

OMNISIFT: SCALE INVARIANT FEATURES IN OMNIDIRECTIONAL IMAGES

Zafer Arican and Pascal Frossard

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

ABSTRACT

We propose a method to compute scale invariant features in omnidirectional images. We present a formulation based on Riemannian geometry for the definition of differential operators on non-Euclidian manifolds that correspond to the particular form of the mirrors in omnidirectional imaging. These operators lead to a scale-space analysis that preserves the geometry of the visual information in omnidirectional images. We eventually build novel scale-invariant omniSIFT features inspired by the planar SIFT framework. We apply our generic solution to omnidirectional images captured with parabolic mirrors. Simple descriptors that use omniSIFT characteristics offer promising performance in the case of image rotation or translation where visual features can be preserved due to the proper handling of the implicit image geometry.

Index Terms— Omnidirectional vision, scale-invariant features, Riemannian geometry

1. INTRODUCTION

Omnidirectional vision has been an active research field in robotics and surveillance where sensors with large fields of view present several advantages for scene analysis, representation or detection, for example. The omnidirectional cameras typically consist of either a fisheye lens or a lens and a mirror system with a smooth surface such as parabolic or hyperbolic mirrors. These sensors collect the light rays from the scene and project them onto a sensor with a regular grid of sensitive cells. The structure of the resulting images is highly dependent on the geometry of the mirror, which should be taken into account for an appropriate processing of the light information.

Applications such as camera calibration, object detection, recognition or tracking generally rely on the localization and matching of particular visual features in multiple images. Scale invariance is an important characteristic of visual features that permits to be less sensitive to imperfect camera settings. The most popular scale invariant feature detection algorithm is certainly the SIFT framework [1] for perspective camera images. Many other methods have been proposed with different distinctive descriptors and feature detection methods [2, 3, 4, 5] for classical cameras. However, omnidirectional images have generally a specific geometry due to the sensor characteristics, which typically causes partial scale changes in different regions of the images. For example, a scene captured with a catadioptric camera using a paraboloid mirror is sampled more densely in the outer parts of the image than in the center. Classical feature detection algorithms do not take into account the implicit geometry of the mirrors; this penalizes the performance of the image analysis applications when they are applied directly on the sensor images [6, 7, 8].

We propose in this paper a novel framework for the computation of scale invariant features on omnidirectional images created by sensors with particular geometries. In particular we build on Riemannian geometry to define differential operators on non-Euclidian manifolds, such that the images can be processed in their native geometry. We then define a scale-space analysis that permits to build scale invariant features that are adapted to the geometry of the omnidirectional images. We illustrate our framework in the case of parabolic omnidirectional images that are commonly used in robotics and surveillance applications. We show in experiments that simple descriptors based on our new features provide invariance to rotation on $SO(3)$ and give good matching performance in the case of translation. Our framework provides a promising solution for calibration and feature detection application in omnidirectional camera networks.

Recent works such as [9, 10, 11] have proposed to process omnidirectional images on the sphere after an inverse stereographic projection that preserves the geometry of the light information [12, 13]. In these works, the scale-space representation is computed with Gaussian kernels on the sphere, while the convolution is performed using the spherical Fourier transform on a equiangular grid. The extra interpolation step between different sampling grids however induces loss of precision on the pixel positions. In addition, the non-uniform sampling grid does not preserve the original sampling density and can cause spurious upsampling and downsampling processes that affect the scale of the computed features. Furthermore, even if the spherical Fourier transform provides an efficient way to perform convolution, its inherent bandwidth limitations can cause aliasing and extra smoothing. In an attempt to better preserve the image geometry, an approximate solution that maps the gaussian functions back to the original image is proposed in [14]. It confirms that processing the images on their original sampling grid has important benefits. In this paper, we adopt a different strategy where the geometry of the imaging system is represented as a Riemannian manifold. We then build on [15] to propose a scale-space analysis that permits to compute scale invariant features with help of differential operators instead of Gaussian kernels. Our generic framework can be applied to different types of omnidirectional imaging systems where it provides scale invariance and preserves the geometry of the light information.

