An Omni-RGB+D Camera Rig Calibration and Fusion using Unified Camera Model for 3D Reconstruction

Ahmad Zawawi Jamaluddin, Osama Mazhar, Cansen Jiang, Ralph Seulin, Olivier Morel, David Fofi

An Omni-RGB+D Camera Rig Calibration and Fusion using Unified Camera Model for 3D Reconstruction.

Ahmad Zawawi Jamaluddin, Osama Mazhar, Cansen Jiang, Ralph Seulin, Olivier Morel, David Fofi
Laboratoire Electronique, Informatique et Image (Le2i), Université Bourgogne Franche-Comté, France

ABSTRACT

The perfect vision system could be a system which can obtain surround images or information at once. We present a vision system which can view the image in 360 with high- resolution depth information. The proposed vision system is compact and rigid with two fisheye cameras that provide a 360-degree field of view(FoV). Alongside, a high-resolution stereo vision camera is mounted to monitor anterior FoV for precise depth perception of the scene. To effectively calibrate the proposed camera system, we offer a novel camera calibration approach taking the advantages of Unified Camera Model representation. The proposed calibration method outperforms the state-of-the-art methods. Moreover, we proposed more affective algorithm in fusing the two fisheye images into a single unified sphere, which offers seamless stitching results. This new omni-vision rig system is designed to obtain sufficient information to be used on a robot for object detection and recognition. A large scale SLAM and dense 3D reconstruction can be achieved taking the advantage of the large FoV.

Keywords: Camera calibration, vision system, fisheye camera, 3D reconstruction

1. INTRODUCTION

The research and development on computer vision are extremely increasing in parallel with the development on robotics technology[1]. The fabrication of hybrid camera systems with wide FoV, combine with the computer vision technique makes this research more interesting. The objectives of this research were to fully utilise vision sensor as the useful kit for robotics navigation. This paper presents the novel camera system which can provide a 360° FoV and the depth information concurrently. The system minimizes the use of equipment and image-data and has the ability to acquire sufficient information on the scene. The applications of omnidirectional cameras are mainly for robotics such as localization and mapping, robot navigation[2], object tracking, visual servoing, structure from motion/motion from structure and virtual reality.

1.1 Concept and Motivation

Most of the living creatures have their own capability to look on the object. The ability of view depends upon the position, form and structure of the eyes. The omni vision cameras rig has been developed by referring to the two categories of animals, prey, and predator. A prey animal has a wide FoV but a small binocular view. They use the eyes for observing the environment from the predator or other threats. While the predator has a large frontal binocular view for targeting and attacking. The same vision system can be realized using multiple camera for the purpose of object observation, recognition and detection[3].

1.2 Proposed System

The proposed vision system consists of two CCD cameras mounted with a fisheye lens with each has FoV 185°. The fisheye cameras are placed back to back so that they cover the whole 360° FoV vertically and horizontally. A high-resolution stereo vision camera is placed in front of the rig, so that its baseline is in parallel with the baseline of the fisheye cameras. The stereo vision camera named ZED camera, provides a high-resolution RGB image with depth information. Fig.1 shows the illustration and the views of camera rig.

Further author information:
Ahmad Zawawi Jamaluddin : E-mail: ahmad.jamaluddin@u-bourgogne.fr, Telephone: +33(0)385731090

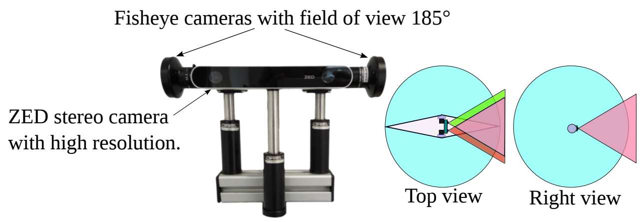

Figure 1. The front view of the proposed system with the desire FoV.

