Scale-Invariant Features and Polar Descriptors in Omnidirectional Imaging

Zafer Arıcan, *Member, IEEE*, and Pascal Frossard, *Senior Member, IEEE*

***Abstract*—We propose a method to compute scale-invariant features in omnidirectional images. We present a formulation based on the Riemannian geometry for the definition of differential operators on non-Euclidian manifolds that adapt to the mirror and lens structures in omnidirectional imaging. These operators lead to a scale-space analysis that preserves the geometry of the visual information in omnidirectional images. We then build a novel scale-invariant feature detection framework for omnidirectional images that can be mapped on the sphere. We further present a new descriptor and feature matching solution for these omnidirectional images. The descriptor builds on the log-polar planar descriptors and adapts the descriptor computation to the specific geometry and the nonuniform sampling density of omnidirectional images. We also propose a rotation-invariant matching method that eliminates the orientation computation during the feature detection phase and thus decreases the computational complexity. Experimental results demonstrate that the new feature computation method combined with the adapted descriptors offers promising detection and matching performance, i.e., it improves on the common scale-invariant feature transform (SIFT) features computed on the unwrapped omnidirectional images, as well as spherical SIFT features. Finally, we show that the proposed framework also permits to match features between images with different native geometry.**

I. Introduction

APPLICATIONS such as camera calibration, object detection, recognition, or tracking generally rely on the localization and the matching of visual features in multiple images. The visual features encoded in descriptors are matched between different images with similar content in order to establish correspondences in image sets. While such a framework has been widely studied for perspective cameras, the recent development of omnidirectional vision in applications such as robotics and surveillance has created the need for adapting the feature detection and matching algorithms to the new geometry of images with large fields of view.

Manuscript received August 24, 2010; revised October 31, 2011; accepted January 01, 2012. Date of publication January 27, 2012; date of current version April 18, 2012. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Hsueh-Ming Hang.

Z. Arıcan was with the Signal Processing Laboratory (LTS4), Institute of Electrical Engineering, Ecole Polytechnique Fédérale de Lausanne, 1015 Lausanne Switzerland. He is currently with Turk Telekom, Orta Doğu Teknik Üniversitesi Teknokent, Ankara 06531, Turkey (e-mail: zafer.arican@turktelekom.com.tr).

P. Frossard is with the Signal Processing Laboratory (LTS4), Institute of Electrical Engineering, Ecole Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland (e-mail: pascal.frossard@epfl.ch).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIP.2012.2185937

In perspective images, feature detection largely relies on scale invariance such that the visual features are less sensitive to imperfect camera settings and transformations. The most popular scale-invariant feature detection algorithm is certainly the scale-invariant feature transform (SIFT) framework [1], but many other methods have been proposed with different descriptors and feature detection approaches [2]–[5]. In order to also achieve scale invariance for features in omnidirectional imaging, the geometry of the sensing system has to be carefully considered in processing the light information. 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; it might lead to inaccurate feature matching if the processing is not adapted to the sensor. Classical scale-invariant feature detection algorithms, unfortunately, do not take into account the implicit geometry of the mirrors. It penalizes the performance of the image analysis applications that directly use the sensor images [6]–[8].

In this paper, we propose a novel framework for the computation of scale-invariant features in omnidirectional images captured by sensors whose geometry permits a mapping to spherical images. In particular, we build on the Riemannian geometry to define differential operators on non-Euclidian manifolds, such that the images can be processed in their native geometry. We then propose a scale-space analysis framework with scale-invariant features that are adapted to the geometry of omnidirectional images. The scale-space representation is computed by a set of difference functions that approximate the heat diffusion equation on the sphere. We illustrate our framework in the case of catadioptric systems with parabolic mirrors that are commonly used in robotics and surveillance applications. The benefits of our approach is threefold. First, the scale-space analysis is performed by considering the true geometry of the omnidirectional system in both the mirror geometry and the sampling grid. This provides a better positioning of the features in the scale-space representation. Second, it performs a local analysis of the omnidirectional images that is important when images are not complete due to the construction of the mirrors in omnidirectional images. Third, as the image is directly processed in its native geometry, there is no need for interpolation prior to feature computation, and inaccuracies are avoided.