2. SCALE-SPACE ANALYSIS ON NON-EUCLIDIAN MANIFOLDS

2.1. Riemannian Geometry Framework

Scale-space analysis is generally performed with help of Gaussian kernels and differences of Gaussians on planar images. Gaussian kernels can however not be used on generic smooth surfaces. One

can however compute a scale-space representation $I(x, y, t)$ on non-Euclidian manifolds with help of the heat equation. It reads

$$\frac{\partial I(x, y, t)}{\partial t} = \Delta I(x, y, t) \quad (1)$$

where Δ is the Laplacian operator and t is the scale level. The initial condition is given as $I(x, y, t_0) = I(x, y)$. It can be noted that the Gaussian function with standard deviation \sqrt{t} is the solution for the heat equation (1) on planar images.

The heat equation permits to develop a scale-space analysis with differential operators. These operators can be defined on smooth manifolds with help of Riemannian geometry, as recalled in [15]. Let \mathcal{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 \mathcal{M} , the components of the gradient of the scalar function I read $\nabla = g^{ij} \frac{\partial}{\partial x^j}$, where g^{ij} is the inverse of g_{ij} . Furthermore, the divergence of a vector field \mathbf{V} on \mathcal{M} is given as $\text{div} \mathbf{V} = \frac{1}{\sqrt{g}} \partial_i (V^i \sqrt{g})$, where g is the determinant of g^{ij} . We can then define the Laplace-Beltrami operator as the second order differential operator on the scalar field I on \mathcal{M} , as

$$\Delta I = -\frac{1}{\sqrt{g}} \partial_j (\sqrt{g} g^{ij} \partial_i I) \quad (2)$$

This operator that corresponds to the Laplace operator on the plane can be used to solve the heat equation (1) on non-Euclidian manifolds. The specific form of the Laplace-Beltrami operator depends on the particular geometry of the manifold \mathcal{M} .

2.2. Parabolic Mirror Systems

We consider here the specific case of omnidirectional imaging systems with parabolic mirrors that are quite common in robotics and surveillance applications. Images from parabolic mirrors can be uniquely mapped on the 2-sphere by inverse stereographic projection, similarly to images from most simple mirrors and catadioptric systems. This enables easier processing of the parabolic images. We then derive the metric necessary to the construction of differential operators on the sphere, in order to perform scale-space analysis and feature detection by properly taking into account the geometry of the images.

First, we can define the Euclidian line element dl on the 2-sphere S^2 in terms of the variables r , θ and ϕ that represent the spherical coordinates. The line element satisfies

$$dl^2 = r^2 (d\theta^2 + \sin^2 \theta d\phi^2). \quad (3)$$

Stereographic projection maps each point on the sphere to a plane \mathbb{R}^2 of coordinate (x, y) . A point in polar coordinates (R, ϕ) on the plane is related to the a point (θ, ϕ) on the sphere by $R = 2r \tan(\frac{\theta}{2})$ and $\phi = \phi$. Using the identities, $R^2 = x^2 + y^2$, $\phi = \tan^{-1}(\frac{y}{x})$ in cartesian coordinates and $r = 1$, the line element reads

$$dl^2 = \frac{16}{(4 + x^2 + y^2)^2} (dx^2 + dy^2) \quad (4)$$

giving the metric

$$g_{ij} = \begin{pmatrix} \frac{16}{(4+x^2+y^2)^2} & 0 \\ 0 & \frac{16}{(4+x^2+y^2)^2} \end{pmatrix} \quad (5)$$

and the inverse metric

$$g^{ij} = \begin{pmatrix} \frac{(4+x^2+y^2)^2}{16} & 0 \\ 0 & \frac{(4+x^2+y^2)^2}{16} \end{pmatrix} \quad (6)$$

Equipped with this metric, we can finally compute the differential operators on the sphere with help of Eq. (2). In particular, the norm of the gradient reads

$$|\nabla_{S^2} I|^2 = \frac{(4 + x^2 + y^2)^2}{16} |\nabla_{\mathbb{R}^2} I|^2 \quad (7)$$

while the norm of the Laplace-Beltrami operator can be written as

$$|\Delta_{S^2} I|^2 = \frac{(4 + x^2 + y^2)^2}{16} |\Delta_{\mathbb{R}^2} I|^2 \quad (8)$$

These operators permit to compute a scale-space representation of the images in the sensor plane, while providing an accurate representation of the geometry in the omnidirectional images through proper Riemannian metrics.