2. HYBRID CAMERA SYSTEM

2.1 Omnidirectional Camera

There are three main types of the omnidirectional camera which each has their own advantages. A dioptric system[4], has a FoV more than 180°. It consists of a single camera mounted with a special lens to increase the FoV. The catadioptric camera is another type of omnidirectional camera[5]. It has more than 180° FoV with a single image. It consists of a camera system with a cone shape reflected mirror. This camera produces also a black spot which is the image of the camera itself. The third types is polydioptric camera. It consists of several overlapping and non-overlapping identical cameras to obtain a panoramic 360° FoV. Fig.2 illustrates the different omnidirectional vision systems.

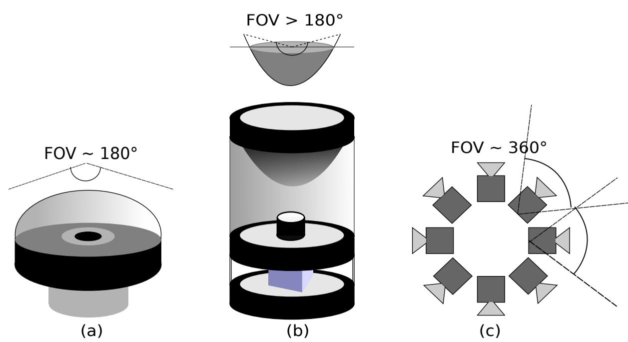

Figure 2. The illustration shows the omniderectional vision systems (a) Dioptric, (b) Catadioptric and (c) Polydioptric.

The linear perspective geometry has been preserved from any camera system with a single viewpoint. Normally, a single viewpoint exists in all omnidirectional camera. It allows the omnidirectional camera to extract a different view, from perspective to panoramic view. However, for polydioptric which consists of several identical cameras, it considered either they have a unique viewpoint[4][6]. The multi cameras system has a stereovision capability or to enhance the FoV, but it is impossible for the camera rig to have a single effective viewpoint.

2.2 Stereo vision Camera

Stereovision is a technique used to build three dimensional description of a scene observed from different viewpoints [7]. This technique is used to perceive depth information through generating disparity maps, which then is used to detect obstacles in the environment [8]. This is a classical technique that helps in the field of robotics for localization, navigation and obstacle detection since several decades. Lately, the release of ZED Stereo Camera[9] enables the researchers to get high-resolution three dimensional depth sensing. Fig. 3 shows the illustration of stereo vision system.

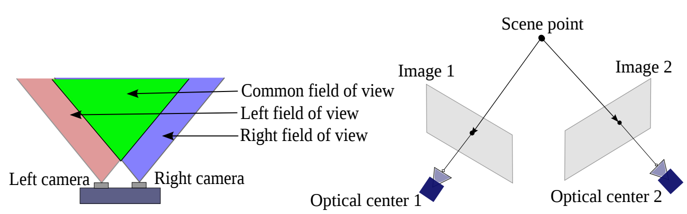

Figure 3. The Illustration shows the stereo vision system. (a): Left and right cameras have their own view, the green colour is the common view or stereovision. (b): A simplified stereovision model.

3. METHODOLOGY

3.1 Unified Spherical Camera Model

The unified spherical camera model has been proposed by Barreto J.P[6][10]. The image formation on dioptric camera was effected by the radial distortion. Due to that, the point on the scene is not linear with the point on the dioptric image.

Then, the model was extended by Christopher Mei[11] and he developed the calibration toolbox to calibrate the omnidirectional camera. The model was enhanced by introducing another type of distortion called tangential distortion which is incorporated with the radial distortion. Fig. 4 shows the Mei's projection model from camera to unified spherical model. The eccentricity, ξ parameter which defines the amount of distortion has been explained. Mei's projection model[11][15] has been used as a reference and it provides procedures to map the image on an unified spherical model.
Let consider a 3D point $\chi = (X,Y,Z)^T$ in the world, and project it to the unit sphere with C_m as a center of the sphere.

$$\chi_s = \frac{\chi}{\|\chi\|} = (X_s, Y_s, Z_s)^T, \tag{1}$$

Then the point χ_s mapped to the new reference frame with the new center C_p

$$(\chi_s)_{F_m} \rightarrow (\chi_s)_{F_p} = (X_s, Y_s, Z_s)^T, \tag{2}$$

Next, the point projected onto the normalised plane

$$m_u = (\frac{X_s}{Z_s+\xi}, \frac{Y_s}{Z_s+\xi}, 1)^T = \hbar X_s, \tag{3}$$

The model of distortion (tangential and radial) are added to the projection model. It consists of three radial and two tangential distortion parameters.

