A study on multi-view calibration methods for RGB-D cameras

Diedre Carmo, Yan Jovita, Rafael Ferrari, Luciano Oliveira
Intelligent Vision Research Lab
Federal University of Bahia
Salvador, Bahia, Brazil
http://www.ivisionlab.eng.ufba.br

***Abstract*—RGB-D cameras became part of our daily life in applications such as human-computer interface and game interaction, just to cite a few. Because of their easy programming interface and response precision, such cameras have also been increasingly used to 3D reconstruction and movement analysis. In view of that, calibration of multiple cameras is an essential task. On that account, the goal of this paper is to present a preliminary study of methods which tackle the problem of multi-view geometry computation using RGB-D cameras. A brief overview of camera geometry is presented, some methods of calibration are discussed and one of them is evaluated in practice; finally, some important points are addressed about practical issues involving the problem.**

***Keywords*-RGB-D cameras; multi-view calibration; cross-talk interference**

I. Introduction

Often, parts of objects being observed by a camera are occluded from the camera viewpoint, making it harder to be detected or tracked. By using a world view from multiple cameras, it is possible to cope adequately with such problem. For example, two cameras can be used as a stereo vision, mimicking the human vision, and capable to reconstruct a scene by means of a triangulation scheme [1].

Nowadays time-of-flight sensors (such as RGB-D cameras) have been largely exploited due to the low cost, presenting an integration of a depth and an RGB cameras, which need to be calibrated to provide a precise depth color map [2]. An inconvenient effect of parallel projection of RGB-D cameras are the occluded parts which create shadows of non-visible regions in the image. A way to circumvent this inherent drawback is to use multiple cameras that demands an extrinsic calibration among them. Applications of such a vision setup span fields like human movement analysis by a composite skeleton [3] and 3D object reconstruction [14] (refer to Fig. 1 for examples of these applications).

Although calibration of RGB-D cameras overcome the inherent problems of a single RGB-D camera, this is not a simple task. Once these sensors are based on a matrix of infra-red (IR) emitters and receptors, cross-talk problems with serious interference on pair of cameras is surely presented. Yet, due to the single sensor low range, there are limitations when pointing cameras to see the same calibration reference (that is, a checkerboard). The goal of this paper is then to present a study about the existing methods of multi-view calibration, discussing important points towards a robust multi-view calibration method.

The reminder of this paper is structured as follows. Section II presents a background on camera geometry, and its intrinsic and extrinsic calibration parameters. Section IV address some issues involved in calibration of multiple RGB-D cameras. In Section V, an implemented method is described and discussed, and, finally, Section VI draws some conclusion and future works.

II. Background on camera geometry

The main goal of camera calibration is to find camera intrinsic and extrinsic parameters, which ultimately describe the camera model and relate the image of one camera to the real world. Focal length, aspect ratio of the image, and principal point are the intrinsic parameters, while rotation and translation parameters are the extrinsic ones.

Points inside a camera pinhole model (see Fig. 2) is usually represented in homogeneous coordinates, where a point represented by $(x, y)^T$ in Cartesian coordinates is represented in homogeneous coordinates by $(x, y, w)^T$. After normalizing all coordinates by w, we have $(X, Y, 1)^T$. A point, P, in world coordinates system is projected into a point, p, in a image coordinate system, according

$$p = K[R|t]P \tag{1}$$

Via a calibration procedure, it is necessary to find the intrinsic and extrinsic parameters presented in the matrices K and $[R|t]$, respectively, where K is given by

$$K = \begin{bmatrix} f.m_w & h & uo \\ f.m_h & ah & vo \\ 0 & 0 & 1 \end{bmatrix} \tag{2}$$

where K is composed of the focal length, f, skew coefficient, h, and the principal point (uo, vo). m_w and m_h are the scaling factors that relate pixels to distance.

