

Still, each of the resulting images contained unwanted artifacts, and, if not chosen properly, the MCDTH and CTH values may diminish the desired arm segmentation. Thus, it was determined that a simple filter would be beneficial in order to remove excess noise prior to quantifying the patient's physical attributes. By performing a Gaussian blur, which is a type of image-blurring filter that uses a Gaussian function for calculating the transformation to apply to each pixel in the image [16], coupled with Suzuki's well-known process for border following [15] to extract and group the contour(s) within each image, the largest contour representing the subject's arm movement can be determined. The resulting image is shown in Figure 4.

B. Finding World Coordinates

Once the subject's upper-limb has been properly segmented, the next step is to determine the correlating world coordinates of each pixel representing the arm's location within the image. Classic computer vision algorithms determine the depth of specific objects using a RGB stereo camera pair. The designer's of the Microsoft Kinect (PrimeSense) have incorporated this aspect using a different approach,

Fig. 4. Resulting image after Gaussian blur and contour size threshold.

namely two cameras (one IR and one RGB) coupled with a laser-based projector. Since the IR-projector pair only returns the depth information of the segmented arm, traditional computer vision techniques have to be applied in order to extract the arm's world coordinates.

For a pinhole camera model, a 3D scene point P with world coordinates (X,Y,Z) is projected to a 3D point Q in the virtual image plane. By rescaling, the coordinates for Q with respect to P are can be determined using the pinhole camera model equations in classic computer vision books, see [16].

The same basic concept applies for a stereo camera pair with the exception that the cameras are shifted along the *x*-axis with a precalculated distance B between the two cameras. Using this known distance coupled with the pinhole camera model allows researchers to determine the depth of a pixel in the world by way of its image coordinates, and, by extension, the X and Y world coordinates are also calculated.

The Kinect disparity is related to a normalized disparity by the relation

$$d = \frac{1}{8}(d_{off} - k_d) \tag{2}$$

where d is a normalized disparity, k_d is the Kinect disparity, and d_{off} is an offset value particular to a given Kinect device. The factor $\frac{1}{8}$ appears because the values of k_d are in $\frac{1}{8}$ pixel units. The value for b is always about 7.5 cm, which is consistent with the measured distance between the IR and projector lenses, and d_{off} is typically around 1090 [17]. Utilizing this information coupled with classic computer vision algorithms, the calculations for finding the (X,Y,Z) coordinates of the arm becomes:

$$\begin{aligned} X &= \tfrac{(u-k_{xd})Z}{f} \\ Y &= \tfrac{(v-k_{yd})Z}{f} \\ Z &= Z \end{aligned} \tag{3}$$

where k_{xd} and k_{yd} are the Kinect disparities in the x and y directions respectively. These values are calculated via the calibration process found in [17]. Once the world coordinates have been calculated, the next step is to negate any variations in subject data that may be caused by the varying physique and location of each subject by normalizing all segmented images.

C. Image Moments and Affine Transforms

The details of the image normalization procedure were adopted directly from [18]. Let $I(x,y)$ denote a digital image

of size M x N. Its geometric moments m_{pq} and central moments $\mu_{pq}, p,q = 0,1,2,...$ are defined, respectively, as:

$$m_{pq} = \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} x^p y^q I(x,y) \tag{4}$$

and

$$\mu_{pq} = \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} (x-\overline{x})^p (y-\overline{y})^q I(x,y) \tag{5}$$

where

$$\overline{x} = \frac{m_{10}}{m_{00}}, \overline{y} = \frac{m_{01}}{m_{00}}. \tag{6}$$

An image $I^*(x,y)$ is said to be an affine transform of $I(x,y)$ if there is a matrix $A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$ and vector $d = \begin{pmatrix} d_1 \\ d_2 \end{pmatrix}$ such that $I^*(x,y) = I(x_a, y_a)$, where

$$\begin{pmatrix} x_a \\ y_a \end{pmatrix} = A \cdot \begin{pmatrix} x \\ y \end{pmatrix} - d. \tag{7}$$

