Scale Space for Camera Invariant Features

Luis Puig\textsuperscript{1,2}, J.J. Guerrero
\textsuperscript{1}Departamento de Informática e Ingeniería de Sistemas
Instituto de Investigación en Ingeniería de Aragón, I3A Universidad de Zaragoza, Spain
Kostas Daniilidis
\textsuperscript{2}GRASP Laboratory, University of Pennsylvania, 3330 Walnut Street, L402, Philadelphia

Abstract—In this paper we propose a new approach to compute the scale space of any central projection system, such as catadioptric, fisheye or conventional cameras. Since these systems can be explained using a unified model, the single parameter that defines each type of system is used to automatically compute the corresponding Riemannian metric. This metric, is combined with the partial differential equations framework on manifolds, allows us to compute the Laplace-Beltrami (LB) operator, enabling the computation of the scale space of any central projection system. Scale space is essential for the intrinsic scale selection and neighborhood description in features like SIFT. We perform experiments with synthetic and real images to validate the generalization of our approach to any central projection system. We compare our approach with the best-existing methods showing competitive results in all type of cameras: catadioptric, fisheye, and perspective.

Index Terms—central projection systems, image smoothing, linear diffusion equation, Laplace-Beltrami operator, scale space.

I. INTRODUCTION

The interest in omnidirectional cameras has increased considerably in recent years. There exist several types of omnidirectional cameras, which can be classified as central or non-central. In this paper, we focus on central omnidirectional cameras and slightly non-central cameras, or catadioptric and fisheye systems, respectively. Catadioptric systems utilize conventional cameras and mirrors, while the fisheye systems use conventional cameras and wide angle lenses. Baker and Nayar [3] demonstrated which catadioptric systems satisfied the central property, among these, the two most popular are the paracatadioptric and the hypercatadioptric systems. These systems use either a paraboloidal shaped mirror coupled with an orthographic camera or a hyperboloidal shaped mirror coupled with a perspective camera. Geyer and Daniilidis [13] proposed a unified model, which encapsulates elliptic, parabolic and hyperbolic mirrors as well as conventional cameras under a single representation. Later, an extension of this model for the fisheye systems was proposed in [32], and the relation between fisheye models and the unified model is presented in [8].

At the same time, the development of feature detection and feature description algorithms have advanced for conventional images. A very important property of features is the ability to be scale-invariant. The earliest work on linear scale-space was developed by Ijima [30] but his work was unfortunately not known around the world. Witkin [31] proposed the idea of scale-space representation by convolving the original image with Gaussians of increasing width. Then Lindeberg studied the linear discrete scale-space in images [17], focusing on detecting features like blobs, edges, and rinses with automatic scale selection [18], [19]. Inspired by this work Lowe developed SIFT [23], which approximates the Laplacian of Gaussians through the difference of Gaussian method [9], from which the extrema are detected. Similarly, Mikolajczyk and Schmid [26] use the trace of the Hessian matrix, which is equal to the Laplacian-of-Gaussians, and the determinant of this matrix to detect more robust scale invariant features but with a higher computational cost than SIFT. Other scale invariant feature detectors are based only on the gray scale intensity and do not require the construction of scale spaces. The maximal stable regions proposed in [24] are invariant to affine transformations of image coordinates and to affine transformations of intensity. In [16], Kadir and Brady present a probabilistic approach, which considers saliency, scale selection and content description as a non-separable task. The SIFT approach has inspired various works attempting to replicate its results on different imagery systems, particularly on omnidirectional cameras. In this paper we compute the Laplace-Beltrami (LB) operator, which is a second order differential operator required to perform the Gaussian smoothing on omnidirectional and conventional images. We combine the unified model proposed in [13] with the partial differential equations framework on manifolds [6] to smooth any image acquired with a central projection system.

A. Related Work

In recent years much of the research has focused on extending the results from conventional images to omnidirectional images. Daniilidis et al. [11] presented the computation of optical flow in omnidirectional images. They state that the signal processing of catadioptric images has to take place in the originally sampled values. They proposed the use of the unitary sphere as an auxiliary non-deformed domain where the image operations for optical flow are written and then applied to the image plane through variable substitution. Although it avoids the mapping from the image to the sphere, it results in an adaptive Gaussian kernel that changes at every pixel position, making its implementation costly. However, the idea of using the unitary sphere as a representation for catadioptric images is relevant and even more so if we consider that any

This work has been supported by project DPI2012-31781
central projection system can be represented easily under the unified spherical camera model [13]. In [7] Bulow makes use of the spherical heat diffusion equation to represent the scale-space for functions defined on the sphere. Based on this idea Hansen et al. [15] studied the generalization of the SIFT algorithm for images on the sphere domain. The formulation of SIFT on the sphere enables full invariance to pure camera rotation motion and implicitly handles perspective deformations. However, undesired aliasing effects are introduced. More recently [14] and [10] propose different strategies to deal with these problems. In [14] a new mapping from the sphere to an intermediate stereographic projection is proposed to approximate the spherical diffusion. This method maps both the omnidirectional image and the spherical Gaussian function to the stereographic image plane. Following this step, the spherical diffusion is approximated by convolving the stereographic image with the stereographic version of the Gaussian kernel. Although this approach is not limited by aliasing problems, it requires the extra resampling of the spherical image to the stereographic one. In [10], the aliasing caused by the down sampling of the images is avoided by the use of an anti-aliasing criterion, which decides whether an image is down-sampled or not. The computation of the spherical Fourier transform and its approximation through a stereographic projection are computationally costly, making these methods slower than those approaches using the information from the original image.