In summary, the intrinsic parameters are defined as follows [4]:

- **Focal length**: Describes how strongly the camera diverges or converges light (see, f, in Fig. 2).
- **Skew coefficient**: Determines the angle between the x and y axes, usually close to zero.

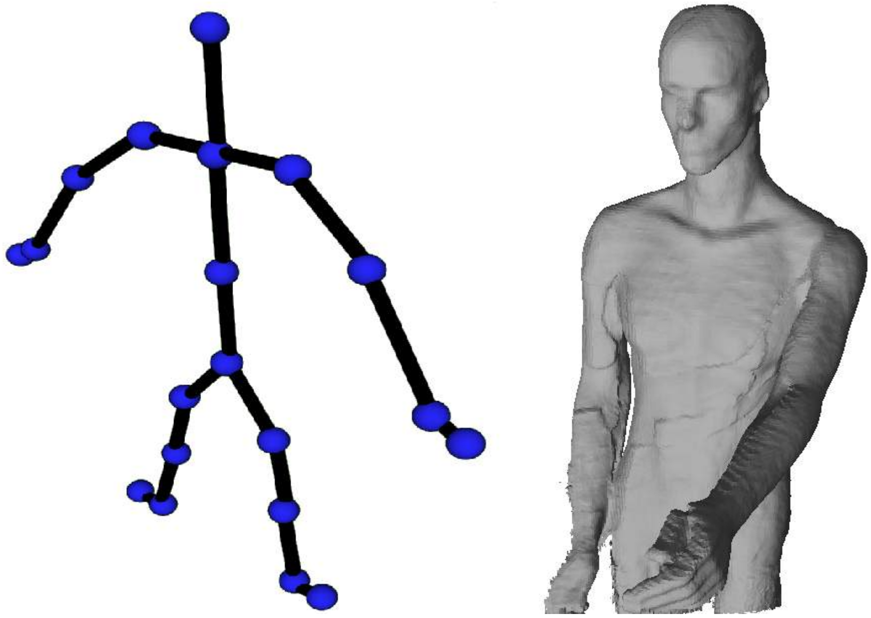

Fig. 1. Examples of multi-view camera applications. Left: A composite skeleton; right: A 3D reconstruction of a person.

- **Principal point**: The intersection of the principal plane and the camera axis, ideally on the center of the image (in Fig. 2, it is the point in the image plane where Z_c intersects).
- **Lens distortion**: A non-linear parameter used in most of the calibration methods. It is a form of optical aberration where straight lines get distorted in some way due to the shape of the lens.

Extrinsic camera calibration consists in achieving a transformation from camera-fixed Cartesian space to an arbitrary 3D coordinate system that whether the camera is located or it is shared by multiple cameras [11]. Extrinsic calibration is required by multiple cameras, and the goal is that a common target is shared by the camera posed in different points of view. $[R|t]$ represents the matrix of extrinsic parameter, where R is a rotation matrix, and t is a translation vector, defined respectively as

$$R = \begin{bmatrix} R_{xx} & R_{yx} & R_{zx} \\ R_{xy} & R_{yy} & R_{zy} \\ R_{xz} & R_{yz} & R_{zz} \end{bmatrix} \quad (3)$$

where $R_{..}$ are the rotation parameter over each one of the pairs of planes formed by x, y and z axes.

$$t = \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix} \quad (4)$$

where $t_.$ is the translation parameter in each one of the x, y and z axes.

t is the translation of the origin of the world coordinate system expressed in coordinates of the camera-centered coordinate system (CCCS) and R represents the rotation of the world coordinate system regards the (CCCS).

More details about camera parameters can be found in [10], [11].

III. Multi-view calibration methods: a review

When using multiple cameras, it is almost mandatory camera calibration in order for the system to be used in some applications. It is an absolute necessary step when a set of camera aims to analyze the same scene to produce a 3D information, for example.