$$x_c = x1 + k_1 r^2 + k_2 r^4 + k_5 r^6 + 2k_3 xy + k_4(r^2 + 2x^2), \tag{4}$$

$$y_c = y1 + k_1 r^2 + k_2 r^4 + k_5 r^6 + 2k_4 xy + k_3(r^2 + 2y^2), \tag{5}$$

where:

$$r = \sqrt{x^2 + y^2}, \tag{6}$$

and the sum of distortion is

$$m_d = m_u + D(m_u, V), \tag{7}$$

where V contains the coefficients of distortion.

$$V = (k_1, k_2, k_3, k_4, k_5, k_6), \tag{8}$$

and finally, the point m_d is projected to the image plane using K, which is a generalized camera projection matrix. The value f and η should be also generalized to the whole system (camera+lens).

$$p = Km_d = \begin{bmatrix} f_1\eta & f_1\eta\alpha & u_0 \\ 0 & f_2\eta & v_0 \\ 0 & 0 & 1 \end{bmatrix} m_d, \tag{9}$$

where the $[f_1, f_2]^T$ is the focal length, (u_0, v_0) is the principal point and α is the skew factor. Finally, by using the projection model, the point on the normalized camera plane can be lifted to the unit sphere by the following equation:

$$\hbar^{-1}(m_u) = \begin{bmatrix} \frac{\xi+\sqrt{1+(1-\xi^2)(x^2+y^2}}{x^2+y^2+1} x \\ \frac{\xi+\sqrt{1+(1-\xi^2)(x^2+y^2}}{x^2+y^2+1} y \\ \frac{\xi+\sqrt{1+(1-\xi^2)(x^2+y^2}}{x^2+y^2+1} - \xi \end{bmatrix}, \tag{10}$$

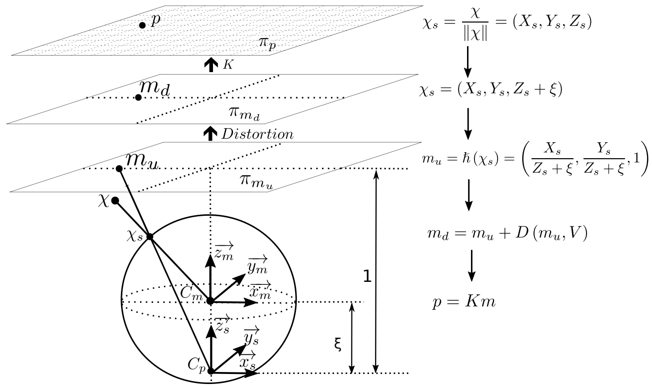

Figure 4. Mei's projection model from camera to unified spherical model. This figure is the courtesy of Christopher Mei [11].

3.2 Fisheye Camera Calibration.

Mei's calibration toolbox has been used to estimate the intrinsic parameters, such as the principle point, focal length and the enfolding parameter, ξ. For the proposed multi-camera setup, there are two fisheye cameras rigidly attached opposite to each other. Since the FoV of the fisheye camera has 185^o, the proposed setup has an overlapping area between the two fisheye cameras periphery. The 2.5^o overlapping area has been used to extract the extrinsic parameters to estimate the transformation matrix between two fisheye cameras. The fusion of right fisheye to the left fisheye image and zed image to the unit sphere are used the rigid 3D transformation by assuming the pure rotation. The Interior Point Optimization Algorithm is used to to estimate the rotation between the projected points[12].

