Journal of Zhejiang University SCIENCE A
ISSN 1673-565X (Print); ISSN 1862-1775 (Online)
www.zju.edu.cn/jzus; www.springerlink.com
E-mail: jzus@zju.edu.cn

Unwrapping and stereo rectification for omnidirectional images*

Jie LEI[1,2], Xin DU[†‡1,2], Yun-fang ZHU[3], Ji-lin LIU[1,2]

([1]*Institute of Information and Communication Engineering, Zhejiang University, Hangzhou 310027, China*)

([2]*Zhejiang Provincial Key Laboratory of Information Network Technology, Hangzhou 310027, China*)

([3]*College of Computer and Information Engineering, Zhejiang Gongshang University, Hangzhou 310027, China*)

[†]E-mail: duxin@zju.edu.cn

Received May 9, 2008; Revision accepted Sept. 2, 2008; Crosschecked May 28, 2009

Abstract: Omnidirectional imaging sensors have been used in more and more applications when a very large field of view is required. In this paper, we investigate the unwrapping, epipolar geometry and stereo rectification issues for omnidirectional vision when the particular mirror model and the camera parameters are unknown in priori. First, the omnidirectional camera is calibrated under the Taylor model, and the parameters related to this model are obtained. In order to make the classical computer vision algorithms of conventional perspective cameras applicable, the ring omnidirectional image is unwrapped into two kinds of panoramas: cylinder and cuboid. Then the epipolar geometry of arbitrary camera configuration is analyzed and the essential matrix is deduced with its properties being indicated for ring images. After that, a simple stereo rectification method based on the essential matrix and the conformal mapping is proposed. Simulations and real data experimental results illustrate that our methods are effective for the omnidirectional camera under the constraint of a single view point.

Key words: Single point of view, Calibration, Catadioptric image unwrapping, Omnidirectional stereo vision, Epipolar geometry, Essential matrix, Conformal mapping

doi:10.1631/jzus.A0820357 **Document code:** A **CLC number:** TP317.4; TP391

INTRODUCTION

In recent years, the omnidirectional camera has received considerable attention, due to its large viewing area, in the computer vision community with its applications in the fields such as video conference, image-based rendering, virtual reality, and robot navigation. A typical and commonly used omnidirectional camera is the catadioptric camera, which consists of a quadratic mirror [paraboloid, hyperboloid, ellipsoid or cone (Lin and Bajcsy, 2006)] and a CCD sensor (perspective or orthogonal) (Baker and Nayar, 1999). The catadioptric camera can obtain a 360° view of the environment at one shot, which is more convenient and has less computational load than that using the method of stitching images induced from a rotating perspective camera around a fixed axis or along a circle locus (Shum and Szeliski, 1995; Shum and He, 1999; Benosman and Kang, 2001; Agarwala *et al*., 2006).

Motivation and contributions

Previous research on omnidirectional vision has shown promising results in the last decade. However, most of them are based on an accurate camera model and the parameters are known in priori. Unfortunately, the parameters of most omnidirectional cameras are unavailable in practice. How to make these cameras useful in vision applications remains a challenging problem.

In this study, we attempt to develop a generalized unwrapping and rectification method for an omnidirectional camera when prior parameters are unknown for solving the counterpart problem of the perspective camera. Some important issues are investigated, such

‡ Corresponding author

* Project supported by the National Natural Science Foundation of China (Nos. 60502006, 60534070 and 90820306) and the Science and Technology Plan of Zhejiang Province, China (No. 2007C21007)

as omnidirectional camera calibration, ring omnidirectional image unwrapping, the epipolar geometry, and the rectification for the omnidirectional stereo pair. Based on the information above, we can reconstruct the 3D environment from two omnidirectional images. The contributions of our work are twofold:

(1) In the case of arbitrary camera configuration, a generalized essential matrix is deduced on a virtual quadratic surface. It can be computed using a new method similar to the classical eight-point algorithm (Hartley and Zisserman, 2000). Some important properties of the epipolar geometry are emphasized. The four epipoles located in a quadratic space, which is the null space of the essential matrix. The intersection of two lines that connect the two null spaces is the projection center. We also investigate the shape of epipolar conics for some kinds of configuration.

(2) After epipolar geometry is developed, a simple rectification method based on the essential matrix and the conformal mapping is proposed. It processes the columns and rows of image separately, and does not need to compute the constant differences between each pair of corresponding conics, which are enforced to lie on the same scan line from the essential matrix.

Related research

As a typical omnidirectional camera, the catadioptric imaging system consists of a hyperboloid [ellipsoid, cone (Lin and Bajcsy, 2006)] mirror with a perspective camera, or a parabolic mirror with an orthogonal camera (Baker and Nayar, 1999). The single view point constraint is verified by a careful selection and assembly of mirrors and imaging devices. In order to acquire the relationship between points in the 3D world and the catadioptric image plane, we shall first derive the model of image formation process and calculate the parameters. The related works can be classified into two different categories: individual model and unified model.

The individual model deals with the different types of catadioptric system separately. The types of quadric mirror surface must be given before the parameters could be calibrated. Yamazawa and Yagi (Yamazawa *et al*., 1993; Yagi *et al*., 1996) described an image sensor with a hyperbolic mirror for vision based navigation of a mobile robot. In (Svoboda and Pajdla, 2002), the epipolar geometry was studied for central catadioptric cameras separately. Each kind of epipolar geometry depended on the known parameters of the mirror. Kang (2000) proposed a technique dealing with the parabolic mirror with the aid of bounding circles in the image plane. In (Geyer and Daniilidis, 2002a), the authors also treated the case of parabolic mirror with two sets of parallel lines to calibrate the intrinsic parameters. However, the properties used in their study cannot be employed to other kinds of mirrors. The spherical mirror (Benosman and Kang, 2001) was also used. It could provide images with fewer blur and had a simple parametric model. However, the field of view was very sensitive to the camera-mirror alignment. All above mentioned techniques allow obtaining accurate calibration results, but primarily focus on particular sensor types and could not generalize to other kinds of sensors.

In recent years, novel calibration techniques have been developed, which can be applied to any kind of catadioptric system. Geyer and Daniilidis (2001) introduced a unified sphere model for all central catadioptric systems, where the conventional perspective imaging appeared as a particular case. It was proved that central catadioptric projection was isomorphic to a projective mapping from the unit sphere to a plane with a projection center on the perpendicular to the plane. Based on this model, some researchers (Ying and Hu, 2004; Ying and Zha, 2008; Barreto and Araujo, 2005) calibrated different catadioptric systems by utilizing certain geometrical invariants of projection when the line or sphere in 3D space was transformed to the image plane. Mei and Rives (2007) took misalignment and lens distortion into account in the calibration procedure. Besides the calibration, the epipolar geometry of the parabolic catadioptric system had also been derived. After the computation of the fundamental matrix and intrinsic parameters (Geyer and Daniilidis, 2002b), the motion of the camera and the structure of the 3D scene could be obtained (Geyer and Daniilidis, 2003b). Unfortunately, we noticed that the model of the parabolic system must be known in advance at this stage. Different from the sphere model, Micusik (2004) studied thoroughly the auto-calibration issues of some kinds of wide angle cameras including the parabolic catadioptric camera, hyperbolic catadioptric camera, and spherical catadioptric camera, and fish-eye lens. For each imaging model, a parametric mapping between

image points and 3D light rays could be obtained after the Taylor expansion at initial values of unknowns. The parametric model depended on different omnidirectional camera parameters. But there were some shortcomings: the first problem is that some parameters of the model have to be known, such as the field of view and the initial values of mirror parameters for the hyperbolic catadioptric camera; the second problem is that the 3D vectors have to be linearized to solve the quadratic eigenvalue problem for the spherical catadioptric camera.

Another simple and interesting unified model was proposed by Scaramuzza *et al*.(2006) (henceforth, called the Taylor model). This technique does not use any specific model of the omnidirectional system. It only assumes that the imaging process could be described by a Taylor series expansion, whose coefficients are the parameters to be estimated. These parameters are estimated by solving a four-step least-squares linear minimization problem, followed by a nonlinear refinement.

Among the above unified models, the sphere model is the most popular one and proves to be equivalent to the central catadioptric image formation accurately. But before the parameters of the unit sphere could be calibrated, the types of quadric mirror must be given in priori. The same problems exist with Micusik's model. This requirement makes them unsuitable for the systems with unknown prior parameters. In practice, the only model that could be used under the condition of unknown parameters is the Taylor model. In this study, we extend the research of the Taylor model to deal with the problem of ring image unwrapping, epipolar geometry and stereo rectification.

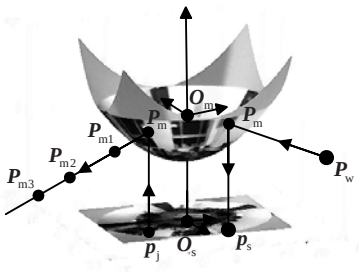
TAYLOR MODEL AND CALIBRATION

As mentioned above, all previous calibration procedures focus on particular sensor types and strongly depend on the model used. In practice, because the camera-mirror misalignments cannot be avoided, it is difficult to get a catadioptric system satisfying the single view point. Considering this problem, we utilize the unified model proposed by Scaramuzza *et al*.(2006), which is suitable for different kinds of omnidirectional vision systems. It models the projection from a pixel in the image plane to a 3D light ray by a Taylor series, whose coefficients constitute the calibration parameters.