Considering the recent advance and low price of RGB-D cameras, calibration process of these devices requires special procedures in order to tackle with inherent characteristics of the sensor. There are two manners to calibrate multi-RGBD cameras, that is, via depth map or infrared (IR) images. Both considering the same camera in the sensor.

In [7], the authors accomplish a RGB-D camera external calibration prior to reconstructing objects in a virtual environment. To do so, it was developed a custom-made calibration object with three planar surfaces, to be registered by all of the five sensors. The object is moved to various positions at a some given time, and, in each captured frame, the planar surfaces are acquired by all of the depth sensors. The pose estimated for the sensors is obtained by refining an appropriate referenced pose with optimization techniques.

The work in [6] attempts to register and reconstruct a scene using four Kinect sensors. For that, it evaluates some calibration methods. The authors ultimately use the IR output of the sensors to perform a checkerboard-based calibration using Bouguet's toolbox [16]. The resulting scenario is capable of real-time rendering of the merged data after an offline calibration setup.

Considering the reference target used for calibration, there are two types found in the literature, presenting each one of

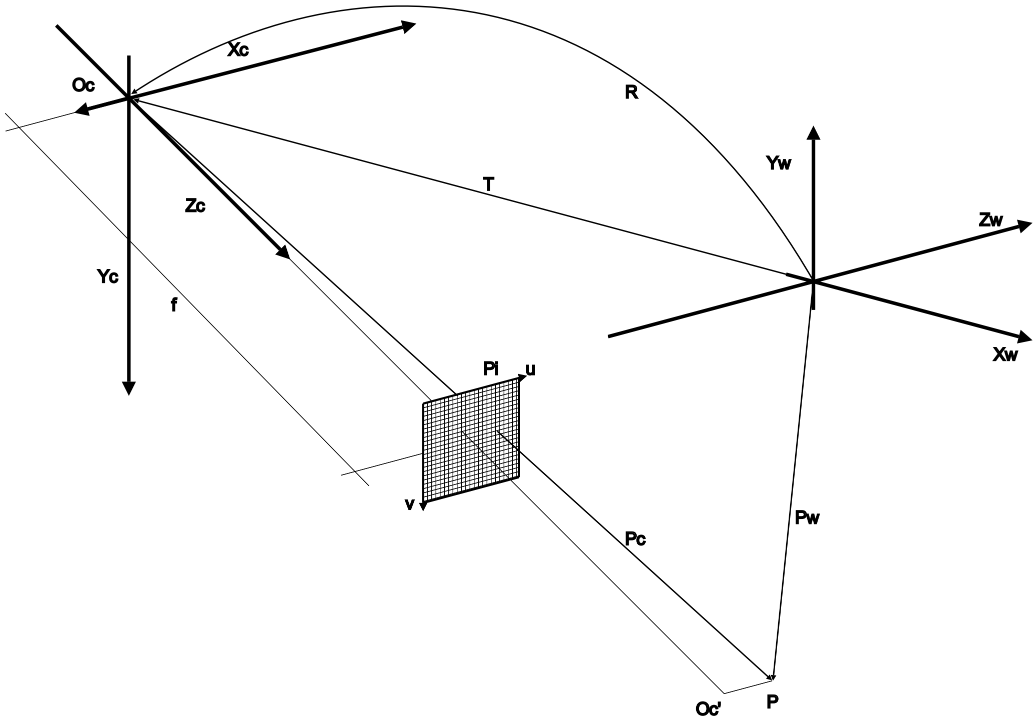

Fig. 2. Pin-hole camera model.

them certain particular aspects. Several of the existing camera calibration works, such as in [6], accomplish calibration with Bouguet's toolbox [16] to calibrate IR images. In [8], the proposed method performs calibration with a single shot, using multiple checkerboards placed at different locations in the image.

