Spherical Imaging in Omnidirectional Camera Networks

Ivana Tošić and Pascal Frossard

Ecole Polytechnique Fédérale de Lausanne (EPFL)
Signal Processing Laboratory (LTS4)
Lausanne - 1015, Switzerland

Abstract

We propose in this chapter to consider the emerging framework of networks of omnidirectional cameras. We first describe how omnidirectional images captured by different types of mirrors or lenses can be uniquely mapped to spherical images. Spherical imaging can then be used for calibration, scene analysis or distributed processing in omni-directional camera networks. The chapter then presents calibration methods that are specific to omnidirectional cameras. We observe the multi-view geometry framework with an omnivision perspective by re-formulating the epipolar geometry constraint for spherical projective imaging. In particular, we describe depth and disparity estimation in networks of omnidirectional cameras. Finally, we discuss the application of sparse approximations methods to spherical images, and we show how geometric representations can be used for distributed coding or scene understanding applications.

Key words: stereographic projection, omnivision, spherical imaging, epipolar geometry, depth estimation
PACS:

1 Introduction

Representation of three-dimensional visual content is nowadays mostly limited to planar projections, which gives the observer only a windowed view of the world. Since the creation of the first paintings, our minds have been strongly bound to this idea. However, planar projections have important limitations for building accurate models of 3D environments, since light has naturally a radial form. The fundamental object underlying dynamic vision and providing a firm mathematical foundation is called the Plenoptic Function (PF) [1]. The plenoptic function simply measures the light intensity at all positions in a scene and for all directions. In static scenes, the function becomes independent on the time, and it is convenient to define the plenoptic function as a function on the product manifold $\mathbb{R}^3 \times S^2$, where S^2 is the 2D sphere, and we drop the chromaticity components for the sake of simplicity. We can consider the plenoptic function as the model for a perfect vision sensor, and hence processing the visual information on the sphere becomes very interesting.

Most state of the art techniques [2] first map the plenoptic information to a set of Euclidean views, in replacing the spherical coordinates by Euclidean ones: $(u, v) \simeq (\theta, \varphi)$. These images are then typically processed by regular image processing algorithms. Such a way of dealing with the plenoptic function readily limits the scope and functionalities of the image processing applications. Indeed the Euclidean approximation is only valid for very small scales, or locally. But it is one of the most important features of the plenoptic function to encompass information *globally*, at all scales and all directions. Thus either one has to deal with a large number of regular images approximating the plenoptic function, or one faces the problem that large discrepancies occur between the Euclidean approximation and the effective plenoptic function.

Therefore, processing the visual information on the sphere is certainly attractive in order to avoid the limitations due to planar projections. The radial organization of photo-receptors in the human fovea also suggests that we should reconsider the way we acquire and sample the visual information, and depart from the classical planar imaging with rectangular sampling. This chapter discusses the potential and benefits of using omnidirectional vision sensors, which acquire a 360 degrees field of view, for image and 3D scene representation. Numerous efficient tools and methods have been developed recently in omnidirectional imaging, for different purposes including image compression, reconstruction, surveillance, computer vision, robotics. In this chapter, the focus is put on omnidirectional cameras with a single point of projection, whose output can be uniquely mapped on a surface of a unit sphere. In this context, we describe image processing methods that can be applied to spherical imaging and we provide a geometrical framework for networks of omnidirectional cameras. In particular, we discuss the calibration issue, and the disparity estimation problems. We finally show how the scene geometry can be efficiently captured by sparse signal approximations with geometrical dictionaries, and we present an application in distributed coding in camera networks.

2 Omnidirectional imaging

2.1 Cameras

Omnidirectional vision sensors are devices that can capture a 360 degrees view of the surrounding scene. According to their construction, these devices can be classified into three types [3]: systems that use multiple images (i.e., image mosaics), devices that use special lenses, and catadioptric devices that employ a combination of convex mirrors and lenses. The images acquired by traditional perspective cameras can be used to construct an omnidirectional image, either by rotating a single camera, or by a construction of a multi-camera system. Obtained images are then aligned and stitched together to form a 360 degree view. Rotating camera system is however limited to capturing static scenes due to long acquisition time of all images. They may also suffer from mechanical problems which lead to maintenance issues. On the other side, multiple camera systems can be used for real-time applications, but they suffer from difficulties in alignment and calibration of cameras.

In this context, true omnidirectional sensors are interesting since each image provide a wide field of view on the scene of interest. Special lenses such as fish-eye lenses probably represent the most popular classes of systems that are capable of capturing omnidirectional images. However, such cameras present the important disadvantage that it does not have a single center of projection. This problem makes omnidirectional image analysis extremely complicated. Alternatively, omnidirectional images can also be generated by catadioptric devices. These systems use a combination of a convex mirror placed above a perspective camera, where the optical axis of the lens is aligned with the mirror's axis. A class of catadioptric cameras with quadric mirrors are of particular interest, since they represent omnidirectional cameras with a single center of projection. Moreover, the images obtained with such cameras can be uniquely mapped on a surface of the sphere [4]. This chapter will focus on the multi-view omnidirectional geometry and image processing for the class of catadioptric systems with quadric mirrors.

2.2 Projective Geometry for Catadioptric systems

Catadioptric cameras achieve almost hemispherical field of view with a perspective camera and a catadioptric system [5], which represents a combination of a reflective (catoptric) and refractive (dioptric) elements [4, 6]. Such a system is schematically shown on the Fig. 1(a) for the case of a parabolic mirror. Fig. 1(b) illustrates one omnidirectional image captured by the described parabolic catadioptric camera.

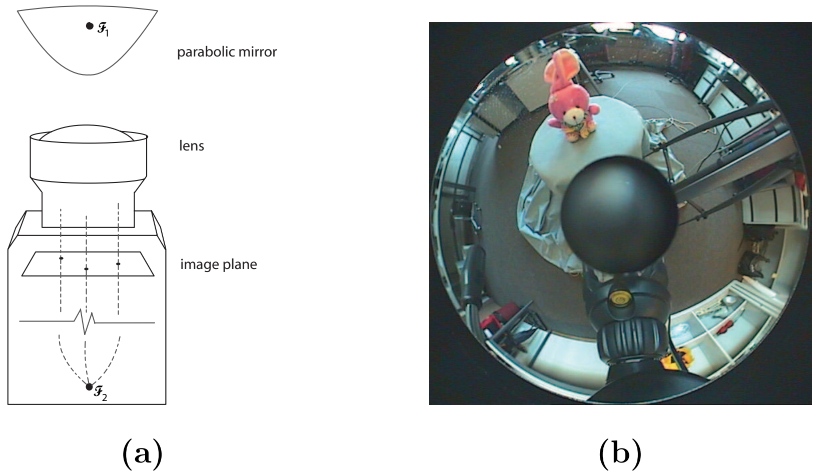


Fig. 1. (a) Omnidirectional system with parabolic mirror: the parabolic mirror is placed at parabolic focus $\mathcal{F}_1$, while the other focus $\mathcal{F}_2$ is at infinity [4]. (b) Omnidirectional image captured by the parabolic catadioptric camera in (a).

Catadioptric image formation has been studied in [4,7,8]. Central catadioptric systems are certainly of special interest since they have a single effective viewpoint. This property is important for easier analysis of the captured scene, but also for performing the multi-view 3D reconstruction. The images captured by the perspective camera from the light reflected by the catadioptric system are not straightforward to analyze. The lines in the 3D space are projected onto conic sections [4, 9, 10]. The unifying model for the catadioptric projective geometry for central catadioptric cameras has however been established by Geyer and Daniilidis in 2001 [4]. Any central catadioptric projection is equivalent to a composition of two mappings on the sphere. The first mapping represents a central spherical projection, with the center of the sphere incident to the focal point of the mirror, and it is independent of the mirror shape. The second mapping is a projection from a point on the principal axis of the sphere to the plane perpendicular to that axis. The position of the projection point on the axis of the sphere depends however on the shape of the mirror. The model developed in [4] includes perspective, parabolic, hyperbolic and elliptic projections. The catadioptric projective framework permits to derive efficient and simple scene analysis from omnidirectional images captured by catadioptric cameras. We describe now in more details the projective model for a parabolic mirror, where the second mapping represents the projection from the north pole to the plane that includes the equator. In particular, the second mapping is known as the stereographic projection and it is conformal.

We consider a cross-section of the paraboloid in a catadioptric system with a parabolic mirror. This is shown on Fig. 2. All points on the parabola are equidistant to the focus $\mathcal{F}_1$ and the directrix d. Let l pass through $\mathcal{F}_1$ and be perpendicular to the parabolic axis. If a circle has center $\mathcal{F}_1$ and radius equal

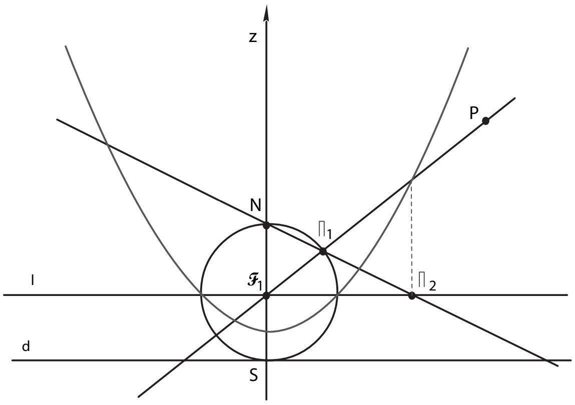


Fig. 2. Cross-section of mapping the omnidirectional image on the sphere [4]

to the double of the focal length of the paraboloid, then the circle and parabola intersect twice the line l and the directrix is tangent to the circle. The North Pole N of the circle is the point diametrically opposite to the intersection of the circle and the directrix. Point P is projected on the circle from its center, which gives Π_1. This is equivalent to a projective representation, where the projective space (set of rays) is represented as a circle here. It can be seen that Π_2 is the stereographic projection of the point Π_1 to the line l from the North pole N, where Π_1 is the intersection of the ray $\mathcal{F}_1 P$ and the circle. We can thus conclude that the parabolic projection of a point P yields point Π_2 which is collinear with Π_1 and N. Extending this reasoning to three dimensions, the projection by a parabolic mirror is equivalent to projection on the sphere (Π_1) followed by stereographic projection (Π_2). The formal proof of the equivalence between parabolic catadioptric projection and the described composite mapping through the sphere can be found in [4]. A direct corollary of this result is that the parabolic catadioptric projection is conformal, since it represents a composition of two conformal mappings.

