Design and Calibration of a Specialized Polydioptic Camera Rig
Osama Mazhar, Ahmad Zawawi Jamaluddin, Cansen Jiang, David Fofi, Ralph Seulin, Olivier Morel

To cite this version:
Osama Mazhar, Ahmad Zawawi Jamaluddin, Cansen Jiang, David Fofi, Ralph Seulin, et al.. Design and Calibration of a Specialized Polydioptic Camera Rig. Frontiers in information and communication technologies, 2017, 4, 10.3389/fict.2017.00019. hal-01620563

HAL Id: hal-01620563
https://u-bourgogne.hal.science/hal-01620563
Submitted on 8 Jun 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution 4.0 International License
Design and Calibration of a Specialized Polydioptric Camera Rig

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

Le2i FRE2005, CNRS, ENSAM, Université de Bourgogne Franche-Comté, Le Creusot, France

It has been observed in the nature that all creatures have evolved highly exclusive sensory organs depending on their habitat and the form of resources availability for their survival. In this project, a novel omnidirectional camera rig, inspired from natural vision sensors, is proposed. It is exclusively designed to operate for highly specified tasks in the field of mobile robotics. Navigation problems on uneven terrains and detection of the moving objects while the robot is itself in motion are the core problems that omnidirectional systems tackle. The proposed omnidirectional system is a compact and a rigid vision system with dioptric cameras that provide a 360° field of view in horizontal and vertical, with no blind spot in their site combined with a high-resolution stereo camera to monitor anterior field of view for a more accurate perception with depth information of the scene. Structure from motion algorithm is adapted and implemented to prove the design validity of the proposed camera rig, and a technique is developed to calibrate similar systems.

Keywords: omnidirectional, polydioptric, calibration, structure from motion, stereo-vision, depth, RGB-D

1. INTRODUCTION

The rapid development in computing systems, and their availability to the consumer market, soon made researchers realize that computational inability may not necessarily be the only handicap in all scientific problems. If observed in nature, all creatures have evolved very unique and highly specified anatomical and physiological traits that depend on the habitat they live and the availability of the resources their survival is dependent on. Moreover, in the field of computer vision, sensors with larger field of view are always appreciated. These are specialized systems that enable to acquire more information with less equipment/image-data used. Omnidirectional or panoramic cameras have become an affordable and popular photographic tool that allows to capture 360° panoramic images (Knill and Ramirez-Herran, 2007).

The applications of the omnidirectional cameras are, but not limited to, robot localization and mapping (Hart, 2014; Kim et al., 2014; Liu and Siegwart, 2014; Lukierski et al., 2015), robot navigation (Zhang et al., 2012; Watanabe et al., 2013; Delgado-Galvan et al., 2015; Hart et al., 2015), object tracking (Cogal et al., 2014; Depraz et al., 2015; Sablak, 2015), visual servoing (Caron et al., 2014; Liu et al., 2013; Marković et al., 2014; Pasteau et al., 2016), structure from motion (Chang and Hebert, 2000; Micusik and Pajdla, 2006; Kawanishi et al., 2008; Kim and Oh, 2008), and virtual reality/visual telepresence (Li et al., 2013; Kawauchi and Rekimoto, 2014; Kasahara and Rekimoto, 2015).

1.1. Proposed System

An omnivision camera rig has been developed (see Figure 1) using two fisheye cameras with field of view of approximately 185° each, which are fixed opposite to each other facing laterally, so as
to cover 360° in horizontal and vertical (Jamaluddin et al., 2016). A depth camera, namely “ZED Camera,” is also mounted in front of the rig that covers the anterior view providing high-resolution RGB + depth image (StereoLabs, 2016).

1.2. Concept and Motivation
It is observed that creatures in nature are divided into two categories: preys and predators (Paine, 1966). The first category, i.e., preys, usually have eyes on lateral sides of their heads covering a larger field of view as to detect any danger easily. They need not to have a higher resolution or larger field with depth information, as their main purpose is to detect and mostly running away to a safer place. However, predators have eyes mostly on their anterior portion of their heads, though with a limited field of view, but have a larger depth information in contrast to earlier. In this reference, a vision system is designed, with capabilities of that of both. This system may be used for several tasks as stated before, including the tasks related to navigating on an uneven terrain or detecting the movements of other objects while the robot with this camera rig is moving itself.