Most of the extrinsic calibration methods is based on checkerboard references, working by comparing a pattern of known geometry presented in the real-world environment from different perspectives. These methods, however, require that the camera setup never changes even slightly, or the calibration parameters obtained are not likely to hold. Recent methods, as in [9], propose automatic calibration; by taking advantage of the perceived environment, the method is able to update the calibration parameters of multiple cameras on-the-go, based on the data that is being acquired. This procedure can be perfectly matched with projects which require real-time feedback and have no information of the position of the cameras. The drawback of the auto calibration is that it often requires more time to find the calibration parameters when compared to manual calibration methods, with a special advantage of being more flexible.

IV. Practical issues about calibration of RGB-D cameras

In this section, some issues about RGB-D sensor calibration are addressed. This is important in order to characterize the multiple problems involved the calibration of RGB-D cameras.

A. Depth calibration

If extrinsic calibration of the depth maps in RGB-D cameras is required, it is not possible to use the checkerboard-based calibration process as it is. There are several workarounds to solve this issue, such as the one described in [6] which uses the IR output of the Kinect instead of the raw depth, in order to make it work with the classical checkerboard-based calibration.

A common approach for depth calibration is the use of different objects which contain depth-acquirable features. A semi-transparent checkerboard, alternating between color squares and transparent ones, is proposed by [17]. This solution allows the checkerboard pattern to be perceived by the depth sensor, therefore, allowing it to estimate the same parameters using the classical checkerboard methods, as well. Other solutions usually consist of modifying the calibration objects, like [15], which uses a cylindrical shape in order to calculate the same extrinsic parameters. Because these approaches depend on different kinds of objects, many other calibration methods can be created, unveiling a range of possibly more efficient algorithms.

B. Cross-talk interference

One significant problem found in the calibration process by using multiple RGB-D devices is the cross-talk interference. Due to nature of RGB-D camera sensing, multiple infra-red emitters may cause noise in the depth map captured. Berger et al. [12] measured a high error in a experiment with motion capturing and body joint calculation by using multiple RGB-D cameras. In practice, errors due to noise can be easily observed

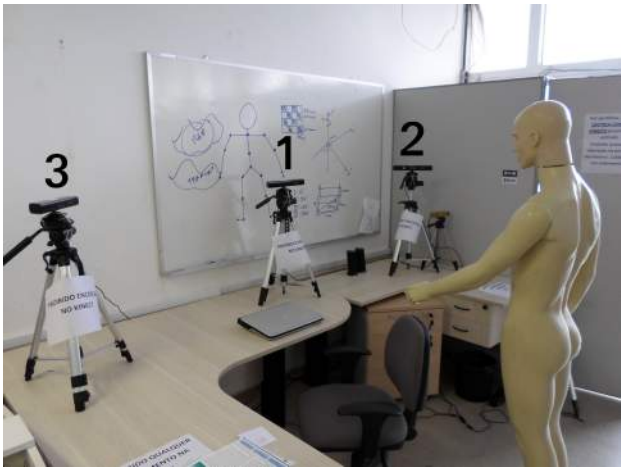

Fig. 3. Our camera setup with enumerated cameras. Camera 1 is considered the ”Master Camera” and the other cameras are the slave ones. Homograph is calculated among slaves and master camera in a pairwise manner.

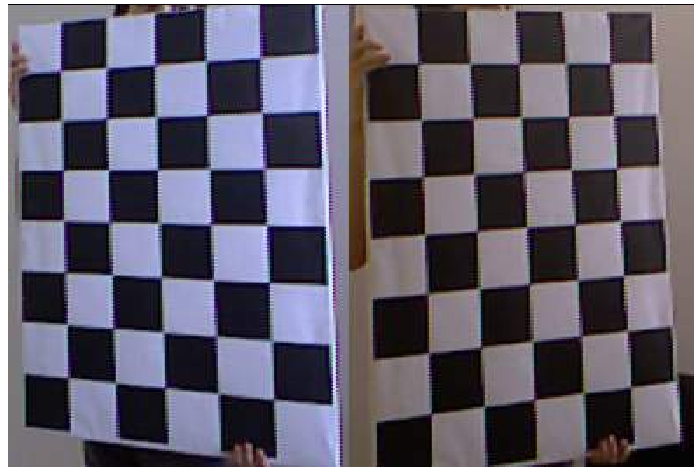

Fig. 4. Sample of the used checkerboard used. Data set .