For the other types of mirrors in catadioptric systems, the position of the point of projection in the second mapping is a function of the eccentricity ϵ of the conic (see Theorem 1 in [4]). For hyperbolic mirrors with $\epsilon > 1$ and elliptic mirrors with $0 < \epsilon < 1$, the projection point lies on the principal axis of the sphere, between the center of the sphere and the north pole. Perspective camera can be also considered as a degenerative case of a catadioptric system with a conic of the eccentricity $\epsilon = \infty$. In this case, the point of projection for the second mapping coincides with the center of the sphere.

2.3 Spherical camera model

We exploit the equivalence between the catadioptric projection and the composite mapping through the sphere in order to map the omnidirectional image through the inverse stereographic projection to the surface of a sphere whose

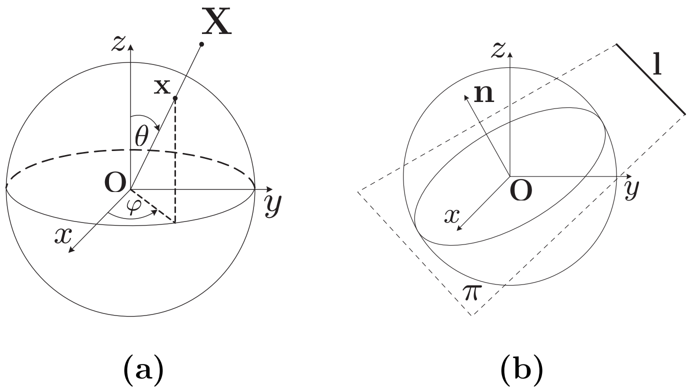


Fig. 3. Spherical projection model: (a) The projection of a point $\mathbf{X} \in \mathbb{R}^3$ to a point $\mathbf{x}$ on the spherical image. (b) The projection of a line $\mathbf{l}$ on the plane $\pi \in \mathbb{R}^3$ to a great circle on the spherical image [11]

center coincides with the focal point of the mirror. It leads to the definition of the Spherical camera model [11], which consists of a camera center and a surface of a unit sphere whose center is the camera center.

The surface of the unit sphere is usually referred to as **spherical image**. The spherical image is formed by a central spherical projection from a point $\mathbf{X} \in \mathbb{R}^3$ to the unit sphere with the center $\mathbf{O} \in \mathbb{R}^3$, as shown on the Figure 3(a). Point $\mathbf{X}$ is projected into a point $\mathbf{x}$ on the unit sphere S^2, where the projection is given by the following relation:

$$\mathbf{x} = \frac{1}{|\mathbf{X}|}\mathbf{X}. \tag{1}$$

The point $\mathbf{x}$ on the unit sphere can be expressed in spherical coordinates:

$$\mathbf{x} = \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} \sin\theta\cos\varphi \\ \sin\theta\sin\varphi \\ \cos\theta \end{pmatrix}$$

where $\theta \in [0, \pi]$ is the elevation angle, and $\varphi \in [0, 2\pi)$ is the azimuth angle. The spherical image is then represented by a function defined on the S^2, i.e., $I(\theta, \varphi) \in S^2$.

We now briefly discuss the consequences of the mapping of three-dimensional information on the 2D sphere. The spherical projection of a line $\mathbf{l}$ in $\mathbb{R}^3$ is a great circle on the sphere S^2, as illustrated on the Figure 3(b). This great circle, denoted as C, is obtained as the intersection of the unit sphere and the plane π that passes through the line $\mathbf{l}$ and the camera center $\mathbf{O}$. Since the great circle C is completely defined by the normal vector $\mathbf{n} \in S^2$ of the plane π, there is a duality of a great circle and a point on the sphere, as pointed out by Torii et al. in [11]. Let us denote with S^2_+ the union of the positive unit hemisphere $H_+ = \{(x,y,z)|z > 0\}$, and the semicircle on the sphere $\mathbf{s} = \{(x,y,z)|z = 0, y > 0\}$, i.e., $S^2_+ = X_+ \cup \mathbf{s}$. Without loss of generality, we can consider the normal vector $\mathbf{n}$ of the plane π, which belongs to S^2_+. The great circle C is then simply defined as:

$$C = \{\mathbf{x}|\mathbf{n}^T\mathbf{x} = 0, |\mathbf{x}| = 1\}. \tag{2}$$

The transform from the great circle C, defined as above, to the point $\mathbf{n} \in S^2_+$ is then formulated as follows:

$$f(C) = \lambda \frac{\mathbf{x} \times \mathbf{y}}{|\mathbf{x} \times \mathbf{y}|}, \quad \mathbf{x}, \mathbf{y} \in S^2, \quad \lambda \in \{-1, 1\}, \tag{3}$$

where λ is selected such that $\mathbf{n} \in S^2_+$. Similarly, we can define the inverse transform, from the point $\mathbf{n} \in S^2_+$ to the great circle C as:

$$f^{-1}(\mathbf{n}) = \{\mathbf{x}|\mathbf{n}^T\mathbf{x} = 0, |\mathbf{x}| = 1, \mathbf{n} \in S^2_+\}. \tag{4}$$

The defined transform in Eq. (3) and its inverse transform given by Eq. (4) represent the duality relations between great circles and points on the sphere.

Finally, there exists a duality between a line $\mathbf{l}$ on a projective plane $P^2 = \{(x,y,z)|z = 1\}$ and a point on S^2_+, as given in [11]. This duality is represented by the transform of a line $\mathbf{l} \in P^2$ to a point $\mathbf{n} \in S^2_+$:

$$g(\mathbf{l}) = \lambda | \frac{\xi \times \eta}{|\xi \times \eta|}, \quad \xi = (\mathbf{x}^T, 1)^T, \quad \eta = (\mathbf{y}^T, 1)^T, \quad \mathbf{x}, \mathbf{y} \in \mathbf{l}, \tag{5}$$

and its inverse transform, from a point $\mathbf{n} \in S^2_+$ to a line $\mathbf{l}$:

$$g^{-1}(\mathbf{n}) = \{\mathbf{x}|\mathbf{n}^T\xi = 0, \xi = (\mathbf{x}^T, 1)^T, \mathbf{n} \in S^2_+\}. \tag{6}$$

The formulated duality relations play an important role in defining the trifocal tensor for spherical cameras [11], which is usually used to express the three-view geometry constraints.

2.4 Image processing on the sphere

As images captured by catadioptric systems can be uniquely mapped on the sphere, it becomes interesting to process the visual information directly in the spherical coordinates system [12]. Similarly to the Euclidian framework, harmonic analysis and multiresolution decomposition represent efficient tools for processing data on the sphere. When mapped to spherical coordinates, the omnidirectional images are re-sampled on an equi-angular grid on the sphere:

$$\mathcal{G}_j = \{(\theta_{jp}, \varphi_{jq}) \in S^2 : \theta_{jp} = \tfrac{(2p+1)\pi}{4B_j}, \varphi_{jq} = \tfrac{q\pi}{B_j}\}, \tag{7}$$

$p, q \in \mathcal{N}_j \equiv \{n \in \mathbb{N} : n < 2B_j\}$ and for some range of bandwidth $B = \{B_j \in 2\mathbb{N}, j \in \mathbb{Z}\}$. These grids allow us to perfectly sample any band-limited function $I \in L^2(S^2)$ of bandwidth B_j.

The omnidirectional images can then be represented as spherical signals, modeled by elements of the Hilbert space of square-integrable functions on the two-dimensional sphere $L^2(S^2, d\mu)$, where $d\mu(\theta, \varphi) = d\cos\theta d\varphi$ is the rotation invariant Lebesgue measure on the sphere. These functions are characterized by their Fourier coefficients $\hat{I}(m, n)$, defined through spherical harmonics expansion :

$$\hat{I}(m,n) = \int_{S^2} d\mu(\theta,\varphi)\, Y^*_{m,n}(\theta,\varphi) I(\theta,\varphi),$$

where $Y^*_{m,n}$ is the complex conjugate of the spherical harmonic of order (m,n) [13]. It can be noted here that this class of sampling grids is associated to a Fast Spherical Fourier Transform [14].

Multiresolution representations are particularly interesting in applications such as image analysis or image coding. The two most successful embodiments of this paradigm that are the various wavelet decompositions [15] and the Laplacian Pyramid (LP) [16] can be extended to spherical manifolds.