3.3 Epipolar Geometry of Omnidirectional Camera.

The epipolar geometry for an omnidirectional camera has been studied and it originally used for a catadioptric camera as a model[13]. The study was extended to the dioptric or fisheye camera system. Fig. 8 shows the epipolar geometry of fisheye camera. Lets consider the two positions of a fisheye camera which observed a point P in the space. The points P_1 and P_2 are the projection point P onto the unit sphere image in two different fisheye's positions[10]. The points P, P_1, P_2, O_1 and O_2 are coplanar, such that:

$$\begin{aligned} \overline{O_1O_2} \times \overline{O_2P_1} \cdot \overline{O_2P_2} &= 0, \\ O_1^2 \times P_1^2 \cdot P_2 &= 0, \end{aligned} \tag{11}$$

where, O_1^2 and P_1^2 are the coordinates of O_1 and P_1 in coordinate system X_2, Y_2, Z_2. The transformation between system X_1, Y_1, Z_1 and X_2, Y_2, Z_2 can be described by rotation R and translation t. The transformation equations are:

$$\begin{aligned} O_1^2 &= R \cdot O_1 + t = t, \\ P_1^2 &= P_1 \cdot O_1 + t, \end{aligned} \tag{12}$$

By substituting (12) in (11) we get,

$$P_2^T E P_1 = 0, \tag{13}$$

where, $E = [t]_\times R$ is the essential matrix which consists of rotation and translation. In order to estimate the essential matrix, the points correspondence pairs on the fisheye images are stacked into the linear system, thus the overall epipolar constraint becomes.

$$Uf = 0\,,\, where\ U = [u_1, u_2, ..., u_n]^T \tag{14}$$

and u_i and f are vectors constructed by stacking column of matrices P_i and E respectively.

$$P_i = P_i P_i^{'T}, \tag{15}$$

$$E = \begin{bmatrix} f_1 & f_4 & f_7 \\ f_2 & f_5 & f_8 \\ f_3 & f_6 & f_9 \end{bmatrix}, \tag{16}$$

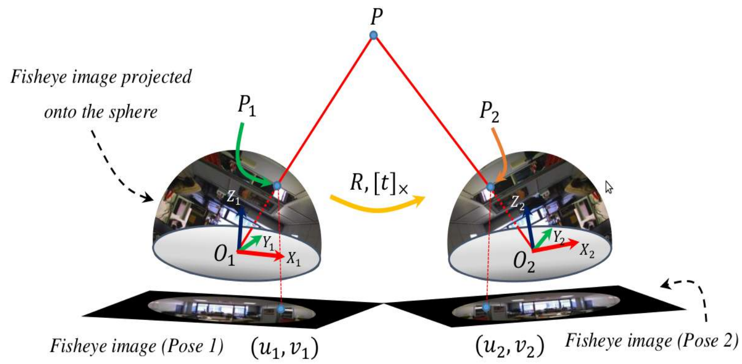

Figure 5. The epipolar geometry of fisheye camera.

The essential matrix can be estimated with linear least square by solving equation (18) and (19), where $P_{i'}^{i}$ is the projected point which corresponds to P_2 of the Fig. 5, U is $n \times 9$ matrix and f is 9×1 vector containing the 9 elements of E. The initial estimate of essential matrix is then utilized for the robust estimation of essential matrix and a modified iteratively re-weighted least square method for omnivision cameras is proposed which is originally explained in [14]. This assigns minimal weights to the outliers and noisy correspondences. The weight assignment is performed by finding the residual r_i for each point.

$$r_i = f_1 x_i' x_i + f_4 x_i' y_i + f_7 x_i' z_i + f_2 x_i y_i' + f_5 y_i y' i + f_8 y_i' z_i + f_3 x_i z_i' + f_6 y_i z_i' + f_9 z_i z_i', \tag{17}$$

$$err \rightarrow \min_f \sum_{i=1}^{n} \left(w_{Si} f^T u_i \right)^2, \tag{18}$$

$$w_{Si} = \frac{1}{\nabla r_i}, \, where \ r_i = (r_{xi}^2 + r_{yi}^2 + r_{zi}^2 + r_{xi'}^2 + r_{yi'}^2 + r_{zi'}^2)^{\frac{1}{2}}, \tag{19}$$

where, w_{Si} s the weight (known as Sampson's weighting) that will be assigned to each set of corresponding point and ∇r_i is the gradient; r_{xi} and so on are the partial derivatives found from equation (17), as $r_{xi} = f_1 x_i' + f_2 y_i' + f_3 y_i'$.
Once all the weights are computed, U matrix is updated as follow,: $U = WU$.
where, W is a diagonal matrix of the weights computed using equation (18). The essential matrix is estimated at each step and forced to be of rank 2 in each iteration. The procrustean approach is adopted here and singular value decomposition is used for this purpose.

