Effects of Gender Mapping on the Perception of Emotion from Upper Body Movement in Virtual Characters

Maurizio Mancini[1], Andrei Ermilov[2], Ginevra Castellano[3], Fotis Liarokapis[4], Giovanna Varni[1], and Christopher Peters[5]

[1] InfoMus Lab, University of Genoa, Italy
{maurizio.mancini,giovanna.varni}@unige.it
[2] Faculty of Engineering and Computing, University of Coventry, UK
ermilova@uni.coventry.ac.uk
[3] School of Electronic, Electrical and Computer Engineering, University of Birmingham, UK
g.castellano@bham.ac.uk
[4] Interactive Worlds Applied Research Group & Serious Games Institute, Coventry University, UK
aa3235@coventry.ac.uk
[5] School of Computer Science and Communication, Royal Institute of Technology (KTH), Sweden
chpeters@kth.se

Abstract. Despite recent advancements in our understanding of the human perception of the emotional behaviour of embodied artificial entities in virtual reality environments, little remains known about various specifics relating to the effect of gender mapping on the perception of emotion from body movement. In this paper, a pilot experiment is presented investigating the effects of gender congruency on the perception of emotion from upper body movements. Male and female actors were enrolled to conduct a number of gestures within six general categories of emotion. These motions were mapped onto virtual characters with male and female embodiments. According to the gender congruency condition, the motions of male actors were mapped onto male characters (congruent) or onto female characters (incongruent) and vice-versa. A significant effect of gender mapping was found in the ratings of perception of three emotions (anger, fear and happiness), suggesting that gender may be an important aspect to be considered in the perception, and hence generation, of some emotional behaviours.

1 Introduction

Several studies explored the perception of behaviour in virtual characters [1,2]. Results from these studies are significant, as they can be used to contribute to the design and development of more efficient and plausible simulations of artificial entities, with applications in important and complex fields such as virtual and

R. Shumaker and S. Lackey (Eds.): VAMR 2014, Part I, LNCS 8525, pp. 263–273, 2014.

augmented reality. Perhaps of equal significance, these studies also have impact on deepening our understanding of how humans perceive other humans and biological versus artificially generated motions.

While studies have started to investigate the relationship between emotion and body movement by using virtual characters [3], little is known about the effect of gender on the perception of emotion from body movement. This issue is important for creating expressive social entities that are able to successfully communicate with humans. For example, if humans are more sensitive to motions of anger from male embodiments, this suggests that such behaviour needs to be moderated in order to create more desirable impressions.

In this paper, we present a pilot experiment investigating the perception of six basic emotions (anger, disgust, fear, happiness, sadness and surprise) from upper body movements that have been mapped onto virtual characters of the same and opposite gender from actors who originally conducted the movements. A corpus of emotional gestures was recorded from the performances of male and female actors using a Microsoft Kinect [4]. These movements were mapped onto virtual character as follows: movements generated by a Female Actor (FA) were mapped onto both Female and Male Characters (FC and MC) and movements generated by a Male Actor were mapped onto both Male and Female Characters (MC and FC).

We expected that when the gender of the virtual character was congruent with the gender of the original actor, the recognition rate of the expressed emotion would be higher. An online experiment was performed to test our hypothesis, in which videos of the virtual characters were shown to twenty-four subjects. The results indicated that when the virtual character's gender is congruent with the gender of the original actor, the recognition rate of the expressed emotion is higher for a subset of the six emotions considered here.

This paper is organised as follows. The next section (Section 2) provides a summary of relevant literature. Section 3 presents the corpus that was recorded with the Microsoft Kinect as part of this study, detailing the mapping process and the virtual stimuli used in the experiment. Section 4 describes a pre-experiment and Section 5 provides details of the mapping from actors to virtual characters. Section 6 describes the online perceptual experiment, summarising and discussing the main results. Finally, Section 7 summarised the contributions and limitations of the work in the context of future studies of impact in this domain.

2 Related Work

In the affective computing and computer animation communities, there has been growing interest in the study of perception of emotion from body movement using virtual characters. McDonnell and colleagues [1], for example, investigated the perception of emotion expressed by virtual characters with different embodiments and, more recently, the ability of humans to determine the gender of conversing characters based on facial and body cues [5]. Ennis and Eggs [6] explored the use of complex emotional body language for a virtual character and found that participants are better able to recognise complex emotions with negative connotations rather than positive. Castellano et al. presented an experiment investigating the perception of synthesized emotional gestures performed by an embodied virtual agent based on actor's movements with manual modulations [3].

