Depth cameras calibration for obstructive sleep apnea diagnostic support

Dusan Koniar, Jozef Volak, Frantisek Jabloncik, Libor Hargas
Department of Mechatronics and Electronics
Faculty of Electrical Engineering, University of Žilina
Žilina, Slovak republic
Email: jozef.volak@fel.uniza.sk

***Abstract*— In our research, we use depth cameras to obtain precise 3D models of patients' heads and faces. Visual system based on these cameras is basic element of developed screening tool for obstructive sleep apnea. Decreasing of scanning time plays key role mainly for pediatric patients and suppresses generation of motion artifacts in 3D models. Used Time of Flight sensors are relatively non-expensive solution, especially in our case (parallel usage of several cameras). To obtain accurate models, the step of camera calibration is very important. In this article, we bring detailed look on geometrical distortion (radial and tangential) of used cameras and also methods for spatial calibration of system composed from three cameras. As software tool, we used Camera Calibrator implemented in MATLAB. As result of mentioned research, we found suitable multi-camera calibration method and set of parameters related to our visual system (given topology and geometrical properties of scanning cabin). Calibration decreased average distance error (measured using Hausdorff distances) from ca. 3 mm before calibration to ca. 2 mm after calibration and made easier the registration of partial models to the resulting one. Designed visual system based on depth cameras with proper calibration and interference suppression seems to have big potential for obstructive sleep apnea screening not only in pediatrics.**

Keywords—3D model, depth cameras, obstructive sleep apnea, camera calibration

I. INTRODUCTION

Obstructive sleep apnea syndrome (OSAS) is a disease that is one of the most common sleep problems in different age groups of people [1]-[3]. A comprehensive diagnostics of this disorder occurs in specialized sleep labs using videopolysomnography. However, this examination is time consuming and costly and causes long waiting periods for patients. An ancillary solution is to diagnose risky patients who may then be prioritized for complex medical examination. In practice, a questionnaire called "Clinical Sleep Record" is used, which determines the probability of the disease in a subjective manner [4]. The results of the questionnaire are often influenced by the subjective assessment of the specialist. This type of screening is based on disease demonstration on patient's head. This is the reason for capturing accurate 3D models of patient's head and face [5].

Objectification of the examination can be increased by a screening tool based on depth cameras that can replace manual measurements of facial parameters predisposing OSAS. At present, specific properties are measured by hand-held contact meters. This method of measurement is slow and stressful especially for patients in the pediatric population. Patient stress often avoids objective access to necessary information. Contactless measurement method would increase the objectivity and accuracy of the measurement. It would also reduce the duration of the examination and allow the measurement of parameters that are not usually recorded.

In addition, RGB-D cameras can also capture color/texture information and join it with depth data. Using this combination, it is then possible to reproduce not only the geometric features of the patient's head and face, but also fully textured 3D model. 3D object reconstruction is a technical problem that is applied in many fields from the gaming industry through engineering to medicine. Just in medicine, 3D modeling is interdisciplinary task. The patient model can be used in various examinations to create "patientless approach". To make the most accurate model, RGB-D cameras need to be calibrated to eliminate the errors caused by distortion of the lens. Result of this research is calibration method and set of parameters suitable for calibration of selected cameras in defined topology.

II. METHODS

Creating of 3D scans of heads and faces can be done by two possible basic ways:

- single-camera system: static camera - dynamic object, static object - dynamic camera,
- multi-camera system: static object - static cameras, static cameras - dynamic object.

In the first point, the main difference is the number of cameras used. In a single-camera system, only one camera is used, which must capture the object in desired angles. This can be done by rotating the camera around a static object or rotating an object around a static camera. Single-camera system is more time consuming against the multi-camera system. The patient (especially pediatric one) could do many unwanted or random movements, therefore the accuracy of the model is decreased by the serious motion artifacts. From these reasons usage of multi-camera system seems to be better solution in the aspect of time and accuracy.

In a multi-camera system, the cameras are located in the space in defined topology to capture all the necessary details in the shortest time. Scanning is faster but more challenging for hardware and software. The key role in this multi-camera system is to do proper calibration.

A. *Mathematical model of camera with lens distortion*

For single-camera systems, the geometric camera calibration estimates the lens and image sensor parameters. These parameters are used to correct lens distortion and measure the size of objects in real-world units. Multi-camera calibration is performed for spatial registration of point clouds obtained from 3D reprojections of each camera. Camera parameters consist of internal (intrinsic), external (extrinsic) coefficients and distortion factors. For their estimation it is required to know 3D points and their corresponding 2D locations. Obtaining these correspondences is possible by extracting easily identifiable points. One of the most used calibration patterns is a chessboard where is a sufficient intensity contrast between the playing fields.