4. EXPERIMENTAL RESULTS

4.1 The estimation of intrinsic parameters

The unknown parameters f_1, f_2, u_0, v_0 and ξ of fisheye cameras are estimated using the omnidirectional camera toolbox provided by Christopher Mei's. The fisheye images are projected onto the unit sphere using the Inverse Mapping Function (10) defined in Christopher Mei's camera. Fig. 7 shows the images from left and right fisheye cameras projected onto the unit sphere.

4.2 The estimation of extrinsic parameters

4.2.1 Rigid transformation between two fisheye images.

Our system has an overlapping features about 2.5^o between the left and right hemispheres. The selected points are projected onto the unit sphere. The rigid 3D transformation matrix are estimated using the overlapping features. The Interior Point Optimization Algorithm is used to estimate the rotation between a set of projected points. Fig. 6 shows that the set of projected points are aligned together. The rotation matrix is parameterized in terms of Euler angles and cost function is developed that minimize the Euclidian distance between the reference (point projections of left camera image) and the three dimensional points from the right camera image.

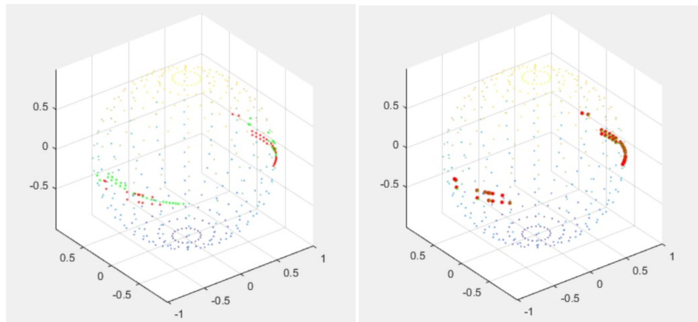

Figure 6. The selected points are aligned together.

The transformation matrix of two fisheye cameras with pure rotation with zero translation.,

$$T_{rftolf} = \begin{bmatrix} -1.0000 & -0.0048 & 0.0085 & 0 \\ -0.0045 & 0.9994 & -0.0335 & 0 \\ -0.0087 & -0.0334 & -0.9994 & 0 \\ 0 & 0 & 0 & 1.0000 \end{bmatrix}, \tag{20}$$

Fig. 7 shows that the two hemispheres are fused together using the estimated transformation matrix. Once the rotation

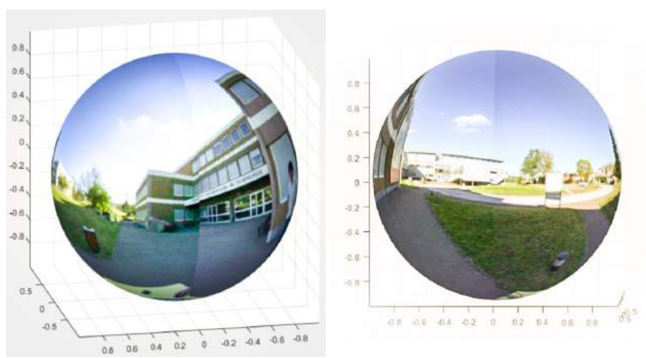

Figure 7. The unit sphere; images front and rear view from the camera rig.

matrix is estimated, fusion of the two hemispheres is only two step procedure. The points on the hemispheres that are beyond the zero plane are first eliminated. Then the transformation (rotation only) is applied on the hemisphere of the right fisheye camera and the point matrices are concatenated to get a full unit sphere.