In previous work on copying behaviour of real motion in virtual agents [7], it was shown that movement expressivity can convey the emotional content of people's behaviour, e.g., if a virtual agent's expressivity is not altered, then emotional content cannot be conveyed effectively. Investigations in [8] considered whether and how the type of gesture performed by a virtual agent affects the perception of emotion, concluding that a combination of type of movement performed and its quality are important for successfully communicating emotions.

In relation to gender perception from body movement, early studies by Kozlowski and Cutting [9] argued that the gender of walkers can be accurately recognised without familiarity cues from dynamic displays of point-light sources placed on upper/lower-body joints, respectively. Further, they also pointed out how changes in the degree of arm-swing or in walking speed can interfere with this recognition. More recently, gender recognition has had an important impact in computer vision, where researchers have developed several approaches based on 2D-3D facial or full-body analysis (see [10] for a survey). However, relatively little is currently known about the effect of gender on the perception of emotion from body movement; the majority of studies within psychology and neurosciences have focused on perception of emotion from facial expressions, showing that females are better at perceiving emotion from facial expression than males (e.g. [11]), and that different brain regions are activated by females and males when viewing facial expressions that are expressing sadness and happiness [12]. Further, Bartneck et al. [13] studied the effect of both culture and gender on the emotion recognition from complex expressions of avatars.

Tracking movement is not a new domain of research and there are a number of approaches that have been used in the past. One of the most common techniques is motion capture using body sensors or markers. These systems are typically expensive and usually require the sensors or markers to be positioned at selected places on the user. Although researchers have developed wearable systems that can capture movement for mapping onto virtual characters [14], there are other alternatives including computer vision approaches. The advances of vision-based human motion capture have been well documented [15] and they are becoming more popular. For our experimental implementation we focused on the later approaches by utilising Microsoft Kinect as a low-cost alternative for capturing motion. The maximum distance Kinect can operate is 5 metres and random error accuracy ranges from a few millimetres at 0.5m to 4cm of error at 5m [16], providing reasonable accuracy for some categories of motion recording.

3 Data Collection from Actors

A corpus of emotional gestures was recorded from the performances of male and female actors using a Microsoft Kinect as follows.

Two amateur actors (1M:1F) were recruited to act six basic emotional states: anger, disgust, fear, happiness, sadness and surprise. The actors were instructed about the final goals of the emotional gesture mapping experiment and provided their written consent for using the recorded video and numerical data for research purposes.

The actors were instructed to stand facing the camera and were asked not to move beyond the boundaries of a square area marked on the center of the stage. This constraint has been imposed to minimize potential tracking problems, such as occlusion, which may occur due to the use of a single Kinect if the actors were free to walk around the stage.

Each performed gesture involved the actor starting from and returning to a *starting* pose (i.e., standing still with arms along the body). Each gesture lasted up to 6 seconds. Actors could decide the gesture starting time by clapping their hands, after which the 6 second interval was measured. At the end of the recording interval, a sound informed the actor that they should return to the starting pose.

For each of the 6 emotional states the actor had to perform 10 repetitions, with a break of approximately 15 seconds between consecutive repetitions. The emotional expressions were performed in the following order: anger, sadness, happiness, surprise, fear, disgust.

Before performing the 10 emotion repetitions, actors were prepared with a short predefined scenario in order to support them. For example, for sadness: “You are listening to your iPod, when it drops, you pick it up, but it won’t work.”

Kinect Studio (part of the Windows Developer Toolkit) was used to record the actors, saving the information required for the mapping of motions to the characters described in Section 5. After each gesture was performed by the actor, a Kinect data file was generated containing colour and depth information (Figures 3a and 3b). Thus the actors were recording at the same session and the virtual characters were mapped at a later stage.

4 Pre-experiment User Study

From the corpus of gestures recorded in the data collection phase (Section 3), three gestures per emotion per actor were selected to form a final set of gestures that were as diverse as possible for each emotion.