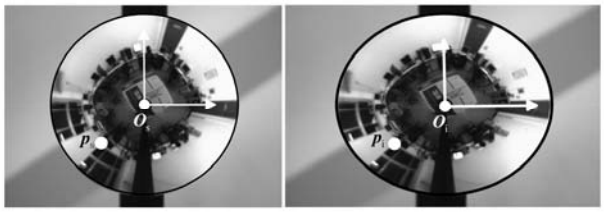
In Fig.1a, the bent surface with the reference frame $\boldsymbol{O}_\mathrm{m}$ is the virtual calibrated mirror which reflects the 3D light ray onto the sensor plane with the reference coordinates $\boldsymbol{O}_\mathrm{s}$, as plotted in Fig.1b. Suppose a 3D point $\boldsymbol{P}_\mathrm{w}=[X, Y, Z]$ has a projection point $\boldsymbol{P}_\mathrm{m}=[X_\mathrm{m}, Y_\mathrm{m}, Z_\mathrm{m}]$ on the mirror surface. The projection of $\boldsymbol{P}_\mathrm{w}$ in the sensor plane is $\boldsymbol{p}_\mathrm{s}=[u_\mathrm{s}, v_\mathrm{s}]$ and $\boldsymbol{p}_\mathrm{i}=[u_\mathrm{i}, v_\mathrm{i}]$ in image plane $\boldsymbol{O}_\mathrm{i}$, as described in Fig.1c. All coordinates are expressed in the mirror reference coordinates with the Z-axis aligned with the reference coordinates of the sensor plane. The complete model of an omnidirectional camera is

$$\begin{aligned}\lambda\cdot\boldsymbol{O}_\mathrm{m}\boldsymbol{P}_\mathrm{m}&=\lambda\cdot g(\boldsymbol{p}_\mathrm{s})=\lambda\cdot g(\boldsymbol{A}\boldsymbol{p}_\mathrm{i}+\boldsymbol{t})\\&=\boldsymbol{R}\boldsymbol{P}_\mathrm{w}+\boldsymbol{T}=\boldsymbol{M}[\boldsymbol{P}_\mathrm{w},\ 1]^\mathrm{T},\end{aligned}\tag{1}$$

where $\boldsymbol{M}=[\boldsymbol{R}, \boldsymbol{T}]$ is the perspective projection matrix; $\boldsymbol{A}$, $\boldsymbol{t}$ are the affine transformation components which encode the small camera-mirror misalignments and digitize the process error between image plane $\boldsymbol{O}_\mathrm{i}$ and sensor plane $\boldsymbol{O}_\mathrm{s}$. The function $g(\cdot)$ captures the relationship between a point $\boldsymbol{p}_\mathrm{s}$ in the sensor plane and the vector $\boldsymbol{O}_\mathrm{m}\boldsymbol{P}_\mathrm{m}$. In the Taylor model, $g(\cdot)$ has the form



(a)



(b) (c)

Fig.1 Taylor model and omnidirectional camera calibration
(a) Taylor-series-based image formation and coordinates system in the catadioptric case; (b) Reference coordinates of the sensor plane in metric coordinates; (c) Reference coordinates of image plane in pixel coordinates. The two reference coordinates are related by affine transformation $\boldsymbol{A}$, $\boldsymbol{t}$, whose effects can be illustrated from the circle in (b) and the ellipse in (c)

$$g(\boldsymbol{p}_s)=[u_s, v_s, f(\rho)]^T = [u_s, v_s, a_N\rho^N + a_{N-1}\rho^{N-1} + ... + a_1\rho + a_0]^T, \quad (2)$$

where $\rho=\sqrt{u_s^2+v_s^2}$ is the radial distance from image point $\boldsymbol{p}_s$ to projection center $\boldsymbol{O}_s$ in the sensor plane. The function of Taylor series $f(\cdot)$ is a generalized parametric polynomial of the imaging model and can compensate for small camera-mirror misalignments. So far the intrinsic parameters to be calibrated are a_N, a_{N-1}, …, a_1, a_0 and the extrinsic parameters $\boldsymbol{M}=[\boldsymbol{R}, \boldsymbol{T}]$, which are the transformation from the world reference coordinates to the mirror reference coordinates. After substituting Eq.(2) into Eq.(1), a four-step least-squares linear minimization problem is solved, followed by a nonlinear refinement based on the maximum likelihood criterion, to estimate these parameters. In most practical cases, the mirror is revolution symmetrical when its axis coincident with the principal axis of the camera. The Taylor series only contains components with even order. According to the minimization of the reprojection error and one-to-one mapping from 3D point to its 2D image projections, higher order terms are suppressed and the imaging model can be approximated by a quadratic polynomial. The error caused by the approximation is in a tolerable range as illustrated by our experimental results.

In the back projection, the image point $\boldsymbol{p}_j=[u_j, v_j]$ has the corresponding projection on the virtual mirror surface $\boldsymbol{P}_m=[\lambda u_j, \lambda v_j, f(\lambda u_j, \lambda v_j)]$, according to Eq.(2), as portrayed in Fig.1a. But only from this $\boldsymbol{P}_m$ one cannot determine the position of the corresponding 3D point in the world, shown as $\boldsymbol{P}_{m1}$, $\boldsymbol{P}_{m2}$, $\boldsymbol{P}_{m3}$, because from the knowledge of stereo vision at least two images of the same scene structure are needed. In other words, the functions $g(\cdot)$ and $f(\cdot)$ model the projection procedure from the 3D point $\boldsymbol{P}_w$ to the sensor point $\boldsymbol{p}_s$, but it can only map the sensor point $\boldsymbol{p}_j$ to a 3D light ray $\boldsymbol{O}_m\boldsymbol{P}_m$, not a particular 3D point, in the mirror reference coordinates in the back projection.

UNWRAPPING OF RING OMNIDIRECTIONAL IMAGES

Calibration is implemented by using ring catadioptric images, which are extremely distorted from the real world compared with conventional perspective images. Furthermore, many classical algorithms of planar perspective images cannot be applied to the ring image directly. It is difficult to develop the counterpart algorithms of conventional perspective cameras for the ring image. Motivated by this observation, we transform the ring image to two kinds of images: cylinder panorama and cuboid panorama.

Roughly speaking, the panorama is a stitched picture from a set of images (Shum and Szeliski, 1995; Shum and He, 1999; Benosman and Kang, 2001; Agarwala *et al*., 2006), which are produced by a perspective camera when it is rotated around a fixed vertical axis or following a circle locus with a certain angle interval. Compared with a single image captured by the conventional perspective camera, the panorama has a wide field of view with 360° in the same local resolution. However, re-sampling and the rectification in stitching can also cause new errors. These off-line operations will also become a bottleneck when the real-time need should be satisfied.

Here we utilize a ring catadioptric image and unwrap it to a cylinder panorama based on the calibrated model as Eq.(2) has shown directly, so that the complexity and the time consumption of rectification and stitching can be avoided. Fig.2 describes the process from a ring omnidirectional image to a cylinder panorama at the vertical cross section and at a horizontal section, respectively. Assuming the radius of the cylinder is R (it depends on the radial distance at the maximum elevation), the cylinder has a circumference of $2\pi R$ and its height is

$$H = R(\tan\theta_1 + \tan\theta_2), \quad (3)$$

where $\tan\theta_1$ and $\tan\theta_2$ are the maximum elevation and maximum depression, respectively. At the vertical cross section in Fig.2a, for a point $\boldsymbol{P}(h, \varphi)$ with a position of height h and rotating angle φ on the cylinder surface, the line through $\boldsymbol{O}_m$ and $\boldsymbol{P}(h, \varphi)$ has the form

$$f(\rho)=\frac{h}{R}\rho. \quad (4)$$

From Eqs.(2) and (4), we obtain a point $\boldsymbol{\rho}_0$ on the virtual mirror. Note that a line intersects a conic with two points. We choose the point that lies between the origin $\boldsymbol{O}_m$ and $\boldsymbol{P}(h, \varphi)$. For a rotating angle φ, $\boldsymbol{\rho}_0$ has its corresponding image point $[u_0, v_0]$ through Eq.(5):

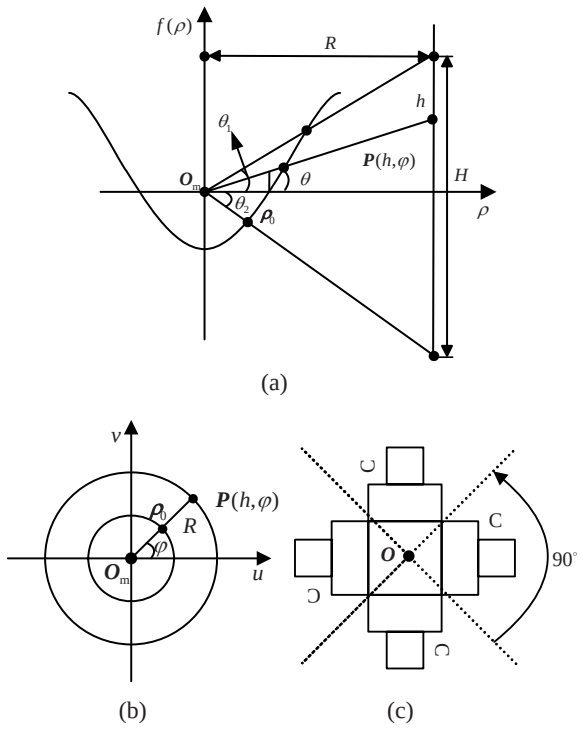


Fig.2 Cylinder panorama unwrapping

(a) In vertical cross section; (b) In horizontal section; (c) The equivalent network of the perspective cameras for the cuboid panorama

$$[u_0, v_0]^{\mathrm{T}} = \left(\boldsymbol{A} \begin{bmatrix} \rho_0 \cos\varphi \\ \rho_0 \sin\varphi \end{bmatrix} + \boldsymbol{t} \right) + \begin{bmatrix} c_x \\ c_y \end{bmatrix}, \qquad (5)$$

where $\boldsymbol{A}$ and $\boldsymbol{t}$ are the affine transformation parameters obtained from the calibration; c_x, c_y are the coordinates of the projection center. Finally, interpolation is required to obtain the intensity or color for the point $\boldsymbol{P}(h, \varphi)$ on the cylinder surface.