with multiple sensors, (see fig 5), occurring even if the devices are not directly pointing to each other. This problem happens due to reflections of the IR from one device being captured by the others [13]. In [18], a solution can be found to cope with this problem.

C. Angle among the calibration cameras

When working with a multi-camera calibration, specially when using a checkerboard pattern to calibrate, the angle among the cameras should be considered. If the textured plane is rotated too much with respect to some of the cameras, the calibration methods are likely to fail in recognizing the pattern.

V. Preliminary evaluation of a practical calibration

This section describes an attempt to realize calibration of multiple Kinects using an RGB calibration. The offset between the Depth and RGB camera was considered zero, and a test using skeleton stream data was realized. With the goal of finding calibration parameters of the three RGB-D cameras in the setup, with respect to the ”master” camera (camera 1 in Fig. 3), the AMCC Toolbox [5] was used considering the RGB part of the RGB-D cameras. The toolbox is a modified version of Bouguet’s Camera Calibration Toolbox for Matlab and the Robust Automatic Detection Of Calibration Chessboards (RADOCC) toolbox. The toolbox extends automatic checkerboard detection to multiple cameras and implements entirely automated monocular and stereo calibration procedures.

A data set creator system was developed, which automatically takes image samples from all cameras at the same time to create synchronized checkerboard images between pair of cameras (see fig. 4). With the data set in hands, the toolbox was used to generate the necessary calibration.

A checkerboard (see fig. 4) with 99.5mm by 100.6mm squares was used. To be a rectangle is not a problem, for the toolbox, as it has configuration parameters for the square dimensions. Using the checkerboard, a calibration was performed between pairs of cameras, with the setup depicted in Fig. 3. Camera 1 was chosen as the ”master camera”, Camera 2 and 3 (slaves) were calibrated in pairs with Camera 1. After achieving the results, an early version of a composite skeleton system was used to transform a point in the left hand position found in the Cameras 2 and 3 coordinate systems into the same point in the master camera coordinate system, using extrinsic calibration parameters obtained by our checkerboard RGB calibration. The system considers the offset between the RGB and depth cameras equal to zero, which was proved to be a poor assumption.

To evaluate the calibration system performance, the points considered in each one of the slave cameras were mapped into the master camera coordinate system (MCCS) and the Euclidean distance was calculated among the mapped points and the reference point in the MCCS. The results were 79 cm and 19 cm, from Cameras 2 and 3, respectively, indicating a high error of point mapping.

A. Discussion

The parameters found in the RGB calibration show to have a correlation with the real world camera setup. According to the setup position, cameras are located in such a way that it is only necessary to be rotated in the Y axis to have a perfect alignment among them. In practice, rotation parameters were close to zero except on the Y axis. Also, the distances given by the translation vectors had some relation with real world observation, with almost no translation in the Y axis, since the cameras are at the same level, and close to one meter distance in the x axis.

The distances measured between transformed points and the ground truth point captured by the master camera were very high, with almost one meter of error in some cases. The calibration parameters were coherent with real world measures, and the calibration was repeated two times with different data sets, resulting in almost the same values. It is very likely that our error comes from the consideration that