A categorisation task was performed in order to identify those gestures from the actors that were most recognisable for each emotion. Twenty participants took part in this pre-experiment user study, which utilised a six-alternative forced choice paradigm: participants were asked to watch the video stimuli of the real actors and in each case indicate which one of the six basic emotions they thought was being expressed. While a forced choice paradigm does not allow participants to deviate from the proposed alternatives, it forces them to select the single closest option matching their judgement, which was desirable for the purposes of this study.

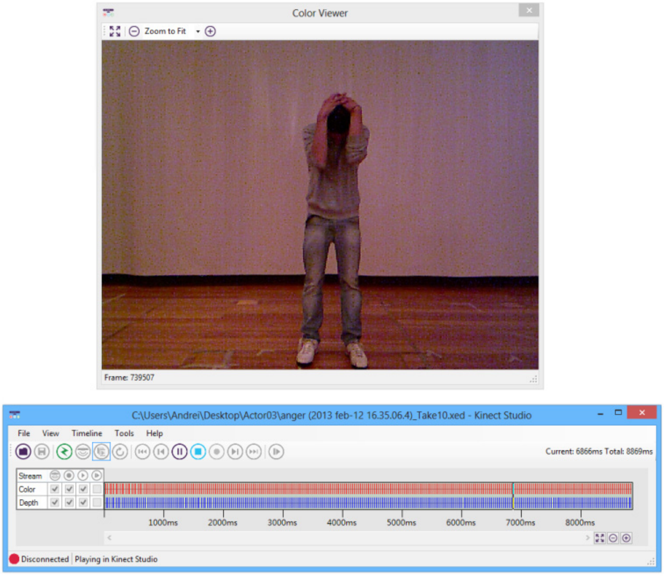

Fig. 1. The interface of Kinect Studio

The face and hands of the actors in the videos were blurred so as not to interfere with the focus of the study, which is upper body movement. For each actor and for each emotion, the gesture with the highest recognition rate was selected (Table 1). A few gestures had the same recognition rate: in these cases, the gesture whose expressed emotion was the most highly rated/less misclassified across all emotions for that specific gesture was selected for the character mapping phase, described in the next Section.

Table 1. Recognition rates of the body movements performed by the actors

	Anger	Disgust	Fear	Happiness	Sadness	Surprise
Female actor	75%	35%	50%	85%	90%	55%
Male actor	85%	75%	65%	85%	90%	70%

5 Character Mapping

Starting from the collected actors' data (Section 3) and the corresponding recognition rates (Section 4), the character mapping stage involved two free pre-made 3D character embodiments, one male and one female, that were chosen from Mixamo[1].

[1] http://www.mixamo.com

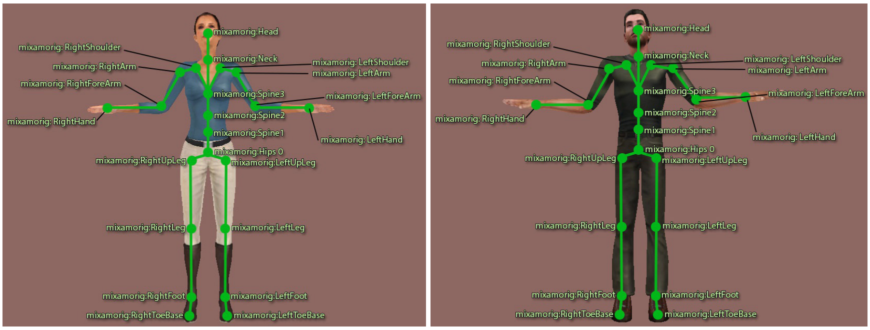

Fig. 2. The virtual character's skeleton data generated by the auto-rigging process

The mapping between the actors' motions and characters was conducted using a custom made real time skeleton motion tracking application developed in Visual Studio 2012 using the Microsoft for Windows SDK and Digital Rune Engine [2], which was used for rendering. This engine extends the XNA functionality making it easy to integrate the Kinect SDK in the engine.

As described in Section 3, movement tracking was conducted using the Kinect SDK that analyses the depth data from the camera and provides skeleton data (see Figure 2). This data consists of the movements of a set of joints. However, these joints differed from those in the characters (Figure 2). An auto-rigging service from Mixamo was used to create the same skeleton system for both characters, which was imported into our application.