The spherical wavelet transform (SCWT) has been introduced by Antoine and Vandergheynst [17]. The SCWT is based on affine transformations on the sphere, namely: rotations, defined by the element ρ of the group $SO(3)$, and dilations D_a, parameterized by the scale $a \in \mathbb{R}^*_+$ [18]. Interestingly enough, it can be proved that any admissible 2-D wavelet in $\mathbb{R}^2$ yields an admissible spherical wavelet by inverse stereographic projection. The action of rotations and dilations, together with an admissible wavelet $\psi \in L^2(S^2)$ permits to write the SCWT of a function $I \in L^2(S^2)$ as :

$$W_I(\rho, a) = \langle \psi_{\rho,a} | f \rangle = \int_{S^2} \mathrm{d}\mu(\theta,\varphi)\, I(\theta,\varphi)\, [R_\rho D_a \psi]^*(\theta,\varphi). \tag{8}$$

This last expression is nothing else but a spherical correlation, i.e., $W_I(\rho, a) = (I * \psi^*_a)(\rho)$.

Since the stereographic dilation is radial around the North Pole $N \in S^2$, an *axisymmetric* wavelet ψ on S^2, i.e., invariant under rotation around N, remains axisymmetric through dilation. So, if any rotation $\rho \in SO(3)$ is decomposed in its Euler angles $\varphi, \theta, \alpha \in S^1$, i.e., $\rho = \rho(\varphi, \theta, \alpha)$, then $R_\rho \psi_a = R_{[\omega]} \psi_a$, where $[\omega] = \rho(\varphi, \theta, 0) \in SO(3)$, is the result of two consecutive rotations moving N to $\omega = (\theta, \varphi) \in S^2$. Consequently, the SCWT is redefined on $S^2 \times \mathbb{R}^*_+$ by

$$W_I(\omega, a) \equiv (I * \psi^*_a)([\omega]) \equiv (I \star \psi^*_a)(\omega), \tag{9}$$

with $a \in \mathbb{R}^*_+$.

In order to process images given on discrete spherical grids, the SCWT can be replaced by *frames* of spherical wavelets [19, 20]. One of their most appealing features is the ability to expand any spherical map into a finite multiresolution hierarchy of wavelet coefficients. The scales are discretized in a monotonic way, namely :

$$a \in A = \{a_j \in \mathbb{R}_+^* : a_j > a_{j+1}, j \in \mathbb{Z}\}, \tag{10}$$

and the positions are taken in an equi-angular grid $\mathcal{G}_j$ described above.

Another simple way of dealing with discrete spherical data in a multiresolution fashion is to extend the Laplacian Pyramid [16] to spherical coordinates. This can be simply done considering the recent advances in harmonic analysis on S^2, in particular the work of Driscol et al [13]. Indeed, based on the notations introduced above, one can introduce a series of downsampled grids $\mathcal{G}_j$ with $B_j = 2^{-j}B_0$. A series of multiresolution spherical images S_j can be generated by recursively applying convolution with a lowpass filter h and downsampling. The filter h could for example take the form of an axisymmetric low-pass filter defined by its Fourier coefficients :

$$\hat{h}_{\sigma_0}(m) = e^{-\sigma_0^2 m^2}. \tag{11}$$

Suppose then that the original data I_0 is bandlimited, i.e, $\hat{I}_0(m, n) = 0$, $\forall m > B_0$, and sampled on $\mathcal{G}_0$. The bandwidth parameter σ_0 is chosen so that the filter is numerically close to a perfect half-band filter $\hat{H}_{\sigma_0}(m) = 0$, $\forall m > B_0/2$. The low pass filtered data is then downsampled on the nested sub-grid $\mathcal{G}_1$, which gives the low-pass channel of the pyramid I_1. The high-pass channel of the pyramid is computed as usual, that is by first upsampling I_1 on the finer grid $\mathcal{G}_0$, low-pass filtering it with H_{σ_0} and taking the difference with I_0. Coarser resolutions are computed by iterating this algorithm on the low-pass channel I_l and scaling the filter bandwidth accordingly, i.e., $\sigma_l = 2^l \sigma_0$.

Frames of spherical wavelets or the Spherical Laplacian Pyramid can efficiently supersede current solutions based on spherical harmonics for example due to their multiresolution and local nature, and also thanks to their covariance under rigid rotations, as already pointed out in [21].

3 Calibration of catadioptric cameras

3.1 Intrinsic parameters

We have assumed up to this point that the camera parameters are perfectly known, and that the cameras are calibrated. However, camera calibration is generally not given in practical systems. One usually has to estimate the intrinsic parameters of the projection, which enable mapping the pixels on the image to corresponding light rays in the space. For catadioptric cameras these parameters include: the focal length of the catadioptric system (combined focal length of the mirror and the camera), the image center, and the aspect ratio and skew factor of the imaging sensor. For the catadioptric camera with the parabolic mirror, the projection of the boundary of the mirror is projected to a circle on the image, which can be exploited for calibration. Namely, one can fit a circle to the image of the mirror's boundary and calculate the image center as the center of that circle. Then, knowing the field of view of the catadioptric camera with respect to the elevation angle, the focal length can be evaluated by simple calculations. This simple strategy for calibration is very advantageous since it does not require the capturing and analysis of the calibration pattern. It can be performed on any image, where the mirror boundary is sufficiently visible.

Another approach for calibration of catadioptric cameras uses the image projections of lines to estimate the intrinsic camera parameters [4, 22]. Unlike for perspective cameras where calibration from lines is not possible without any metric information, Geyer and Daniilidis have shown that it is possible to calibrate central catadioptric cameras only from a set of line images. They first consider the parabolic case and assume that the aspect ratio is one and skew is zero, so the total number of unknown intrinsic parameters is three (focal length and two coordinates of the image center). In the parabolic case, a line in the space projects into a circle on the image plane. Since a circle is defined by three points, each line gives a set of three constraints, but also introduces two unknowns that specify the orientation of the plane containing the line. Therefore, each line contributes with one additional constraint, leading us to the conclusion that three lines are sufficient to perform calibration. In the hyperbolic case, there is one additional intrinsic parameter to be estimated, i.e., the eccentricity, so in total there are four unknowns. The line projects into a conic, which is defined by five points, giving five constraints. Therefore, for the calibration of a hyperbolic catadioptric camera, only two line images suffice.

Based on this reasoning, Geyer and Daniilidis have proposed an algorithm for calibration of parabolic catadioptric cameras from line images. We will here briefly explain the main steps of their algorithm, while for the details we refer the interested reader to [4, 22]. The only assumption of the algorithm is that images of at least three lines are obtained. In the first step, the algorithm obtains points from the line images, where the number of points per line is $M \geqslant 3$ if the aspect ratio and skew factor are known, or $M \geqslant 5$ if they are not known. An ellipse is fitted to each set of points on the same line image in the second step. This gives a unique affine transformation, which transforms these ellipses whose axis are parallel and aspect ratios identical into a set of corresponding circles. When the skew factor is equal to zero, the aspect ratio can be derived from the obtained affine transformation in the closed form [22]. However, in the presence of skew, aspect ratio and skew factor are solutions of the polynomial equation, and have to be evaluated numerically. The evaluated affine transformation is applied on the line image points, which are then fitted to circles in the third step of the algorithm. For each line i, the center $\mathbf{c}_i$ and the radius r_i of the corresponding circle are calculated. In the final step, the algorithm finds the image center $\xi = (\xi_x, \xi_y)$ and the focal length f, using the knowledge that a sphere constructed on each image circle passes through the point $(\xi_x, \xi_y, 2f)^T$. To understand this fact, we need first to define the fronto-parallel horizon, which represents a projection of a plane parallel to the image plane and is a circle centered at the image center with radius $2f$. The fronto-parallel horizon is depicted on the Fig. 4. It was shown in [22] that a line image intersects the fronto-parallel horizon antipodally, i.e., at two points both at distance $2f$ from the image center. Therefore, by symmetry, a sphere constructed on each image circle passes through the point $(\xi_x, \xi_y, 2f)^T$. All three coordinates of this point are however unknown, and one needs three defined spheres in order to find this point in the intersection of the three spheres, as illustrated on the Fig. 4.

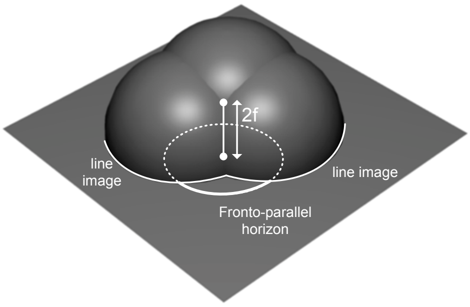


Fig. 4. Intersection of three spheres constructed from line images yields a point on the mirror's axis a distance $2f$ above the image plane. [22]

Due to the presence of noise this intersection most probably does not exist, hence Geyer and Daniilidis proposed to find instead a point in space that minimizes the distance to all of the spheres. This reduces to minimizing the following objective function:

$$\chi(\xi, f) = \sum_{i=1}^{L} \left((\xi - \mathbf{c}_i)^T (\xi - \mathbf{c}_i) + 4f^2 - r_i^2 \right)^2, \tag{12}$$

where L is the number of obtained line images. By solving $(\partial\chi/\partial f)(\xi, f) = 0$, we get the solution for f:

$$f_0^2 = \frac{1}{4L} \sum_{i=1}^{L} \left(r_i^2 - (\xi - \mathbf{c}_i)^T (\xi - \mathbf{c}_i) \right). \tag{13}$$

With the obtained solution for f, minimizing $\chi(\xi, f)$ over ξ yields:

$$\xi_0 = -\frac{1}{2} \mathbf{A}^{-1} \mathbf{b}, \tag{14}$$

where $\mathbf{A}$ and $\mathbf{b}$ are given with:

$$\begin{aligned} \mathbf{A} &= \sum_{i,j,k}^{L} (\mathbf{c}_k - \mathbf{c}_i)^T (\mathbf{c}_j - \mathbf{c}_i) \\ \mathbf{b} &= \sum_{i,j,k}^{L} (\mathbf{c}_i^T \mathbf{c}_i - r_i^2 - \mathbf{c}_j^T \mathbf{c}_j + r_j^2)(\mathbf{c}_k - \mathbf{c}_i). \end{aligned} \tag{15}$$

While we have focused on calibration using the projection of lines, camera calibration can also be performed by the projection of spheres [23], which actually offers improved robustness. Besides calibration of paracatadioptric cameras, researchers have also investigated the calibration of cameras with different types of mirrors, such as hyperbolic/elliptical mirrors [24]. For example, Scaramuzza et al. propose in [25] a calibration method that uses a generalized parametric model of the single viewpoint omnidirectional sensor and can be applied to any type of mirror in the catadioptric system. The calibration requires two or more images of the planar pattern at different orientations. Calibration of non-central catadioptric cameras was proposed by Mičušík and Pajdla [26], based on epipolar correspondence matching from two catadioptric images. Epipolar geometry has also been exploited in [27] for calibration of paracatadioptric cameras.