3. FEATURE DETECTION

We get inspiration from the SIFT framework to construct features and design simple descriptors for omnidirectional images, based on the scale-space analysis presented in the previous section. For planar images, it has been shown that differences of gaussian can approximate scale-normalized Laplacian of Gaussian if the scale levels are separated by a constant multiplicative factor [1], since this satisfies Lindeberg's normalization condition for scale invariance [16]. In order to benefit from scale invariance, we adopt a similar method and define a multiplicative factor k that controls the scaling in the heat equation. We set $t_i = k^{2i} \sigma^2$ in Eq. (1) and we compute the heat equation successively at times t_i defined in terms of the normalization and scale factors k and σ . We form the scale levels such that scale normalized difference images are obtained after scale-space analysis. Similarly to the SIFT framework, we select $k = 2^{1/3}$ and use 4 successive octaves.

Note that we use discrete operators for the computation of the scale-space representation. The timesteps in the heat equations are discrete, and we use discrete differential operators on the plane for the computation of the gradient (i.e., $[-1 \ 1]/d_s$) and laplacian (i.e., $[-1 \ 2 \ 1]/d_s$). Smoothing is performed by updating $I(x, y, t)$ with the differences that have been computed at previous steps. Finally, the images are downsampled for each octave in order to reduce the computation time. However, since the induced metric is dependent on the position, after downsampling, the sampling factor, ds is doubled for each octave in the differential operators.

Once the scale space images are formed, we use a similar strategy as the SIFT framework [1] in order to detect the most important visual features. First, we detect local extremum points by checking 26 neighbor points in windows of 3×3 pixels in the current and adjacent difference images. Differences between neighbors permit to remove low contrast features by thresholding on the magnitude of these differences. In [1], edge responses are then removed by checking the ratio between maximum and minimum principle curvatures of the difference image at the pixel position and features with ratio greater than 10 are deleted. Finally, a 3D quadratic function is fit to the pixel position and scale for additional refinement of the set of features.

4. EXPERIMENTAL RESULTS

We propose experiments in order to test the invariance of our novel features to rotation and translation in omnidirectional images. We build simple descriptors from the visual features, and we compute the performance of matching features in transformed images.

We use two simple descriptors inspired from planar SIFT descriptors [1]. The first type of descriptor (i.e., Planar descriptor) is simply computed on the planar sensor image. The window size in the definition of the descriptor is however adapted to the position on the plane by using the metric g^{ij} for handling the geometry of the omnidirectional images. In addition, all the gradient computations necessary for building histograms of gradients in the descriptor are performed on the original manifold using the same metric g^{ij} . The second type of descriptor (i.e., Spherical descriptor) is computed using planes tangent to the sphere. An image patch on the tangent plane at a given pixel position is formed by mapping the neighboring pixels onto the image patch. Then a classical SIFT descriptor is computed on that planar image patch.

We compare the matching performance of these simple descriptors built on our new visual features, with those using a planar SIFT framework [17] on the sensor image and the spherical SIFT method [9].

4.1. Invariance to rotation

In a first experiment, we test the invariance of the different features to rotation in $SO(3)$. Recall that the catadioptric image from the parabolic mirror can be mapped uniquely onto the sphere. A rotation in $SO(3)$ should not change the scale of the features, but it might change the resolution of the image in different regions; this can cause problems for methods that do not consider the geometry of the images.

We first create a test set of five planar test images (see Figure 1). These images are mapped onto a 10×10 unit plane. The omnidirectional camera captures this image plane from two different positions that are set to be 8 and 10 units away from the plane. We create transformed images with 9 different rotations by moving the camera around its center with angles ranging from -40 to 40 degrees by intervals of 10 degrees.



Fig. 1: Test Images mapped on the synthetic plane

For each rotation value, we measure the matching performance by comparing the image to all other rotated versions and then average the results for all images and each capture position (i.e., $8 \times 2 \times 5$). We measure the number of correct matches, the number of features and the localization errors for all methods under comparison. Since the transformation between images is known, the correct matches are computed by checking the distance between the feature points and by comparing the scales of the features. We use a decision threshold that is adaptive to the scale levels and to the descriptor framework.