Calibration uses the pinhole camera model and lens distortion. The pinhole camera model does not contain lens distortion because the ideal camera does not have a lens. The calculated distortion must be added to the equations describing ideal camera model [6], [7].

Calibration provides information about:

- the relative position of the camera and the subject being captured,
- error of 3D object replication in 2D plane,
- parameter estimation error.

Equation 1 represents the interconnection of the 2D plane and the 3D scene. Parameter s is a ratio factor, points [u, v, 1] are 2D plane points and points [X, Y, Z, 1] are 3D scene points. Camera parameters are represented by a 4x3 matrix called the camera matrix P [8]:

$$S\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = P\begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \rightarrow P = \begin{bmatrix} R \\ t \end{bmatrix} K = \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \end{bmatrix} \begin{bmatrix} f_x & 0 & c \\ 0 & f_y & c \\ 0 & 0 & 1 \end{bmatrix} \quad (1)$$

The camera matrix P consists of intrinsic and extrinsic parameters. Matrix R represents rotation and *t* translation, which are extrinsic parameters [8]. Matrix K represents the intrinsic parameters:

- c_x,c_y: the principal point, usually in the center of the image, the intrinsic parameters,
- f_x, f_y: focal lengths expressed in pixel units, intrinsic parameters,
- r_{xxx}: camera rotation parameters,
- t_x: camera translation parameters,
- *s*: scaling factor.

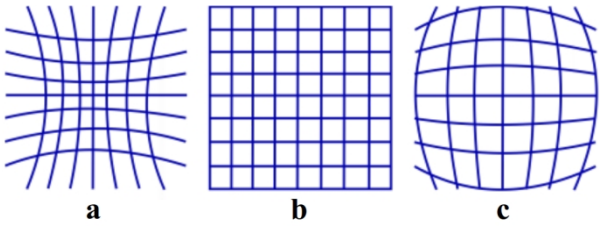

Fig. 1. Radial distortion: (a) negative, (b) zero, (c) positive.

Therefore, if the camera image is changed by factor *s*, all parameters should be multiplied or divided by the same factor. The matrix of internal parameters does not depend on the displayed scene, so it will be constant at static focal length. Radial and tangential distortion occurs in real cameras. Most commonly, radial distortion is demonstrated by the formation of a barrel or pincushion image as shown in Fig. 1. Tangential distortion occurs when image plane and lens are not parallel (Fig. 2). Equations of radial and tangential distortion are denoted in 2a-2d [8]:

$$x_{radial} = x\left(1 + k_1 r^2 + k_2 r^4 + k_3 r^6\right) \quad (2a)$$

$$y_{radial} = y(1 + k_1 r^2 + k_2 r^4 + k_3 r^6) \quad (2b)$$

$$x_{tangential} = \left[2 p_1 xy + p_2\left(r^2 + 2 x^2\right)\right] \quad (2c)$$

$$y_{tangential} = \left[2 p_2 xy + p_1\left(r^2 + 2 y^2\right)\right] \quad (2d)$$

where:

- k_1, k_2, k_3: radial distortion coefficients of the lens,
- p_1, p_2: tangential distortion coefficients of the lens,
- *x*, *y*: undistorted pixel locations in normalized image coordinates,
- $r^2 = x^2 + y^2$.

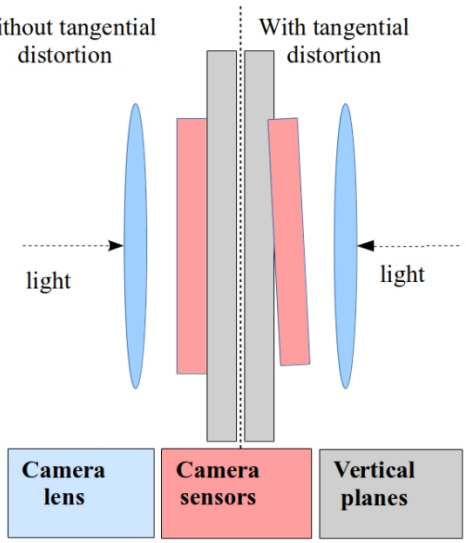

Fig. 2. Camera sensor and camera lens placement with and without tangential distortion.