3.2 Extrinsic parameters

In multi-camera systems, the calibration process includes the estimation of extrinsic parameters in addition to the intrinsic ones. Extrinsic parameters include the relative rotation and translation between cameras, which are necessary in applications like depth estimation, structure from motion, etc. For non-central catadioptric cameras, Mičušík and Pajdla [26] have addressed the problem of intrinsic parameters calibration (as mentioned in Sec. 3.1), and also the estimation of extrinsic parameters. They extract possible point correspondences from two camera views, and validate them based on the approximated central camera model. The relative rotation and translation are evaluated by linear estimation from the epipolar geometry constraint, and the estimation robustness against outliers is improved by the RANSAC algorithm. Antone and Teller [28] consider the calibration of extrinsic parameters for omnidirectional camera networks, while the intrinsic parameters are assumed to be known. Their approach decouples the rotation and translation estimation in order to obtain a linear time calibration algorithm. The Expectation maximization (EM) algorithm recovers the rotation matrix from the vanishing points, followed by the position recovery using feature correspondence coupling and the Hough transform with Monte Carlo EM refinement. Robustness of extrinsic parameters recovery can be improved by avoiding commitment to point correspondences, as presented by Makadia and Daniilidis in [29]. For a general spherical image model, they introduce a correspondenceless method for camera rotation and translation recovery based on the Radon transform. They define the Epipolar Delta Filter (EDF) which embeds the epipolar geometry constraint for all pairs of features with a series of Diracs on the sphere. Moreover, they define the similarity function on all feature pairs. The main result of this work is the realization that the Radon transform is actually a correlation on the SO(3) group of rotations of the EDF and a similarity function, and as such it can be efficiently evaluated by the fast Fourier transform on the SO(3). The extrinsic parameters are then found in the maximum of the five-dimensional space given by the Radon transform.

4 Multi-camera systems

With the development of panoramic cameras, epipolar or multi-view geometry has been recently formalized for general camera models, including non-central cameras [30]. However, since single viewpoint cameras have some advantages, like simple image rectification at any direction, we describe in this section the epipolar geometry only for central catadioptric cameras. Due to the equivalence between the catadioptric projection and the composite mapping through the sphere, the epipolar geometry constraint can be formulated through the spherical camera model. We consider in particular the case of calibrated paracatadioptric cameras, where the omnidirectional image can be uniquely mapped through the inverse stereographic projection to the surface of the sphere whose center coincides with the focal point of the mirror. Two- and three- view geometry for spherical cameras has been introduced by Torii et al. [11], and we overview the two-view case in this section. We refer the interested reader to [11] for the three-view geometry framework. Epipolar geometry has very important implications in the representation or the understanding of 3D scenes. As an example, we discuss the use of geometry constraints for the estimation of disparity in networks of omnidirectional cameras.

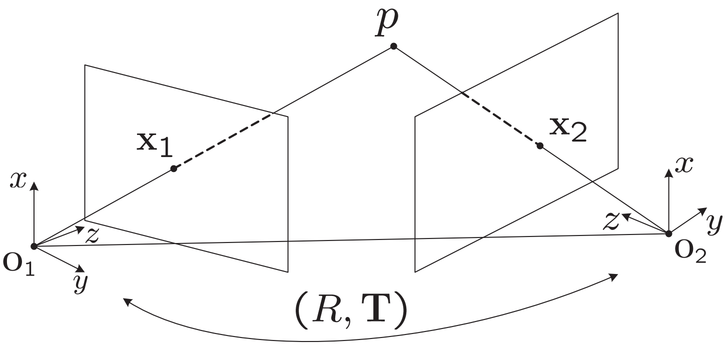


Fig. 5. Epipolar geometry for the pinhole camera model.

4.1 Epipolar geometry for paracatadioptric cameras

Epipolar geometry relates the multiple images of the observed environment with the 3-dimensional structure of that environment. It represents a geometric relation between 3D points and their image projections, which enables 3-dimensional reconstruction of the scene using multiple images taken from different viewpoints. Epipolar geometry has been first formulated for the pinhole camera model, leading to the well-known epipolar constraint. Consider a point p in $\mathbb{R}^3$, given by its spatial coordinates $\mathbf{X}$, and two images of this point given by their homogeneous coordinates $\mathbf{x}_1 = [x_1 \; y_1 \; 1]^T$ and $\mathbf{x}_2 = [x_2 \; y_2 \; 1]^T$ in two camera frames. The epipolar constraint gives the geometric relationship between $\mathbf{x}_1$, $\mathbf{x}_2$, as given in the following theorem [31]:

Theorem 1 *Consider two images $\mathbf{x}_1$ and $\mathbf{x}_2$ of the same point p from two camera positions with relative pose $(R, \mathbf{T})$, where $R \in SO(3)$ is the relative orientation and $\mathbf{T} \in \mathbb{R}^3$ is the relative position. Then $\mathbf{x}_1$, $\mathbf{x}_2$ satisfy:*

$$\langle \mathbf{x}_2, \mathbf{T} \times R\mathbf{x}_1 \rangle, \qquad \text{or} \qquad \mathbf{x}_2^T \hat{T} R \mathbf{x}_1 = 0. \tag{16}$$

The matrix $\hat{T}$ is obtained by representing the cross product of $\mathbf{T}$ with $R\mathbf{x}_1$ as matrix multiplication, i.e., $\hat{T}R\mathbf{x}_1 = \mathbf{T} \times R\mathbf{x}_1$. Given $\mathbf{T} = [t_1 \; t_2 \; t_3]^T$, $\hat{T}$ can be expressed as:

$$\hat{T} = \begin{pmatrix} 0 & -t_3 & t_2 \\ t_3 & 0 & -t_1 \\ -t_2 & t_1 & 0 \end{pmatrix}$$

The matrix $E = \hat{T}R \in \mathbb{R}^{3\times3}$ is called the **essential matrix**. The epipolar geometry constraint is derived from the coplanarity of the vectors $\mathbf{x}_2$, $\mathbf{T}$ and $R\mathbf{x}_1$, as shown on the Figure 5.

Epipolar geometry can be also used to describe the geometrical constraints in systems with two spherical cameras. Let $\mathbf{O}_1$ and $\mathbf{O}_2$ be the centers of two spherical cameras, and let the world coordinate frame be placed at the center of the first camera, i.e., $\mathbf{O}_1 = (0, 0, 0)$. A point $p \in \mathbb{R}^3$ is projected to unit spheres corresponding to the cameras, giving projection points $\mathbf{x}_1, \mathbf{x}_2 \in S^2$, as illustrated on the Figure 6. Let X_1 be the coordinates of the point p in the coordinate system of camera centered at $\mathbf{O}_1$. The spherical projection of a point p to the camera centered at $\mathbf{O}_1$ is given as:

$$\lambda_1 \mathbf{x}_1 = \mathbf{X}_1, \quad \lambda_1 \in \mathbb{R} \tag{17}$$

If we further denote the transform of the coordinate system between two spherical cameras with R and $\mathbf{T}$, where R denotes the relative rotation and $\mathbf{T}$ denotes the relative position, the coordinates of the point p can be expressed in the coordinate system of the camera at $\mathbf{O}_2$ as $\mathbf{X}_2 = R\mathbf{X}_1 + \mathbf{T}$. The projection of the point p to the camera centered at $\mathbf{O}_2$ is then given by:

$$\lambda_2 \mathbf{x}_2 = \mathbf{X}_2 = R\mathbf{X}_1 + \mathbf{T}, \quad \lambda_2 \in \mathbb{R}. \tag{18}$$

Similarly as for pinhole camera model, vectors $\mathbf{x}_2$, $R\mathbf{x}_1$ and $\mathbf{T}$ are coplanar, and the epipolar geometry constraint is formalized as:

$$\mathbf{x}_2^T \hat{T} R \mathbf{x}_1 = \mathbf{x}_2^T E \mathbf{x}_1 = 0. \tag{19}$$

The epipolar constraint is one of the fundamental relations in the multi-view geometry because it allows the estimation of the 3-dimensional coordinates of the point p from its images $\mathbf{x}_1, \mathbf{x}_2$, given R and $\mathbf{T}$, i.e., it allows scene geometry reconstruction. However, when the point p lies on the vector $\mathbf{T}$, which connects camera centers $\mathbf{O}_1$ and $\mathbf{O}_2$, it leads to a degenerative case of the epipolar constraint because vectors $\mathbf{x}_2$, $R\mathbf{x}_1$ and $\mathbf{T}$ are collinear, and the coordinates of the point p cannot be determined. The intersection points of unit spheres of both cameras and the vector $\mathbf{T}$ are called the **epipoles**, and are denoted as $\mathbf{e}_1$ and $\mathbf{e}_2$ on the Figure 6. In other words, when the point p is projected to epipoles of two cameras, the reconstruction of p is not possible from these cameras.