Other examples of affine transforms include: 1) shearing in the x direction, which corresponds to $A = \begin{pmatrix} 1 & \beta \\ 0 & 1 \end{pmatrix} \triangleq A_x$; 2) shearing in the y direction, $A = \begin{pmatrix} 1 & 0 \\ \gamma & 1 \end{pmatrix} \triangleq A_x$; 3) scaling in both x and y directions, which corresponds to $A = \begin{pmatrix} \alpha & 0 \\ 0 & \delta \end{pmatrix} \triangleq A_s$. Where β, γ, α, and δ are arbitrarily defined shearing and scaling factors.

D. Image Normalization

The general concept of image normalization using moments is well-known in pattern recognition problems [19], where the idea is to extract image features that are invariant to affine transforms). In this application, a normalization procedure is applied to the image so that it meets a set of predefined moment criteria. By applying the same concepts in Section II-C, which are shearing and scaling in the x and y directions, we can also normalize each image to a standard location and size.

The scale factor for each subject was created by taking the average pixel width and height of each subject's segmented movements and determining the factor of difference between the ground truth width and height averages obtained *apriori*. The pixel width and height of the segmented movements were determined by using a bounding rectangle based upon the top-most, bottom-most, left-most, and right-most pixels within an isolated segmentation. Also, calculating the average image centroid remains the same. This information can be used to scale and shift each image representing the subject to a standard size and location.

E. Kinematic Metrics

The motion for determining proper adherence consisted of an upward reaching movement with the elbow bent 90° as the starting point for one repetition as shown in Figure 5. Reaching is typically analyzed by five kinematic characteristics: movement time (MT), total displacement (TD), peak velocity (PV), movement units (MU), and normalized jerk score (NJS) [20]. MT is calculated by dividing the number of frames collected during movement by the known frame rate of the Kinect. The TD can be computed because world coordinate information has already been calculated as described earlier. To determine the TD, one only needs to calculate the distance between the beginning and ending position of the segmented arm during the motion. The researchers chose to calculate the Manhattan distance between the left-most non-zero pixel's beginning and ending position.

(a) (b)

Fig. 5. Illustrations of the a)starting and b)ending positions of the reaching exercise.

PV can be calculated by dividing the TD by the time at each frame and taking the maximum value. However, a peak angular velocity (PAV) would allow better tracking and an easier calculation of the MU and NJS. To calculate the PAV, we first used a previous approach from earlier work found in [11] to determine the position or Active Range of Motion for each subject given the normalized segmentation. As a result, finding the PAV involves taking the derivative of the position data. MU are obtained from the acceleration and deceleration data, obtained by taking the derivative of the PAV data, during the motion; one acceleration and deceleration phase comprise one movement unit. The NJS is calculated as follows:

$$NJS = \sqrt{\frac{1}{2} \cdot \int j^2(t)dt \cdot \frac{d^5}{l^2}} \tag{8}$$

where j is the third time derivative of position data, d is the movement duration, and l is the movement amplitude [21].

III. EXPERIMENTAL PROTOCOL

A. Procedure

Testing took place in the Motor Development Lab in the Division of Physical Therapy at Georgia State University. Each subject was given an IRB approved consent form informing him or her of the testing procedure. After consent was given, the testing procedure began. The procedure was conducted as follows:

Task 1:

Subjects were shown an on-screen simulated robot. Subjects were told to watch the movement of the simulated robot and to repeat its movement once it has stopped.

Task 2:

Subjects were shown an embodied robot. Subjects

were told to watch the movement of the embodied robot and to repeat its movement once it has stopped.

Each task consisted of one action, reaching, as shown in Figure 5. The testing procedure was a between-subjects study with each subject interacting with (mimicking) both agents in a random order (i.e. the simulated agent performed the action first or the embodied agent performed the action first). Each subject mimicked each agent only once for a total of two mimicking actions.