Next, joints were mapped from the Kinect skeleton matching the actors' motions to the skeleton used in the virtual characters, i.e., the Hip_Center joint was mapped to the mixamorig:Hips0. This enabled the skeleton of the virtual character to copy the movements of the actors standing in front of the camera. Kinect Studio was used to connect the recording to our application for rendering, which was recorded using the Fraps [3] video recording utility. Female and Male Actors (FA and MA) were mapped onto both Female and Male Characters (FC and MC): a total of 24 videos were created during this phase (6 emotions x 2 actors x 2 characters).

There were a number of limitations to recording with a single Kinect: In particular, there is no finger tracking so the hands of the character are not mapped and remain in the same position in the generated videos. As in the case of the videos of the original performing actors, a blurring effect was placed on the head and hands of the virtual characters.

For each video, the process consisted of (1) a recording played in Kinect studio (Figure 3a and b), (2) a Fraps recording of the output of our application (Figure 3c) and (3) application of the blurring effect applied to the video (Figure 3d). The outputs of this process were the set of final mapped videos of virtual character movements to be used in the experiments, described in the next Section.

[2] http://www.digitalrune.com
[3] http://www.fraps.com

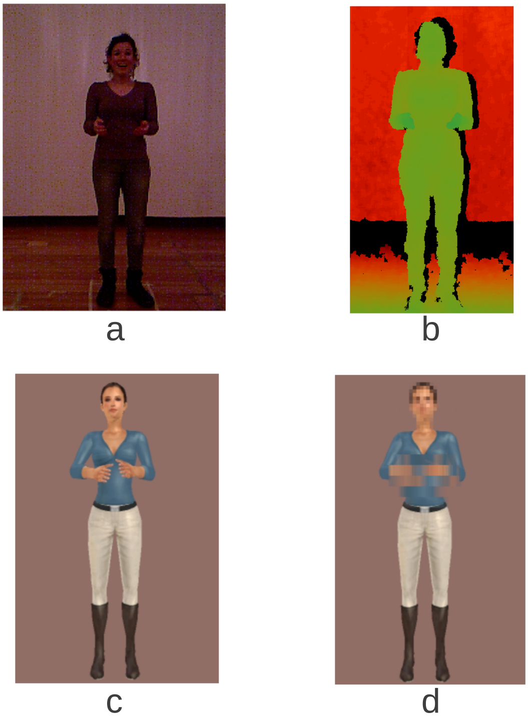

Fig. 3. RGB data output from the camera (a); depth data output from the camera (b); mapped avatar using the recorded data (c); blurred video using Sony Vegas (d)

6 Experiment

This section describes the experiment and methodology employed to assess the following hypothesis: *when the gender of a virtual character is congruent with the gender of an original actor, the recognition rate of the expressed emotion is higher.*

6.1 Participants

Twenty-four anonymous participants took part in the experiment. They were contacted by sending an invitation letter to a number of research mailing lists.

The order of the 24 videos the participants had to evaluate was randomised by using a 24 24 latin square. Each user was assigned to a different row of the square, consisting in a randomised sequence of 24 videos. This approach allowed for a balancing of practice effects across participants.

In order to ensure that participants were assigned all 24 sequences from the latin square, a PHP script[4] automatically assigned the first element of a pool

[4] The script is freely available for research purposes at the address: `https://sourceforge.net/projects/phptestproject/`.

containing all the available sequences (i.e., the sequences not yet evaluated by participants) to each participant as they started to perform the experiment. If the experiment was not completed in 3 hours, the assigned sequence was put back in the pool of the available sequences. Otherwise the sequence was marked as "completed".

To play back videos, the script uses the free version of JW Player [5]. Also, it performs a web browser version check to reduce the risk of compatibility issues with the experiment web-based interface.

6.2 Procedure

The participants were instructed to navigate to a web page specified in an email message as explained in previous section. The web page contained a brief description of the experiment goals and duration. Participants were also informed that no personal data was collected during the experiment and that they could leave the experiment at any time just by closing their web browser.

After commencing the experiment, a graphical interface was displayed. Participants were instructed to click on the play button to start the video playback. They were asked to provide an answer to the proposed question ("Do you think that the emotional state expressed by the person is:") after watching each video entirely. They could also watch the video again as many times as they wished. The answer to the question was constrained to one of 6 emotional states listed in alphabetical order: anger, fear, disgust, happiness, sadness, surprise. Participants were not allowed to provide a blank answer (e.g., "no emotion" or "other").