4.2 Disparity estimation

Geometry constraints play a key role in scene representation or understanding. In particular, the problem of depth estimation is mainly based on geometrical constraints, in order to reconstruct the depth information with images from multiple cameras. Dense disparity estimation on omnidirectional images has become a part of localization, navigation and obstacle avoidance research. When several cameras capture the same scene, the geometry of the scene can

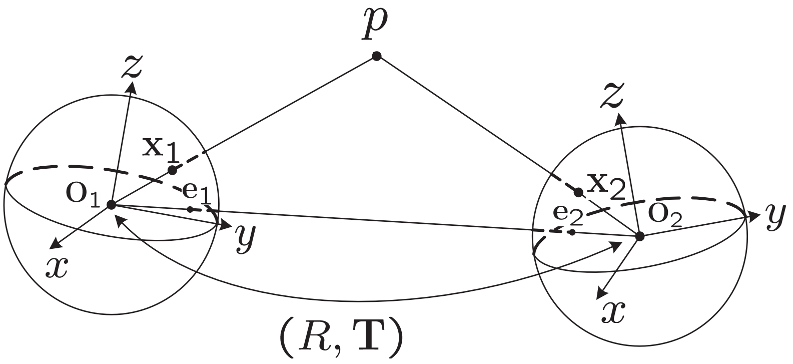


Fig. 6. Epipolar geometry for the spherical camera model

be estimated by comparing the images from the different sensors. The differences between the respective positions of 3D points on multiple 2D images represent disparities that can be estimated by stereo matching methods. In general, the strategies that have been developed for dense disparity estimation from standard camera images, are also applicable to omnidirectional images. The algorithms are generally based on re-projection of omnidirectional images on simpler manifolds. For example, Takiguchi *et al* [32] re-project omnidirectional images onto cylinders, while Gonzalez-Barbosa *et al* [33] and Geyer *et al* [34] rectify omnidirectional images on the rectangular grid. Both cylindrical and rectangular projections are however not sufficient to represent neighborhood and correlations among the pixels. The equiangular grid on the sphere has a better representation for the omnidirectional images and it is possible to map omnidirectional images onto the 2D sphere by inverse stereographic projection as discussed earlier.

We therefore discuss here the problem of disparity estimation in systems with two omnidirectional cameras, and we refer the reader to [35] for depth estimation in systems with more cameras. In order to perform disparity estimation directly in a spherical framework, rectification of the omnidirectional images is performed in the spherical domain. Then, global energy minimization algorithm based on the graph-cut algorithm can be implemented to perform dense disparity estimation on the sphere. Interestingly, disparities can be directly processed on the 2D sphere in order to better exploit the geometry of omnidirectional images and to improve the accuracy of the estimation.

Rectification is an important step in stereo vision using standard camera images. It aims at reducing the stereo correspondence estimation to a one-dimensional search problem. It basically consists in image warping, which is computed such that epipolar lines coincide with the scan lines. It does not only ease the implementation of disparity estimation algorithms, but also makes computations faster. In the spherical framework, rectification can be performed directly on spherical images. The following observations about epipoles on spherical images can be used here: (i) epipoles resemble the coordinate poles

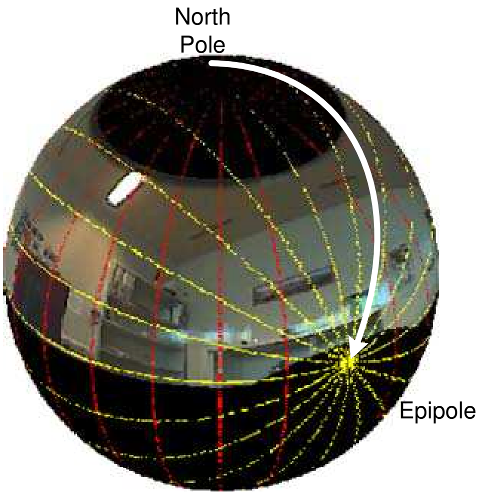


Fig. 7. Resemblance between longitudes and epipolar great circles and corresponding rotation.

and (ii) epipolar great circles intersecting on epipoles are like longitude circles. Spherical image pairs can thus undergo rotation in the spherical domain such that epipoles coincide with the coordinate poles. In this way, epipolar great circles coincide with the longitudes and disparity estimation becomes a mono-dimensional problem. Figure 7 illustrates the rectification strategy, and Figure 8 shows original and rectified spherical images, represented as rectangular images with latitude and longitude angles as axes that correspond to an equiangular grid on the sphere. Rectification permits to extend the disparity estimation algorithms developed for standard images to spherical images with fast computations.

In the spherical framework, disparity can be defined as the difference in angle values between the representation of the same 3D point, in two different omnidirectional images. Since pixel coordinates are defined with angles, we define the disparity γ, as the difference between the angles corresponding to pixel coordinates on the two images, i.e., $\gamma = \alpha - \beta$, as illustrated on the Fig. 9.

Figure 9 shows the 2D representation of the geometry between cameras and the 3D point. The depth R_1 is defined as the distance of the 3D point to the reference camera center. The relation between the disparity γ, the depth R_1 and baseline distance d, is given as

$$\gamma = arcsin \frac{dsin\beta}{R_1} \tag{20}$$

This relation holds for all epipolar great circles on rectified stereo images.

The disparity estimation problem can now be casted in an energy minimization problem. Let $\mathcal{P}$ denote the set of all pixels on two rectified spherical images. Let further $\mathcal{L} = \{l_1, l_2, \ldots, l_{max}\}$ represent the finite set of discrete labels

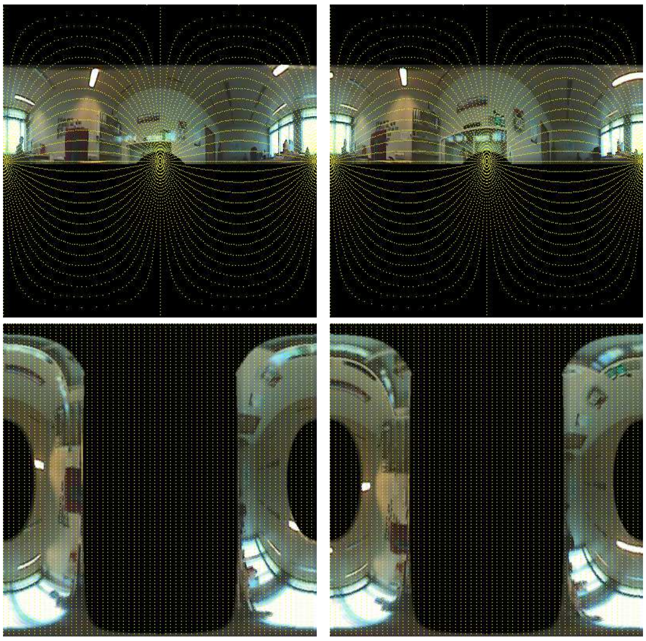


Fig. 8. Original images (top), and rectified ones (bottom). Epipolar great circles turn into straight vertical lines in rectified images.

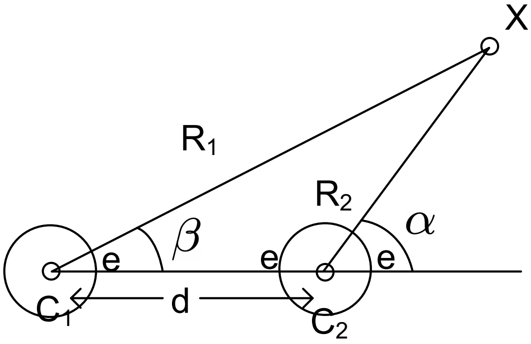


Fig. 9. 2D representation of the geometry between cameras and the 3D point

corresponding to disparity values. Recall that the disparity estimation problem is mono-dimensional, due to the image rectification step. A single value per pixel is therefore sufficient to represent the disparity, and the disparity values for each pixel together form the disparity map. If $f : \mathcal{P} \to \mathcal{L}$ is a mapping so that each pixel is assigned a disparity label, our aim is to find the optimum mapping f^* such that the disparity map is as accurate and smooth as possible.

The computation of the optimum mapping f^* can be formulated as an energy minimization problem, where the energy function $E(f)$ is built on two components E_d and E_σ, which respectively represent the data and smoothness functions.

$$E(f) = E_d(f) + E_\sigma(f) \tag{21}$$

The data function first reports the photo-consistency between the omnidirectional images. It can be written as :

$$E_d(f) = \sum_{(p,q)\in\mathcal{P}^2} D(p,q), \tag{22}$$

where p and q are corresponding pixels in two images under a mapping function f. $D(.,.)$ is non-positive cost function, which can be expressed as :

$$D(p,q) = \min\{0, (I(p) - I(q))^2 - K\}, \tag{23}$$

where $I(i)$ represents the intensity or luminance of pixel i and K is a positive constant. The intensity $I(i)$ can be defined as in [36], where it presents the advantage to be insensitive to image sampling, which is useful since equiangular grid on sphere causes non-uniform sampling.