A different approach that enhances the classic SIFT, enabling it to deal with images with radial distortion is presented by Lourenço et al. [20], [22]. They developed an adaptive filter that accounts for radial distortion at each image pixel. For computational efficiency, this filter is simplified as a separable Gaussian kernel, whose standard deviation varies with the image radius, adapting to the radial distortion present in that position. This approach makes use of the division model to explain the radial distortion, making it less generalizable than other approaches, although, still capable of handling images with high distortion such as fisheye and catadioptric images [21], by developing a particular adaptive kernel for each camera system.

Following the idea that the signal processing of omnidirectional images should be performed on the original image values, Bogdanova et al. [6] presented a new approach to solve the diffusion (heat) equation on the sphere using the original image data. They applied the partial differential equations (PDE) framework on Riemannian manifolds to generate a Laplace-Beltrami operator suitable for the smoothing of the catadioptric images. This operator computes weights for each pixel of the first and second derivatives of the image. Similarly to [22], these weights depend on the image radius and the projection model used to represent the omnidirectional image. This information is encoded on the corresponding Riemannian metric. Notice that this approach naturally avoids the bandwidth selection and aliasing problems since no warping of the image to the sphere is required. Inspired by this work Arican and Frossard [1], [2] presented a scale invariant feature detector for paracatadioptric images, which similar to [6] can be generalized to any projection system, as long as the corresponding Riemannian metric is provided.

These two last works were the inspiration for our previous contribution, in which we were able to generalize the computation of the Riemannian metric for central catadioptric systems [28]. In that paper we presented a generic approach to automatically compute the corresponding Riemannian metric for any catadioptric system described by the unified spherical camera model [13], [4], in contrast to [2], where the Riemannian metric has to be provided by the user. In this current paper we extend our previous work [28] to include perspective and fisheye cameras. Perspective cameras are already explained under the spherical camera model, while the relation of fisheye cameras with the spherical camera model is shown in [8]. This result endows all existing algorithms for catadioptric cameras to be used directly on fisheye systems. This extension to the spherical camera model combined with the partial differential equations on manifolds framework, enables the automatic computation of the corresponding Riemannian metric for any central projection system and fisheye cameras. With this metric we calculate the corresponding Laplace-Beltrami operator, which is used to perform the Gaussian smoothing directly on images acquired with central projection systems and fisheye cameras. With the smoothed images the construction of the scale-space for any central projection system is straightforward, allowing for the computation of SIFT-like features for any type of camera.

The rest of the paper is organized as follows. In section II, we present the unified camera model, which considers all central projection systems and fisheye cameras. In section III, the computation of the Laplace-Beltrami operator for a generic projection system is developed. Section V describes experiments computing the scale space for different projection systems using synthetic and real images. Finally, in section VI, we summarize the results of this paper.

II. GENERIC PROJECTION MODEL

In this section we present the spherical camera model, which considers all central catadioptric systems [13] and the extension to this model introduced by Ying and Hu [32]. This extension allow us to calibrate fisheye systems using the methods originally designed for central catadioptric systems.

A. Central Catadioptric Model

According to [13], [4], all central catadioptric cameras can be modeled by a unit sphere and a perspective projection, such that the projection of 3D points can be performed in two steps (Fig. 1). First, one projects the 3D point \( \mathbf{X} = (X, Y, Z) \) onto the unit sphere, obtaining the intersection of the sphere and the line joining its center and the 3D point. There are two such intersection points, which are represented as \( s_\pm \). These points are then projected in the second step, using a perspective projection \( M \) resulting in two image points \( x_\pm \), one of which is physically true.

This model covers all central catadioptric cameras, encoded by \( \xi \), which is the distance between the perspective projection and the center of the sphere. We have \( \xi = 0 \) for perspective, \( \xi = 1 \) for paracatadioptric and \( 0 < \xi < 1 \) for hypercatadioptric...
A point on the unitary sphere \( s = (X, Y, Z) \) can also be represented by two angles \((\theta, \varphi)\) (Fig. 2a). Depending on the location of the projection point (mirror parameter \( \xi \)) we have different radial projection functions [12]. In the paracatadioptric projection case (\( \xi = 1 \)), which coincides with the inverse stereographic projection (Fig. 2c), its radial projection function is \( 2 \tan \left( \frac{\theta}{2} \right) \). In the perspective case where \( \xi = 0 \) the corresponding projection function is \( \tan(\theta) \). In the hypercatadioptric case (\( 0 < \xi < 1 \)) the radial projection function becomes \( (1 + \xi) \frac{\sin \theta}{\xi - \cos \theta} \) (Fig. 2d). This function will be useful later to generalize the computing of the metric for any mirror shape in central catadioptric systems.

C. Relation between Spherical Camera Model and Projections on the Sphere

In this section we integrate the spherical camera model (section II) that models any central projection system and the techniques developed to compute the differential operators on non-Euclidean manifolds [5] such as the mirror surfaces present in catadioptric systems.

III. Scale Space of Central Projection Systems

In this section we integrate the spherical camera model (section II) that models any central projection system and the techniques developed to compute the differential operators on non-Euclidean manifolds [5] such as the mirror surfaces present in catadioptric systems.

A. Differential Operators on Riemannian Manifolds

The scale space representation \( I(x, y, t) \) is computed using the heat diffusion equation and differential operators on non-Euclidean manifolds. It is defined as

\[
\frac{\partial I(x, y, t)}{\partial t} = \Delta I(x, y, t)
\]

where \( \Delta \) is the Laplacian-Beltrami operator and \( t \) is the scale level. The initial condition is \( I(x, y, t_0) = I(x, y) \) where \( I(x, y) \) is the original image.

We briefly define the differential operators on the manifolds which make possible the computation of the Laplace-Beltrami operator. Let \( M \) be a parametric surface on \( \mathbb{R}^3 \) with an induced Riemannian metric \( g_{ij} \) that encodes the geometrical properties of the manifold.