Due to the image projection onto a cylinder surface, this kind of panorama has a great deal of systematic distortions. Thus we consider projecting the omnidirectional image onto a plane to obtain an equivalent perspective image that is easy to interpret; in such a case, most classical vision algorithms, which are valid only for perspective cameras, can be applied. However, it is a data lossy transformation when the ring omnidirectional image is projected only onto a single plane. We use a cuboid to encompass the whole omnidirectional image. The cuboid panorama can be produced by an equivalent system that contains four perspective cameras (Fig.2c), each of them has a field of view of 90° and they have a common principal point $\boldsymbol{O}$. For clarity, Fig.3 gives the geometrical projection relationship of cuboid panorama, which is similar to the cylinder panorama except that the corresponding line of Eq.(4) is $f(\rho)=\rho h(\cos\omega)/R$, where ω is the angular position on each profile of the cuboid. The geometry derivation is omitted due to space limitation.

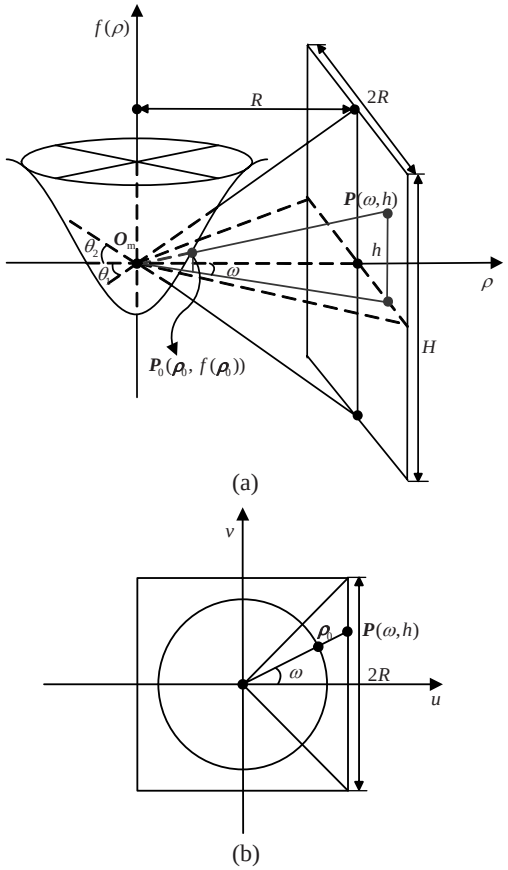


Fig.3 Cuboid panorama unwrapping

(a) In vertical cross section; (b) In horizontal section

EPIPOLAR GEOMETRY

It is well known that stereo matching is difficult and time consuming. For a stereo pair of perspective cameras, the epipolar geometry constraint reduces the search range of correspondences from 2D on a plane to 1D on a line, so it is indispensable to find an epipolar geometry rule for all of stereo systems. Svoboda and Pajdla (2002) derived a closed-form solution for an ideal central catadioptric system of a single view point and the conclusion resembles the rule of the perspective camera: for a catadioptric stereo system, a

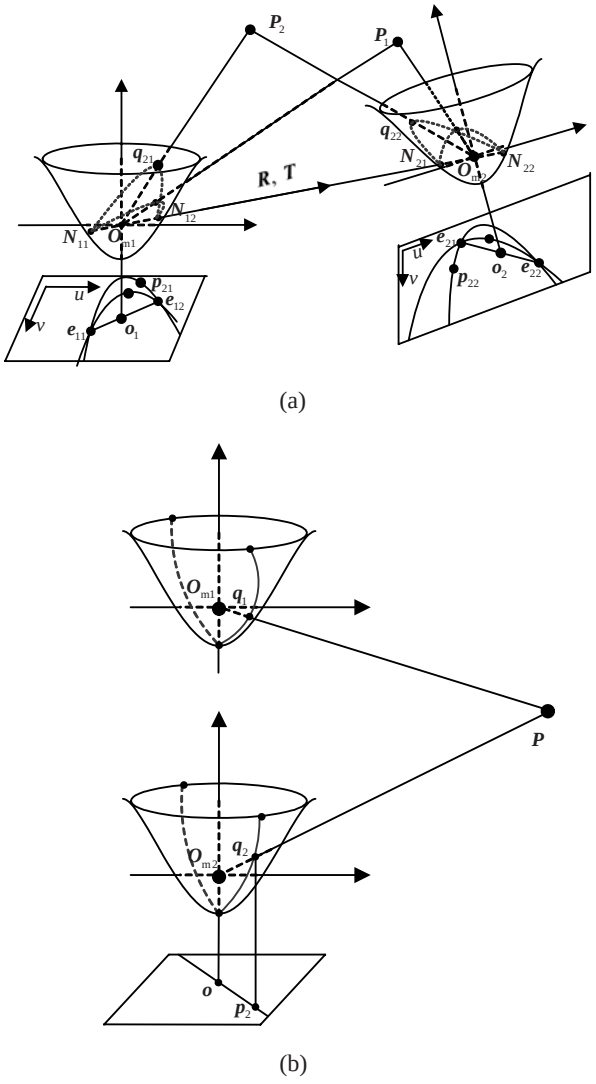
point in one view defines a conic in the other, and the corresponding point in the other view lies on the defined conic. Unfortunately, their solution shares the same problem of strongly depending on the prior parameters and is not suitable for the systems with unknown prior parameters. Based on the Taylor model introduced in Section 2, the epipolar geometry of arbitrary camera configuration is deduced through some geometrical derivations.

Fig.4a illustrates a generalized epipolar geometry relationship between two views of the catadioptric camera. The origins of the mirrors are $\boldsymbol{O}_{\mathrm{m1}}$ and $\boldsymbol{O}_{\mathrm{m2}}$, respectively. The corresponding reference coordinates of image planes are $\boldsymbol{o}_1$ and $\boldsymbol{o}_2$. Assume $\boldsymbol{p}_{21}=[u_{21}, v_{21}]$ and $\boldsymbol{p}_{22}=[u_{22}, v_{22}]$ are two image points of the same 3D point $\boldsymbol{P}_2=[X, Y, Z]$. $\boldsymbol{q}_{21}=[u_{21}, v_{21}, f(\boldsymbol{\rho}_{21})]^{\mathrm{T}}$ and $\boldsymbol{q}_{22}=[u_{22}, v_{22}, f(\boldsymbol{\rho}_{22})]^{\mathrm{T}}$ are the corresponding points on the mirror surfaces according to the calibrated model. Suppose the world reference coordinates coincide with the reference coordinates of the left mirror, and the relative pose is given by a rotation $\boldsymbol{R}$ and a translation $\boldsymbol{T}$. In the frame of $\boldsymbol{O}_{\mathrm{m1}}$, the normal $\boldsymbol{l}_{21}=\boldsymbol{T}\times\boldsymbol{q}_{21}$ of epipolar plane $\boldsymbol{O}_{\mathrm{m1}}\boldsymbol{P}\boldsymbol{O}_{\mathrm{m2}}$ is rotated to the frame of $\boldsymbol{O}_{\mathrm{m2}}$, $\boldsymbol{l}_{22}=\boldsymbol{R}\boldsymbol{l}_{21}=\boldsymbol{R}(\boldsymbol{T}\times\boldsymbol{q}_{21})$. The vector $\boldsymbol{q}_{22}\in\boldsymbol{O}_{\mathrm{m1}}\boldsymbol{P}_2\boldsymbol{O}_{\mathrm{m2}}$ is orthogonal to the normal; hence $\boldsymbol{q}_{22}^{\mathrm{T}}\cdot\boldsymbol{l}_{22}=\boldsymbol{q}_{22}^{\mathrm{T}}\cdot\boldsymbol{R}(\boldsymbol{T}\times\boldsymbol{q}_{21})=0$ becomes a bilinear equation after some rearrangements:

$$\boldsymbol{q}_{22}^{\mathrm{T}}\boldsymbol{E}\boldsymbol{q}_{21}=0, \tag{6}$$

where $\boldsymbol{E}=\boldsymbol{R}\hat{\boldsymbol{T}}$, $\hat{\boldsymbol{T}}=\begin{bmatrix}0 & -T_z & T_y\\ T_z & 0 & -T_x\\ -T_y & T_x & 0\end{bmatrix}$ is the skew-symmetric form of the translation $\boldsymbol{T}=[T_x, T_y, T_z]$. Eq.(6) has a similar constraint form to the essential matrix of the perspective cameras, but the vectors included are defined on a quadratic surface, which is different from the vectors in 3D projective space of the perspective camera. From Eq.(6), we can find that a point $\boldsymbol{q}_{21}$ on the surface of the left mirror corresponds to a conic on the right mirror surface, i.e., $\boldsymbol{C}=\boldsymbol{E}\boldsymbol{q}_{21}=[a, b, c]^{\mathrm{T}}$. The corresponding point, i.e., $\boldsymbol{q}_{22}=[u_{22}, v_{22}, f(\boldsymbol{\rho}_{22})]^{\mathrm{T}}$, lies on this conic. Hence, the conic has the following form:

$$\boldsymbol{q}_{22}^{\mathrm{T}}\boldsymbol{C}=au_{22}+bv_{22}+cf(\boldsymbol{\rho}_{22})=0. \tag{7}$$



(a)

(b)

Fig.4 Epipolar geometry of the catadioptric stereo system
(a) An arbitrary configuration of the catadioptric cameras, with its epipolar line as a conic; (b) A typical configuration for the catadioptric stereo system, with its epipolar line as a straight line through the projection center

After projecting onto the image plane, we can obtain the image form of the conic in Eq.(7). Similarly, the conic on the left mirror corresponding to $\boldsymbol{q}_{22}$ in the right mirror can be computed. A special stereo configuration of the catadioptric cameras is plotted in Fig.4b, which brings a simple epipolar geometry relationship. Due to the up-down configuration of the cameras, the epipolar line is actually a family of lines incident with the projection center. This configuration has also been used to construct a rectified stereo system by (Gluckman *et al.*, 1998; Lin and Bajcsy, 2003).

