International Journal of Information, Control and Computer Sciences
ISSN: 2517-9942
Vol:8, No:11, 2014

Motion Capture Based Wizard of Oz Technique for Humanoid Robot

Rafal Stegierski, Krzysztof Dmitruk

Abstract—The paper focus on robotic telepresence system build around humanoid robot operated with controller-less Wizard of Oz technique. Proposed solution gives possibility to quick start acting as a operator with short, if any, initial training.

Keywords—Robotics, Motion Capture, Wizard of Oz, Humanoid Robots, Human Robot Interaction.

I. Introduction

EASY way, eg. gesture or manipulator-less, teleoperated humanoid robot could be very good, if not the best, way to build up robotic telepresence system. We know how humanoid robot should appear and act. Creating of robot which will be accepted by human partner has to take place on many different levels. We should consider many aspects from mechanical one, eg. limbs with similar degrees of freedom and functionality as human joints, through facial expressions and voice system corresponded with emotions, to, least but not last, motivation system associated tightly with behaviour [1]. The last part should correspond to general purpose and appearance of a robot. We expect different actions and reactions from a child than from an adult [19].

To overcome lack of possibility to build up fully autonomous humanoid robot with wide range of reactions to examine dyadic or triadic interaction between a human and a robot Wizard Of Oz technique with human operator is usually used [14],[16]. In simple variant there is of course also possibility to implement pseudo random, but reasonable behaviours [18].

Of course all the time we should remember that any kind of interaction with human is biased not only by social but also ethical problems. There is all the time thin line which should not be crossed. It is especially disturbing when the fact of teleoperation is hidden with technological facade and nature of experiment is unclear for participants. In some cases possibility to ”lifting the curtain” change the result but hiding the fact is ethically doubtful [13]. Of course the is also the other side of coin and we could point situations where this way of communication could be lesser evil like has it place in interaction with autistic people [14], [20].

R. Stegierski is with the Institute of Computer Science, Faculty of Mathematics, Physics and Computer Science, Maria Curie Sklodowska University, Lublin, Poland (e-mail: rafal.stegierski@umcs.pl, ORCID: 0000-0001-7225-3275).

K. Dmitruk is with the Institute of Computer Science, Faculty of Mathematics, Physics and Computer Science, Maria Curie Sklodowska University, Lublin, Poland (e-mail: krzysztof.dmitruk@umcs.pl, ORCID: 0000-0003-1464-5822).

Fig. 1. Aldebaran NAO H25.

When interaction takes place between a humanoid robot and a human elements such a eye contact or response to sound and also proper limbs movements are necessary to create illusion of participation of the machine in the communication [3].

In this paper we try to put aside all moral and ethical concerns and focus on technical aspect of teleoperation based on motion capture techniques [5], [11].

II. Background

In most of cases robot teleoperation is realized with some kind of controller. We could point out solutions where operator use, in simplest situation, just keyboard, joystick or kind of gamepad. In more advanced systems force feedback or haptic interface is realized. Next group obtain information from movement data collected with kind of motion capture suit [7].

In proposed method we gave up any dedicated equipment in place of maker less motion capture optical system based on Microsoft Kinect.

III. Hardware configuration

At hardware level system is build up around few main components with two different way of communication depending on the direction. On direction operator – participant or participants there is robotic telepresence with teleoperated humanoid robot. In our case it is Aldebaran Nao H25. Other direction is based on teleconfenrence presentation, in perfect situation, with 1:1 scale projection or at least large (preferably 40 inches diagonal dimension or more) screen. To reduce problems of walking robot movements semi static situation with only head and arms movements is arranged.