After the detection of the visual features, descriptors are computed so that features can be matched in different images. The descriptors are expected to be distinctive, as well as invariant to scale, rotation, and affine transformations, and robust to illumination changes. We design a novel polar descriptor for scale-invariant features on the sphere that builds on log-polar descriptors used for planar images. This descriptor exploits the geom-

etry of the sphere and takes into account the different sampling densities caused by the equiangular grid on the sphere. We further extend these descriptors by exploiting the relation between the orientation bins and gradient orientation histograms in order to get rid of the orientation in the descriptor computation. This leads to a novel matching strategy that permits to reduce the computational complexity in the construction of the descriptors. Experiments on both synthetic and natural images show that the proposed feature detection and matching method outperforms the SIFT method that is applied on the planar omnidirectional images, as well as spherical SIFT features. Finally, we also show that it is possible to match features computed in different types of omnidirectional images, which is of particular interest in networks of hybrid cameras.

A few recent works [9]–[11] have also addressed the problem of feature detection and matching in omnidirectional images. They propose to process omnidirectional images on the sphere after an inverse stereographic projection that preserves the geometry of the light information [12], [13]. The scale-space representation is computed with Gaussian kernels on the sphere where the convolution is performed using the spherical Fourier transform on an equiangular grid. An extra interpolation step between different sampling grids however induces the loss of precision on the pixel positions. In addition, the nonuniform sampling grid does not preserve the original sampling density and can cause spurious upsampling and downsampling phenomena that affect the scale of the computed features. The inherent bandwidth limitations of equiangular grids can also cause aliasing and extra smoothing when working in the frequency domain. 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.

The design of effective descriptors have also attracted interest from the research community. Descriptors based on histogram computation generally provide the highest robustness to transformations that might happen between images. The SIFT descriptor [1] is based on gradient orientation histograms computed in the region around feature points. Descriptors with different histograms such as SURF [15] and CHoG [16] have been recently proposed to improve the matching performance. The GLOH framework [2] provides a log-polar descriptor that computes histograms in spatial bins by radial division of the support region of the feature. All these algorithms are however designed for planar images and do not consider the specific geometry of omnidirectional images. Furthermore, they assume that the sampling is uniform in the image. For omnidirectional images, the sampling density however differs from region to region, and this should be taken into account in the computation of the descriptors. One approach [10] maps the image around the feature to the tangent plane to form a planar image patch and computes SIFT descriptors on the mapped image patch. This simple approach implicitly deals with different sampling densities, but an extra interpolation step is performed, which may change the true scale of the feature and the value of the gradient in the descriptors. In another approach [9], the SIFT descriptors are directly computed on the spherical surface. This approach is however affected by different sampling densities and does not take into account the different number of samples in the diverse spatial bins.

The rest of this paper is organized as follows. It first introduces in Section II the scale-space analysis framework for non-Euclidean manifolds with an example on parabolic omnidirectional images. The feature detection method based on the novel scale-space representation is presented in Section III. The polar descriptor with a feature matching criteria for both oriented and nonoriented descriptors is explained in Section IV. Section V discusses the experimental results and further discusses applications of feature matching between images from different cameras.

II. Scale-Space Analysis on Non-Euclidian Manifolds

The scale-space analysis is the vital element of scale-invariant feature detection. The scale-space representation of images obtained by smoothing is widely studied for planar images. The smoothing is typically performed by convolution with Gaussian kernels, which satisfy the scale-space axioms [17]. For non-Euclidian manifolds corresponding to nonplanar sensors, the Gaussian kernels can not be however directly applied as the shift-invariance requirements are not satisfied. The heat diffusion equation can be still used to smooth these images and form the scale-space representation after proper adaptation to the image geometry. In this section, we define the scale-space representation for parabolic mirror systems as a specific case of non-Euclidian manifolds using the heat diffusion equation after a mapping on the sphere. The spherical representation is obtained by using the property that any central catadioptric omnidirectional image can be uniquely mapped onto the sphere [12]. The methodology is not limited to parabolic mirror systems and can be extended to other mirror systems that present the same property, and we briefly discuss the example of hyperbolic mirrors at the end of this section.

We describe the scale-space representation for a specific case of omnidirectional imaging systems with parabolic mirrors that are commonly used in robotics and surveillance applications. Images from parabolic mirrors can be uniquely mapped on the 2-sphere, similar to images from most simple mirrors and catadioptric systems. This enables easier processing of the parabolic images with the capture of light rays from a central point that is the center of the sphere. We now derive the metric specific to the mirror that is necessary for the construction of differential operators on the sphere similar to [18]; it permits to the scale-space analysis and the feature detection by properly taking into account the geometry of the images.