In a local system of coordinates \( x^i \) on \( M \), the components of the gradient read

\[
\nabla^i = g^{ij} \frac{\partial}{\partial x^j}
\]

where \( g^{ij} \) is the inverse of \( g_{ij} \) [6]. A similar reasoning is used to obtain the expression of the divergence of a vector field \( X \) on \( M \)

\[
\text{div} X = \frac{1}{\sqrt{g}} \partial_i (X^i \sqrt{g}),
\]

where \( g \) is the determinant of \( g_{ij} \). Finally, combining these two operators we obtain the Laplace-Beltrami operator, which is the second order differential operator defined on scalar fields on \( M \) by

\[
\Delta I = -\frac{1}{\sqrt{g}} \partial_i (\sqrt{g} g^{ij} \partial_i I).
\]

B. Computing a Generic Metric on the Sphere

As explained in section II omnidirectional images are formed in two steps. The first one projects a 3D point to the unitary sphere. Then this point is projected from the unitary sphere to the image plane through a variable projection point, which is determined by the geometry of the mirror (parameter \( \xi \)). If the system is calibrated [27], it is also possible like in any conventional camera, to map the catadioptric/fisheye image to the unitary sphere.

In [6], [2], the mapping from paracatadioptric images to the sphere is used for the computation of the differential operators explained before. This allows to process the spherical image directly using the image coordinates. In this paper we extend that approach to all central projection systems.

Consider the unitary sphere \( S^2 \) (Fig. 2a). A point on \( S^2 \) can be represented in Cartesian and polar coordinates as

\[
(X, Y, Z) = (\sin \theta \cos \varphi, \sin \theta \sin \varphi, \cos \theta)
\]

The Euclidean element in Cartesian and polar coordinates is defined as

\[
dl^2 = dX^2 + dY^2 + dZ^2 = d\theta^2 + \sin^2 \theta d\varphi^2
\]
Under the spherical camera model, a point on the sphere \((\theta, \varphi)\), is mapped to a point in polar coordinates \((R, \varphi)\) in the image plane. The \(\theta\) angle depends on the central catadioptric system we are dealing with, while \(\varphi\) remain the same. For example, a conventional perspective system is described with \(\theta = \arctan(R)\) and a paracatadioptric system with \(\theta = 2 \arctan \left(\frac{R}{2}\right)\). In the general case (see Fig. 2d) we have

\[
\theta = \arctan \left( \frac{R + \sqrt{(1+\xi)^2 - R^2(\xi^2 - 1)}}{1 + R^2 \xi + \sqrt{(1+\xi)^2 - R^2(\xi^2 - 1)}} \right)
\]  

(7)

In terms of these new coordinates the metric becomes

\[
d\ell^2 = \frac{\left(\xi + \xi^2 + \sqrt{(1+\xi)^2 - R^2(\xi^2 - 1)}\right)^2}{(R^2 + (1+\xi)^2)^2} \left( R^2 d\varphi^2 + \frac{(1+\xi) R^2}{1 - R^2(\xi - 1)} d\xi \right)
\]  

(8)

Let \((x, y) \in \mathbb{R}^2\) on the sensor plane define Cartesian coordinates, where \(R^2 = x^2 + y^2\) and \(\varphi = \arctan \left(\frac{y}{x}\right)\)

\[
d\ell^2 = \lambda \left(4(ydx - xdy)^2 - \frac{4(1+\xi)(dx^2 + dy^2)}{(x^2 + y^2)(\xi - 1)}\right)
\]  

(9)

where

\[
\lambda = \frac{\left(\xi + \xi^2 + \sqrt{(1+\xi)^2 - (x^2 + y^2)(\xi^2 - 1)}\right)^2}{4(x^2 + y^2)(x^2 + y^2 + (1+\xi)^2)}
\]  

To verify the correctness of this approach we substitute \(\xi\) to the values for which we already know the metric. For \(\xi = 1\) the result is the same as that one presented in [2]. The case of \(\xi = 0\) that corresponds to the perspective, however we obtain the following instead of the Cartesian metric

\[
d\ell^2 = \frac{(1 + y^2)dx^2 - 2xydxdy + (1 + x^2)dy^2}{(1 + x^2 + y^2)^2}
\]  

(10)

This is expected since there is no map from the sphere to the plane being both conformal and area-preserving. However, from the generic metric we have already computed, we are able to compute the Laplace-Beltrami operator. This result is particularly relevant, since it has been shown that the perspective projection only works well for narrow-angle images and where the lens distortion is minimal. When the field of view increases other projection models have to be considered [12]. Our intuition indicates that this metric might consider those deformations in the image plane introduced when the field of view increases, where by object deformations we mean the one introduced by the mapping from the sphere to the image plane and not the distortion produced by the lens construction.

C. Generic Laplace-Beltrami Operator

From (9) we can compute the generic metric in matrix form \(g^{ij}\) and its corresponding inverse matrix \(g^{ij}\)

\[
g^{ij} = \gamma_2 \begin{pmatrix}
-x^2(\xi - 1) + \xi + 1 & xy(\xi - 1) \\
xy(\xi - 1) & -y^2(\xi - 1) + \xi + 1
\end{pmatrix}
\]  

(11)

with

\[
\gamma_2 = \left(\frac{x^2 + y^2 + (1+\xi)^2}{(1+\xi)(\xi + \xi^2 + \sqrt{(1+\xi)^2 - (x^2 + y^2)(\xi^2 - 1)} + 2\xi + \xi^2)}\right)^2
\]  

The determinant of (11) is

\[
\det(g^{ij}) = -\frac{(x^2 + (1+\xi)^2 + y^2)^3((x^2 + y^2)(\xi - 1) - \xi - 1)}{(1+\xi)(\xi + \xi^2 + \sqrt{(1+\xi)^2 - (x^2 + y^2)(\xi^2 - 1)} + 2\xi + \xi^2)^3}
\]  