The smoothness function then captures the variations of the disparity between neighboring labels. The goal of the smoothness function is to penalize the estimation of labels that are different from their neighborhood, in order to obtain a smooth disparity field. The neighborhood $\mathcal{O}$ is generally represented by the 4 surrounding labels. The smoothness function under a mapping f can for example be expressed as :

$$E_\sigma(f) = \sum_{(p,q)\in\mathcal{O}} V_{p,q}(f(p), f(q)). \tag{24}$$

The term $V_{p,q} = \min\{|l_p, l_q|, K\}$ represents a distance metric proposed in [37], where K is a constant. It reports the difference between labels l_p amd l_q attributed to neighboring pixels p and q in the neighborhood $\mathcal{O}$.

The dense disparity estimation problem now consists in minimizing the energy function $E(f)$, in order to obtain an accurate and smooth disparity map, and such a global minimization can typically be performed efficiently by graph-cut algorithms [35, 37], or belief propagation methods. In order to illustrate the performance of the disparity estimation algorithm, we show in Figure 10 the dense disparity map computed by the graph-cut algorithm. The results correspond to the images illustrated in Figure 8, where a room has been captured from two different positions with a catadioptric system and parabolic mirrors.

5 Sparse approximations and geometric estimation

5.1 Correlation estimation with sparse approximations

This section presents a geometry-based correlation model between multi-view images that relates image projections of 3D scene features in different views,

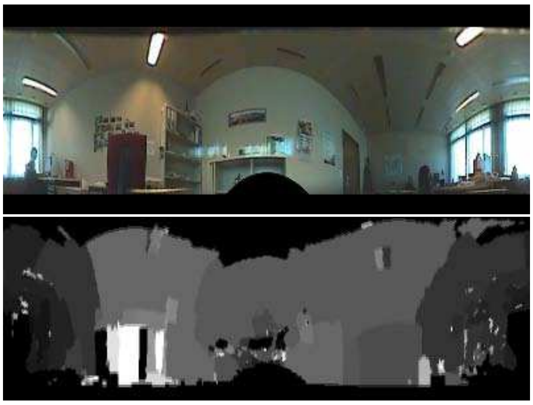


Fig. 10. Reference image and disparity images a graph-cut method on the sphere [35].

assuming that these features are correlated by local transforms, such as translation, rotation or scaling [38]. These features can be represented by a sparse image expansion with geometric atoms taken from a redundant dictionary of functions. This provides a flexible and efficient alternative for autonomous systems, which cannot afford manual feature selection.

The correlation model between multi-view images introduced in [38] relates image components that approximate the same 3D object in different views, by local transforms that include translation, rotation and anisotropic scaling. Given a redundant dictionary of atoms $\mathcal{D} = \{\phi_k\}, k = 1, ..., N$, in the Hilbert space H, we say that image I has a *sparse* representation in $\mathcal{D}$ if it can be approximated by a linear combination of a small number of vectors from $\mathcal{D}$. Therefore, sparse approximations of two[1] multi-view images can be expressed as $I_1 = \mathbf{\Phi}_{\mathbf{\Omega_1}} c_1 + \eta_1$ and $I_2 = \mathbf{\Phi}_{\mathbf{\Omega_2}} c_2 + \eta_2$, where $\Omega_{1,2}$ labels the set of atoms $\{\phi_k\}_{k \in \Omega_{1,2}}$ participating in the sparse representation, $\mathbf{\Phi}_{\mathbf{\Omega_{1,2}}}$ is a matrix composed of atoms ϕ_k as columns, and $\eta_{1,2}$ represents the approximation error. Since I_1 and I_2 capture the same 3D scene, their sparse approximations over the sets of atoms Ω_1 and Ω_2 are correlated. The geometric correlation model makes two main assumptions in order to relate the atoms in Ω_1 and Ω_2:

1. The most prominent (energetic) features in a 3D scene are present in sparse approximations of both images, with high probability. The projections of these features in images I_1 and I_2 are represented as subsets of atoms indexed by $Q_1 \in \Omega_1$ and $Q_2 \in \Omega_2$ respectively.
2. These atoms are correlated, possibly under some local geometric transforms. We denote by $F(\phi)$ the transform of an atom ϕ between two image decompositions that results from a viewpoint change.

Under these assumptions the correlation between the images is modeled as a set of transforms F_i between corresponding atoms in sets indexed by Q_1 and

[1] Two images are taken for the sake of clarity, but the correlation model can be generalized to any number of images.

Q_2. The approximation of the image I_2 can be rewritten as the sum of the contributions of transformed atoms, remaining atoms in Ω_2, and noise η_2:

$$I_2 = \sum_{i \in Q_1} c_{2,i} F_i(\phi_i) + \sum_{k \in \Omega_2 \setminus Q_2} c_{2,k} \phi_k + \eta_2. \tag{25}$$

The model from Eq. (25) is applied in [38] to atoms from the sparse decompositions of omnidirectional multi-view images mapped onto the sphere. The approach is based on the use of a structured redundant dictionary of atoms that are derived from a single waveform subjected to rotation, translation and scaling. More formally, given a generating function g defined in H (in the case of spherical images g is defined on the 2-sphere), the dictionary $\mathcal{D} = \{\phi_k\} = \{g_\gamma\}_{\gamma \in \Gamma}$ is constructed by changing the atom index $\gamma \in \Gamma$ that defines rotation, translation and scaling parameters applied to the generating function g. This is equivalent to applying a unitary operator $U(\gamma)$ to the generating function g, i.e.: $g_\gamma = U(\gamma)g$. As an example, Gaussian atoms on the sphere are illustrated on the Figure 11, for different motion, rotation and anisotropic scaling parameters.

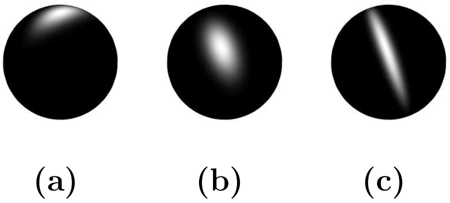


Fig. 11. Gaussian atoms: a) on the North pole ($\tau = 0, \nu = 0$), $\psi = 0, \alpha = 2, \beta = 4$; b) $\tau = \frac{\pi}{4}, \nu = \frac{\pi}{4}, \psi = \frac{\pi}{8}, \alpha = 2, \beta = 4$; c) $\tau = \frac{\pi}{4}, \nu = \frac{\pi}{4}, \psi = \frac{\pi}{8}, \alpha = 1, \beta = 8$.

The main property of the structured dictionary is that it is transform-invariant, i.e., the transformation of an atom by any of the combination of translation, rotation and anisotropic scaling transforms results in another atom in the same dictionary. Let $\{g_\gamma\}_{\gamma \in \Gamma}$ and $\{h_\gamma\}_{\gamma \in \Gamma}$ respectively denote the set of functions used for the expansions of images I_1 and I_2. When the transform-invariant dictionary is used for both images, the transform of the atom g_{γ_i} in image I_1 to the atom h_{γ_j} in image I_2 reduces to a transform of its parameters, i.e., $h_{\gamma_j} = F(g_{\gamma_i}) = U(\gamma')g_{\gamma_i} = U(\gamma' \circ \gamma_i)g$. Due to the geometric constraints that exist in multi-view images, only a subset of all local transforms between $\{g_\gamma\}$ and $\{h_\gamma\}$ are feasible. This subset can be defined by identifying two constraints between corresponding atoms, namely the *epipolar* constraint and the *shape similarity* constraint. Given the atom $\{g_\gamma\}_{\gamma \in \Gamma}$ in image y_1 these two constraints give the subset of possible parameters $\Gamma_i \subseteq \Gamma$ of the correlated atom $\{h_{\gamma_j}\}$. Pairs of atoms that correspond to the same 3D points have to satisfy the epipolar constraints that represent one of the fundamental relations in multi-view analysis. Two corresponding atoms are said to match when their epipolar atom distance $d_{EA}(g_{\gamma_i}, h_{\gamma_j})$ is smaller than a certain threshold κ (for more details on this distance we refer the reader to [38]). The set of possible candidate atoms in I_2, that respect epipolar constraints with the atom g_{γ_i}

in I_1, called the *epipolar candidates set*, is then defined as the set of indexes $\Gamma_i^E \subset \Gamma$, with:

$$\Gamma_i^E = \{\gamma_j | h_{\gamma_j} = U(\gamma')g_{\gamma_i}, d_{EA}(g_{\gamma_i}, h_{\gamma_j}) < \kappa\}. \tag{26}$$

The shape similarity constraint assumes that the change of viewpoint on a 3D object results in a limited difference between shapes of corresponding atoms since they represent the same object in the scene. From the set of atom parameters γ, the last three parameters (ψ, α, β) describe the atom shape (its rotation and scaling), and they are thus taken into account for the shape similarity constraint. We measure the similarity or coherence of atoms by the inner product $\mu(i,j) = |\langle g_{\gamma_i}, h_{\gamma_j}\rangle|$ between centered atoms (at the same position (τ, ν)), and we impose a minimal coherence between candidate atoms, i.e., $\mu(i,j) > s$. This defines a set of possible transforms $V_i^\mu \subseteq V_i^0$ with respect to atom shape, as:

$$V_i^\mu = \{\gamma' | h_{\gamma_j} = U(\gamma')g_{\gamma_i}, \mu(i,j) > s\}. \tag{27}$$

Equivalently, the set of atoms h_{γ_j} in I_2 that are possible transformed versions of the atom g_{γ_i} is denoted as the *shape candidates set*. It is defined by the set of atoms indexes $\Gamma_i^\mu \subset \Gamma$, with

$$\Gamma_i^\mu = \{\gamma_j | h_{\gamma_j} = U(\gamma')g_{\gamma_i}, \gamma' \in V_i^\mu\}. \tag{28}$$

Finally, we combine the epipolar and shape similarity constraints to define the set of possible parameters of the transformed atom in I_2 as $\Gamma_i = \Gamma_i^E \cap \Gamma_i^\mu$.