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

$$\mathrm{dl}^2 = (d\theta^2 + \sin^2\theta d\phi^2). \tag{1}$$

The corresponding angles for θ and ϕ are given in Fig. 1. The stereographic projection maps each point on the sphere to plane $\mathcal{R}^2$ of coordinate (x, y). A point in the polar coordinates (R, ϕ) on the stereographic plane is related to point (θ, ϕ) on the sphere

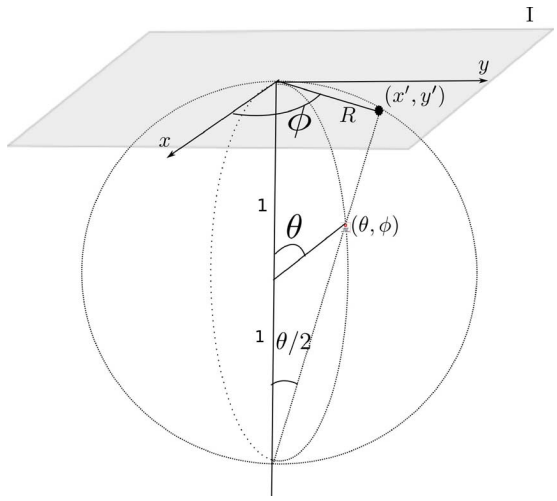


Fig. 1. Geometry of the 2-sphere with spherical polar coordinates. In the stereographic projection, the plane corresponds to the image plane, and θ is the corresponding colatitude angle on the sphere.

via relation $R = 2\tan(\theta/2)$. Then, the terms in the line element of (1) read

$$d\theta^2 = \frac{16}{(4+R^2)^2} dR^2$$
$$\sin^2(\theta) = \frac{16R^2}{(4+R^2)^2} \tag{2}$$

and the line element dl^2 is

$$\mathrm{dl}^2 = \frac{16}{(4+R^2)^2}(dR^2 + R^2 d\phi^2). \tag{3}$$

Let $(x, y) \in \mathbb{R}$ on the sensor plane define Cartesian coordinates, where $R^2 = x^2 + y^2$ and $\phi = \tan^{-1}(y/x)$. The line element then reads

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

giving the induced Riemannian metric, i.e.,

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

and the inverse metric, i.e.,

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

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

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

while the norm of the Laplace–Beltrami (LB) operator can be written as

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

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. It also enables the fast computation of the LB operator by first computing the Laplacian of the image as if it is a planar image and then weighting by the proper metric.

In order to illustrate that the aforementioned method is quite generic, we briefly discuss the case of hyperbolic systems. These systems use perspective projection instead of an orthographic projection used in parabolic systems. The sampling becomes different and more dense in the center of the images [18]. However, similar to the aforementioned development, one can still define a metric that differs from the Euclidean metric by the following scaling factor [18]:

$$|\Delta_{\mathcal{S}^2} I|^2 = \frac{\left(1 - (x^2+y^2)\right)}{4} |\Delta_{\mathcal{R}^2} I|^2. \tag{9}$$

This metric can be used in the computation of the heat diffusion equations and gradient computations for proper handling of omnidirectional images captured by hyperbolic mirror systems.

III. Feature Detection

Equipped with the scale-space representation, we now present the feature detection method in omnidirectional images. In the classical case of planar images, it has been shown that maxima and minima of the scale-normalized Laplacian of Gaussian images provide the most stable scale-invariant features [19]. The difference of Gaussian images can approximate scale-normalized Laplacian of Gaussian images if the scale levels are separated by a constant multiplicative factor [1]. 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 diffusion equation. We thus compute the heat diffusion equation at successive time steps t_i, where $t_i = k^{2i}\sigma_o^2$ is defined in terms of the normalization and scale factors k and the base smoothing level σ_o. We form the scale levels such that scale-normalized difference images correspond to the scale-space analysis.