B. *Reprojecting images to 3D*

By reprojecting, we consider transferring the 2D plane to the 3D space. Transformation equations:

$$X = \frac{(x - c_x)\,depth}{f_x\left(X_{tangential} + X_{radial}\right)} \quad (3a)$$

$$Y = \frac{(y - c_y)\, depth}{f_y (Y_{tangential} + Y_{radial})} \quad (3b)$$

$$Z = depth \quad (3c)$$

are derived from the pinhole camera model, equations of lens distortion and perspective transform. The input data to the equations are images obtained from the RGB-D camera [9], [10]. The RGB-D camera provides information about the color and depth of the scene being shot. The depth image is represented by a gray-scale image, where the value of individual pixels represents the absolute distance of the cameras from the object (depth in equations).

B. Camera calibration

To obtain intrinsic, extrinsic camera parameters and lens distortion coefficients, the camera calibration in MATLAB was performed using the Camera Calibrator [11]. For the geometrical calibration, we used RGB and IR images. Each such image contains the calibration patterns, and images are captured from 30 different angle views and distances (Fig. 3a, 3c). Generally, chessboard pattern is very useful for calibration for many reasons: sharp edges between black and white fields, good contrast, easily detectable key points such as corners of fields etc. In our case, we used chessboard with field size 36 mm x 36 mm. Undistorted images are presented on Fig. 3b and Fig 3d. Original images are positive distorted.

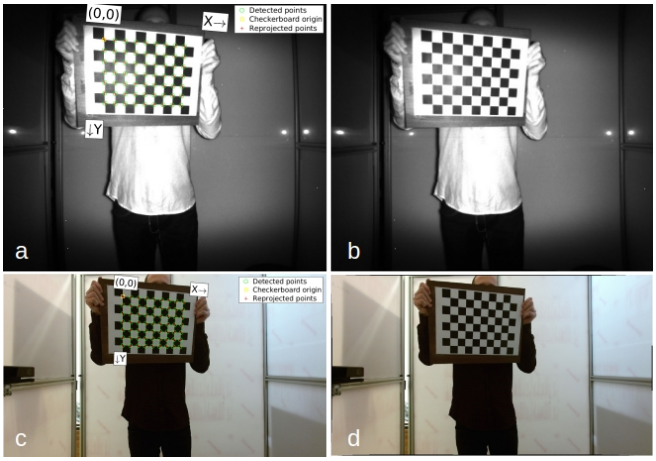

Fig. 3. IR and RGB calibration images and undistorted images: (a) input IR images with calibration pattern, (b) undistorted IR image, (c) input RGB images with calibration pattern, (b) undistorted RGB image.

In our research work we use several depth cameras working simultaneously to decrease scanning time. Decreasing time is very important to minimize the motion artifacts caused by movement of patients. Cooperation with pediatric patient is often very complicated. Calibration of given multi-camera visual system plays crucial role in accuracy of resulting 3D model (in the process of registration of point clouds).

The principle of multi-camera system calibration is used to identify the relative positions of cameras in space. In our visual system implemented in scanning cabin we used 3 cameras Microsoft Kinect v2. The first, each camera was corrected using intrinsic calibration parameters described above. Then, the next calibration process was focused on calculating extrinsic parameters depending on defined camera topology. In the first step, frontal view camera (camera 1 in Fig. 4) was set as reference camera. In second step mutual calibration between camera 1 and camera 2 and also camera 1 and camera 3 was made. The matrix of extrinsic parameters for camera 2 and camera 3 (rotation and translation coefficients) is the result of this step.

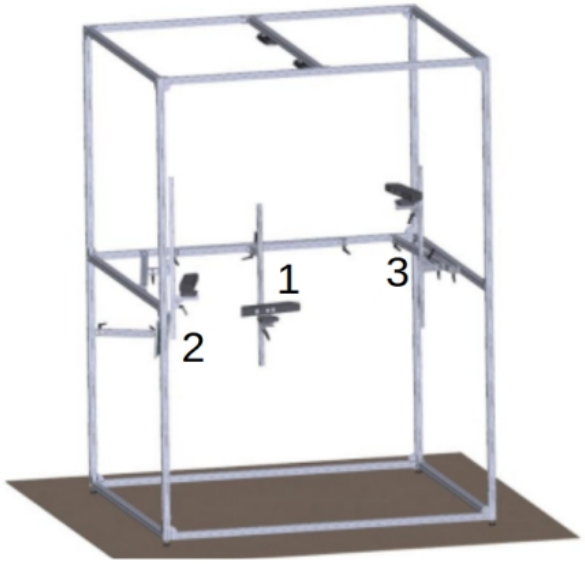

Fig. 4. Scanning cabin with defined topology of cameras.