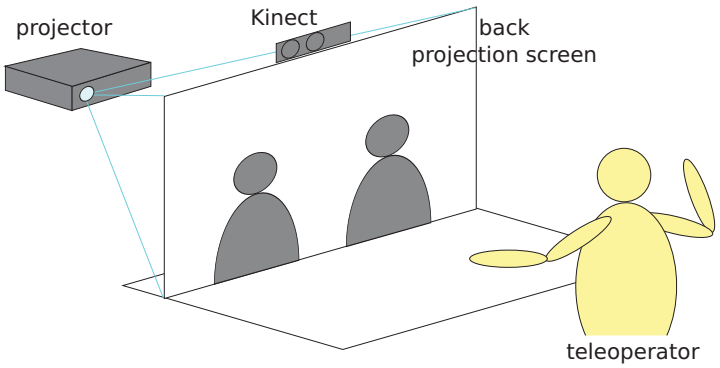

Fig. 2. Teleoperator equipment configuration.

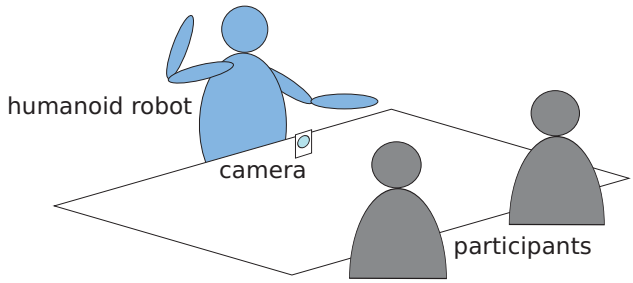

Fig. 3. Robot position and equipment configuration.

IV. Calibration

In perfect scenario with 1:1 teleconference presentation we could try to preserve all distances between teleoperator and Kinect and participants and camera. But even then calibration of all systems seems to be necessary.

In most cases in real environment we couldn't provide guidelines how to put all equipment up and calibration procedure should be first step in preparation of the environment. We suggest to use simple and based on calculated multiplication factors values for pitch and yaw head movements. In this case we are independent from any distance requirements.

Calibration is performed in five steps with red dot or similar easy distinguishable marker placed in front of camera in center and all corners of field of view. Every time new position of the marker is set posture of the head of the operator is find out. Parallel to this operation head of the robot is also set in direction of marker what is quite easy thanks to preview from camera in the head of humanoid robot. Head joints position is obtained and stored.

To reduce the impact of imperfect align of camera and teleconference system multiplication factors are interpolated for each possible angle with inverse distance weighted interpolation. We decided to use well known Shepard's method with distance expressed in angle between vectors and pitch and yaw angles value calculate in spherical coordinates with r component equal 1 as [15]:

$$p(\Theta) = \begin{cases} \dfrac{\sum_{i=1}^{N} w_i(\Theta) p_i}{\sum_{i=1}^{N} w_i(\Theta)} & \textbf{if } d(\Theta, \Theta_i) \neq 0 \\ \mathbf{p_i} & \textbf{if } d(\Theta, \Theta_i) = 0 \end{cases} \quad (1)$$

Where N is total number of measured positions, p_i are angles of measured positions, and weight w_i is calculated like in modified Shepard's method [12]:

$$w_i = \left(\frac{R_\Theta - d(\Theta - \Theta_i)}{R_\Theta d(\Theta, \Theta_i)} \right)^2 \quad (2)$$

with value of R_Θ expressed as

$$R_\Theta = max\left\{ d(\Theta, \Theta_i), \Theta_i = \Theta_0, \cdots, \Theta_N \right\} \quad (3)$$

V. Motion capture

Robot's ability to mimic human motion and gestures requires position and movement of head and arms data. For our experiment Microsoft Kinect sensor has been chosen.

Its most significant advantage over default camera is ability to read depth data trivializing segmentation task. Sensor works using infrared projector and two cameras – standard RGB and monochromatic operating in infrared wavelength spectrum.

Projector casts predefined pattern of dots on observed area and by comparing distortions of IR camera frame with reference image is able to compute depth of a pixel [8]. We have:

$$\frac{D}{b} = \frac{Z_0 - Z_k}{Z_0} \quad (4)$$

where D is object width, b is distance between IR camera and projector on the sensor, Z_0 is the reference plane, in our case, wall and Z_k is object plane distance we want to compute. We also know:

$$\frac{d}{f} = \frac{D}{Z_k} \tag{5}$$

where d is width of the disparity observed on the registered image and f is focal length of IR camera. By substitution we finally get [8]:

$$Z_k = \frac{Z_0}{1 + \frac{Z_0}{fb}d}. \tag{6}$$

Effect of this measurements provided to software in Y10B (10 bit packed grayscale) format.