(12)
With all these elements we are now able to compute the Laplace-Beltrami operator (4) which is represented by the differential operators

\[
\Delta I = \lambda_1 \frac{\partial^2 I}{\partial x^2} + \lambda_2 \frac{\partial^2 I}{\partial y^2} + \lambda_3 \frac{\partial^2 I}{\partial xy} + \lambda_4 \frac{\partial I}{\partial x} + \lambda_5 \frac{\partial I}{\partial y}
\]

\[
\lambda_1 = \frac{(x^2(x^2-1)-1)(1+x^2+y^2)^2}{-1-x^2(x^2-1)(x^2+y^2)}
\]

\[
\lambda_2 = \frac{(y^2(x^2-1)-1)(1+x^2+y^2)^2}{-1-x^2(x^2-1)(x^2+y^2)}
\]

\[
\lambda_3 = \frac{2xy(x^2-1)(1+x^2+y^2)^2}{-1-x^2(x^2-1)(x^2+y^2)}
\]

\[
\lambda_4 = \frac{4x(1+x^2+y^2)(x^2-2(x^2-1)(x^2+y^2)-\xi\sqrt{1-(x^2-1)(x^2+y^2)})}{-1-x^2(x^2-1)(x^2+y^2)}
\]

\[
\lambda_5 = \frac{4y(1+x^2+y^2)(x^2-2(x^2-1)(x^2+y^2)-\xi\sqrt{1-(x^2-1)(x^2+y^2)})}{-1-x^2(x^2-1)(x^2+y^2)}
\]

An analogous process has been performed for the computation of the metric in the normalized plane, but the equations are too big to fit in these pages. The code for these equations can be checked on the author’s website. In the rest of the paper we use the metric computed on the sphere.

IV. SMOOTHING IMAGES USING THE GENERIC LAPLACE-BELTRAMI OPERATOR

We compute the smoothing of the catadioptric images using the heat diffusion equation. The differentiation with respect to time in the heat diffusion equation is discretized with time intervals, \( d_t \). To compute the discrete differential representations of the image \( \frac{\partial I}{\partial x}, \frac{\partial I}{\partial y}, \frac{\partial^2 I}{\partial x \partial y}, \frac{\partial^2 I}{\partial x^2}, \frac{\partial^2 I}{\partial y^2} \) and \( \frac{\partial I}{\partial y} \) we convolve the catadioptric image with different kernels

\[
I_x = \begin{bmatrix} -1 & 1 \end{bmatrix}, \quad I_y = \begin{bmatrix} -1 & 0 \end{bmatrix}, \quad I_{xx} = \begin{bmatrix} 1 & -2 & 1 \end{bmatrix}, \quad I_{yy} = \begin{bmatrix} -2 & 0 \end{bmatrix}, \quad I_{xy} = \begin{bmatrix} 0 & 1 & 0 \end{bmatrix}
\]

\[
\begin{equation}
\begin{aligned}
& I_x = [ -1 \ 1 ] \quad I_y = \begin{bmatrix} -1 \\ 1 \end{bmatrix} \quad I_{xx} = [ 1 \ -2 \ 1 ] \\
& I_{yy} = \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix} \quad I_{xy} = \begin{bmatrix} 0 & 1 & 0 \end{bmatrix}
\end{aligned}
\end{equation}
\]

For a particular catadioptric system defined by \( \xi \) we compute the corresponding coefficients (14-18) which multiply their corresponding convolved image. Then we compute the specific Laplace-Beltrami operator. Finally, smoothing is performed by updating \( I(x, y, t) \) with the differences that have been computed at previous time steps. We summarize the process in the following algorithm.

\[
\begin{align*}
\text{Input} & : I(x, y), t, d_t, \xi, I_x, I_y, I_{xy}, I_{xx}, I_{yy} \\
\text{Output} & : I(x, y, t)
\end{align*}
\]

\[
1 \text{ /*Compute } \lambda_i \text{ based on the Riemannian metric*/} \\
2 \text{ [cIc, cIy, cIxx, cIyy]} \leftarrow \text{ CDCoeff } (I(x, y), \xi) \\
3 \text{ ntimes } \leftarrow t^2/d_t \\
4 I(x, y, t) \leftarrow I(x, y) \\
5 \text{ for } i \leftarrow 1 \text{ to ntimes do} \\
6 \text{ ImIx } \leftarrow \text{ cxI } \ast \text{ Convolve } (I(x, y, t), Ix) \\
7 \text{ ImIy } \leftarrow \text{ cIy } \ast \text{ Convolve } (I(x, y, t), Iy) \\
8 \text{ ImIxy } \leftarrow \text{ cIxy } \ast \text{ Convolve } (I(x, y, t), Ixy) \\
9 \text{ ImIxx } \leftarrow \text{ cIxx } \ast \text{ Convolve } (I(x, y, t), Ixx) \\
10 \text{ ImIy } \leftarrow \text{ cIy } \ast \text{ Convolve } (I(x, y, t), Iy) \\
11 \text{ LBO } \leftarrow \text{ ImIx } + \text{ ImIy } + \text{ ImIxy } + \text{ ImIxx } + \text{ ImIy} \\
12 I(x, y, t) \leftarrow I(x, y, t) + d_t \ast \text{ LBO}
\end{align*}
\]

Algorithm 1: Iterative algorithm to smooth omnidirectional images using the Laplace-Beltrami operator.

V. EXPERIMENTS

In this section we perform several experiments using synthetic and real images. In order to have ground truth synthetic catadioptric and fisheye images are generated using the ray tracing software POV-ray\(^2\). Two real hypercatadioptric systems with mirror parameters of \( \xi = 0.9662 \) (m1) and \( \xi = 0.70544 \) (m2) and a fisheye lens with \( \xi = 1.78 \) are modeled. The real images are acquired using the hypercatadioptric system designed by Neovision.