4.2.2 Rigid transformation between a fisheyes and ZED camera

The same procedures are used to estimate the transformation matrix between the image from ZED camera and two hemispheres. The transformation matrix image ZED camera refers to left fisheye.

$$T_{ztolf} = \begin{bmatrix} -0.0143 & -0.0290 & -0.9995 & 0 \\ -0.0062 & 0.9996 & -0.0289 & 0 \\ -0.9999 & 0.0058 & -0.0145 & 0 \\ 0 & 0 & 0 & 1.0000 \end{bmatrix}, \tag{21}$$

As shown in fig. 8, the RGB and depth images from ZED camera are overlapped onto the unit sphere. It is also recovered the scale between the ZED and fisheye cameras. The high-resolution RGB image and the depth image are fused onto the unit sphere by the help of transformation matrix estimated. It should be noted that, the fusion of zed image onto the unit sphere is an approximation. The result obtained is reasonable after handling the strong distortion on the fisheye images.

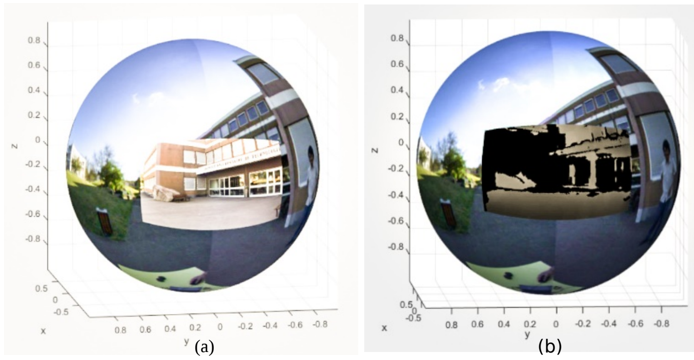

Figure 8. The unit sphere overlaps with image from ZED camera (a: RGB image, b: Depth image).

4.3 3D reconstruction using the camera rig.

The 3D scene can be reconstructed using the triangulation, rotation and translation. The goal of triangulation is to minimized the distance between the two lines toward point P in 3D world. This problem can be expressed as a least square problem.

$$\min_{a,b} \|aP_1 - bRP_2 - t\|, \tag{22}$$

$$\begin{bmatrix} a^* \\ b_* \end{bmatrix} = \left(A^T A\right)^{-1} A^T t, A = [P_1 - RP_2], \tag{23}$$

By referring to fig. 5, looking from the first pose, point P, the line passing through O_1 and P_1 can be written as aP and the line passing through O_2 and P_2 can be written as $bRP_2 + t$, where $a, b \in \mathbb{R}$ and;

- P is the observed point.
- O_1, P_1 are the camera center and the point on the unit sphere at first position.
- O_2, P_2 are the camera center and the point on the unit sphere at second position.
- R and t are the rotation and translation.

The 3D point P is reconstructed by finding the middle point of the minimal distance between the two lines. It can be computed by;

$$P = \frac{a^* P_1 + b^* RP_2 + t}{2}, \tag{24}$$

Fig. 9 and fig.10 show the results of feature matching and scene reconstruction algorithm developed following the spherical model of the camera.

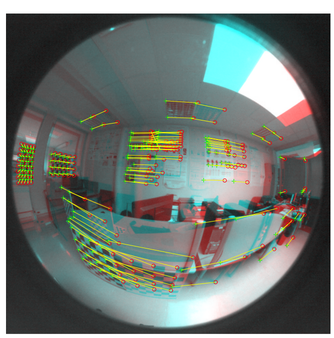

Figure 9. The features matching points between two different poses of fisheye cameras.

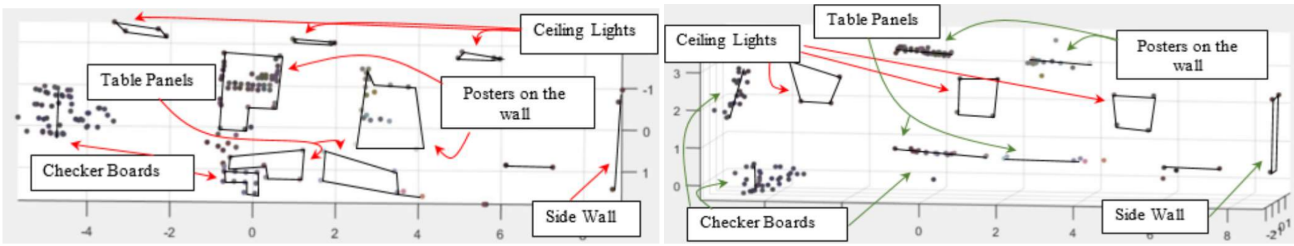

Figure 10. The front view(left) and top view(right) of the three dimensional reconstruction scenes.