Note that we use discrete operators for the computation of the scale-space representation. The differentiation with respect to time in the heat diffusion equation is discretized with time intervals d_t. We further use discrete differential operators on the plane for the computation of the gradient and the Laplacian. We use $[-1 \quad 1]/d_s$ as the gradient operator and $[-1 \quad 2 \quad -1]/d_s$ as the Laplacian operator. Smoothing is finally performed by updating $I(x, y, t)$ with the differences that have been computed at previous time steps. In addition, the images are downsampled for each octave in order to reduce the computation time. However, since the induced metric is dependent on the position, the sampling factor d_s is doubled for each octave in the differential

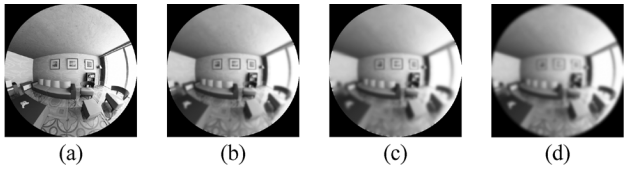


Fig. 2. Original parabolic omnidirectional image. (b) Smoothing $t = 2.25$. (c) Smoothing $t = 5.76$. (d) Smoothed image with a Gaussian kernel on the planar image.

operators after downsampling. The smoothing process is summarized in Algorithm 1.

Algorithm 1 Smoothing of omnidirectional images with heat diffusion equation

1: **Initialization**:

$\bar{I}(x, y, t_i)$: The intermediate smoothed image during iterations

$I(x, y)$: The original image

t: The smoothing level

d_t: The time interval

n_i: The max number of iterations

$\bar{I}(x, y, 0) = I(x, y), n_i = t/d_t$

2: **repeat**

3: Compute gradient $\nabla_{\mathcal{R}^2}$ and Laplacian $\Delta_{\mathcal{R}^2}$ on the planar image $\bar{I}(x, y)$.

4: Compute Laplacian $\Delta_{\mathcal{S}^2}$ on the manifold using the induced metric g_{ij} and Laplacian $\Delta_{\mathcal{R}^2}$.

5: Update $\bar{I}(x, y)$ by $\bar{I}(x, y) = \bar{I}(x, y) + d_t * \Delta_{\mathcal{S}^2}$

6: **until** n_i is reached.

7: $I(x, y, t) = \bar{I}(x, y)$.

Fig. 2 illustrates some smoothed images for a parabolic omnidirectional image. Note that the central regions are smoothed less than the outer regions as the sampling is more dense around the outer regions for parabolic omnidirectional images.

As we formulate the problem in the spherical framework in order to preserve the geometry, the smoothing is expected to be uniform on the sphere. Gaussian smoothing used in the SIFT method does not take the geometry into account and causes nonuniform smoothing on the sphere. The smoothing method used in SIFT on the sphere [10], which is based on spherical Fourier transform, provides a uniform smoothing, but it requires that the image is first mapped onto an equiangular grid on the sphere. This requires an extra interpolation step and causes spurious scaling due to change in sampling densities. Fig. 3 illustrates the smoothing for these two methods and the proposed smoothing based on the Riemannian geometry. For the same smoothing level, the effect should be equivalent around the poles and around the equatorial region. The smoothing using the Gaussian kernels used in the planar SIFT

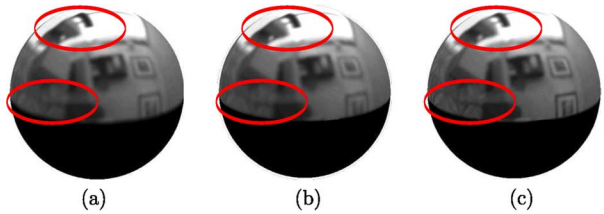


Fig. 3. (a) Smoothed spherical image using spherical Fourier transform [10] after mapping onto equiangular grid. (b) Smoothed spherical image using the proposed smoothing directly on the image plane. (c) Smoothed spherical image using Gaussian smoothing [1] in the image plane, without geometry adaptation. The regions in the upper ellipses all have the same smoothing levels. The images smoothed on the image plane are mapped onto the sphere for illustration purposes.

method does not provide this equivalence, and the equatorial region remains sharp. The proposed method processes the image directly on the image plane as the planar SIFT, but it can provide a uniform smoothing in the image's native geometry without any interpolation step.

Once the scale-space images are formed, the scale-normalized Laplacian of Gaussian images are approximated by the difference of Gaussian images similar to the SIFT framework [1]. This permits to detect the most important visual features by computing the difference of images of successive smoothing levels. However, note that Gaussian images are formed by the heat diffusion equation in our proposed framework. The difference image $L(x, y, t_i)$ is computed as

$$L(x, y, t_i) = I(x, y, t_i) - I(x, y, t_{i-1}) \tag{10}$$

where $I(x, y, t_i)$ and $I(x, y, t_{i-1})$ are the smoothed images at successive scale levels.