B. Apparatus

All participants stood directly in front of the simulated agent, depicted on a projector screen, and the embodied agent. A 100Hz six-camera Vicon (370) Motion Capture System (VMCS) recorded position data of the subjects' joints. Three small reflective markers (9 mm in diameter) were attached to each subject's scapula (shoulder), lateral epicondyle of the humerus (elbow), and ulnar styloid process (wrist). VMCS cameras were placed around the participant at a distance of 5m in order to track the reflective markers simultaneously. The VMCS cameras in relation to the testing were initially calibrated by using a calibration rod. In addition to the kinematic data, movements were recorded throughout the collection session with a Microsoft Kinect, which allows collection of depth information for the purpose of comparing it with the kinematic data. Thus, motion was captured via two methods.

C. Simulated Agent

The simulated agent depicted a model version of the Manoi humanoid introduced by Kyosho [22] and was projected onto a projector screen. The simulated agent was constructed in-house using the Robot Operating System (ROS) in conjunction with Rviz (a graphics simulator used to depict model robots). When given the command via a laptop, the agent performed a prescripted upper-limb reaching motion (as described in Section II-E) for one-half a repetition.

D. Embodied Agent

The embodied agent that was utilized was the Darwin-OP (Dynamic Anthropomorphic Robot with Intelligence - Open Platform) [23]. It is 455mm (17.9in) in height and weighs 2.8 kg (6.2 lbs). It is equipped with Robotis Dynamixel motors, a FitPC for computing, and LiPo batteries for power. The Darwin-OP was also prescripted to perform the same upper-limb reaching motion (as described in Section II-E) for one-half a repetition using a ROS package designed in-house by the Human Automation Systems (HumAnS) Laboratory at Georgia Tech. The entire testing setup, which includes each apparatus and both agents, can be seen in Figure 6.

IV. RESULTS

Eleven healthy young adults (10 females, one male) between the ages of 19 and 22 performed each task – once per agent mimicking. The first step was to show the legitimacy of the image processing approach for calculating the kinematic metrics. This was achieved by calculating the ablsolute error between the VMCS and the image processing data for each subject. The absolute error is calculated by computing the absolute value of the difference between the image processing data and the VMCS data and dividing by the VMCS data, which again is the ground truth data of each subject. Stated explicitly, here we compare the VMCS exercise data of Subject X with the image processing data of the same subject (Subject X). In other words, the capturing systems are different, but the movements are exactly the same. This is considered the ground truth data set.The average error of all subjects for each kinematic metric is shown in Table I.

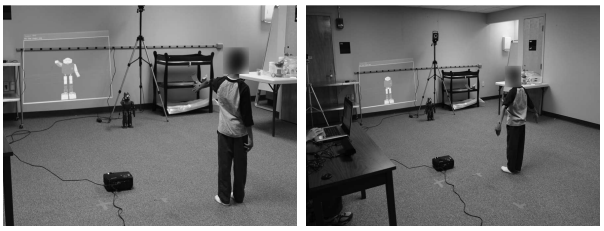


Fig. 6. Illustrations of the test setup.

TABLE I
AVERAGE PERCENTAGE ERROR KINEMATIC DATA FOR VMCS VERSUS IMAGE PROCESSING.

Parameter	Error	Std
MT	2.54	1.21
TD	19.69	6.46
PAV	12.06	0.29
MU	0.00	0.00
NJS	9.22	0.12

In order to determine proper correlation with regard to interaction with the simulated agent versus the embodied agent, the average error of all subjects for the kinematic metric data was calculated for each interaction. The proper movement for the metrics was obtained from the prescripted motion that each agent mimicked during testing (as stated earlier), which stemmed from the pre-recording of an expert's kinematic metric data obtained via the same method. In other words, each agent performed actions with the same kinematic metrics as an expert and each subject mimicked those actions. Stated explicitly, here we compare the image processing data of Subject X with the image processing data of Subject Y (a human expert's kinematic data set). Table II is a presentation of the resulting data.