Figure 2 shows the percentage of the correct matches over number of features. The proposed algorithm is not affected by the rotations and provides the best percentage. Although the planar SIFT has the highest peak percentage, it is affected by rotations and the performance drops sharply at the high rotation values. It is a result of feature detection mechanism that does not take geometry into consideration. The spherical SIFT is not affected by rotation but the percentage of correct matches is low. The same behavior is observed for the our visual features and the spherical descriptor. This is due to the computation of the descriptors on the sphere where the nec-

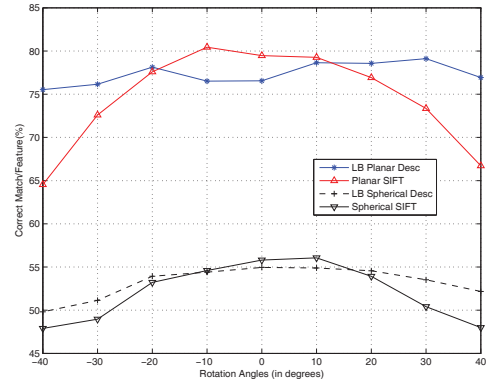


Fig. 2: Percentage of correct matches over number of features for different rotation values.

essary extra interpolation step in both algorithms generates incorrect descriptors.

We also measure the mean localization error for the matched features. This error is computed by measuring the error on the localization of features that provide correct matches. The matching parameters for the different methods are selected such that the average number of features in each case is approximately constant. Figure 3 shows the mean localization error for different rotation values for all methods. The spherical SIFT descriptor has a significant localization error while the proposed features and the planar SIFT permit to keep the localization error low. This performance confirms that the proposed framework can be used for calibration and scene reconstruction with omnidirectional images.

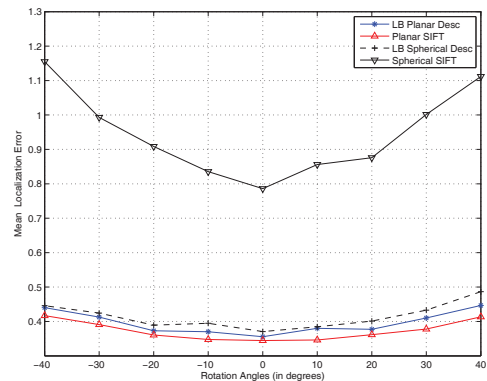


Fig. 3: Mean localization error for different rotations

4.2. Invariance to translation

We then propose experiments that include translation of the cameras. We use the same test images as for the first experiments, and we set the distance of the camera to the image plane to be of 10 units. We select the matching parameters so that the number of features are approximately equal.

We consider two types of translation. First, the camera moves perpendicularly to its optical axis, which itself stays perpendicular to the test image plane. Figure 4 shows the percentage of the correct matches with respect to the number of detected features in

each method. Although the proposed method gives high percentage, the planar SIFT method is better for this type of translation. For large translation, the test plane moves towards the outer region of the image where more important smoothing is performed in our scale-space analysis, and this degrades the feature detection performance.

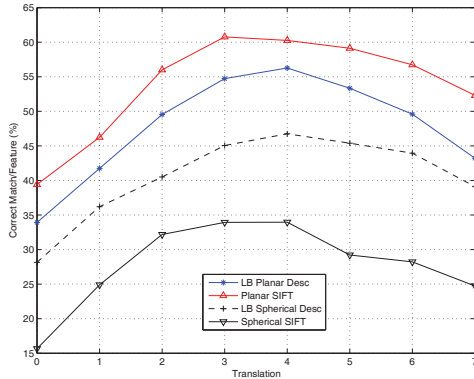


Fig. 4: Percentage of correct matches with respect to the number of features for different displacements for the first translation test.

Second, the camera moves along its optical axis with some rotations. The camera makes 5 units of displacement with 1 unit intervals. The axis of the rotation is the same as for the above rotation test. The values are averaged for different translations. This test checks the response of the proposed method to combinations of translation and scaling transformations. Figure 5 shows percentage of correct matches with respect to the number of features detected by each method. The proposed features with Planar descriptor lead to the best performance when translation is combined with scaling, which represents probably the most common type of transformation in practice.