Given successive levels of difference images $L(x, y, t_i)$, we then detect local extremum points by checking 26 neighbor points in windows of 3×3 pixels in the current difference image $L(x, y, t_i)$ and the adjacent difference images $L(x, y, t_{i-1})$ and $L(x, y, t_{i+1})$ as in [1]. The detected minima and maxima points are the stable points in terms of scale invariance [19]. The detected feature point is assigned as the scale level $\sigma = \sqrt{t_i}$. Note that the Laplacian of the smoothed images are already computed at each time step as required by the heat diffusion equation. However, it is not practical to check for the extrema at each time step. Thus, the difference images are formed for discrete smoothing levels.

The detected extrema points are the potential feature points, but some refinement and elimination are performed to increase stability and accuracy. First, the unstable points are removed. Generally, extrema points in low contrast regions and edges are affected by the localization errors. That is why the detected feature points in low textured areas and at edges have to be removed. The magnitude of the difference of Gaussian images is used to remove the low-contrast feature points, the features for which the value of $L(x, y, t_i)$ is below a predefined threshold are removed. This is a result of the smoothing based in heat diffusion where high textured regions will have higher difference between smoothed images compared with the low textured regions. Difference images are also used to decide about the edge responses. Again, as in [1], edge responses are removed by

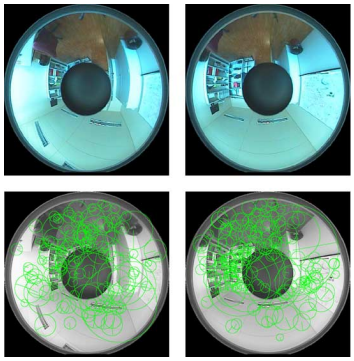


Fig. 4. (Top) Original omnidirectional images and (bottom) the detected features in these images. The diameter of the circles denote the scale level.

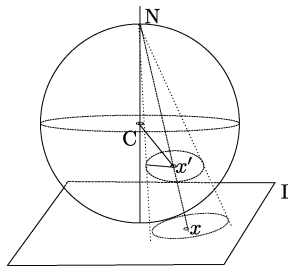


Fig. 5. Circular region around the mapped feature point x' on the sphere is mapped to the omnidirectional image plane I by stereographic projection from the north pole N.

checking the ratio between maximum and minimum principle curvatures of the difference image at the feature position; features with a ratio greater than a predefined threshold are deleted. Finally, the position of the feature point is refined to a subpixel level by fitting a 3-D quadratic function to the pixel position and scale. At the feature point, the 3-D quadratic function is fitted on the difference of Gaussian functions computed on the same level and the upper level. The extrema of the fitted function specify the refined position of the feature as in [1]. Fig. 4 illustrates the detected features in two omnidirectional images that undergo a translation.

IV. Feature Descriptors and Matching

Once scale-invariant features are computed by the difference of smoothed images, they have to be characterized by a descriptor so that they become discriminative enough for proper matching between different images. In this section, we describe two new descriptors, a polar descriptor for omnidirectional images, and a nonoriented version of this polar descriptor. We also describe a feature matching criterion based on the Kullback–Leibler (KL) divergence and propose two matching criteria for nonoriented polar descriptors based on L_2 distance and KL divergence.

A. Sampling-Aware Polar Descriptors on the Sphere

We propose a new feature descriptor that takes into account the specific geometry and sampling structure of omnidirectional images. We use the property that central catadioptric images can be uniquely mapped onto the sphere. Thus, the new descriptor is computed based on the sphere representation. The structure of the sphere and the pixel densities on the surface of the sphere are taken into account. The descriptor is computed using the pixels around the feature point. Considering the scale-space analysis on the sphere, a circular support region around the feature point on the sphere is computed based on the scale of the visual features, i.e., σ, and the radius of this region is proportional to $\sin(\sigma)$, similar to the method proposed in [10]. As illustrated in Fig. 5, if features are detected in the omnidirectional image, this circular region is mapped on the omnidirectional image plane, and image points inside this region are considered.