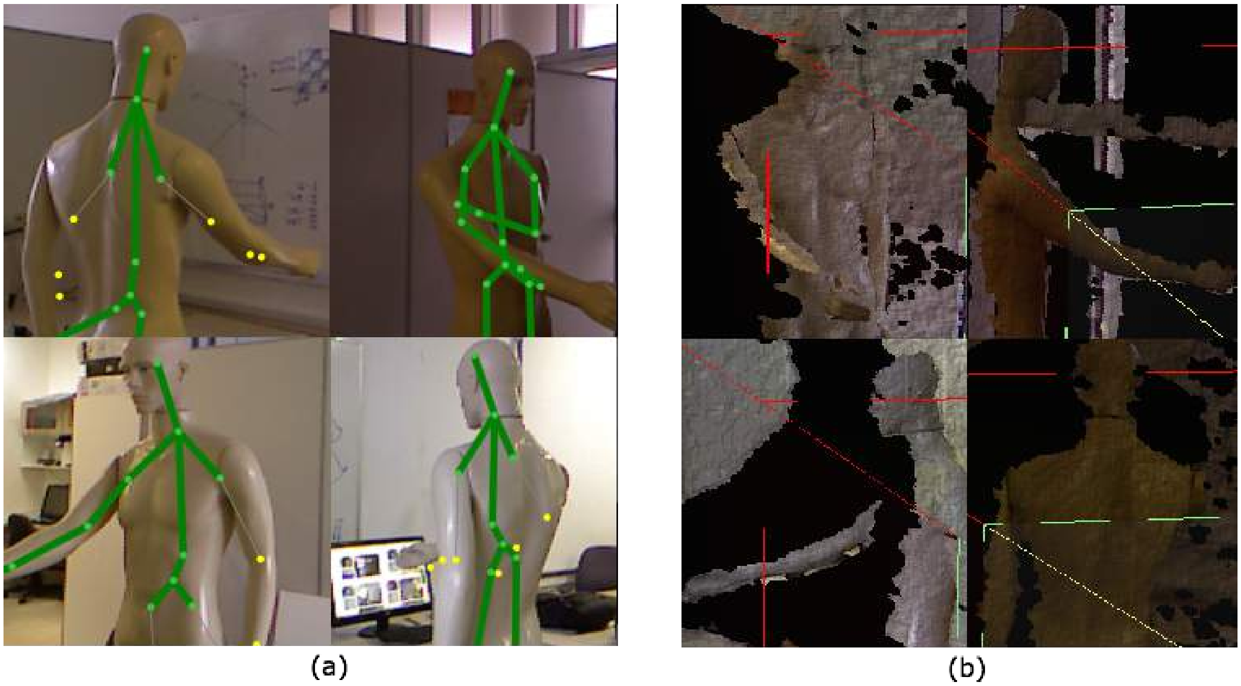

Fig. 5. Cross-talk interference among the RGB-D cameras. (a) deformed skeleton with wrongly placed joint position, and (b) pervasive holes in a point cloud reconstruction.

the offset within the RGB and depth sensors is zero. The small offset between the RGB and depth camera was considerable propagated into high measurement error when transforming depth points using RGB calibration.

VI. Conclusion

A brief camera calibration survey and a implementation of an RGB camera calibration method, using AMCC toolbox, was presented in this paper. An RGB calibration was proved not to be adequate for the composite skeleton construction. High error in hand position measurements occurred between transformed points and the ground truth point in Camera 1 coordinate system were found. A depth calibration will be necessary by using the depth map or IR images to continue our work.

A. Future Work

We are working to create a composite skeleton, made of the fusion of body tracking data from three cameras, following the work of [3]. This composite skeleton would be immune to self-occlusion and much more accurate than the one generated by only one RGB-D camera.

Acknowledgment

All the authors contributed equally. This work has been supported by Fundação de Amparo à Pesquisa do Estado da Bahia (FAPESB) under Grant JCB0062/2013.

References

[1] Seitz, S.; Curless, B.; Diebel, J.; Scharstein, D.; Szeliski, R. *A comparison and evaluation of multi-view stereo reconstruction algorithms*, Proceedings of IEEE Computer Society Conference on Computer Vision and pattern recognition, pp. 519–528, 2006.