V. DISCUSSION

The results of the study seem to indicate that subjects relate better to the speed of movement with the simulated agent than with the embodied agent. However, each subject was only asked to mimic the overall motion of each agent. Subjects were not told that they should also try to mimic the movement speed and displacement of each agent. While, each subject did indeed perform the basic actions that were demonstrated, most subjects did not focus on certain intricacies. More specifically, the subjects were not attentive of

TABLE II

AVERAGE PERCENTAGE ERROR FOR KINEMATIC DATA FOR SIMULATION VERSUS EMBODIED INTERACTION.

Parameter	Simulation %	Std	Embodied %	Std
MT	14.70	1.87	31.37	1.14
TD	9.26	7.80	6.72	6.11
PAV	7.77	0.63	18.38	1.07
MU	0.00	0.00	0.00	0.00
NJS	17.27	0.26	5.29	0.33

the exact height and speed of the demonstrated movements that were created from a human expert's kinematic data set, which, based upon observation, effected the successive kinematic matching. Perhaps more important in this scenario are simply the magnitudes of difference between the interactions with the simulated and embodied agents.

Perhaps the most interesting result is that of the NJS. It appears that the amount of jerk was more closely related to the embodied interaction than with the simulated counterpart. It is suspected that the ability of humans to decipher movements in general may lead to some degree of proper mimicking regardless of the agent. It is possible, however, that more difficult exercises would require the instruction of an agent capable of providing a real-world perspective rather than a virtual view, such as the perspective provided by the embodied agent.

VI. CONCLUSIONS AND FUTURE WORKS

In this paper, we have presented a novel approach for extracting and quantifying reaching movements using a single depth camera. The researchers have also utilized this process to delve into the topic of simulated versus embodied agents in the realm of physical therapeutic assistance, specifically focusing on the ability of humans to properly adhere to exercise demonstrations. The future direction of this research is to incorporate more subjects as well as more complicated and longer exercises. More specifically, the researchers intend to include testing from child subjects in order to gauge the effectiveness and overall appeal regarding one agent versus the other. The long-term goal is to ensure proper correlation during self-directed exercises for in-home activities.

VII. ACKNOWLEDGMENTS

This work was supported in part by the National Science Foundation under Grant CNS-0958487. Any opinion, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

REFERENCES

[1] H. Tuzun, M. Yilmaz-Soylu, T. Karakus, Y. Inal, and G. Kizilkaya, "The effects of computer games on primary school students' achievement and motivation in geography learning," *Computers & Education*, vol. 52, no. 1, pp. 68–77, 2009.

[2] M. Cameirão, S. Badia, E. Oller, P. Verschure, *et al.*, "Neurorehabilitation using the virtual reality based rehabilitation gaming system: methodology, design, psychometrics, usability and validation," *Journal of neuroengineering and rehabilitation*, vol. 7, no. 1, p. 48, 2010.

[3] S. Park, S. Ji, D. Ryu, and H. Cho, "A smart and realistic chatting interface for gaming agents in 3-d virtual space," in *2008 International Conference on Games Research and Development 2008 (Cyber Games), To appear*, 2008.

[4] J. Takeuchi, K. Kushida, Y. Nishimura, H. Dohi, M. Ishizuka, M. Nakano, and H. Tsujino, "Comparison of a humanoid robot and an on-screen agent as presenters to audiences," in *Intelligent Robots and Systems, 2006 IEEE/RSJ International Conference on*. IEEE, 2006, pp. 3964–3969.

[5] A. Powers, S. Kiesler, S. Fussell, and C. Torrey, "Comparing a computer agent with a humanoid robot," *Human Robot Interaction 2007*, pp. 145–152, 2007.