A. Smoothing Images using Laplace-Beltrami operator

In order to verify the validity of our approach we perform an experiment where we compare the generic LB operator computed using our approach to the Cartesian Laplace operator. To obtain this operator we smooth the omnidirectional image to the same scale \( t \), using the corresponding Gaussian kernel. Then, we compute the Laplacian of this image using the classical Cartesian operator. We select the hypercatadioptric (m1) image for which we compute scales \( t = \{1, 3.5 \} \) and their corresponding generic LB and Cartesian Laplacian. In Fig. 3 we show the results. We observe that the generic LB operator considers a difference between the pixels close to the center and those on the periphery, while the Cartesian Laplacian has the same effect on all pixels, without taking into account their location in the image.

To quantify the last statement we sum the values of the generic LB operator inside circles with different radii and origin at the image center. We perform the same procedure with the Cartesian Laplacian. The last row of Fig. 3 shows the comparison of these sums for the different scales. We observe that the pixels in the periphery have smaller values in the generic Laplace-Beltrami operator than those on the Cartesian Laplace operator. In the center where the resolution

\[1\text{ http://webdiis.unizar.es/~lpuig/Mathematica_code/}
\[2\text{ http://www.povray.org}
\[3\text{ http://www.neovision.cz}
\[4\text{ http://www.accowle.com}

is performed, the size of the Gaussian kernel is usually defined as two times the covariance $\sigma$. In the Laplace-Beltrami approach we define the smoothing level to be applied to the image with the scale $t$. Similar to the Cartesian case we have observed that the scale factor $t$ requires an extra factor we called $\tau$. This factor defines the final number of times we have to iteratively apply the Laplace-Beltrami operator over the image to obtain the correctly smoothed image. To evaluate this factor we conduct the following experiment. We generate a synthetic scene with spheres of the same size and which are located at the same distance from the mirror (Fig. 4a). We use the m1 catadioptric system to generate the corresponding omnidirectional image (Fig. 4b). For this image we compute a different scale space for each $\tau$ factor.

We observed that using $t$, the scale space extrema were located mostly in the borders of the circles (Fig. 5b), in opposition to the extrema obtained with the Cartesian approach (Fig. 5a). We tried different $\tau$ factors in order to get the correct scale of the circles present in the image. With $\tau = 2$ we obtain the results shown in Fig. 5c. We repeat the same experiment for different central projection systems to select the factor giving the best fitting. We observe that there is no direct relation between the mirror parameter $\xi$ and the best $\tau$ factor, which values range is $[1.2 - 2]$. In the following experiments we use the corresponding $\tau$ factor to the projection system we are dealing with.

B. Impact of $\xi$ parameter on the scale space computation

In this experiment we analyze the impact of the parameter $\xi$ on different types of images. As mentioned before, there is a $\xi$ value which describes a particular projection system under the unified spherical camera model. Perspective ($\xi = 0$), hypercatadioptric ($0 < \xi < 1$), paracatadioptric ($\xi = 1$) and fisheye ($\xi > 1$) are the type of cameras considered under this model. This parameter can be obtained either from a calibration process or directly from the manufacturer. When a calibration process is used the estimated value of $\xi$ can be inaccurate. In this experiment we vary the mirror parameter $\xi$ around the correct value and observe its influence on the number of extrema detected, as well as their repeatability ratio with respect to an image obtained by rotating the camera thirty degrees. We perform this evaluation on the four types of images mentioned previously.

1) Hypercatadioptric images: In this case the correct mirror value is $\xi = 0.7054$, which is obtained from the manufacturer. We use the set of values $\xi = \{0.5, 0.6, 0.7054, 0.8, 0.9\}$. In Fig. 7a we observe the repeatability ratio on the left y-axis and the number of extrema on the right y-axis. We observe that as the mirror parameter $\xi$ increases the number of detected extrema also increases. Despite this positive effect we observe that the improvement of the repeatability ratio is minimal.

2) Paracatadioptric images: In this case the correct mirror value is $\xi = 1$ and corresponds to an stereographic projection. We vary the mirror parameter in the interval $\xi = \{0.8, 0.9, 1.0, 1.1, 1.2\}$. The results for this experiment are shown in Fig. 7b. We observe the same behavior as in
the hypercatadioptric case. The number of detected extrema increases but the repeatability ratio slightly improves.

3) Fisheye images: In this experiment we use a fisheye image with mirror parameter $\xi = 1.78$ obtained from calibration. The set of values used are $\xi = \{1.6, 1.7, 1.78, 1.9, 2.0\}$. In this case we observe a bigger increment in the number of detected extrema than in previous cases. In this case we observe that the highest repeatability ratio is obtained with a smaller $\xi$ parameter than the calibrated and with a smaller number of detected extrema. But again this improvement is minimal.

4) Conventional images: For this type of images which corresponding mirror parameter is $\xi = 0$ we use only two more values, $\xi = 0.1, 0.2$, since negative values have no meaning on the spherical camera model. Similar to wide angle cameras we observe an increment on the number of detected extrema but the repeatability ratio in this case showed no improvement.

As a general conclusion we observe that the perturbations on the mirror parameter $\xi$ have a small impact on the repeatability ratio of the analyzed systems. An interesting result is the fisheye case, which has the biggest mirror parameter $\xi$, where the number of detected extrema increased considerably with respect to other systems.

C. Repeatability and Matching Experiment

In this experiment we test the repeatability of the extrema points detected with the scale space computed using our approach. We compute a scale space composed of four octaves and four scales per octave. We take as reference the scales given by Vedaldi’s implementation [29] in order to generate equivalent representations of the scale space. The initial sigma is defined as $t_0 = 1$ and the smoothing interval is defined as $k = 2$. We compute the extrema from the approximation to the scale-normalized Laplacian of Gaussian through differences of Gaussians, similar to [23]. We compare our approach with the following methods:

- SIFT [23]. This is the classical SIFT method which assumes as input a conventional image with homogeneous resolution. It computes the scale space using differences of Gaussians as an approximation to the Laplacian of Gaussians. This approach is designed for conventional images and it does not require calibration parameters. It can also be applied to other type of images.