For captured image analysis two external libraries: OpenNI and NiTE has been used. OpenNI simplifies access to depth sensor data by making an interface between hardware driver, allowing to easily view depth frame as 8 bit or 16 bit image.

Using information about sensor's optical parameters – focal length and lens distortion it can measure real distance between pixel and camera [4].

OpenNI uses distance data to create a point cloud – sparse three dimensional matrix of points. By pairing them with RGB camera data point cloud can be coloured. Due to complexity of point cloud, recreating skeleton from this format in real time would be impossible with consumer level workstation.

Skeleton is generated using NiTE library. NiTE is closed software and all informations about it's algorithms comes from official specifications and cannot be confronted with code.

NiTE strongly related to OpenNI and requires it's input to run. Generated skeletal model has 24 connected joints. Each joint have information about position and rotation in three dimensional space with sensor placement as reference system (which is fixed through the experiment), though, due to hardware limitations not every joint has reliable rotation data.

Joints position are determined by an estimated pose found in detected frame. NiTE uses random forests [17]. A tree votes if some depth pattern is present. Trees are connected in classes representing certain poses. In available space of time, random trees are chosen and checked. Pose with most votes is chosen and displayed [2].

Both robot and examined person are sitting, so only joints that has been used were upper and middle body parts: head, neck, torso and left/right, collar bone, shoulder, elbow, wrist, hand and fingertip.

During development process it turned out that exact position of arm joints were not important, because human flexibility is unreachable by the robot. Placement of hands turned out to be enough to simulate tested person's upper limbs movements and robot could mimic that with inverse kinematics. Information about hand's rotation were also dropped due to inaccuracy, especially when hands were shadowed by torso.

Before tracking can be started character skeleton must be recognized and connected to user's body. Because of algorithm's iterative construction, it is important to get clear initial skeleton. The best way is to perform a psi pose: standing with separated legs and hands reaching up.

This pose cannot be obtained in our case, because tested user is sitting, though lack of lower body part did not obscured our perception of upper part skeleton, but arms and hands should be separated from the torso. For best results, user should be located between 0.5 to 2.5 meter from the sensor.

VI. Robot movements operation

Because position of hands is more important for interlocutor than mutual position of upper and lower arm we could use inverse kinematic to position of each arm according to position of hand. It is also simpler to proper track only hands. Even in case of sign language movement and relative position of hands to each other and to key body locations are three of four most important high level features [6].

To calculate proper position of each joint in arms implementation of Inverse Kinematics solver for Aldebaran NAO was used. Precision and efficient computation offered by NAOKinematics library is good enough for our solution [10], [9].

One of the main problems of robot teleoperation is lag between operator and humanoid movement. In semi static situation as has place in presented solution even delay around one second are still acceptable. When similar solution use whole body control this times should be reduced.

VII. Conclusions

At this moment system is implemented and working but need more profound research. Series of tests on group of participants with proposed solution and control group with classic controller–based Wizard of Oz technique is necessary. There is also need to get involved psychologist to proper interpret results of test and answers from post test survey.