Using a similar derivation process as described above, sinusoid epipolar geometry can be obtained for cylinder panoramas, similar to what proposed originally by McMillan and Bishop (1995).

For the form of the epipolar geometry constraint shown in Eq.(6), the essential matrix in a quadratic space can be recovered through the method similar to the linear eight-point algorithm (Hartley and Zisserman, 2000). The differences are that our correspondences are defined in a quadratic space and nine pairs of points are required. Due to the skew-symmetry of the translation matrix $\hat{\boldsymbol{T}}$ and the rank-3 property of the rotation matrix $\boldsymbol{R}$, the essential matrix $\boldsymbol{E}$ is rank-2. Hence, all of the epipolar conics in one of the ring images intersect each other at two common points, called epipoles, shown as $\boldsymbol{e}_{ij}$ (i=1, 2; j=1, 2) in Fig.4a. The projections of these epipoles $\boldsymbol{e}_{ij}$ on the mirror, i.e., $\boldsymbol{N}_{1j}$ and $\boldsymbol{N}_{2j}$, lie in the null spaces of the essential matrix:

$$\boldsymbol{E}\boldsymbol{N}_{1j}=\boldsymbol{0},\quad \boldsymbol{E}^{\mathrm{T}}\boldsymbol{N}_{2j}=\boldsymbol{0}. \tag{8}$$

The center of projection lies on the direction of two mirrors' centers in the image plane and is the intersection of the two lines that connect the two null spaces. These lines coincide with the orientation of the mirrors' centers. Given the essential matrix of the two views for the catadioptric camera, the rotation $\boldsymbol{R}$ and the translation $\hat{\boldsymbol{T}}$ can be extracted (Hartley and Zisserman, 2000). As soon as we know the epipolar geometry of catadioptric cameras, the 3D scene can be reconstructed through the correspondences in two views. The correspondences are constrained on the epipolar conics. In order to make the correspondences search simpler and faster in the 2D image plane, the rectification for an omnidirectional stereo pair is required.

RECTIFICATION AND RECONSTRUCTION

For a stereo system of perspective cameras, rectification is employed when a dense match or reconstruction is required. Here we investigate the stereo rectification for the omnidirectional images including panoramas induced in Section 3 and the two arbitrary ring catadioptric images.

Rectification for panorama stereo

A pair of ring images is unwrapped to produce a pair of panoramas in Section 3. These two ring images are captured by two up-down omnidirectional cameras, similar to the typical configuration of omnidirectional stereo in Fig.4b. This kind of stereo configuration does not have the problem of occlusion, which often exists in other configurations. It also brings a simple kind of epipolar geometry and can generate naturally two rectified omnidirectional images if the mirrors and the cameras are placed properly (Gluckman *et al.*, 1998; Lin and Bajcsy, 2003; Spacek, 2005). However, since pre-defined ideal up-down configuration is often violated in practice, the rectification procedure is needed to make the epipolar line coincident with the columns. The epipolar constraint in columns of unwrapped panoramic images is similar to that of perspective cameras. Here we adopt the rectification algorithm proposed in (Ma *et al.*, 2003). It computes two projective transformations $\boldsymbol{H}_1$, $\boldsymbol{H}_2\in\mathbb{R}^{3\times3}$ which satisfy

$$\boldsymbol{H}_1\boldsymbol{e}_1\sim[1,0,0]^{\mathrm{T}},\quad \boldsymbol{H}_2\boldsymbol{e}_2\sim[1,0,0]^{\mathrm{T}}, \tag{9}$$

where $\boldsymbol{e}_1$ and $\boldsymbol{e}_2$ are the two epipoles corresponding to the fundamental matrix in the panoramas, and $\mathbb{R}^{3\times3}$ is the space of real 3×3 matrices. After the transformations of $\boldsymbol{H}_1$ and $\boldsymbol{H}_2$, the corresponding epipolar lines lie in the same columns.

Rectification for ring omnidirectional stereo

Now we consider the situation of cameras in the arbitrary configurations. In this case, the two epipoles are located in the image plane. The projection of the intersection of the epipolar plane and the mirror is an epipolar conic. With the case of the Taylor model, the epipolar conics happen to be a family of circles according to the minimization of the reprojection error and one-to-one mapping from 3D points to its image projections. However, there is not much work concerning the rectification for omnidirectional stereo. Based on the simple epipolar geometry of a vertical camera configuration, Gluckman and Nayar (2000) used folded mirrors combined with a camera to build a rectified stereo system, which can produce a rectified stereo pair directly. Lin and Bajcsy (2003) also obtained two epipolar-line-aligned stereo images with the help of a beam splitter. Geyer and Daniilidis

(2003a) introduced an interesting rectification method for the parabolic catadioptric cameras. In their method, the image plane was transformed onto a complex plane, and then the constant differences between pairs of epipolar conics were estimated. In practice, their method is very complex and time consuming because the constants need to be calculated every time. If the parameters in their sphere model are not known, the rectification could not be implemented. Inspired by Geyer and Daniilidis (2003a), we propose a new rectification method based on the conformal mapping and the essential matrix, which extends the applications range of ring omnidirectional stereo pair rectification.

1. Bipolar coordinates

Two-center bipolar coordinates (Lockwood, 1967) are one of the bipolar coordinates based on Apollonian circles that contain two families of circles with each circle in the first family intersecting any circle in the second family at a right angle. These two families of circles are the two bases of bipolar coordinates, τ and σ. There are two fixed points $\boldsymbol{e}_1$ and $\boldsymbol{e}_2$ in τ-σ coordinates with the corresponding positions of $(-r, 0)$ and $(r, 0)$ in Cartesian coordinates. The mapping between these two systems is

$$x = r_0 \frac{\sinh\tau}{\cosh\tau - \cos\sigma}, \tag{10}$$

$$y = r_0 \frac{\sin\sigma}{\cosh\tau - \cos\sigma}, \tag{11}$$

where $\sigma\in(0, 2\pi)$, $\tau\in(-\infty, +\infty)$. When τ and σ are constants, the two families of the orthogonal circles have the forms in Eqs.(12) and (13):

$$x^2 + (y - r_0\cot\sigma) = \frac{r_0^2}{\sin^2\sigma}, \tag{12}$$

$$y^2 + (x - r_0\coth\tau) = \frac{r_0^2}{\sinh^2\tau}. \tag{13}$$

These circles are plotted in Fig.5a. The dashed circles intersect each other at two common points $\boldsymbol{e}_1$, $\boldsymbol{e}_2$, and the centers of the solid circles lie on a line connecting these two fixed points. Actually, the former is the locus of the points satisfying $\boldsymbol{P}=\{\boldsymbol{P}_i|\angle\boldsymbol{e}_1\boldsymbol{P}_i\boldsymbol{e}_2=\sigma_0\}$, and the latter is the set with the constraint of $\boldsymbol{P}=\{\boldsymbol{P}_i|\ln(|\boldsymbol{P}_i\boldsymbol{e}_1|/|\boldsymbol{P}_i\boldsymbol{e}_2|)=\tau_0\}$.

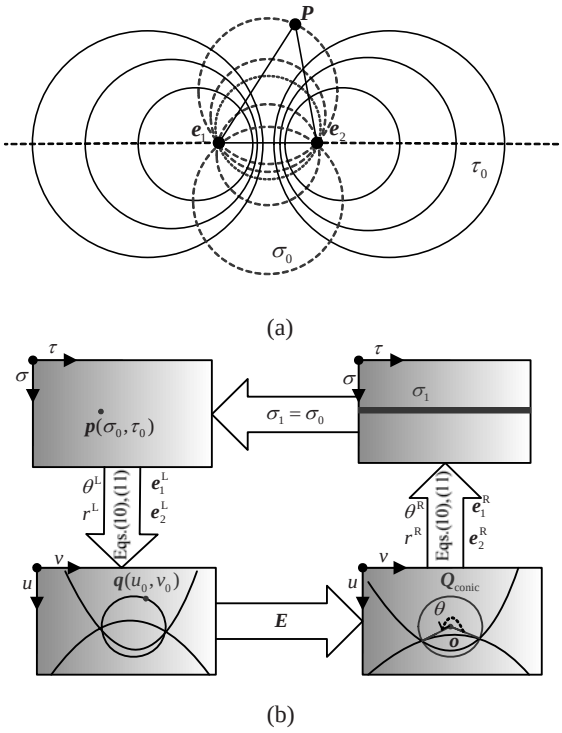


Fig.5 Illustration of the stereo rectification for the ring catadioptric images

(a) The bipolar coordinates with two families of the orthogonal circles plotted; (b) The rectification flow based on the essential matrix and the conformal mapping, with $\boldsymbol{p}$ and σ_1 being the corresponding elements

2. Stereo rectification based on the essential matrix and the conformal mapping

The epipolar circles under the Taylor model have a similar form to the dashed circles in Fig.5a. We can map the image plane u-v to the bipolar plane τ-σ, and the mapping is conformal. The bases of the two coordinate frames are orthogonal to each other, which satisfy the Cauchy-Riemann equations:

$$\frac{\partial x}{\partial\sigma} = \frac{\partial y}{\partial\tau} = -\frac{\sin\sigma\sinh\tau}{(\cosh\tau - \cos\sigma)^2}, \tag{14}$$

$$\frac{\partial x}{\partial\tau} = -\frac{\partial y}{\partial\sigma} = \frac{1 - \cos\sigma\cosh\tau}{(\cosh\tau - \cos\sigma)^2}. \tag{15}$$