[2] Rusu, R; Cousins, S. *3D is here: Point cloud library (pcl)*, Proceedings of IEEE International Conference on Robotics and Automation (ICRA), pp. 1–4, 2011.

[3] Kaenchan, S.; Mongkolnam, P.; Watanapa, B.; Sathienpong, S. *Automatic multiple kinect cameras setting for simple walking posture analysis*, Proceedings of International Conference on Computer Science and Engineering (ICSEC), pp. 245–249, 2013.

[4] Herrera, C.; Kannala, J; Heikkila, J. *Joint depth and color camera calibration with distortion correction*, Proceedings of IEEE Transactions Pattern Analysis and Machine Intelligence, vol. 34, issue 10, pp. 2058–2064, 2012.

[5] Warren, M.; McKinnon, D.; Upcroft, B. *Online calibration of stereo rigs for long-term autonomy*, Proceedings of International Conference on Robotics and Automation (ICRA), pp. 3692–3698, 2013.

[6] Yang, C; Gong, N.; Strozzi, D; Gimel'farb, A. *Multi-kinect scene reconstruction: calibration and depth inconsistencies*, Proceedings of International Conference on Image and Vision Computing, pp. 47–52, 2013.

[7] Alexiadis, Zarpalas, Daras. *Real-Time, full 3-D reconstruction of moving foreground objects from multiple consumer depth cameras*, IEEE Transactions on Multimedia, Vol. 15, No. 2, pp. 339–358, 2013.

[8] Geiger, A.; Moosmann, F.; Car, O.; Schuster, B. *Automatic camera and range sensor calibration using a single shot*, Proceedings of IEEE International Conference on Robotics and Automation (IROS), pp. 3936–3943, 2012.

[9] Miller, S.; Teichman, A.; Thrun, S. *Unsupervised extrinsic calibration of depth sensors in dynamic scenes*, Proceedings of IEEE International Conference on Intelligent Robots and Systems (IROS), pp. 2695–2702, 2013.

[10] Jenkins, F.; White, H. *Fundamentals of optics*, 4th ed., McGraw-Hill, 1976.

[11] Hartley, R.; Zisserman, A. *Multiple view geometry in computer vision*, Cambridge University Press, 2003.

[12] Berger, K.; Ruhl, K.; Schroeder, Y.; Bruemmer, C.; Scholz, A.; Magnor, M. *Markerless motion capture using multiple color-depth sensors*, Proceedings of Vision, Modeling and Visualization (VMV), pp. 317-324, 2011.

[13] Khoshelham, K.; Elberink, S. *Accuracy and resolution of kinect depth data for indoor mapping applications*, Sensors, vol 12, pp. 1437–1454, 2012.

[14] Tong, J.; Zhou, J.; Liu L.; Pan Z.; Yan, H. *Scanning 3D Full Human Bodies Using Kinects*, IEEE Transactions on Visualization and Computer Graphics, Vol 18, pp. 643–650, 2012.

[15] Auvinet, E.; Meunier, J.; Multon, F. *Multiple depth cameras calibration and body volume reconstruction for gait analysis*, Information Science, Signal Processing and their Applications (ISSPA), pp. 478–483, 2012.

[16] Bouguet, J.-Y. *Camera Calibration Toolbox for Matlab*, retrieved online 2015-05-29. [Online]. Available: http://www.vision.caltech.edu/bouguetj/calib_doc.

[17] Kreylos, O. *Kinect Camera Calibration*, retrieved online 2015-05-29. [Online]. Available: http://doc-ok.org/?p=289, 2013.

[18] Butler, D.; Izadi, S.; Hilliges, O.; Molyneaux, D.; Hodges, S.; Kim, D. *Shake'n'sense: reducing interference for overlapping structured light depth cameras*, ACM annual conference on Human Factors in Computing Systems, pp. 1933-1936, 2012.