6.3 Results

In order to investigate the effects of gender mapping on the perception of emotion in the virtual stimuli, a statistical analysis was performed. Two Wilcoxon tests ($N = 24$) were performed for each emotion, one for each actor, for a total of twelve tests. As the data is not normally distributed, the Wilcoxon test was preferred to the t-test.

In each Wilcoxon test, the gender of the virtual character used to map the original actor's movement was considered as the independent variable (two levels), while the emotion rating (i.e., the accuracy performance) was considered as the dependent variable (one level). Each Wilcoxon test aimed to explore whether there is a significant difference in the rating of a specific emotion between stimuli in which the body movement expresses that same emotion when the virtual character's gender is congruent with that of the original actor and stimuli in which the virtual character's gender is not congruent with that of the original actor.

Results showed a significant effect of the virtual character's gender on the emotion rating in three cases. When anger is expressed by the female actor, the ratings of anger are significantly higher for the female virtual character

[5] http://www.longtailvideo.com/jw-player

($z = -2.236, N = 5, p = 0.01$). When fear is expressed by the male actor, the ratings of fear are significantly higher for the male virtual character ($z = -1.633, N = 6, p = 0.05$). When happiness is expressed by the female actor, the ratings of happiness are significantly higher for the female virtual character ($z = -2.121, N = 8, p = 0.017$).

While the ratings of anger are substantially high, the ratings of fear and happiness are not, although they are above chance level (see Table 2 for a summary of the means and standard deviations for the Wilcoxon test). This is in contrast to the results from the pre-experiment user study, which showed higher recognition accuracies for happiness and fear in the real stimuli. This discrepancy is possibly due to tracking and mapping issues related to the Kinect data, since some movements are not tracked with a high accuracy.

Table 2. Means and standard deviations for the Wilcoxon tests: significant differences between mean values are highlighted with a star (FA-to-FC = female actor to female character; FA-to-MC = female actor to male character; MA-to-MC = male actor to male character; MA-to-FC = male actor to female character)

	Female actor	Male actor
Accuracy anger	*FA-to-FC $\mu = 71\%, \sigma = 0.46$ *FA-to-MC $\mu = 50\%, \sigma = 0.51$	MA-to-MC $\mu = 54\%, \sigma = 0.51$ MA-to-FC $\mu = 71\%, \sigma = 0.46$
Accuracy disgust	FA-to-FC $\mu = 17\%, \sigma = 0.38$ FA-to-MC $\mu = 8\%, \sigma = 0.28$	MA-to-MC $\mu = 75\%, \sigma = 0.44$ MA-to-FC $\mu = 71\%, \sigma = 0.46$
Accuracy fear	FA-to-FC $\mu = 13\%, \sigma = 0.34$ FA-to-MC $\mu = 17\%, \sigma = 0.38$	*MA-to-MC $\mu = 29\%, \sigma = 0.46$ *MA-to-FC $\mu = 13\%, \sigma = 0.34$
Accuracy happiness	*FA-to-FC $\mu = 29\%, \sigma = 0.46$ *FA-to-MC $\mu = 4\%, \sigma = 0.20$	MA-to-MC $\mu = 29\%, \sigma = 0.46$ MA-to-FC $\mu = 38\%, \sigma = 0.50$
Accuracy sadness	FA-to-FC $\mu = 87\%, \sigma = 0.34$ FA-to-MC $\mu = 79\%, \sigma = 0.42$	MA-to-MC $\mu = 92\%, \sigma = 0.28$ MA-to-FC $\mu = 87\%, \sigma = 0.34$
Accuracy surprise	FA-to-FC $\mu = 25\%, \sigma = 0.44$ FA-to-MC $\mu = 29\%, \sigma = 0.46$	MA-to-MC $\mu = 46\%, \sigma = 0.51$ MA-to-FC $\mu = 42\%, \sigma = 0.50$

7 Conclusion

This paper presented an initial experiment studying the effects of gender mapping on the perception of emotion from virtual behaviour. We mapped emotional body movements generated by real actors onto virtual characters using data tracked with a Kinect. Movements generated by a female actor were mapped onto a female and a male character, and movements generated by a male actor were mapped onto a male and a female character.

Our hypothesis was that when the gender of the virtual character is congruent with the gender of the original actor the perception of emotion is higher. Results indeed showed that when the virtual character's gender is congruent with the gender of the original actor, perception of emotion is higher, but only for the emotions anger, fear and happiness, in three out of the twelve cases considered in our analysis.