For a point $\boldsymbol{p}(\sigma_0^{\mathrm{L}}, \tau_0^{\mathrm{L}})$ of the left rectified image τ-σ in Fig.5b, the position in standard Cartesian coordinates can be obtained using Eqs.(10) and (11). The position of two epipoles in this coordinates are $(-r^{\mathrm{L}}, 0)$ and $(r^{\mathrm{L}}, 0)$, where $r^{\mathrm{L}} = \| \boldsymbol{e}_1^{\mathrm{L}} - \boldsymbol{e}_2^{\mathrm{L}} \| /2$. This Cartesian coordinates system is different from the ring

image coordinates by a rotation $\hat{\boldsymbol{R}}^{\mathrm{L}}$ and a translation $\hat{\boldsymbol{t}}^{\mathrm{L}}$, which can be computed from two sets of epipoles, i.e., $\boldsymbol{e}_j^{\mathrm{L}}, \boldsymbol{e}_j^{\mathrm{R}}$ $(j=1, 2)$, in the image plane:

$$\hat{\boldsymbol{R}}^{\mathrm{L}}=\begin{bmatrix}\cos\theta^{\mathrm{L}} & -\sin\theta^{\mathrm{L}}\\ \sin\theta^{\mathrm{L}} & \cos\theta^{\mathrm{L}}\end{bmatrix}, \tag{16}$$

$$\hat{\boldsymbol{t}}^{\mathrm{L}}=\frac{\boldsymbol{e}_1^{\mathrm{L}}+\boldsymbol{e}_2^{\mathrm{L}}}{2}, \tag{17}$$

where θ^{L} is the angle from the x-axis of the image plane to vector $\boldsymbol{e}_1^{\mathrm{L}}-\boldsymbol{e}_2^{\mathrm{L}}$, and the superscript 'L' represents the left image. The transformation can be obtained in the same manner for the right image. After that, we obtain the corresponding image point $\boldsymbol{q}(u_0^{\mathrm{L}}, v_0^{\mathrm{L}})$. The corresponding conic (which is a circle under the Taylor model) in the right image is $\boldsymbol{Q}_{\mathrm{conic}}=\boldsymbol{Eq}$. After the central angle (θ) for two epipoles of this circle is computed, we can determine the σ_0^{R} in the following way:

$$\sigma_0^{\mathrm{R}}=\begin{cases}\theta/2, & \sigma_0^{\mathrm{L}}\in(0,\pi/2]\cup(\pi,3\pi/2],\\ \theta/2-\pi, & \sigma_0^{\mathrm{L}}\in(\pi/2,\pi],\\ \pi-\theta/2, & \sigma_0^{\mathrm{L}}\in(3\pi/2,2\pi].\end{cases} \tag{18}$$

Then for each point $(\sigma_0^{\mathrm{R}}, \tau_0^{\mathrm{R}})$ along a row of σ_0^{R} in the right rectified image, we compute the image correspondence using Eqs.(10) and (11) and the transformation $\hat{\boldsymbol{R}}^{\mathrm{R}}$, $\hat{\boldsymbol{t}}^{\mathrm{R}}$. Because the same rows of the two rectified images are processed, the correspondences along a same pair of epipolar circles lie on the same rows of the rectified images only from epipolar geometry.

After the mapping, the transformed points of the corresponding epipolar circles lie on the rows of σ_0^{L} and σ_0^{R}, respectively. The rectification flow illustrated in Fig.5b and the pseudo-codes presented in Algorithm 1 encode the rectification algorithm based on the essential matrix and the conformal mapping. The algorithm processes the rows and columns of the image plane separately, and it also avoids the constant difference calculation between σ_0^{L} and σ_0^{R}, which are enforced to be equivalency through the essential matrix.

Algorithm 1 Stereo rectification for ring omnidirectional images

Input: $\boldsymbol{I}^{\mathrm{L}}$ and $\boldsymbol{I}^{\mathrm{R}}$: a pair of omnidirectional images,
$\boldsymbol{E}$: essential matrix.
Output: $\boldsymbol{I}^{\mathrm{Lrect}}$ and $\boldsymbol{I}^{\mathrm{Rrect}}$: a pair of rectified images.

Step 1: Find two sets of epipoles, $\boldsymbol{e}_j^{\mathrm{L}}, \boldsymbol{e}_j^{\mathrm{R}}$ $(j=1, 2)$ according to Eq.(8).

Step 2: Compute r^{L} and r^{R}, $\hat{\boldsymbol{R}}^{\mathrm{L}}$ and $\hat{\boldsymbol{R}}^{\mathrm{R}}$, $\hat{\boldsymbol{t}}^{\mathrm{L}}$ and $\hat{\boldsymbol{t}}^{\mathrm{R}}$ according to Eqs.(16) and (17).

Step 3: Select any image as the reference, e.g., the left image, and $\tau^{\mathrm{L}}\in[1, \tau_{\max}^{\mathrm{L}}]$, $\sigma^{\mathrm{L}}\in(0, 2\pi]$.

Step 4:

```
for σ^L∈(0, 2π]
  for τ^L ∈[1, τ^L_max]
    Compute the position of standard coordinates, (x^L, y^L),
      using Eqs.(10) and (11);
    Transform (x^L, y^L) into ring image coodinates:
      [u^L, v^L]^T = R̂^L [x^L, y^L]^T + t̂^L + [c_x^L, c_y^L]^T;
    Calculate the intensity for (σ^L, τ^L) using bilinear
      interpolation;
    for a certain τ_n^L
      Compute the corresponding epipolar conic using
        the Taylor model and essential matrix;
      Find the central angle θ = (∠e_1^R O e_2^R)/2, and ob-
        tain σ^R according to Eq.(18);
      for τ^R ∈[1, τ^R_max]
        Compute the position of standard coordinates,
          (x^R, y^R), using Eqs.(10) and (11);
        Transform (x^R, y^R) into ring image coodinates:
          [u^R, v^R]^T = R̂^R [x^R, y^R]^T + t̂^R + [c_x^R, c_y^R]^T;
        Calculate the intensity for (σ^R, τ^R) using bilinear
          interpolation;
      end
    end
  end
end
```

The size of the rectified images is determined by the position of two epipoles. This rectification is an information lossy transformation when two epipoles are in the ring image simultaneously due to its singularity [see Eqs.(10) and (11)]. The algorithm can be applied to arbitrary pairs of omnidirectional images except those produced by the perspective camera combined with a cone mirror, because the cone is a degenerate conic and the Taylor expansion of this catadioptric system does not contain the term of order two.

Triangulation

Given a rectified pair of omnidirectional images, numerous matching techniques for perspective images reviewed in (Scharstein and Szeliski, 2002) can be used to compute the correspondence. The search is now restricted to one dimension, in the same horizontal scan line in the case of ring catadioptric stereo and the same columns in panorama stereo. Because stereo matching is outside the scope of this study, here we do not discuss the detailed algorithms.

After matching, a set of image correspondences $\{\boldsymbol{x}_i^{\mathrm{L}}, \boldsymbol{x}_i^{\mathrm{R}} = \boldsymbol{x}_i^{\mathrm{L}} - \boldsymbol{d}_i\}_{i=1, 2, ..., N}$ of the same scene $\boldsymbol{X}_i$ and the disparity $\boldsymbol{d}_i$ can be obtained. Assume the world coordinates is coincident with the left mirror coordinates, and two projective matrices are $\boldsymbol{P}^{\mathrm{L}}=[\boldsymbol{I}, \boldsymbol{0}]$ and $\boldsymbol{P}^{\mathrm{L}}=[\boldsymbol{R}, \boldsymbol{t}]$, calculated from the epipolar geometry in Section 4. The 3D light rays $\boldsymbol{l}_i^{\mathrm{L}} = [\boldsymbol{x}_i^{\mathrm{L}}, f(\| \boldsymbol{x}_i^{\mathrm{L}} \|)]^{\mathrm{T}}$ and $\boldsymbol{l}_i^{\mathrm{R}} = [\boldsymbol{x}_i^{\mathrm{R}}, f(\| \boldsymbol{x}_i^{\mathrm{R}} \|)]^{\mathrm{T}}$ on the mirror can be estimated based on the calibrated Taylor model. In each view of the mirror we obtain the following projection equations:

$$\lambda_i^{\mathrm{L}} \boldsymbol{l}_i^{\mathrm{L}} = \boldsymbol{P}^{\mathrm{L}} \boldsymbol{X}_i, \quad \lambda_i^{\mathrm{R}} \boldsymbol{l}_i^{\mathrm{R}} = \boldsymbol{P}^{\mathrm{R}} \boldsymbol{X}_i, \tag{19}$$

where λ_i^{L} and λ_i^{R} are two scalars. In Eq.(19), we have six equations with five unknowns, which can be solved by the linear least-squares method. After that, the range information of $\boldsymbol{X}_i$ is estimated. Here we should address that the estimated $\boldsymbol{X}_i$ is determined only up to scale which can be upgraded to a metric scale given the distance between the two cameras.

EXPERIMENTAL RESULTES

A set of experiments has been conducted to test the described method using both simulations and real data. Experimental results and numerical statistics are presented in this section.