Fig. 5. Multi-camera calibration, example of image pair between reference and side camera.

In second step of calibration, we must ensure that camera 2 and 3 can see entire calibration pattern in the same time (as we can see in Fig. 5). That was done by software triggering of cameras. In this step we used 20 registered RGB to D images from different views for each camera pair (1-2, 1-3).

III. RESULTS

In our results we compare difference between calibrated and uncalibrated intrinsic and extrinsic parameters of cameras used in our visual system. This comparison is divided into two steps: Hausdorff distance between model and reference after intrinsic calibration; accuracy of partial point clouds alignment to resulting model after extrinsic calibration. In Table I we can see the difference between factory settings of intrinsic parameters of RGB and IR sensors for each camera Microsoft Kinect v2 used in our multi-camera visual system. Values from cameras were acquired via Libfreenect 2 (and also written to the cameras after calibration).

TABLE I. CALIBRATED AND UNCALIBRATED INTRINSIC PARAMETERS OF CAMERAS

	Camera 1		**Camera 2**		**Camera 3**	
Parameter	*Original*	*Calibrated*	*Original*	*Calibrated*	*Original*	*Calibrated*
RGB f_x	1081.37	1060.62	1081.37	1052.59	1081.37	1056.74
RGB f_y	1081.37	1063.05	1081.37	1054.43	1081.37	1057.52
RGB c_x	959.5	926.11	959.5	930.29	959.5	943.65
RGB c_y	539.5	559.79	539.5	536.04	539.5	547.36
IR f_x	363.34	364.72	366.13	375.52	363.71	365.98
IR f_y	363.34	365.43	366.13	375.65	363.71	366.26
IR c_x	259.91	258.82	257.53	257.29	254.79	259.24
IR c_y	206.37	208.97	204.53	202.27	209.70	212.10
IR k_1	0.1	0.11	0.088	0.12	0.12	0.09
IR k_2	-0.269	-0.33	-0.27	-0.31	-0.34	-0.23
IR k_3	0.08	0.15	0.098	0.03	0.15	0.03

The impact of intrinsic calibration is evaluated using Hausdorff distance between reference object (model was captured by accurate laser scanner) and models captured by calibrated and uncalibrated cameras. We captured 5 calibrated and 5 uncalibrated models (Fig. 6) using camera 1 (frontal view in our topology).

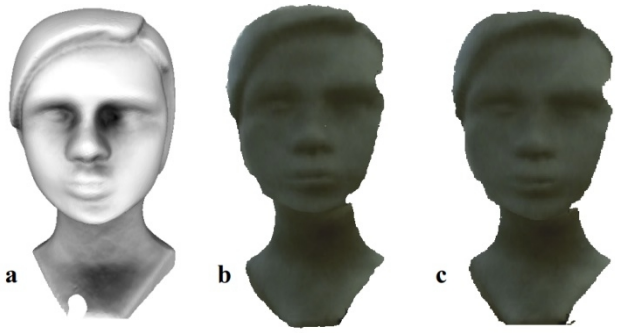

Fig. 6. 3D models of child's head: a) accurate reference, b) uncalibrated model, c) calibrated model.

Each model was captured in constant ambient conditions and for 3D reconstruction 14 RGB-D was used. Average distance of object from camera was 0,5 m. Acquired models was aligned with reference one using Iterative Closest Point (ICP) and Hausdorff distance was computed [12], where target was reference accurate model and sampled mesh was acquired by Kinect camera. Statistical assessment of measurement is in Table II. In this table Average points represents the number of points used for computing Hausdorff distance, Mean error represents average absolute Hausdorff distance and RMS represents RMS error.

TABLE II. STATISTICAL ASSESSMENT OF MEASURMENT AFTER INTRINSIC CALIBRATION

Mesh	**Camera 1 Uncalib - Reference**	**Camera 1 Calib - Reference**
Average points [pts]	185187	296585
Mean [mm]	3,0206	2,1832
RMS [mm]	3,8444	2,9506

In the Table II we can see that average error was decreased by approximately 30% after intrinsic calibration. Difference between models based on Hausdorff distance is also shown as color map in Fig. 7. In uncalibrated models we can see that maximal error is on the model boundaries what is probably caused by positive radial distortion of cameras. Error of calibrated model has uniform character of distribution and it is mostly caused by noise processes during scanning.