- sRDSIFT [22], [21]. This is an extension to SIFT that considers images with distortion. It adapts the smoothing kernel at each pixel position to consider its corresponding distortion level, which depends on the pixel image radius. This approach requires an estimation of the distortion level ($\psi$) which can be computed from a single calibration image$^5$. This approach can also be used on images with high distortion, e.g., fisheye and catadioptric images.

$^5$http://arthronav.isr.uc.pt/easycamcalib/
for which a corresponding adaptive kernel should be computed. In the case of catadioptric systems the full calibration of the system is required.

- **sSIFT** [10]. This approach maps the original image into the sphere where the smoothing is performed through a spherical Gaussian filter. This type of approaches suffer from bandwidth and aliasing problems [15]. In this particular implementation an anti-aliasing criterion is considered. The performance of this approach depends on the size of the spherical image, i.e., the sampling of the sphere, which is defined by the user. This approach can deal with any type of image that can be mapped to the sphere.

- **pSIFT** [14]. In order to avoid the aliasing and bandwidth problems of previous approaches, this method approximates the diffusion on the sphere by using a stereographic projection. It maps the input image to a stereographic image via the sphere, as well as the spherical Gaussian kernel. This kernel is simplified by considering only its middle row and column. Then the stereographic image is convolved with this kernel. However, this approximation introduces different artifacts. This approach can deal with any type of central projection system that can be represented on the sphere. Therefore this approach requires the full calibration of the system. The parameters required are the focal length $\phi$ and the mirror parameter $\xi$.

Since not all the approaches are valid for all type of cameras, they are applied where possible.

**Computation time.** An important aspect with most of these approaches is the computation time. In table I we show the computation times of a scale space of four octaves and four scales per octave for each approach of a fisheye image of resolution $768 \times 768$ pixels. We also show the programming language used in each approach. In the case of pSIFT we show the offline time which is used to compute the approximation of the spherical kernels through the stereographic projection, which takes more than five minutes, while the online process, the detection of extrema takes five seconds. The fastest approach is SIFT with less than a second to obtain features and descriptors. sRDSIFT is the second fastest, it spends extra time in adapting the Gaussian kernel with respect to the image center to consider radial distortion at that radius. sSIFT takes more than one minute to process a single image, considering only the detection of extrema. Our approach Laplace-Beltrami takes almost nine seconds to compute the scale space and obtain the extrema from the image. This is because our approach is iterative and the number of iterations depends on the scale factor $t$ and as we observed before this value can increase due to the iterator factor $\tau$.

**Repeatability criteria.** Since the transformations between the images are known we map the extrema detected in the original image to the other images. The repeatability criteria is the following

$$||\hat{x} - \bar{x}|| \leq \delta_d$$  \hspace{1cm} (20)

where, $\hat{x}$ is the mapped point. We use the Euclidean distance in the image. The distance threshold $\delta_d$ has to be adapted to capture the matches with different scale levels and computed as $\delta_d = \delta_0 \cdot \xi_k$ with $\delta_0$ the distance threshold parameter. The matching condition on the scale ratio is

$$\frac{\min(t_1, t_2)}{\max(t_1, t_2)} \geq \delta_s.$$  \hspace{1cm} (21)

In all the experiments we set $\delta_d = 3$ pixels and $\delta_s = 0.85$, unless otherwise stated. In Tables II-VII we present the number of correct mapped points (in bold) and the extrema detected in parenthesis for each experiment.

**Matching criteria.** We also automatically match catadioptric, fisheye and perspective images. We use the corresponding descriptors of each approach. In the case of the Laplace-Beltrami operator we use the SIFT descriptor from Vedaldi’s implementation [29]. We measure the performance of the matching using recall versus 1-precision which is defined as

$$\text{recall} = \frac{\text{correct matches}}{\text{ground truth matches}} \quad \text{1-precision} = \frac{\text{false matches}}{\text{all matches}}.$$  \hspace{1cm}

We vary the ambiguity threshold from 1 to 2 (in Vedaldi’s implementation). We average the results taking into account all the images in the corresponding experiment.

1) **Catadioptric System with Synthetic Images**: In this experiment we perform the repeatability experiment using the central catadioptric system $m2$ with an iteration factor $\tau = 1.5$. From an initial position the camera is rotated ten, twenty and thirty degrees in the vertical axis where the images are acquired. First row of Fig. 6 shows these images. pSIFT is the approach that detects more extrema, but they are mostly
clustered on image edges (Fig. 8(b)). SIFT, sSIFT and LB approaches detect a similar number of features but their distribution over the image is different. These extrema are shown in Fig. 8. sSIFT extrema also show some clusters but only at the image center. LB and SIFT do not show evident clusters. It should be noticed that SIFT detects a considerable number of extrema in areas with poor texture. The results of the repeatability experiment are presented in Fig. 9a. In Table II we show the number of detected extrema detected by each approach. The best performance is shown by pSIFT and LB approaches with similar number of correct correspondences. pSIFT shows also a good relative performance but this is due to the high number of clustered extrema, each cluster provides several correct matches. Although SIFT has the worst performance among the approaches, it should be observed that it still has a considerable number of correct correspondences, higher than 40%.

In the automatic matching experiment, we observe that our approach shows an improvement over the rest of the approaches considering recall. SIFT comes second with similar results. This is explained since both approaches use the same descriptor and that SIFT has a lower number of ground truth matches than sSIFT and pSIFT (see Table II). sSIFT comes next with a higher number of false matches and also a lower number of correct matches. pSIFT has the lowest number of incorrect matches but also a smaller amount of correct matches than the other approaches.