1.3. Paper Outline
A brief background of omnivision systems is discussed in Section 2, camera model and the epipole geometry adapted for omnidirectional cameras are explained in Section 3, calibration of the camera rig and scene reconstruction are presented in Section 4, and Section 5 concludes the article.

2. BACKGROUND

2.1. Omnidirectional Cameras
Omnidirectional cameras can be mainly designed as either dioptric systems [assisting vision by refracting and focusing (Neumann et al., 2003)] or catadioptric systems (dioptric imaging that incorporates reflecting surfaces (Gluckman and Nayar, 1998) or mirrors; see Figure 2). However, polydioptric cameras may also be considered as another category of the vision systems which will be discussed in a section to come. The camera systems can be classified into two categories determined by whether they do have a unique effective viewpoint or not (Geyer and Daniilidis, 2000).

For any vision system, it is highly anticipated that there exists a single effective viewpoint. It allows to extract pure perspective images from the image taken by such a camera system (see Figure 3). The images taken from these camera systems preserve the linear perspective geometry. The uniqueness of the viewpoint is equivalent to a purely rotating planar camera all of whose viewpoints coincides (Geyer and Daniilidis, 2000). Hence for omnidirectional cameras having a single viewpoint allows to extract not only perspective but also panoramic images without the knowledge of depth of the scene (Nayar, 1997). Also the single viewpoint allows to map the images produced by the vision system onto a sphere (Gluckman and Nayar, 1998).

Conical, spherical and most fisheye cameras do not have a unique effective viewpoint. Nayar and Baker have investigated the cases of catadioptric cameras by Nayar (1997), which led to a conclusion that only two practical solutions exist for the reflecting surfaces, the hyperbolic and elliptical systems. However by Nayar (1997), it is shown that the parabolic mirrors with orthographic projections (paracatadioptric system) can also be used to achieve a single viewpoint. Meanwhile elliptical mirrors in fact only reduce the field of view, so they are not used in...
practice. Hyperbolic and parabolic catadioptric cameras are able to capture at least a hemisphere of the scene from a unique viewpoint.

2.2. Polydioptic Cameras
In order to enhance the capability of the vision system, designs with multiple cameras were proposed by Neumann et al. (2003), thus a term evolved namely a polydioptic camera. The authors describe polydioptic camera as "a generalized camera that captures a multiperspecti­ve subset of the space of light rays." There are several examples for such modalities emerged especially with the development of portable and computational devices such as mobile phones and wireless communication systems.

2.3. Spherical Approximation of Multiple Cameras
Although polydioptic/multiple cameras are used either to have a stereo-vision capability or to enhance the vision system field of view, it is impossible for such a camera rig to have a single effective viewpoint because of physical real-life constraints, e.g., sensor size and bulky camera mounting. In this case, any motion estimation or structure from motion algorithm might simply fail. In the study by Kim et al. (2010), authors not only proposed a solution for this problem by spherical approximation of multiple cameras but also demonstrated that this approximation outperforms the generalized camera model when the features of interest are sufficiently farther when compared with the baseline of the cameras.

3. METHODOLOGY

3.1. Camera Model
The efforts to design a vision system with larger field of view arose several different modalities as discussed in Section 2.1, whose geometry cannot be described using the conventional pinhole model. A unified projection model for the entire class of central catadioptric systems (catadioptric vision system with single effective viewpoint) has been proposed. It explains that all central catadioptric systems can be modeled by a central projection of a point in the space onto the unit sphere followed by another central projection from a point lying between the sphere center and its
north pole onto the image plane (Geyer and Daniilidis, 2000). The fisheye cameras, as said earlier, do not have a single effective viewpoint (see Figure 4) instead have a locus of projection centers (dialectic). Nevertheless in Barreto (2006a,b), it is demonstrated that the small locus of dioptic systems can be approximated by a single viewpoint and thus these systems also underlie a common central projection model.