Results need to be further validated by collecting a larger corpus of emotional gestures, also performed by actors from different cultures. However, despite the current limitations of Kinect, we believe that our approach should be further probed, as the possibility to use inexpensive equipment is desirable for developing, for example, more portable and wearable virtual and augmented reality systems.

References

1. McDonnell, R., Jörg, S., McHugh, J., Newell, F.N., O'Sullivan, C.: Investigating the role of body shape on the perception of emotion. ACM Transactions on Applied Perception (TAP) 6(3), 14 (2009)
2. Courgeon, M., Clavel, C., Tan, N., Martin, J.-C.: Front view vs. Side view of facial and postural expressions of emotions in a virtual character. In: Pan, Z., Cheok, A.D., Müller, W. (eds.) Transactions on Edutainment VI. LNCS, vol. 6758, pp. 132–143. Springer, Heidelberg (2011)
3. Castellano, G., Mancini, M., Peters, C., McOwan, P.W.: Expressive copying behavior for social agents: A perceptual analysis. IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans 42(3), 776–783 (2012)
4. Shotton, J., Fitzgibbon, A., Cook, M., Sharp, T., Finocchio, M., Moore, R., Kipman, A., Blake, A.: Real-time human pose recognition in parts from single depth images. In: Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2011, pp. 1297–1304. IEEE Computer Society, Washington, DC (2011)
5. Zibrek, K., Hoyet, L., Ruhland, K., McDonnell, R.: Evaluating the effect of emotion on gender recognition in virtual humans. In: Proceedings of the ACM Symposium on Applied Perception, SAP 2013, pp. 45–49. ACM, New York (2013)
6. Ennis, C., Egges, A.: Perception of complex emotional body language of a virtual character. In: Kallmann, M., Bekris, K. (eds.) MIG 2012. LNCS, vol. 7660, pp. 112–121. Springer, Heidelberg (2012)
7. Mancini, M., Castellano, G., Peters, C., McOwan, P.W.: Evaluating the communication of emotion via expressive gesture copying behaviour in an embodied humanoid agent. In: D'Mello, S., Graesser, A., Schuller, B., Martin, J.-C. (eds.) ACII 2011, Part I. LNCS, vol. 6974, pp. 215–224. Springer, Heidelberg (2011)
8. Castellano, G., Mancini, M., Peters, C.: Emotion communication via copying behaviour: A case study with the greta embodied agent. In: Proceedings of the AFFINE Workshop, Hosted by the ACM ICMI 2011 Conference (2011)
9. Kozlowski, L.T., Cutting, J.E.: Recognizing the sex of a walker from a dynamic point-light display. Perception and Psychophysics 21(6), 575–580 (1977)
10. Ng, C., Tay, Y., Goi, B.-M.: Recognizing human gender in computer vision: A survey. In: Anthony, P., Ishizuka, M., Lukose, D. (eds.) PRICAI 2012. LNCS, vol. 7458, pp. 335–346. Springer, Heidelberg (2012)
11. Babchuk, W.A., Hames, R.B., Thompson, R.: Sex differences in the recognition of infant facial expressions of emotion: the primary caretaker hypothesis. Ethology and Sociobiology 6, 89–101 (1985)
12. Lee, T., Liu, H., Liao, W., Wu, C., Yuen, K., Chan, C., Fox, P., Gao, J.: Gender differences in neural correlates of recognition of happy and sad faces in humans assessed by functional magnetic resonance imaging. Neuroscience Letters 333(1) (2002)

13. Bartneck, C., Takahashi, T., Katagiri, Y.: Cross-cultural study of expressive avatars. In: Third International Workshop on Social Intelligence Design (2004)
14. New Scientist. Cheap sensors could capture your every move, http://www.newscientist.com/article/dn12963-cheap-sensors-could-capture-your-every-move.html (accessed at: April 8 2013)
15. Moeslund, T.B., Hilton, A., Krüger, V.: A survey of advances in vision-based human motion capture and analysis. Comput. Vis. Image Underst. 104(2), 90–126 (2006)
16. Khoshelham, K., Elberink, S.O.: Accuracy and resolution of kinect depth data for indoor mapping applications. Sensors 12(2), 1437–1454 (2012)