2) Fisheye Images: In this experiment we use the fisheye projection given by PovRay with an equivalent mirror parameter of $\xi = 1.78$. The iteration factor in this case is $\tau = 1.5$ (which decreases the number of extrema detected). Second row of Fig. 6 shows the images used in this experiment. In Fig. 11 we observe the distribution of extrema detected by each approach. Similar to the catadioptric case, we observe that SIFT (Fig. 11a) detects features all over the image. pSIFT (Fig. 11b), is again the approach that detects more extrema, which are clustered and located on edges. sSIFT (Fig. 11c), which is not designed to deal with this type of images detects the smallest number of extrema. The extrema detected by sRDSIFT are mostly located on the image center and some of them clustered around corners (see Fig. 11d). Our approach shown in Fig. 11e detects extrema all over the image and as

<table>
<thead>
<tr>
<th>Reference</th>
<th>Rot 10°</th>
<th>Rot 20°</th>
<th>Rot 30°</th>
</tr>
</thead>
<tbody>
<tr>
<td>SIFT</td>
<td>1215</td>
<td>543 (1169)</td>
<td>506 (1124)</td>
</tr>
<tr>
<td>pSIFT</td>
<td>2015</td>
<td>1131 (1892)</td>
<td>979 (1756)</td>
</tr>
<tr>
<td>sSIFT</td>
<td>1177</td>
<td>709 (1098)</td>
<td>633 (1048)</td>
</tr>
<tr>
<td>LB</td>
<td>1186</td>
<td>795 (1194)</td>
<td>630 (1096)</td>
</tr>
</tbody>
</table>
SIFT and pSIFT around corners.

In Fig. 9b we observe the results of the repeatability experiment. sRDSIFT is the approach with the best repeatability score, followed closely by our approach, while surprisingly sSIFT improves pSIFT in two cases. As expected, conventional SIFT has the lowest repeatability ratio. In Table III we present the number of extrema detected and the number of correctly mapped points.

The results of the automatic matching experiment are shown in the second row of Fig. 9b. We observe that with fisheye images all approaches had less false matches. The lower number of ground truth matches of SIFT makes it to have a better behavior in this experiment. Our approach and sRDSIFT show similar results, our approach with less number of false matches and sRDSIFT with a slightly higher recall. SIFT shows similar and even higher recall values than sRDSIFT but with a much larger number of false matches. pSIFT shows a slightly inferior performance. The advantage of sRDSIFT over our approach is due to that the spherical camera model is only an approximation to fisheye lenses.

3) Conventional Images: In this experiment we use two conventional synthetic images, one rotated with respect to the other by thirty degrees around the vertical axis. We also scaled these two images to test the scale invariance of the approaches. These images are shown in the last row of Fig. 6. We omit sSIFT in this experiment since the repeatability rate was very low due to the inaccurate mapping from the sphere to the image plane. In Table IV we observe the number of extrema detected by each approach. As in the previous cases, pSIFT detects a lot of extrema but obtains a low repeatability ratio (see Fig. 9c). In this particular case, the low performance of this approach might be caused by the different steps involved in the extrema computation, which include a mapping from the original image representation (plane) to a stereographic representation, where the extrema are detected and refined by interpolation, then these extrema and their scales are mapped back to the original plane representation. In this type of images SIFT and sRDSIFT must have a similar performance, since no distortion ($\psi = 0$) is considered. This is verified by this experiment. Surprisingly, sRDSIFT performs better on the rotation+scale experiment, but this is explained due to its lower number of features detected. Our approach obtains a repeatability ratio lower than SIFT and sRDSIFT but with the advantage of a higher number of correctly mapped extrema. In the automatic matching experiment, SIFT and sRDSIFT must have a similar performance, since no distortion ($\psi = 0$) is considered. This is verified by this experiment. Surprisingly, sRDSIFT performs better on the rotation+scale experiment, but this is explained due to its lower number of features detected. Our approach obtains a repeatability ratio lower than SIFT and sRDSIFT but with the advantage of a higher number of correctly mapped extrema. In the automatic matching experiment, SIFT and sRDSIFT must have a similar performance, since no distortion ($\psi = 0$) is considered. This is verified by this experiment. Surprisingly, sRDSIFT performs better on the rotation+scale experiment, but this is explained due to its lower number of features detected. Our approach obtains a repeatability ratio lower than SIFT and sRDSIFT but with the advantage of a higher number of correctly mapped extrema.

4) Catadioptric System with Real Images: In the case of real images, these were acquired with the calibrated hyper-
Fig. 12. Real hypercatadioptric images. (a) Reference image. (b) Rotated 20°. (c) Rotated 30°. (d) Results of the experiment.

Table V

<table>
<thead>
<tr>
<th></th>
<th>Reference</th>
<th>Rot 10°</th>
<th>Rot 20°</th>
<th>Rot 30°</th>
</tr>
</thead>
<tbody>
<tr>
<td>SIFT</td>
<td>876</td>
<td>483 (871)</td>
<td>428 (865)</td>
<td>436 (849)</td>
</tr>
<tr>
<td>pSIFT</td>
<td>415</td>
<td>283 (448)</td>
<td>233 (475)</td>
<td>199 (507)</td>
</tr>
<tr>
<td>sSIFT</td>
<td>440</td>
<td>236 (449)</td>
<td>235 (465)</td>
<td>210 (456)</td>
</tr>
<tr>
<td>LB</td>
<td>732</td>
<td>484 (668)</td>
<td>400 (640)</td>
<td>360 (657)</td>
</tr>
</tbody>
</table>

camera has been mounted on a precision head allowing us to accurately rotate it around its principal axis. Since calibration of real systems is not perfect we increase the threshold of the matching criteria. In this case we use $\delta_d = 5$ pixels. In Fig. 12 we show the images used for this experiment. In Table V we show the number of extrema detected as well as the number of correct mapped extrema. We observe that our approach has the highest repeatability ratio followed by pSIFT. In the most deformed image SIFT, pSIFT and sSIFT show a similar behavior, but it is considerable worse than our approach. An automatic matching experiment is not included for this experiment due to the lack of ground truth information.