References

[1] Breazeal, C.: Designing Sociable Robots. A Bradford Book, The MIT Press, Cambridge, 2002.
[2] Breiman, L. Random forests, Machine Learning 45, 2001, pp. 5–32.
[3] Cassell, J.: Nudge nudge wink wink: Elements of face-to-face conversation for embodied conversational agents. Embodied Conversational Agents. The MIT Press, Cambridge, 1999, pp. 1–27.
[4] Fanelli, G., Weise, T., Gall, J., Van Gool, L., Real Time Head Pose Estimation from Consumer Depth Cameras, 33rd Annual Symposium of the German Association for Pattern Recognition (DAGM'11), 2011.
[5] Fraser, N. M., Gilbert, G. N.: Simulating speech systems. Computer Speech & Language, 5(1), 1991, pp. 81–99.
[6] Kadir T., Bowden R., Ong E. J., Zisserman A.: Minimal Training, Large Lexicon, Unconstrained Sign Language Recognition, In Proc. BMVC, 2004.
[7] Kelley, J. F.: An iterative design methodology for user-friendly natural language office information applications. ACM Transactions on Information Systems, 2(1), 1984, pp. 26–41.
[8] Khoshelham, K., Accuracy analysis of Kinect depth data, International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XXXVIII-5/W12, 2011.
[9] Kofinas N., Orfanoudakis E., Lagoudakis M.: Complete Analytical Inverse Kinematics for NAO, Proceedings of the 13th International Conference on Autonomous Robot Systems and Competitions (ROBOTICA), Lisbon, Portugal, 2013, pp. 1–6.
[10] Kofinas N.: Forward and Inverse Kinematics for the NAO Humanoid Robot, Diploma Thesis, Department of Electronic and Computer Engineering, Technical University of Crete, 2012.
[11] Miller, K. W.: Its Not Nice to Fool Humans. IT Professional, 12(1), 2010, pp. 51–52.

[12] Renka, R. J.: Multivariate interpolation of large sets of scattered data. ACM Transactions on Mathematical Software, vol. 14 Issue 28, New York, 1988, pp. 139–14.
[13] Riek, L. D., Watson, R.: The Age of Avatar Realism. IEEE Robotics & Automation Magazine, 17(4), 2010, pp. 37–42.
[14] Robins, B., Dautenhahn, K., Te Boekhorst R., Billard A.: Robotic assistants in therapy and education of children with autism: can a small humanoid robot help encourage social interaction skills? Universal Access in the Information Society, vol. 4. Issue 2. Springer, 2005, pp. 105–120.
[15] Shepard, D.: A two-dimensional interpolation function for irregularly-spaced data. Proceedings of the 1968 ACM National Conference, 1968, pp. 517–524.
[16] Shimada M., Yoshikawa Y., Asada M., Saiwaki N., Ishiguro H.: Effects of Observing Eye Contact between a Robot and Another Person. International Journal of Social Robotics, vol. 3. Issue 2. Springer, 2011, pp. 143–154.
[17] Shotton, J., et al., Real-time human pose recognition in parts from single depth images, In In CVPR, 2011.
[18] Stegierski R., Kuczyski K.: The Perception of Humanoid Robot by Human. Image Processing and Communications Challenges 5, Advances in Intelligent Systems and Computing vol. 233, Springer-Verlag, Berlin Heidelberg, 1st Edition, 2014, pp. 65–72.
[19] Yoshikawa, Y., Noda, T., Ikemoto, S., Ishiguro, H., Asada, M.: CB2: A child robot with biomimetic body for cognitive developmental robotics, Humanoid Robots, 2007 7th IEEE-RAS International Conference on Humanoid Robots, Pittsburg, 2007, pp. 557–562.
[20] Colton, M., Ricks, D., Goodrich, M., Dariush, B., Fujimura, K., Fujiki, M.: Toward Therapist-in-the-Loop Assistive Robotics for Children with Autism and Specific Language Impairment. AISB 2009 Syposium on New Frontiers in Human-Robot Interaction, Edinburgh, 2009.

Rafal Stegierski *is assistant professor in Institute of Computer Science, Faculty of Mathematics, Physics and Computer Science, Maria Curie Sklodowska University, Lublin. In 2000 he earned a master's degree in physics in speciality computational physics. He received his Ph.D. in computer science from Silesian University of Technology in 2008 upon completion of his doctoral thesis in image processing.*

Krzysztof Dmitruk *is an assistant in Institute of Computer Science, Faculty of Mathematics, Physics and Computer Science, Maria Curie Sklodowska University, Lublin. His work focuses on computer vision, video analysis and pattern recognition.*