Mei and Rives (2007) proposed a model for all single viewpoint omnidirectional cameras and developed a toolbox for camera calibration using planar grids as for conventional cameras. This model is an extension of the model proposed by Barreto and Geyer (Geyer and Daniilidis, 2000; Barreto, 2006b) and differs only with respect to some minor differences in conventions regarding the direction of the camera, addition of tangential distortion parameters, and projecting the three-dimensional points in space onto a sphere instead of a paraboloid. This model anticipates the distortions in the fisheye image and describes a parameter \( E \) that determines the amount of radial distortion. This also considers the projection of the scene onto a unit sphere, thus allows to exploit the spherical approximation for a polydioptic camera rig, as in our case. We opt for this model to be followed in this project.

3.2. Omnidirectional Epipolar Geometry

The epipolar geometry for panoramic cameras is studied by Svo-boda et al. (1998) for hyperbolic mirror. In this case, a simpler mathematical formulation proposed by Chang and Hebert (2000), originally for catadioptric cameras, is adapted for dioptic systems. The epipolar constraints for fisheye cameras are shown in Figure 5. \( P \) is a point in the scene, \( P_1 \) and \( P_2 \) are its projections onto the unit spheres, and \((u_1, v_1)\) and \((u_2, v_2)\) are its corresponding image coordinates on the two fisheye images. The projections onto the unit sphere are computed as by Mei and Rives (2007). From the figure, it can be observed that \( P, P_1, P_2, O_1, \) and \( O_2 \) are coplanar, hence we get,

\[
\begin{align*}
O_2O_1 \times O_2P_1 \cdot O_2P_2 &= 0 \\
O_1^2 \times P_1^2 \cdot P_2 &= 0
\end{align*}
\] (1)

FIGURE 8 | Finding the overlapping image pixels (field of view) in fisheye images and in high-resolution image of stereo camera. (A) High-resolution image from the stereo camera. (B) Overlapping field of view is defined by the area inside the red boundaries (informed and written consent has been obtained from the subjects for the publication of this image).
where $O_1^2$ and $P_1^2$ are the coordinates of $O_1$ and $P_1$ in coordinate system $X_2$, $Y_2$, and $Z_2$. The rigid motion between $X_1$, $Y_1$, and $Z_1$ and $X_2$, $Y_2$, and $Z_2$ can be described by rotation $R$ and translation $t$. The transformation equations can be written as,

$$O_1^2 = R.O_1 + t$$
$$P_1^2 = R.P + t$$

(2)

Substituting (2) in (1) we get,

$$P_2^T EP_1 = 0$$

(3)

where $E = [t] \times R$ is the essential matrix which may be decomposed into the motion parameters. If a set of point correspondence $(m_i, m'_i)$ is given such that $i = 1, \ldots, n$, where $n \geq 7$ and $m_i = [u_i, v_i]$ are the image coordinates of fisheye camera, essential matrix can be computed by minimizing the epipolar errors. In order to estimate the essential matrix, the point correspondence pairs are stacked into one linear system, thus the overall epipolar constraint becomes,

$$Uf = 0$$

(4)

where

$$U = [u_1, u_2, \ldots, u_n]^T$$

and $f$ and $u_i$ are vectors constructed by stacking columns of matrices $P_i$ and $E$, respectively.

$$P_i = P_i^T P_i^T$$

$$E = \begin{bmatrix} f_1 & f_2 & f_3 \\ f_2 & f_3 & f_4 \\ f_3 & f_4 & f_5 \end{bmatrix}$$

(5)

The essential matrix can be estimated with linear least square by solving Equations (4) and (5), where $P_1^2$ is the projected point, which corresponds to $P_2$ of the Figure 5. $U$ is a $n \times 9$ matrix and $f$ is $9 \times 1$ vector containing the 9 elements of $E$. The initial estimate of essential matrix is then exploited for the robust estimation of essential matrix. A modified iteratively reweighted least square method for omnivision cameras, originally explained by Torr and Murray (1997), is then proposed. This assigns minimal weights to the outliers and noisy correspondences. The weight assignment is performed by finding the residual $r_i$ for each point.

$$r_i = f_i x'_i x_i + f_i x'_i y_i + f_i x'_i z_i + f_i x y'_i + f_i y y'_i + f_i y'_i z_i$$

$$+ f_i x z'_i + f_i y z'_i + f_i z z'_i$$

(6)

$$\text{err} \rightarrow \min_{f} \sum_{i=1}^{n} (w_{ij} f_i^T u_i)^2$$

(7)

$$w_{ij} = \frac{1}{\nabla r_i}$$

(8)

$$\nabla r_i = (r_{x_i}^2 + r_{y_i}^2 + r_{z_i}^2 + r_{x'_{y_i}}^2 + r_{x'_{z_i}}^2)^{1/2}$$