Simulations

A catadioptric system consisting of a hyperbolic mirror and a perspective camera was simulated. The expression of the hyperbolic mirror and the intrinsic parameters of the perspective camera is shown in Eq.(20) with meter and pixel as units respectively.

$$\begin{cases} \dfrac{(z+\sqrt{0.04^2+0.02^2})^2}{0.04^2} - \dfrac{x^2+y^2}{0.02^2} = 1, \\ \boldsymbol{K} = \begin{bmatrix} 1000 & 0 & 320 \\ 0 & 1000 & 240 \\ 0 & 0 & 1 \end{bmatrix}. \end{cases} \tag{20}$$

After the calibration, we obtained a Taylor expansion $f(\rho)=0.0045\rho^2-55.728$, the center of projection, (320.0021, 240.0011), and the affine transformation $\boldsymbol{A} = \begin{bmatrix} 0.999\,96 & 0.4260\times10^{-6} \\ 0.2138\times10^{-6} & 1 \end{bmatrix}$. These parameters encoded the mapping from a point in the ring image to 3D light rays in the mirror. Then a set of random points was generated in the 3D space, and the corresponding points in the image plane can be produced by this simulated catadioptric system for the camera's two views (Figs.6a and 6b). The ring catadioptric image in Fig.6b was unwrapped to the cuboid and cylinder panorama in Fig.6c and Fig.6d. The mirror was also plotted, and the reference coordinate of the mirror was coincident with the panoramas for simplicity. At least nine points were needed to compute the essential matrix in the quadratic space. Two pairs of epipoles lay in the left and right null spaces of this essential matrix. In Fig.6e, some sampled epipolar conics were plotted with the two epipoles marked with crosses. The intersection of two lines is the projection center. Given the essential matrix, the rotation and translation can be extracted by the method proposed in (Hartley and Zisserman, 2000). The coordinates of the mirror in the left view was defined as the world coordinates. After linear triangulation, we obtained the 3D points corresponding to the points in the images. The results were plotted in Fig.6f with two poses of the camera and the line connecting the focuses of two views.

In order to test the robustness of our method, Gaussian noise $N\sim(0, \sigma^2)$ was added to the simulated ring image points coordinates. Under each level of noise, we measured the Euclidian distance (*Qtern.Dist* in Eq.(21a)) between the quaternions $\boldsymbol{q}_{\mathrm{est}}$ and $\boldsymbol{q}_{\mathrm{grth}}$, which represent the estimated rotation and the true rotation respectively, the ratio (*Trans.Ratio* in Eq.(21b)) between the estimated translation $\boldsymbol{T}_{\mathrm{est}}$ and the ground truth $\boldsymbol{T}_{\mathrm{grth}}$, the average distance in pixels (*Av.EpPts.Dist* in Eq.(21c)) from ring image points $\boldsymbol{p}_i$ to their epipolar conic $\boldsymbol{C}_i$, the Sampson distance

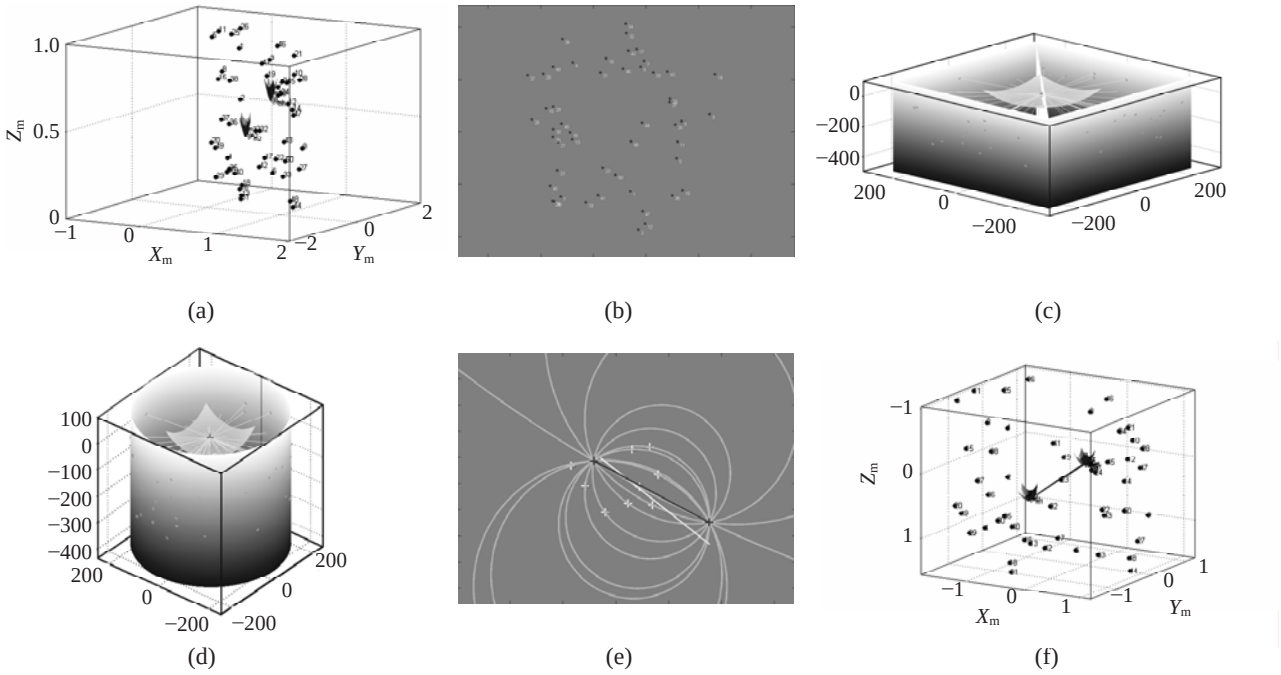


Fig.6 Simulation results of the proposed algorithm

(a) The simulated scene with two poses of the omnidirectional camera; (b) The projection points on the image plane; (c) The cuboid panorama; (d) The cylinder panorama; (e) Some epipolar conics with their corresponding points lying on it (two lines connect two null spaces of the essential matrix and intersect at the projection center); (f) Reconstruction of the simulated scene with the estimated poses of omnidirectional cameras

(Yamazawa *et al*., 1993) in pixels (*Av.Samp.Dist* in Eq.(21d)), the average reprojection error in pixels (*Av.Rpj.Err* in Eq.(21e)), and the average difference of rows (in pixels) between the two corresponding scan lines after rectification, i.e., *Av.Rect.Diff* in Eq.(21f), in which $\boldsymbol{L}_i^{\text{row}}$ and $\boldsymbol{R}_i^{\text{row}}$ represent the row coordinates of the rectified image points in the left and the right images, respectively. The above measurements under the noise levels with standard derivation from 0 to 1 pixel were plotted in Fig.7. *Av.EpPts.Dist* and *Av.Samp.Dist* had a similar trend, and they were both less than one pixel, but the latter was smaller. The reprojection error was less than 0.4 pixels. The distance between the estimated quaternion and its ground truth increased with the noise. The same situation existed with the *Trans.Ratio*. With the ground truth ratio of 2.6493, *Trans.Ratio* fluctuated in the range from 2.64 to 2.66. Fig.7f illustrated the average row difference between corresponding epipolar conics after the stereo rectification. *Av.Rect.Diff* increased mildly at low levels of noise compared with noise levels of larger than 0.8 pixels. But the difference was limited in 0.8 rows.

$$Qtern.Dist = \frac{1}{N}\sum_{i=1}^{N}\left\|\boldsymbol{q}_{\text{est}}^{i} - \boldsymbol{q}_{\text{grth}}^{i}\right\|, \tag{21a}$$

$$Trans.Ratio = \frac{1}{3}\sum_{i=x,y,z}\left(\left\|\boldsymbol{T}_{\text{est}}^{i}\right\| / \left\|\boldsymbol{T}_{\text{grth}}^{i}\right\|\right), \tag{21b}$$

$$Av.EpPts.Dist = \frac{1}{N}\sum_{i=1}^{N} d\left(\left\|\boldsymbol{p}_i - \boldsymbol{C}_i\right\|\right), \tag{21c}$$

$$Av.Samp.Dist = \frac{1}{N}\sum_{i=1}^{N}\frac{(\boldsymbol{p}_i^{\mathrm{T}}\boldsymbol{E}\boldsymbol{q}_i)^2}{\left\|\dfrac{\partial}{\partial(\boldsymbol{p}_i,\boldsymbol{q}_i)}\boldsymbol{p}_i^{\mathrm{T}}\boldsymbol{E}\boldsymbol{q}_i\right\|^2}, \tag{21d}$$

$$Av.Rpj.Err = \frac{1}{N}\sum_{i=1}^{N}\left\|\boldsymbol{p}_i - p_{\text{rpj}}(\boldsymbol{R},\boldsymbol{T}, f(\rho), \boldsymbol{X}_i)\right\|, \tag{21e}$$

$$Av.Rect.Diff = \frac{1}{N}\sum_{i=1}^{N}\left\|\boldsymbol{L}_i^{\text{row}} - \boldsymbol{R}_i^{\text{row}}\right\|. \tag{21f}$$

Real data

The camera we used in the real data experiments is the SONY RPU-C251 (http://www.visioncom.co.il/m4_Cameras_RPU-C2512.asp). Its output is a standard NTSC analog video signal and the maximum resolution is 640×480. The captured image was shown in Fig.8a. The camera is very convenient to use,