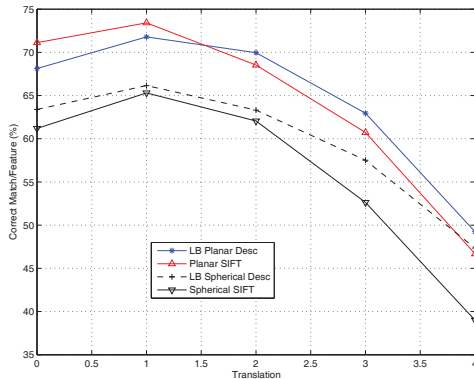


Fig. 5: Percentage of correct matches with respect to the number of features for different displacements for the second translation test

5. CONCLUSION & DISCUSSION

We have proposed a scale invariant feature computation method for omnidirectional images from imaging systems with particular geometry. We have exploited the foundations of the Riemannian geometry to formulate the scale-space analysis and a feature detection framework that works directly on the original image plane without need for

any interpolation. We have derived and tested the proposed method for parabolic omnidirectional images, where experiments show that an accurate exploitation of the geometry leads to invariance of the features to rotations in $SO(3)$, and to competitive performance in the case of translation. The proposed method can be extended to more generic manifolds and computation of a good descriptor with full geometry integration is one of our future works.

6. REFERENCES

- [1] D.G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, Jan 2004.
- [2] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 10, pp. 1615 – 1630, Oct 2005.
- [3] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L.V. Gool, "A comparison of affine region detectors," *International Journal of Computer Vision*, vol. 65, no. 1, pp. 43–72, 2005.
- [4] E. Tola, V. Lepetit, and P. Fua, "A fast local descriptor for dense matching," in *Proc. of CVPR*, 2008, pp. 1–8.
- [5] T. Lindeberg, "Detecting salient blob-like image structures and their scales with a scale-space primal sketch - a method for focus-of-attention," *International Journal of Computer Vision*, vol. 11, no. 3, pp. 283–318, Jan 1993.
- [6] T. Geodeme, T. Tuytelaars, G. Vanacker, M. Nuttin, and L. Van Gool, "Omnidirectional sparse visual path following with occlusion-robust feature tracking," in *Proc. of OMNIVIS Workshop*, 2005.
- [7] Y. Bastanlar, L. Puig, P. Sturm, J. Guerrero, and J. Barreto, "Dit-like calibration of central catadioptric cameras," in *Proc. of OMNIVIS Workshop*, 2008.
- [8] T. Mauthner, F. Fraundorfer, and H. Bischof, "Region matching for omnidirectional images using virtual camera planes," in *Proc. of Computer Vision Winter Workshop*, 2006.
- [9] J. Cruz-Mota, I. Bogdanova, B. Paquier, M. Bierlaire, and J.P. Thiran, "Scale invariant feature transform on the sphere: Theory and applications," *EPFL Technical Report*, pp. 1–44, May 2009.
- [10] P. Hansen, P. Corke, W. Boles, and K. Daniilidis, "Scale invariant feature matching with wide angle images," in *Proc. of IROS*, 2007, pp. 1689–1694.
- [11] P. Hansen, P. Corke, W. Boles, and K. Daniilidis, "Scale-invariant features on the sphere," in *Proc. of ICCV*, Oct 2007, pp. 1 – 8.
- [12] C. Geyer and K. Daniilidis, "Catadioptric projective geometry," *International Journal of Computer Vision*, vol. 45, no. 3, pp. 223–243, Jan 2001.
- [13] X.G. Ying and Z.Y. Hu, "Can we consider central catadioptric cameras and fisheye cameras within a unified imaging model?," *Lecture Notes in Computer Science*, vol. 3021, pp. 442–455, Jan 2004.
- [14] P. Hansen, W. Boles, and P. Corke, "Spherical diffusion for scale-invariant keypoint detection in wide-angle images," in *Proc. of DICTA*, 2008, pp. 525–532.
- [15] I. Bogdanova, X. Bresson, J.P. Thiran, and P. Vandergheynst, "Scale space analysis and active contours for omnidirectional images," *IEEE Transactions on Image Processing*, vol. 16, no. 7, pp. 1888–1901, 2007.
- [16] T. Lindeberg, *Scale-space theory in computer vision*, Springer, 1994.
- [17] A. Vedaldi, "Sift implementation," *Software Documentation*, pp. 1–6, Sep 2007.