(9)

where $w_{ij}$ is the weight (known as Sampson's weighting) that will be assigned to each set of corresponding point and $\nabla r_i$ is the gradient; $r_{x_i}, r_{y_i}$ and so on are the partial derivatives found from Equation (6) as $r_{x_i} = f_i x'_i + f_i z'_i$. Once all the weights are computed, $U$ matrix is updated as follow,

$$U = WU$$

(10)

where $W$ is a diagonal matrix of the weights computed using Equation (8). 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. RESULTS

4.1. Intrinsic Camera Calibration
The estimation of the intrinsic parameters of the fisheye cameras is performed using the study by Mei and Rives (2007), and the images are projected onto the unit sphere using the equations provided by the aforementioned camera model. The value of $\xi$ (eccentricity) plays an important role to determine the extent of enfolding the image onto the unit sphere. It leads to the need to develop a set-up for robust estimation of $\xi$ up to the required accuracy.

4.1.1. Re-estimation of $\xi$
The baseline of the camera rig (distance between the centers of the left and right fisheye lens) is measured, and two parallel lines with the same distance to each other as well as a center line are drawn on a pattern. The rig is faced and aligned in front of the pattern such that the center line touches the edges of both circular fisheye camera images (see Figures 6 and 7). The parallel line corresponding to the edge of fisheye lens of each camera is then forced to the zero plane when the fisheye image is projected onto the unit sphere. This is done by developing a cost function to estimate $\xi$ that minimizes the z-component of pixels on the selected line using internal point optimization algorithm.

4.2. Estimation of Extrinsic Parameters of the Rig
4.2.1. Rigid Transformation Estimation
The estimation of rigid transformation of non-overlapping cameras has been discussed by Lébraly et al. (2010). However, since our camera system has an approximately 5° of overlap along the periphery of two hemispheres, we decided to use the overlapping features to estimate the transformation between two cameras. The matching points in two fisheye camera images are selected manually. The selected points are then projected onto the unit sphere with the calibration results obtained earlier. Interior point optimization algorithm from Kim et al. (2007) is used to iteratively estimate pure rotation between the set of projected points; thus a transformation with pure rotation is obtained.

$$T = \begin{bmatrix} -1.0000 & -0.0048 & 0.0085 & 0 \\ -0.0045 & 0.9994 & -0.0335 & 0 \\ -0.0087 & -0.0334 & -0.9994 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$ (11)
The same procedure is repeated to find pure rotation between the sphere from fisheye cameras and the stereo camera. The transformation result is shown in Equation (12).

$$T = \begin{bmatrix}
-0.0143 & -0.0290 & -0.9995 & 0 \\
-0.0062 & 0.9996 & -0.0289 & 0 \\
0.9999 & -0.0058 & -0.0145 & 0 \\
0 & 0 & 0 & 1
\end{bmatrix}$$  \hspace{1cm} (12)

4.2.2. Estimation of Overlap between Fisheye and Stereo Camera

To perform tasks such as object detection and object tracking, estimating the portion of omnidirectional unit sphere which is mutually covered by high-resolution RGB-D information from the ZED camera, is highly substantial. The boundary pixels of the ZED camera image are first projected onto the unit sphere and the k-Nearest-Neighbor (kNN) algorithm is used to find the overlapping pixels of the fisheye sphere with them. The overlapping pixels from the unit sphere projection of the fisheye images are then projected back onto the image plane using the camera model as in Mei and Rives (2007). Figure 8 shows the approximate overlapping field of view of fisheye cameras and ZED camera.

4.2.3. Fusion of Images (Spherical Approximation)

The transformation matrices (actually rotation matrices) obtained as in Equations (11) and (12) are used to fuse the two hemispheres from the fisheye camera and the projection of the high-resolution
RGB + D image from stereo camera onto the unit sphere (see Figure 9).