[6] K. Lee, Y. Jung, J. Kim, and S. Kim, "Are physically embodied social agents better than disembodied social agents?: The effects of physical embodiment, tactile interaction, and people's loneliness in human-robot interaction," *International Journal of Human-Computer Studies*, vol. 64, no. 10, pp. 962–973, 2006.

[7] J. Pransky, "Aibo-the no. 1 selling service robot," *Industrial robot: An international journal*, vol. 28, no. 1, pp. 24–26, 2001.

[8] A. Koller-Hodac, D. Leonardo, S. Walpen, and D. Felder, "A novel robotic device for knee rehabilitation improved physical therapy through automated process," in *Biomedical Robotics and Biomechatronics (BioRob), 2010 3rd IEEE RAS and EMBS International Conference on*. IEEE, 2010, pp. 820–824.

[9] M. Murphy, C. Willén, and K. Sunnerhagen, "Kinematic variables quantifying upper-extremity performance after stroke during reaching and drinking from a glass," *Neurorehabilitation and neural repair*, vol. 25, no. 1, p. 71, 2011.

[10] E. Stone and M. Skubic, "Evaluation of an inexpensive depth camera for passive in-home fall risk assessment," in *Pervasive Computing Technologies for Healthcare (PervasiveHealth), 2011 5th International Conference on*. IEEE, 2011, pp. 71–77.

[11] D. Brooks and A. Howard, "A computational method for physical rehabilitation assessment," in *Biomedical Robotics and Biomechatronics (BioRob), 2010 3rd IEEE RAS and EMBS International Conference on*. IEEE, 2010, pp. 442–447.

[12] H. Graf, S. Yoon, and C. Malerczyk, "Real-time 3d reconstruction and pose estimation for human motion analysis," in *Image Processing (ICIP), 2010 17th IEEE International Conference on*. IEEE, 2010, pp. 3981–3984.

[13] J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman, and A. Blake, "Real-time human pose recognition in parts from single depth images," in *CVPR*, vol. 2, 2011, p. 3.

[14] S. Apewokin, B. Valentine, D. Forsthoefel, L. Wills, S. Wills, and A. Gentile, "Embedded real-time surveillance using multimodal mean background modeling," *Embedded Computer Vision*, pp. 163–175, 2009.

[15] S. Suzuki *et al.*, "Topological structural analysis of digitized binary images by border following," *Computer Vision, Graphics, and Image Processing*, vol. 30, no. 1, pp. 32–46, 1985.

[16] L. G. Shapiro and G. C. Stockman, *Computer Vision*. Prentice Hall, 2001, retrieved from Wikipedia.

[17] K. Konolige and M. P., "Kinect calibration - technical - ros wiki," dec 2010, http://www.ros.org/wiki/kinect_calibration/technical.

[18] P. Dong, J. Brankov, N. Galatsanos, Y. Yang, and F. Davoine, "Digital watermarking robust to geometric distortions," *Image Processing, IEEE Transactions on*, vol. 14, no. 12, pp. 2140–2150, 2005.

[19] D. Shen, H. Ip, K. Cheung, and E. Teoh, "Symmetry detection by generalized complex (gc) moments: a close-form solution," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 21, no. 5, pp. 466–476, 1999.

[20] J. Chang, T. Wu, W. Wu, and F. Su, "Kinematical measure for spastic reaching in children with cerebral palsy," *Clinical Biomechanics*, vol. 20, no. 4, pp. 381–388, 2005.

[21] J. Alberts, M. Saling, C. Adler, and G. Stelmach, "Disruptions in the reach-to-grasp actions of parkinson's patients," *Experimental brain research*, vol. 134, no. 3, pp. 353–362, 2000.

[22] "1/5 scale athlete humanoid series manoi," mar 2012, http://www.kyosho.com/jpn/products/robot/at01/at01.html.

[23] Robotis, "Robotis - [darwin-op] open platform humanoid project," dec 2010, http://www.robotis.com/xe/37937.