The support regions are first used for computing the orientation of the feature. The orientation of the feature is determined by first forming a histogram of gradient orientations weighted by the magnitude of the gradients. The peak of this histogram is selected as the orientation of the feature. If there are more than one peak, new features are added with the same position and scale information, but with different orientations. The gradient computation is the main part of the orientation assignment. For omnidirectional images mapped onto the sphere and for spherical images with equiangular grid, the sampling density is not constant on the surface of the sphere and thus affects the gradient computation. On a unit sphere with a $N \times N$ equiangular grid for example, the sampling distance between two neighbor pixels differing in colatitude angles θ is constant and equal to π/N. The sampling distance between two neighbor pixels with different longitude angles ϕ is however not constant. It changes with respect to θ and is equal to $\sin(\theta)2\pi/N$. We thus adapt the gradient computation with a multiplicative factor that depends on this distance. Recall that, for omnidirectional images, the gradients are computed using the induced Riemannian metric explained in the previous section.

Once an orientation is assigned to a visual feature, a support region around the same feature point but with a bigger size is again proportionally formed to $\sin(\sigma)$. As discussed in [1], the human visual system is sensitive to the orientation of the gradients and the spatial positions of the gradients. That is the reason why a descriptor formed in this way provides good discrimination. The spatial position of the gradients are determined by grouping the pixels inside the support region into spatial bins. The proposed descriptor divides the support region by discretizing the polar coordinates of the regions around the feature point similar to the GLOH descriptor [2]. Fig. 6 (left) shows the descriptor structure with eight orientation bins and three radial bins. The center bin is not divided into orientation bins in order to increase the robustness to feature localization errors. As one may observe, the subdivision structure is similar to longitudes and latitudes around the poles of a sphere with an equiangular grid. In addition, the selection of increasing radius values for the radial bins is achieved by dividing the latitude angles uniformly. Fig. 6 (right) illustrates this relation. Similar to the orientation computation, the support region and the subdivisions for the descriptor computation are mapped onto the omnidirectional plane using stereographic projection.

For rotation invariance, the support region and the subdivisions are rotated on the sphere around the feature point using the orientation assigned to the feature. For each spatial bin, the

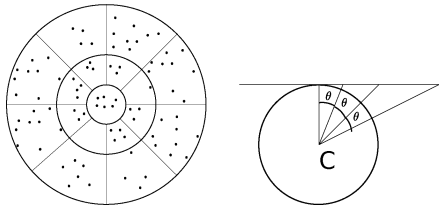


Fig. 6. (Left) Log-polar descriptor and nonuniform samples inside spatial bins. (Right) Radial bin divisions correspond to equal division of θ on the sphere.

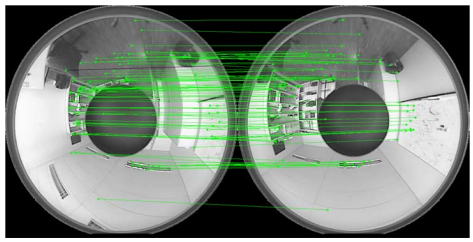


Fig. 7. Matched features using the KL divergence metric and polar descriptors.

orientation and the magnitude of the gradients for pixels inside the spatial bin are computed, and an orientation histogram weighted by the magnitude of the gradients is eventually constructed. Note that the orientation of the feature is subtracted from the computed orientation of the gradients after the rotation of the support region. Similar to the feature orientation computation, the sampling structure on the sphere is taken into account while computing the gradients. Another factor to consider is that the number of samples into a spatial bin also depends on the position of the feature. Thus, a normalization is performed by dividing the histogram values by the number of the samples inside the spatial bin. Both gradient computation and sampling number normalization provides robustness to sampling density differences on the sphere.

Finally, a descriptor vector is formed by concatenating the histograms and normalizing this vector. For the histogram-comparison-based method described later in this paper, the histograms are normalized for each spatial bin independently. It implicitly performs normalization with respect to the number of samples. Note that the proposed descriptor is similar to the log-polar descriptors (GLOH) [2] for planar images. Our descriptor is however improved in order to deal with both the geometry and sampling differences on the surface of the sphere. Finally, the descriptor computation previously described also applies to spherical images defined on equiangular grids, but the mapping using stereographic projection is obviously absent in this case.