4.3. Scene Reconstruction Using Omnidirectional Cameras

The scene can be reconstructed and point $P$ can be triangulated using the parameters $P_1$, $P_2$, $R$, and $t$ as discussed by Ma et al. (2015). Observing Figure 5, the line passing through $O_1$ and $P_1$ can be defined as $aP_1$ and the line passing through $O_2$ and $P_2$ can be defined as $bRP_2 + t$ in the reference frame $O_1$; where $a, b \in \mathbb{R}$. The goal of triangulation is to minimize the distance between two lines, thus the problem can be expressed as a least square problem.

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

$$[\hat{a} \ \hat{b}] = (A^T A)^{-1} A^T t, A = [P_1, -RP_2]$$

The three-dimensional point $P$ is reconstructed by finding the middle point of the minimal distance between the lines $aP_1$ and $bRP_2 + t$ as follows:

$$P_k = \frac{\hat{a}P_1 + \hat{b}RP_2 + t}{2}$$

where $k = 1, 2, 3, 4$  

This triangulation method is applicable for all possible solutions. The conventional criteria to select the optimal solution which gives the positive depth of the reconstructed points is unsuitable in this case because of special geometrical properties of this method. A novel method is proposed to solve the problem of ambiguity of the results which is described as follows.

The scene is reconstructed using each of the four possible solutions and we get four sets of reconstructed points $P_i^*$; where $i = 1, 2, 3, 4$. As already mentioned, the first step for reconstruction of the scene is to capture at least two real images, select the matching points in two-dimensional images and “project the selected points onto the unit sphere” using the equations proposed in the spherical camera model. So for scene reconstruction, we already have at least two set of points projected onto the unit sphere. The projection of points onto the unit sphere at the first pose $P_s$ will be used in the later steps.

All four reconstructed sets are considered one at a time and each point in each set is divided by its norm. As a result we get more unit sphere projections, one for each reconstructed set. Ideally, the distance between the normalized true reconstructed points and the points projected onto the unit sphere of the first pose of camera $P_s$ should be zero, denoted as:

$$\min_{i} \left( \frac{P_i^*}{||P_i^*||} - P_s \right)$$
This distance is evaluated for each solution and the one that gives the least distance is considered the optimal solution. This technique is implemented on synthetic as well as real data, and the result is optimum in all cases.

4.4. Error Analysis
The quantification of error is done by creating synthetic data and comparing the reconstructed points with the synthetic input. Gaussian noise is introduced with multiples of standard deviation which is equal to the largest pixel to pixel distance over the spherical projection. Several patterns of input data were generated and two of them with there reconstructed points are shown in Figure 10.

The error plot is shown in Figure 11, where average Euclidean distance of 50 trials is plotted for each value of standard deviation (0.005) with 50 multiples of 0.05 for pyramid reconstruction. This error pattern is consistent with other synthetic three-dimensional patterns also.

5. CONCLUSION
A technique is developed for the extrinsic calibration of the proposed camera rig and similar vision systems with overlapping field of view. The spherical model of camera is exploited because several image processing algorithms have already been developed based on this model for catadioptric cameras, which can be implemented on the results obtained by our camera rig. The images from all the cameras are projected onto the unit sphere and fused to follow spherical approximation for polypodioptric cameras. The estimation of essential matrix and scene reconstruction by the proposed adaptation of epipolar geometry and triangulation method is performed on synthetic data with Gaussian noise and standard deviation equal to a multiple of the largest pixel to pixel distance over the spherical projection. The results are compared with the ground-truth that is computed from known rotation and translation, and the adaptation of epipolar geometry to the spherical model is found promising. The results are also computed for real images with automatic feature extraction as described by Harris and Stephens (1988), Lowe (2004), Bay et al. (2006), and Donoser and Bischof (2006) and with manual point selection also (see Figures 12 and 13).

AUTHOR CONTRIBUTIONS
DF, RS, and OM jointly conceived and supervised the proposed design. AZJ collected the data. OM performed necessary experiments, developed the code, and wrote the article. CJ helped in developing several parts of the code and in realization of Section 4.1.1. All authors discussed the results and observations and commented on the manuscript at all stages.

FUNDING
This work is part from a project entitled VIPeR (Polarimetric Vision Applied to Robotics Navigation) funded by the French National Research Agency ANR-15-CE22-0009-VIPeR.

REFERENCES


Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Mazhar, Jamaluddin, Jiang, Fofi, Seulin and Morel. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.