5.2 Distributed coding of 3D scenes

Interpreting and compressing the data acquired by a camera network represents a challenging task as the amount of data is typically huge. Moreover, in plenty of cases the communication among cameras is limited because of the bandwidth constraints or the introduced time delay. Development of distributed processing and compression techniques in the multi-camera setup thus becomes necessary in a variety of applications. Distributed coding of multi-view images captured by camera networks recently attained great interest among researchers. Typically, the distributed coding algorithms are based on the disparity estimation between views under epipolar constraints [39, 40], in order to capture the correlation between the different images of the scene. Alternatively, the geometrical correlation model described in the previous section can also be used for the design of a distributed coding method with side information. This model is particularly interesting in scenarios with multi-view omnidirectional images mapped to spherical images, as described in previous sections.

Based on the above geometric correlation model, we can build a Wyner-Ziv coding scheme [41] for multi-view omnidirectional images. The coding scheme illustrated in Fig. 12 is mostly based on the algorithm proposed in [38], with the shape similarity metric described above. The scheme is based on coding with side information, where image I_1 is independently encoded, while the Wyner-Ziv image I_2 is encoded by coset coding of atom indices and quantization of their coefficients. The approach is based on the observation that when atom h_{γ_j} in the Wyner-Ziv image I_2 has its corresponding atom g_{γ_i} in the reference image I_1, then γ_j belongs to the subset $\Gamma_i = \Gamma_i^E \; \cap \Gamma_i^\mu$. Since Γ_i is usually much smaller than Γ, the Wyner-Ziv encoder does not need to send the whole γ_j, but can transmit only the information that is necessary to identify the correct atom in the transform candidate set given by Γ_i. This is achieved by coset coding, by partitioning Γ into distinct cosets that contain dissimilar atoms with respect to their position (τ, ν) and shape (ψ, α, β). Two types of cosets were constructed: Position cosets, and Shape cosets. The encoder eventually sends only the indexes of the corresponding cosets for each atom (i.e., k_n and l_n in Fig. 12). The Position cosets are designed as VQ cosets [38], which are constructed by 2-dimensional interleaved uniform quantization of atom positions (τ, ν) on a rectangular lattice. Shape cosets are designed by distributing all atoms whose parameters belong to Γ_i^μ into different cosets.

The decoder matches corresponding atoms in the reference image and atoms within the cosets of the Wyner-Ziv image decomposition using the correlation model described earlier. The atom pairing is facilitated by the use of quantized coefficients of atoms, which are sent directly. Each identified atom pair contains the information about the local transform between the reference and Wyner-Ziv image, which is exploited by the decoder to update the transform field between them. The transformation of the reference image with respect to the transform field provides an approximation of the Wyner-Ziv image that is used as side information for decoding the atoms without a correspondence in the reference image. These atoms are decoded based on the minimal mean square error between the currently decoded image and the side information. Finally, the WZ image reconstruction $\hat{I}_2$ is obtained as a linear combination of the decoded image I_d, reconstructed by decoded atoms from $\mathbf{\Phi_{\Omega_2}}$, and the projection of the transformed reference image I_{tr} to the orthogonal complement of $\mathbf{\Phi_{\Omega_2}}$ [38].

The decoding procedure does not give any guarantee that all atoms are correctly decoded. In a general multi-view system, occlusions are quite probable, and clearly impair the atom pairing process at the decoder. The atoms which approximate these occlusions cannot be decoded based on the proposed correlation model since they do not have a corresponding feature in the reference

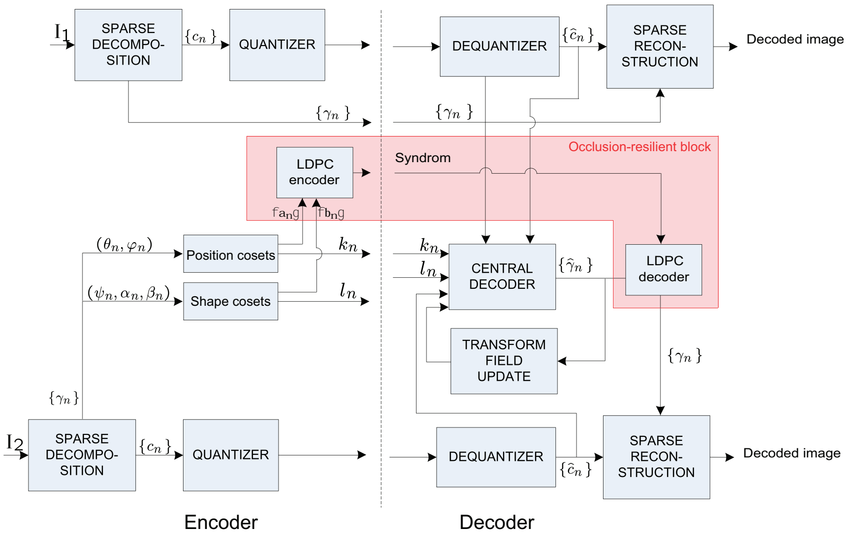


Fig. 12. Occlusion-resilient Wyner-Ziv coder

image. An occlusion-resilient coding block (shaded area in Fig. 12) can be built by performing channel coding on all a_n and b_n, $n = 1, ..., N$, together. The encoder thus sends a unique syndrome for all atoms, which is then used by the decoder to correct the erroneously decoded atom parameters.

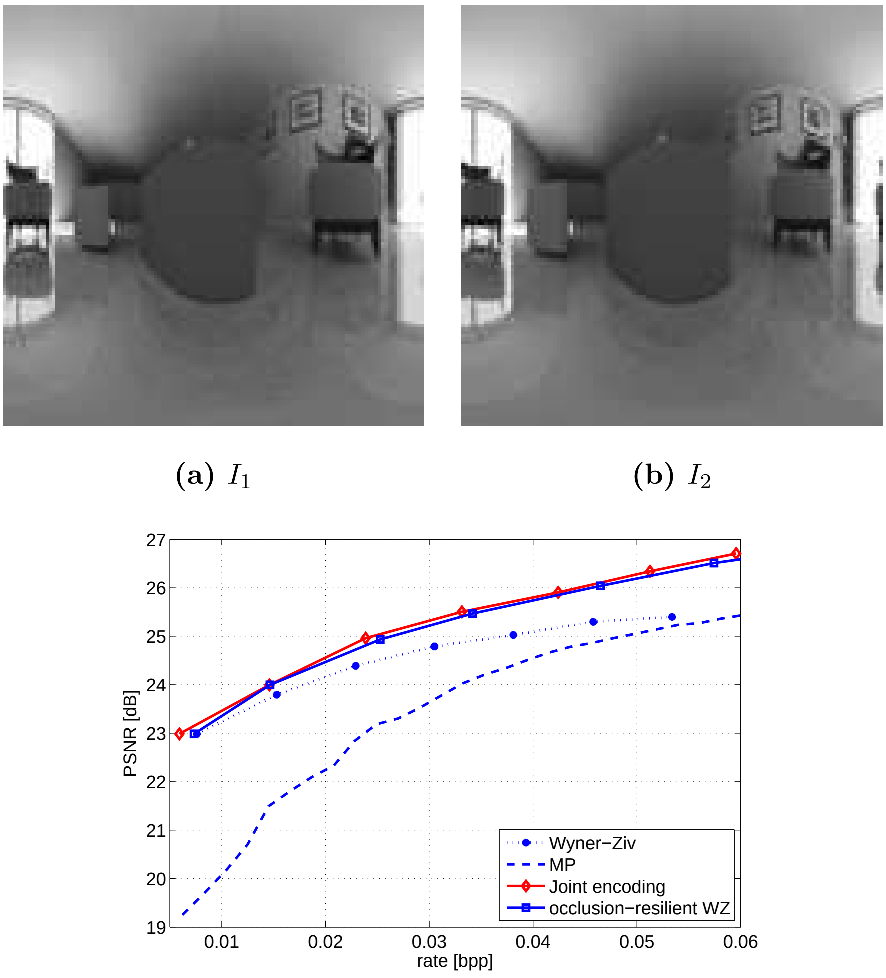
Finally, we discuss briefly the performance of the distributed coding scheme with the geometrical correlation model described above. A deeper analysis is given in [38]. The results are presented for the synthetic Room image set that consists of two 128 × 128 spherical images I_1 and I_2. The sparse image decomposition is obtained using the Matching Pursuit (MP) algorithm on the sphere with the dictionary used in [38]. It is based on two generating functions: a 2D Gaussian function, and a 2D function built on a Gaussian and the second derivative of a 2D Gaussian in the orthogonal direction (i.e., edge-like atoms). The position parameters τ and ν can take 128 different values, while the rotation parameter uses 16 orientations. The scales are distributed logarithmically with 3 scales per octave. This parametrization of the dictionary enables the use of fast computation of correlation on SO(3) for the full atom search within the MP algorithm. In particular, we used the *SpharmonicKit* library[2], which is part of the *YAW toolbox*[3]. The image I_1 is encoded independently at 0.23bpp with a PSNR of 30.95dB. The atom parameters for the expansion of image I_2 are coded with the proposed scheme. The coefficients are obtained by projecting the image I_2 on the atoms selected by MP, in order to improve the atom matching process, and they are quantized uniformly.