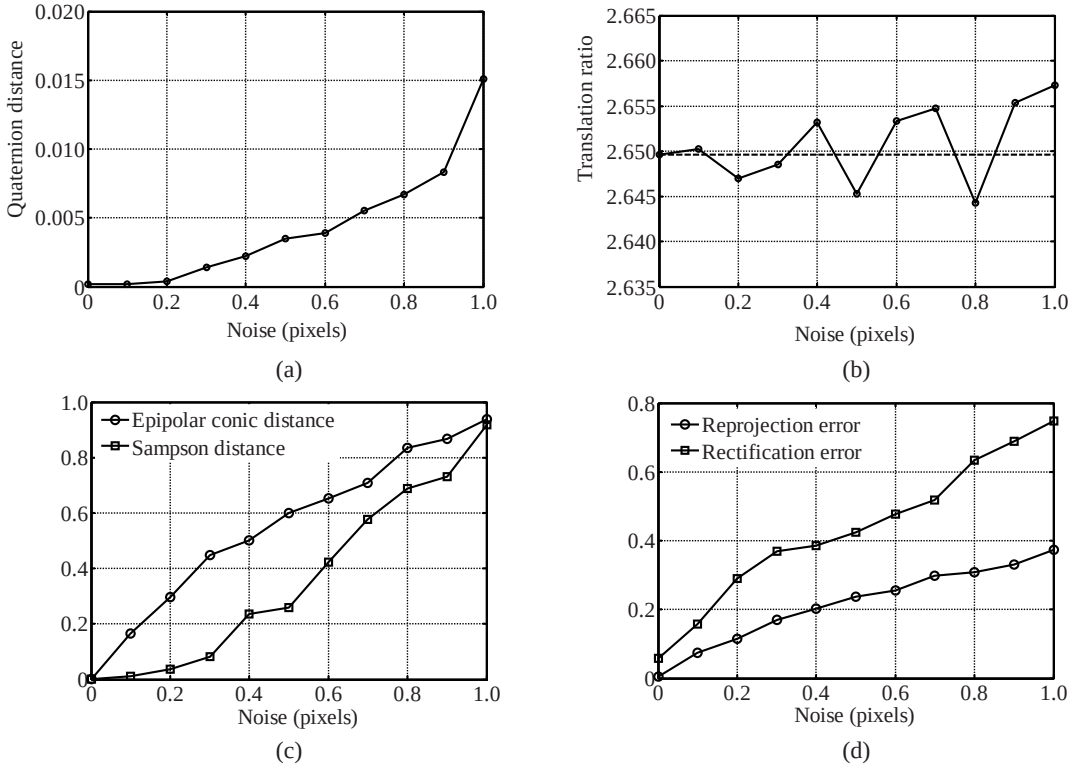
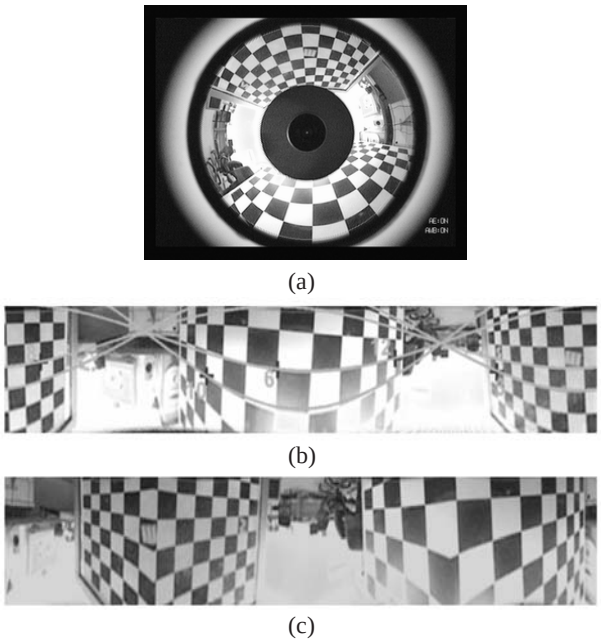


Fig.7 The algorithm performance evaluation under different levels of noise

(a) The distance between the quaternions, Eq.(21a); (b) The ratio between the translations, Eq.(21b); (c) The distance in pixels from image points to the conic (Eq.(21c)) and the Sampson distance in pixels (Eq.(21d)); (d) The reprojection error in pixels (Eq.(21e)) and the rectification difference in pixels between two corresponding epipolar conics (Eq.(21f))



(a)

(b)

(c)

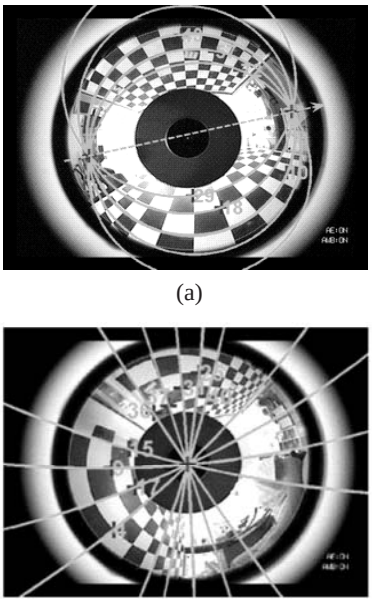
Fig.8 Omnidirectional image unwrapping

(a) Ring omnidirectional image; (b) The cylinder panorama with its sinusoid epipolar curves overlaid; (c) The cuboid panorama

but none of the parameters were available. After calibration, we obtained the Taylor expansion $f(\rho)=0.0032\rho^2-105.3535$, the center of projection (320.3386, 240.0196), and the affine transformation $\boldsymbol{A}=\begin{bmatrix}0.9931 & 3.0928\times10^{-4}\\ 1.4047\times10^{-4} & 1\end{bmatrix}$. These parameters encoded the relationship from the ring image point to the 3D light ray in the reference coordinates of the mirror. The unwrapped panoramas are illustrated in Fig.8b and Fig.8c, from which we can see that a line in the world coordinates corresponded to a sinusoid in the cylinder surface, but it became a straight line in the cuboid, as shown in Fig.8c.

In the computation of the essential matrix, at least nine corresponding points were needed, but we selected more and solved a linear least-squares problem to improve accuracy. The rotation and translation were also estimated simultaneously. In Fig.9, the epipolar conics for two styles of configurations were plotted, with the crosses denoting the points corresponding to the epipolar conics. All the conics

intersected at two fixed points, named epipoles, and the line connecting two epipoles indicated the motion direction of two cameras. Measurement results of Eqs.(21c) and (21d) were 0.41, 0.28 pixels for Fig.9a, and 0.35, 0.21 pixels for Fig.9b.



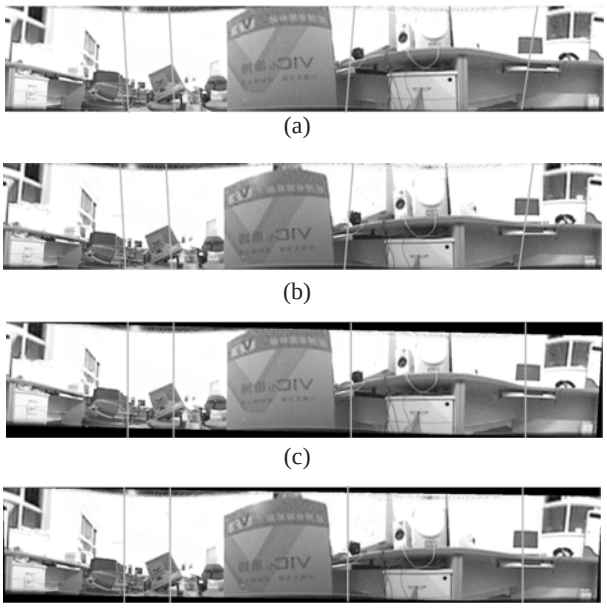
(a)

(b)

Fig.9 The epipolar geometry

(a) Arbitrary cameras configuration with a nonlinear epipolar geometry; (b) Up-down configuration with a simple epipolar geometry that contains a family of lines incident with the projection center

For the epipolar geometry in the panorama case, points were projected onto the panorama surface using the Taylor model, and a method similar to the normalized linear eight-point algorithm was applied to the computation of the essential matrix. The epipolar curves were plotted in Fig.8b. We can see that the epipolar curve for the cylinder panorama was a sinusoid [which was already proofed in (McMillan and Bishop, 1995)]. On the whole, the two sets of epipoles for the panorama in Fig.8b and for the ring images in Fig.9a were in the same position. The errors were mainly caused by panorama unwrapping and manual selection of points. The epipolar conics in Fig.9b were simpler than those in Fig.9a due to the special camera configuration of the up-down style. These epipolar conics were straight lines incident with the projection center, and their projections were a family of vertical lines in the panorama image as shown in Fig.10. Our ring images were produced by the up-down cameras configuration. It was not constrained in the vertical direction exactly due to the misalignment of the cameras. The epipolar line was not a vertical line in practice, as Figs.10a and 10c show. Based on this consideration, a rectification was employed before the matching to speed up the search for correspondences. A pair of rectified images after using the algorithm in (Ma *et al.*, 2003) was shown in Figs.10c and 10d, in which we can see that the epipolar lines were coincident with the columns.