5. CONCLUSIONS

The intrinsic parameter of fisheye and ZED cameras are estimated from the camera calibration toolbox. The distortion of fisheye camera is determined by parameter ξ. The rigid 3D transformation matrix are estimated using the overlapping features which are assuming that it is a pure rotation. The Interior Point Optimization Algorithm is used to estimate the rotation between a set of projected points. The same procedure is used to fuse ZED camera onto the unit sphere. The overlap area between fisheye and ZED cameras is estimated for the purpose of object tracking and detection. The camera rig is applied to reconstruct 3D scene using features matching and an algorithm based on a spherical model of camera.

REFERENCES

[1] Li, Shigang. "Full-view spherical image camera." 18th International Conference on Pattern Recognition (ICPR'06). Vol. 4. IEEE, 2006.

[2] Zhang, Chi, et al. "Development of an omni-directional 3D camera for robot navigation." 2012 IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM). IEEE, 2012.

[3] Marefat, F., A. Partovi, and A. Mousavinia. "A hemispherical omni-directional bio inspired optical sensor." 20th Iranian Conference on Electrical Engineering (ICEE2012). IEEE, 2012.

[4] Nayar, Shree K. "Catadioptric omnidirectional camera." Computer Vision and Pattern Recognition, 1997. Proceedings., 1997 IEEE Computer Society Conference on. IEEE, 1997.

[5] Gluckman, Joshua, and Shree K. Nayar. "Ego-motion and omnidirectional cameras." Computer Vision, 1998. Sixth International Conference on. IEEE, 1998.

[6] Geyer, Christopher, and Kostas Daniilidis. "A unifying theory for central panoramic systems and practical implications." European conference on computer vision. Springer Berlin Heidelberg, 2000.

[7] Hartley, Richard I., and Peter Sturm. "Triangulation." Computer vision and image understanding 68.2 (1997): 146-157.

[8] Kumar, Saurav, Daya Gupta, and Sakshi Yadav. "Sensor fusion of laser and stereo vision camera for depth estimation and obstacle avoidance." International Journal of Computer Applications 1.25 (2010): 20-25

[9] ZED Stereo Camera from https://www.stereolabs.com/zed/specs/.

[10] Barreto, João P. "A unifying geometric representation for central projection systems." Computer Vision and Image Understanding 103.3 (2006): 208-217.

[11] Mei, Christopher, and Patrick Rives. "Single view point omnidirectional camera calibration from planar grids." Proceedings 2007 IEEE International Conference on Robotics and Automation. IEEE, 2007.

[12] Othmani, A.A., Jiang, C., Lomenie, N. et al. Machine Vision and Applications (2016) 27: 751, Alice Ahlem, et al. "A novel Computer-Aided Tree Species Identification method based on Burst Wind Segmentation of 3D bark textures." Machine Vision and Applications (2015): 1-16.

[13] Lourakis, Manolis IA, and Rachid Deriche. Camera self-calibration using the singular value decomposition of the fundamental matrix: From point correspondences to 3D measurements. Diss. INRIA, 1999.

[14] Chang, Peng, and Martial Hebert. "Omni-directional structure from motion." Omnidirectional Vision, 2000. Proceedings. IEEE Workshop on. IEEE, 2000.

[15] Ahmad Zawawi Jamaluddin, Osama Mazhar, Olivier Morel, Ralph Seulin, David Fofi. Design and calibration of an omni-RGB+D camera. 13TH INTERNATIONAL CONFERENCE ON UBIQUITOUS ROBOTS AND AMBIENT INTELLIGENCE, Aug 2016, Xian, China. pp.386 - 387, 2016,