We see that the performance of the Wyner-Ziv scheme with and without the occlusion resilient block are very competitive with joint encoding strategies.

[2] http://www.cs.dartmouth.edu/~geelong/sphere/

[3] http://fyma.fyma.ucl.ac.be/projects/yawtb/

The RD curve for the occlusion-resilient Wyner-Ziv scheme using LDPC coding is given on the Fig. 13 with the blue solid line. We can clearly see that the occlusion resilient coding corrects the saturation behavior of the previous scheme, and performs very close to the joint encoding scheme. The geometrical model based on sparse approximations of the ominidirectional images is therefore able to capture efficiently the correlation between images. This proves to be very beneficial for the distributed representation of 3D scenes.



(a) I_1

(b) I_2

(c)

Fig. 13. (a) and (b) Original Room images (128x128). (c) Rate distortion performance for the image I_2

6 Conclusions

This chapter has presented the spherical camera model that can be used for processing visual information directly in its natural radial form. As images from catadioptric camera systems can be mapped directly on the sphere, we have presented image processing methods that are well adapted to process omnidirectional images. We have then discussed the framework of multi-camera systems, from a spherical imaging perspective. In particular, the epipolar geometry constraints and the disparity estimation problem have been discussed for the case of images lying on the 2-sphere. Finally, we have shown how sparse approximations of spherical images can be used to build geometrical correlation models, which lead to efficient distributed algorithms for the representation of 3D scenes.

7 Acknowledgments

This work has been partly funded by the Swiss National Science Foundation, under grant no 20001-107970/1.

The authors would like to thank Zafer Arican for providing the results on disparity estimation.

References

[1] E. H. Adelson and J. R. Bergen, *Computational Models of Visual Processing.* MIT Press, 1991, ch. The Plenoptic Function and the Elements of Early Vision.

[2] C. Zhang and T. Chen, “A Survey on image-based rendering - representation, sampling and compression,” *Signal Processing: Image Communication*, vol. 19, pp. 1–28, January 2004.

[3] Y. Yagi, “Omnidirectional sensing and its applications,” *IEICE Transactions On Information And Systems*, vol. E82D, no. 3, pp. 568–579, March 1999.

[4] C. Geyer and K. Daniilidis, “Catadioptric projective geometry,” *International Journal of Computer Vision*, vol. 45, no. 3, pp. 223 – 243, December 2001.

[5] S. Nayar, “Catadioptric omnidirectional camera,” in *IEEE Conf. on Computer Vision and Pattern Recognition*, Puerto Rico, June 1997, pp. 482–488.

[6] E. Hecht and A. Zajac, *Optics.* Addison-Wesley: Reading, MA, 1997.

[7] S. K. Nayar, “Omnidirectional vision,” in *Proc. ISRR1997*, Japan, 1997.

[8] S. Baker and S. K. Nayar, “A theory of single-viewpoint catadioptric image formation,” *International Journal of Computer Vision*, vol. 35, no. 2, pp. 1–22, 1999.

[9] T. Svoboda, T. Pajdla and V. Hlaváč, “Epipolar geometry for panoramic cameras,” in *5th European Conference on Computer Vision*, June 1998, pp. 218–231.

[10] S. Nene and S. K. Nayar, “Stereo with mirrors,” in *Sixth International Conference on Computer Vision*, January 1998, pp. 1087–1094.

[11] A. Torii, A. Imiya and N. Ohnishi, “Two- and three- view geometry for spherical cameras,” in *OMNIVIS 2005, The 6th Workshop on Omnidirectional Vision, Camera Networks and Non-classical cameras*, 2005.

[12] P. Schröder and W. Sweldens, “Spherical wavelets: efficiently representing functions on the sphere,” in *Proc. ACM SIGGRAPH*, 1995, pp. 161 – 172.

[13] J. R. Driscoll and D. M. Healy, “Computing fourier transform and convolutions on the 2-sphere,” *Advances in Applied Mathematics*, vol. 15, pp. 202–455, 1994.

[14] D. Healy Jr., D. Rockmore, P. Kostelec and S. Moore, “Ffts for the 2-sphere - improvements and variations,” *Journal of Fourier Analysis and Applications*, vol. 9, no. 3, pp. 341 – 385, 2003.

[15] S. Mallat, *A Wavelet Tour of Signal Processing.* Academic Press, 1998.

[16] P. J. Burt and E. H. Adelson, “The laplacian pyramid as a compact image code,” *IEEE Transactions on Communications*, vol. COM-31,4, pp. 532–540, 1983. [Online]. Available: citeseer.ist.psu.edu/burt83laplacian.html

[17] J. Antoine and P. Vandergheynst, “Wavelets on the n-sphere and related manifolds,” *Journal of Mathematical Physics*, vol. 39, no. 8, pp. 3987–4008, 1998.

[18] ——, “Wavelets on the 2-sphere : a group theoretical approach,” *Applied and Computational Harmonic Analysis*, vol. 7, pp. 1–30, November 1999.

[19] I. Bogdanova, P. Vandergheynst, J. Antoine, L. Jacques and M. Morvidone, “Discrete wavelet frames on the sphere,” in *In proceedings of EUSIPCO 2004, Vienne, Austria*, EUSIPCO. EUSIPCO, September 2004.

[20] ——, “Stereographic Wavelet Frames on the Sphere,” *Applied and Computational Harmonic Analysis*, vol. 19, no. 2, pp. 223–252, 2005.

[21] A. Makadia and K. Daniilidis, “Direct 3d-rotation estimation from spherical images via a generalized shift theorem,” in *Proc. IEEE CVPR*, 2003.

[22] C. Geyer and K. Daniilidis, “Paracatadioptric camera calibration,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 5, pp. 687–695, May 2002.

[23] X. Ying and Z. Hu, “Catadioptric camera calibration using geometric invariants,” *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 26, no. 10, pp. 1260 – 1271, Oct 2004.

[24] J. Barreto and H. Araujo, “Geometric properties of central catadioptric line images and their application in calibration,” *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 27, no. 8, pp. 1327 – 1333, Aug 2005.

[25] D. Scaramuzza, A. Martinelli and R. Siegwart, "A flexible technique for accurate omnidirectional camera calibration and structure from motion," in *Fourth IEEE International Conference on Computer Vision Systems (ICVS'06)*, 2006, p. 45.

[26] B. Mičušík and T. Pajdla, "Autocalibration & 3d reconstruction with non-central catadioptric cameras," in *Computer Vision and Pattern Recognition (CVPR)*, vol. 1, June 2004, pp. I–58 – I–65.

[27] S. B. Kang, "Catadioptric self-calibration," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'00)*, vol. 1, 2000, p. 1201.

[28] M. Antone and S. Teller, "Scalable extrinsic calibration of omni-directional image networks," *International Journal of Computer Vision*, Jan 2002.

[29] A. Makadia, C. Geyer, and K. Daniilidis, "Correspondenceless structure from motion," *International Journal of Computer Vision*, Jan 2007.

[30] P. Sturm, "Multi-view geometry for general camera models," *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, vol. 1, pp. 206–212 vol. 1, June 2005.

[31] Y. Ma, S. Soatto, J. Košeckà and S.S. Sastry, *An Invitation to 3-D Vision: From Images to Geometric Models.* Springer, 2004.

[32] J. Takiguchi, M. Yoshida, A. Takeya, J. Eino, and T. Hashizume, "High precision range estimation from an omnidirectional stereo system," *Intelligent Robots and System, 2002. IEEE/RSJ International Conference on*, vol. 1, pp. 263–268, 2002.

[33] J. Gonzalez-Barbosa and S. Lacroix, "Fast Dense Panoramic Stereovision," *Robotics and Automation, 2005. Proceedings of the 2005 IEEE International Conference on*, pp. 1210–1215, 2005.

[34] C. Geyer and K. Daniilidis, "Conformal Rectification of Omnidirectional Stereo Pairs," *Omnivis 2003: Workshop on Omnidirectional Vision and Camera Networks*, 2003.

[35] Z. Arican and P. Frossard, "Dense Disparity Estimation from Omnidirectional Images," in *Proceedings of the IEEE International Conference on Advanced Video and Signal based Surveillance*, 2007.

[36] S. Birchfield and C. Tomasi, "A pixel dissimilarity measure that is insensitive to image sampling," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 4, pp. 401–406, 1998.

[37] V. Kolmogorov and R. Zabih, "Multi-camera scene reconstruction via graph cuts," *European Conference on Computer Vision*, vol. 3, pp. 82–96, 2002.

[38] I. Tosic and P. Frossard, "Geometry-based distributed scene representation with omnidirectional vision sensors," *IEEE Trans. on Image Processing*, vol. 17, no. 7, pp. 1033–1046, July 2008.

[39] X. Zhu, A. Aaron and B. Girod, “Distributed compression for large camera arrays,” in *Proceedings of IEEE SSP*, September 2003.

[40] N. Gehrig, P. L. Dragotti, “Distributed Compression of Multi-View Images using a Geometrical Coding Approach,” in *Proceedings of IEEE ICIP*, vol. 6, September 2007, pp. VI –421–VI –424.

[41] A. D. Wyner and J. Ziv, “The rate-distortion function for source coding with side-information at the decoder,” *IEEE Trans. on Inform. Theory*, vol. 22, no. 1, pp. 1–10, January 1976.