(a)

(b)

(c)

(d)

Fig.10 The rectification results for panoramas

An unrectified panorama stereo pair is shown in (a) and (b). For the up-down configuration, the simple epipolar geometry is a family of vertical lines. These lines are not vertical exactly for the misalignment of the cameras. (c) and (d) present the rectified panoramas using a method similar to the perspective image

The rectification in Fig.10 was the case of camera up-down configuration, which brings linear epipolar geometry. But it was not constrained to the special trace for the structure from motion as in Fig.9b. Hence, the nonlinear epipolar geometry made the correspondences lie on a conic and the two epipoles existed in the ring image simultaneously, as shown in Fig.9a. In this case, the rectification method for the panorama stereo used in Fig.10 cannot be applied to the ring omnidirectional stereo pair. To solve this problem, the algorithm based on the essential matrix

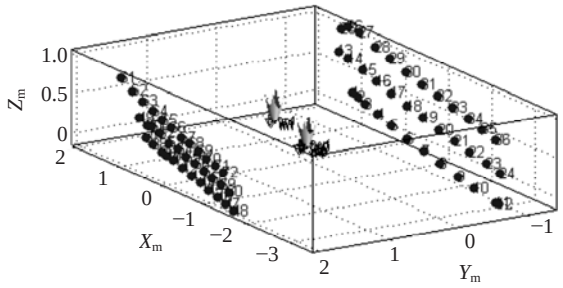
and the conformal mapping presented in Section 5 was used to rectify the ring stereo pair in Figs.9a and 9b. The results were shown in Figs.11a and 11b. For Fig.9a, in which the neighborhood of epipoles was unwrapped into a part with great distortion due to the singularity of Eqs.(10) and (11) at the epipoles. For Fig.9b, the produced rectified images had very large sizes because one of the two epipoles did not lie in the ring image. So we cropped the useful part to display (Fig.11b). In order to test the correctness, we manually selected some matched features in stereo pairs, and connected them with a straight line. As we can see in Figs.11a and 11b, these lines were coincident with the scan lines and were correctly aligned. The error measured by Eq.(21f) was 0.3754 pixels for Fig.11a and 0.2966 pixels for Fig.11b. After rectification, we can reconstruct the scene more efficiently. In our experiments, we reconstructed two chessboards in the scene by manually assigning the disparity. The results were given in Figs.11c and 11d. In order to measure the planeness of the reconstructed plane, we first fitted a plane using all the reconstructed 3D points, and then computed the distance from these points to the plane by

$$d_{\text{non-plane}} = \frac{1}{N}\sum_{i=1}^{N}\frac{\boldsymbol{K}\cdot\boldsymbol{X}}{\|\boldsymbol{K}\|}, \tag{22}$$

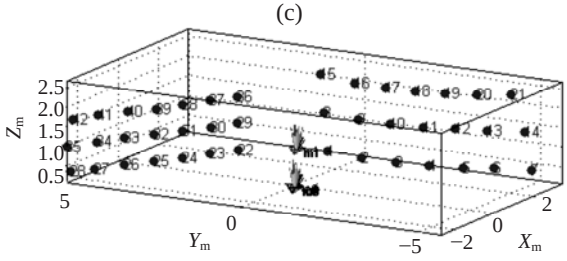
where $\boldsymbol{K}=[k_1, k_2, k_3, k_4]$ is the estimated norm of the plane and $\boldsymbol{X}=[X, Y, Z, 1]^{\mathrm{T}}$ is the homogeneous coordinates. $d_{\text{non-plane}}$ gives this error measurement. The resulting errors for Figs.11a and 11b were 5.21 and 4.85 pixels, respectively.

(a)

(b)



(c)



(d)

Fig.11 (a) Rectification result of Fig.9a; (b) Rectification result of Fig.9b. Two images connected by lines are a pair of rectified images in which the corresponding points are denoted. As we can see, the corresponding points lie in the same row/column in the rectified image. (c) and (d) are the reconstruction results of the two chessboards for Fig.9a and Fig.9b, respectively

CONCLUSION

We developed a generalized unwrapping and rectification method for omnidirectional cameras when the camera parameters are not available. After calibration, we obtained the mirror model based on the Taylor series. These parameters encoded the relationship from the ring image point to a 3D light ray. Then, the ring image was unwrapped to two kinds of panoramas: the cylinder panorama and the cuboid panorama. And the epipolar geometry was deduced for arbitrary camera configuration. We also investigated the stereo rectification problem. For the case of up-down configuration, the simple epipolar geometry eases the rectification process and is similar to that in the perspective cameras. For the arbitrary configuration, we proposed a new method based on the essential matrix and the conformal mapping. Experiments with simulated and real data were conducted. The results showed that our methods are effective.

Future work will focus on the dense matching and reconstruction based on the rectified stereo pairs and the error correction to improve the robustness and the accuracy.

References

Agarwala, A., Agrawala, M., Cohen, M., Salesin, D., Szeliski, R., 2006. Photographing Long Scenes with Multi-viewpoint Panoramas. Proc. SIGGRAPH, p.853-861.

Baker, S., Nayar, S.K., 1999. A theory of single-viewpoint catadioptic image formation. *Int. J. Comput. Vis.*, **35**(2):175-196. [doi:10.1023/A:1008128724364]

Barreto, J.P., Araujo, H., 2005. Geometric properties of central catadioptric line images and their application in calibration. *IEEE Trans. Pattern Anal. Mach. Intell.*, **27**(8):1327-1333. [doi:10.1109/TPAMI.2005.163]

Benosman, R., Kang, S.B., 2001. Panoramic Vision: Sensors, Theory and Applications. Monographs in Computer Science. Springer-Verlag, New York.

Geyer, C., Daniilidis, K., 2001. Catadioptric projective geometry. *Int. J. Comput. Vis.*, **45**(3):223-243. [doi:10.1023/A:1013610201135]

Geyer, C., Daniilidis, K., 2002a. Paracatadioptric camera calibration. *IEEE Trans. Pattern Anal. Mach. Intell.*, **24**(5):687-695. [doi:10.1109/34.1000241]

Geyer, C., Daniilidis, K., 2002b. Properties of the Catadioptric Fundamental Matrix. Proc. European Conf. on Computer Vision, p.140-154.

Geyer, C., Daniilidis, K., 2003a. Conformal Rectification of Omnidirectional Stereo Pairs. Computer Vision and Pattern Recognition Workshop, **7**:73-78. [doi:10.1109/CVPRW.2003.10082]

Geyer, C., Daniilidis, K., 2003b. Mirror in Motion: Epipolar Geometry and Motion Estimation. Proc. Ninth IEEE Int. Conf. on Computer Vision, **2**:766-773. [doi:10.1109/ICCV.2003.1238426]

Gluckman, J.M., Nayar, S.K., 2000. Rectified Catadioptric Stereo Sensors. IEEE Conf. on Computer Vision and Pattern Recognition, **2**:224-236. .

Gluckman, J.M., Thoresz, K., Nayar, S.K., 1998. Real Time Panorama Stereo. DARPA Image Understanding Workshop, p.299-303.

Hartley, R., Zisserman, A., 2000. Multiple View Geometry in Computer Vision. Cambridge University Press, Cambridge, UK.

Kang, S.B., 2000. Catadioptric Self-calibration. Proc. IEEE Conf. on Computer Vision and Pattern Recognition, **1**:201-207.

Lin, S.S., Bajcsy, R., 2003. High Resolution Catadioptric Omni-directional Stereo Sensor for Robot Vision. IEEE Int. Conf. on Robotics and Automation, p.1694-1699.

Lin, S.S., Bajcsy, R., 2006. Single-view-point omnidirectional catadioptric cone mirror imager. *IEEE Trans. Pattern Anal. Mach. Intell.*, **28**(5):840-845. [doi:10.1109/TPAMI.2006.106]

Lockwood, E.H., 1967. A Book of Curves. Cambridge University Press, Cambridge, England, p.186-190.

Ma, Y., Soatto, S., Kosecka, J., Sastry, S.S., 2003. An Invitation to 3-D Vision: From Images to Geometric Models. Springer-Verlag, New York, USA.

McMillan, L., Bishop, G., 1995. Plenoptic Modeling: An Image-based Rendering System. Proc. SIGGRAPH, p.39-46.

Mei, C., Rives, P., 2007. Single View Point Omnidirectional Camera Calibration from Planar Grids. Proc. IEEE Int. Conf. on Robotics and Automation, p.3945-3950. [doi:10.1109/ROBOT.2007.364084]

Micusik, B., 2004. Two-view Geometry of Omnidirectional Cameras. PhD Thesis, Czech Technical University, Prague, Czech Republic.

Scaramuzza, D., Martinelli, A., Siegwart, R., 2006. A Toolbox for Easy Calibrating Omnidirectional Cameras. IEEE/RSJ Int. Conf. on Intelligent Robots and Systems, p.5695-5701. [doi:10.1109/IROS.2006.282372]

Scharstein, D., Szeliski, R., 2002. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *Int. J. Comput. Vis.*, **47**:7-42. [doi:10.1023/A:1014573219977]

Shum, H.Y., He, L.W., 1999. Rendering with Concentric Mosaics. Proc. SIGGRAPH, p.299-306.

Shum, H.Y., Szeliski, R., 1995. Stereo Reconstruction from Multiperspective Panoramas. Proc. Int. Conf. on Computer Vision, p.14-21.

Spacek, L., 2005. A catadioptric sensor with multiple viewpoints. *Rob. Auton. Syst.*, **51**(1):3-15. [doi:10.1016/j.robot.2004.08.009]

Svoboda, T., Pajdla, T., 2002. Epipolar gometry for central catadioptric cameras. *Int. J. Comput. Vis.*, **49**(1):23-37. [doi:10.1023/A:1019869530073]

Yagi, Y., Nishii, W., Yamazawa, K., Yachida, M., 1996. Rolling Motion Estimation for Mobile Robot by Using Omnidirectional Image Sensor HyperOmniVision. Proc. 13th Int. Conf. on Pattern Recognition, **1**:946-950. [doi:10.1109/ICPR.1996.546163]

Yamazawa, K., Yagi, Y., Yachida, M., 1993. Omnidirectional Imaging with Hyperboloidal Projection. Proc. IEEE/RSJ Int. Conf. on Intelligent Robots and Systems, **2**:1029-1034.

Ying, X.H., Hu, Z.Y., 2004. Catadioptric camera calibration using geometric invariants. *IEEE Trans. Pattern Anal. Mach. Intell.*, **26**(10):1260-1271. [doi:10.1109/TPAMI.2004.79]

Ying, X.H., Zha, H.B., 2008. Identical projective geometric properties of central catadioptric line images and sphere images with applications to calibration. *Int. J. Comput. Vis.*, **78**(1):89-105. [doi:10.1007/s11263-007-0082-8]