5) Big distortion experiment: In this experiment we use a planar patch with texture, which is rotated around the axis where the major distortion produced by the conic shape of the mirror is observed. We use the hypercatadioptric system m1 with a iteration factor $\pi = 1.2$. Nine synthetic images with rotations from zero degrees to eighty degrees with steps of ten degrees are generated. The two extreme images are shown in Fig. 13. We map the points from the bigger image to the smaller ones. In Fig. 14 we show the results of this experiment. We observe that the Laplace-Beltrami approach outperforms the rest of the approaches, keeping the highest repeatability ratio under the highest amount of distortion. We observe that pSIFT outperforms conventional SIFT with a small amount of distortion, but decreases dramatically with the increase on the image distortion. This can be, as we commented before, due to the extra mapping from the stereographic projection to the original catadioptric image. In the case of extreme deformations (angles 60-80), pSIFT outperforms SIFT and sSIFT. We also observe that the scale threshold $\delta_s = 0.85$, makes sSIFT to only consider as correct features those with the exactly same scale. In order to also consider the second closest scale as valid, we lowered the scale threshold to $\delta = 0.7$. In Fig. 14 we observe that this change makes sSIFT to outperform all of the approaches in the images with less distortion. However, our approach is still superior in the cases of extreme distortion. The classical SIFT approach observed a rapid decrease in performance as distortion increases and in the case of extreme distortion its repeatability ratio went lower than 10%.

UTR80, Manufactured by Newport with goniometer of 1/60° resolution

We have observed that our approach shows a competitive performance in all types of central cameras. In the case of catadioptric images our approach shows the best performance in the automatic matching experiment and in one case of the repeatability experiment, while sSIFT showed a slightly better performance in the other two cases with difference of less than 5%. In the case of extreme deformation and real hypercatadioptric images our approach showed a clear improvement over SIFT, pSIFT and sSIFT, with average improvements of 20%, 11% and 9%, respectively. With respect to fisheye images our approach shows a slightly lower performance (5%) than sRDSIFT, which has the best performance in this type of images. This is explained due to that the spherical camera model used in our approach is only an
approximation to fisheye lenses, while the division model is a more accurate representation of these cameras. In the case of conventional images, our approach shows a good performance in the repeatability experiment, improving the rest of the approaches in the rotation case. With respect to the matching experiment our approach is located in third place after SIFT and sRDSIFT, which are approaches specifically designed to deal with this type of images, while pSIFT and LB consider the extra mapping from the sphere to the corresponding plane used by each approach.

### D. Mixing different cameras

In order to show the invariance to camera of our approach we perform experiments mixing different types of cameras. We match perspective images with fisheye and hypercatadioptric images. As in previous experiments we map the extrema points from one image to the other and count the number of correct matches.

1) **Mixing fisheye and perspective images:** In this experiment we use fisheye and conventional images. We perform rotations and translations between the images. In order to avoid mismatches due to occlusion or similar phenomena we use the SIFT descriptor as auxiliary to identify the correct matches. The maximum distance between the matched points is 2 pixels. In Fig. 15 we observe three different experiments between three perspective images and a fisheye image. Notice the dramatic change of view point and the difficulty of the scene. The number of correct matches and the number of extrema detected obtained by the different approaches is shown in Table VI. We decide to show the absolute matches since the repeatability ratio in these cases is very low.

We observe that our approach has the highest number of repeatability matches in two scenes, while sRDSIFT has the highest number of repeatability matches in the remaining scene. pSIFT detected again a high number of extrema but the correspondences between images are very low.

2) **Mixing catadioptric and perspective images:** We also perform experiments mixing catadioptric with perspective images. In Fig. 16 we present three different experiments. Notice the complexity of the scene in the second and third configurations. In Table VII we observe the absolute matches for each experiment. The first observation is that our approach detected more extrema in the catadioptric image than in the perspective ones, in contrast to SIFT and pSIFT which detected more extrema in the perspective images. In this case as it has been observed during the experiments, pSIFT detected a high number of extrema, most of them located in small areas and very close to each other. However, its number of correct repeatability matches is low compared to the number of extrema detected. In contrast our approach with less extrema but better distributed on the images, obtained almost the same number of correct repeatability matches in one case and less than pSIFT in the other two cases.

### VI. Conclusions

We have presented a general method to compute the scale-space of any central projection system. We integrate the spherical camera model with its extension to consider all central projection systems with the partial differential equation framework on manifolds to compute a generic version of the second order differential operator Laplace-Beltrami. This operator is used to perform the Gaussian smoothing on catadioptric, fisheye and even conventional perspective images. We perform experiments using synthetic images generated with parameters coming from actual manufactured mirrors and with real images. We observe that the Laplace-Beltrami operator considers correctly the geometry of the corresponding projection systems, since the pixels at the periphery have a different weight than those at the center. This situation explains the natural non-homogeneous resolution inherent to the wide angle central projection systems. We compare our proposal to three different approaches designed to deal with omnidirectional images and the conventional SIFT approach. We perform repeatability and automatic matching experiments using different types of cameras and combinations of cameras. In the case of catadioptric images our approach showed the best performance in most of the experiments. With respect to fisheye systems our approach is slightly behind sRDSIFT, which is an approach specifically designed to deal with this type of images. In the case of perspective images our approach performed similar to the rest of the approaches being superior in the rotation experiment. In general our approach showed a good performance with all the projection systems, being superior to the conventional SIFT and showing a competitive performance with the state-of-the-art methods.

### References


Fig. 15. Hybrid repeatability experiment mixing fisheye images with conventional images at three different positions.
Fig. 16. Hybrid repeatability experiment mixing catadioptric images with conventional images at three different positions.