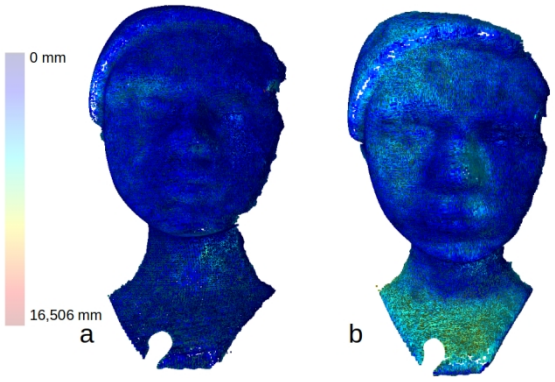

Fig. 7. Hausdorff distance as color map: a) for calibrated model, b) for uncalibrated model.

After multi-camera (extrinsic) calibration we get a group of rotation and translation parameters for camera 2 and 3 (in our topology, Fig. 4). Computed extrinsic parameter for our systems are listed in Table III.

TABLE III. MULTI-CAMERA SYSTEM EXTRINSIC PARAMTERES

	r_{11}	r_{12}	r_{13}	r_{21}	r_{22}	r_{23}	r_{31}	r_{32}	r_{33}	t_1	t_2	t_3
Camera 1	1	0	0	0	1	0	0	0	1	0	0	0
Camera 2	0.09	0.16	-0.98	-0.14	0.98	0.15	0.99	0.13	0.11	0.56	-0.07	0.5
Camera 3	0.02	-0.1	0.99	0.15	0.98	0.09	-0.99	0.15	0.04	-0.74	-0.09	0.62

Comparison of resulting model before and after extrinsic calibration is shown in Fig. 8. Partial models from each camera are acquired using 14 RGB-D images. In Fig. 8a we can see the position of partial models without extrinsic calibration and in Fig. 8b with extrinsic calibration. In the first sight we can see that final registration of resulting model is much easier after extrinsic calibration. The final registration without extrinsic calibration is often impossible.

Fig. 8. Partial models position in resulting model: a) without extrinsic calibration, b) with extrinsic calibration.

Head from resulting model acquired with our system after extrinsic calibration was cropped and registered with reference object using ICP method (as in case of intrinsic calibration). The accuracy was assessed again using Hausdorff distance (Table IV).

TABLE IIV. STATISTICAL ASSESSMENT OF MEASURMENT AFTER EXTRINSIC CALIBRATION

Mesh	**Extrinsic Calibrated - Reference**
Points [pts]	1071662
Mean [mm]	2.364
RMS [mm]	2.950
Max [mm]	15.92

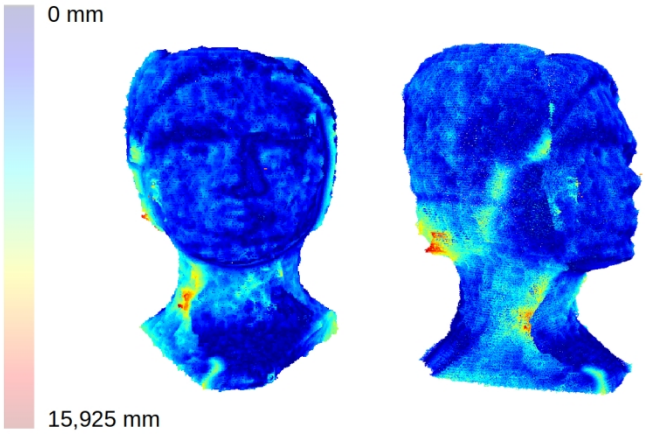

Fig. 9. Hausdorff distance map: reference model vs. model after extrinsic calibration and final registration.

Mean and RMS error are very similar to results in Table I. Extreme values of errors are present in the place of overlapping of partial models (Fig. 9).

IV. Discussion

As mentioned in Introduction, usage of multi-camera visual system is a potential way how to decrease scanning times for obtaining 3D scan of object and increase the model accuracy. The important consequence of this approach is that the cameras must be calibrated. The first, intrinsic calibration is done to repair distortion caused by lens and

chip. Extrinsic calibration computes the set of parameters which cause that the same point seen by several cameras from different distances and angles is localized correctly in 3D space. Accurate 3D models are subsequently used to find key-points on the models (cranio-facial cephalometric features) and the important distances are measured. These values are base for scoring system evaluating the risk factor of OSA syndrome.

In our 3-camera system we used typical chessboard calibrating pattern and we computed parameters for cameras. These parameters were the used in scanning phase.

For comparison of acquired models and accurate reference model we used Hausdorff metric to compute the error. The average distance errors after calibration was reduced to 70% and extrinsic calibration made the registration of partial models (point clouds) much easier. We used ICP method for registration.

Noise and interference processes are the source of remaining distance errors. In the future the suitable methods of noise filtering and some methods for interference suppression will be verified during the scanning phase. System is now experimentally used in clinic environment in Slovakia to check the scanning algorithms and methodology on pediatric patients.

V. Conclusion

In our article we presented a method and need for calibrating cameras as part of visual system for capturing 3D models of heads. Due to usage of multi-camera visual system we must calibrate not only the single cameras (intrinsic calibration) but also extrinsic calibration (computation of rotation and translation parameters of cameras) must be provided. Multi-camera system is implemented to decrease scanning time to the possible minimum to prevent models from motion artifacts. The second reason is that system will be used as predictive screening tool for medical diagnostics and cooperation with patients (especially pediatric ones) is often very complicated - reducing the scanning time reduces the stressing of patient. Considering our visual system composed of parallel depth cameras, we implement absolutely non-invasive diagnostic tool and acquired models support “patientless approach”. Proposed diagnostic method also would be maximally automated to reduce human errors during measurement phase. The first and crucial assumption is to have accurate models. Last but not least we respect the financial aspect of entire system and replacing expensive laser scanners by the cheaper solutions.

References

[1] W. T. McNicholas and M. R. Bonsignore, “Sleep apnoea as an independent risk for cardiovascular disease: Current evidence, basic mechanisms and research priorities,” Eur. Respir. J., vol. 29, no. 1, pp. 156–178, 2007.

[2] C. L. Marcus et al., “Diagnosis and management of childhood obstructive sleep apnea syndrome,” Pediatrics, vol. 130, no. 3, pp. 576–584, 2012.

[3] E. O. Bixler et al., “Sleep disordered breathing in children in a general population sample: Prevalence and risk factors,” Sleep, vol. 32, no. 6, pp. 731–736, 2009.

[4] M. P. Villa et al., “Sleep clinical record: An aid to rapid and accurate diagnosis of paediatric sleep disordered breathing,” Eur. Respir. J., vol. 41, no. 6, pp. 1355–1361, 2013.

[5] J. Volák, D. Koniar, F. Jablončík, and L. Hargaš, “A Study on OSAS Diagnostics Supported by RGB-D Imaging,” 2018 41st Int. Conf. Telecommun. Signal Process. TSP 2018, pp. 497–500, 2018.

[6] Y. M. Wang, Y. Li, and J. B. Zheng, “A camera calibration technique based on OpenCV,” Proc. - 3rd Int. Conf. Inf. Sci. Interact. Sci. ICIS 2010, pp. 403–406, 2010.

[7] A. Fetić, D. Jurić, and D. Osmanković, “The procedure of a camera calibration using Camera Calibration Toolbox for MATLAB,” MIPRO 2012 - 35th Int. Conv. Inf. Commun. Technol. Electron. Microelectron. - Proc., pp. 1752–1757, 2012.

[8] J. Heikkila, O. Silvcn, and I. Oulu, “A Four-step Camera Calibration Procedure with Implicit Image Correction” pp. 1106–1112, 1997.

[9] Z. Zhengyou, “A Flexible New Technique for Camera Calibration,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 22, no. 11, pp. 1330–1334, 2000.

[10] G.-S. Huang and Y.-Y. Tseng, “Application of Stereo Vision 3D Target Recognition Using Camera Calibration Algorithm,” in Proceedings of the 2015 AASRI International Conference on Circuits and Systems, 2015.

[11] “Camera Calibration Toolbox for Matlab”. [Online]. Available: http://www.vision.caltech.edu/bouguetj/calib_doc/. [Accessed: 21-Oct-2019].

[12] P. Cignoni, C. Rocchini, and R. Scopigno, “Metro: Measuring Error on Simplified Surfaces,” Comput. Graph. Forum, vol. 17, no. 2, pp. 167–174, 1998.