B. Feature Matching With Oriented Polar Descriptors

The most common method to compare descriptors is the computation of the L_2-norm between descriptor vectors. It is a simple and fast method, but it is shown that it does not necessarily provide the best matching scores [20]. Histogram-comparison-based methods such as the KL divergence [16] or Earth Mover's distance [20], [21] are shown to provide better matching performance at the price of more computation. The KL divergence method still provides a good tradeoff between accuracy and computation cost. The symmetric KL divergence between histograms h_1 and h_2 is computed as

$$\mathrm{KL}(h_1,h_2)=\sum_{n}^{M}\left(\left(h_1(n)\log\frac{h_1(n)}{h_2(n)}\right)\right)+\sum_{n}^{M}\left(\left(h_2(n)\log\frac{h_2(n)}{h_1(n)}\right)\right) \quad (11)$$

where M is the number of bins in the histograms. We propose to use the KL-divergence for matching the polar descriptors used in this paper. If we denote by D^{ik} the spatial bin of the polar descriptor D for the ith radial division and the kth orientation division, the KL divergence between descriptors D_p and D_q is written as

$$\sum_{ik}\mathrm{KL}\left(D_p^{ik},D_q^{ik}\right).$$

Fig. 7 illustrates some matched features for two omnidirectional images captured by a camera with a parabolic mirror. Note that the geometry adaptive smoothing permits the matching of features with different scaling in different regions of the images.

Finally, it has been shown that the ratio of the best matching score to the second best matching score gives better matching performance compared with checking only the best matching score [2]. This ratio is called the ambiguity factor parameter [10] and is used also in this paper.

C. Rotation-Invariant Matching Criteria for Polar Descriptors

For rotation invariance, the orientation of the features is first estimated, and then, the descriptor kernel for spatial bins is rotated to cope with the orientation differences. This step adds an additional computation that might be particularly critical for low-end sensors. We now exploit an interesting characteristic of our descriptors in order to derive an alternative descriptor that necessitates smaller computational complexity. As the central bin of the descriptor is not divided into orientation bins, the gradient histogram in this bin already captures the orientation of the feature. In addition, for a polar descriptor, the rotation of the descriptor simply corresponds to a shift of the spatial orientation bin indexes and the gradient histogram bin indexes. Hence, the descriptor can capture the distinctive characteristics of the feature even without rotation. A nonoriented version of the polar descriptor is computed as if the orientation of the feature is zero. The correct orientation is then computed by the matching algorithm. We now propose a matching algorithm in order to pair these nonoriented descriptors.

The descriptors are assumed to be computed using our new descriptor but without any orientation so that the first radial bin always points to the north pole. The approximate relative orientation is computed from the central spatial bins of the descriptor. A circular shift is applied to one of the two central bin

Fig. 8. Illustration of the natural test images that are mapped on the synthetic planar patch.

histograms, and the L_1 distance between the histograms is computed. The L_1 distance is particularly efficient for finding a peak orientation. The shift that leads to the smallest distance is chosen as the relative orientation. Formally, shift α is expressed as

$$\alpha = \arg\min_{\alpha} \left[\sum_{n} \left(\left\| D_p^{11}(n) - D_q^{11}(n+\alpha) \right\|_1 \right) \right] \quad (12)$$

where D^{11} denotes the central spatial bin and n is the number of orientation histogram bins. After shift α has been computed, any matching criteria for polar descriptors can be accordingly modified. The L_2 distance, for example, is computed between the descriptors by shifting the indexes of both spatial and orientation bins. In other words, we have

$$\|D_p - D_q\|_2 = \sqrt{\sum_{ik} \sum_{n}^{M} \left(D_p^{ik}(n) - D_p^{i(k-\alpha)}(n+\alpha) \right)} \quad (13)$$

where M is the number of bins in the histograms. If we denote by D^{ik}, the spatial bin of the polar descriptor D for the ith radial division and the kth orientation division.

Note that the number of radial bins are equal to the number of bins in gradient orientation histograms; this leads to easy index shifting. Finally, the KL divergence used for polar descriptors is computed as

$$\sum_{ik} \mathrm{KL} \left(D_p^{ik}, \hat{D}_q^{i(k-\alpha)} \right)$$

where $\hat{D}_q(n) = D_q(n+\alpha)$ and the shift is circular.

V. Experimental Results

A. Feature Detection Performance

We carefully analyze the performance of our new feature detection algorithm when images undergo different rotations or translations in order to highlight the importance of properly handling the sensor geometry. We test the proposed feature detection and the descriptor computation methods on synthetic and natural omnidirectional images. For synthetic images, we create a simple synthetic scene with a 10×10 unit planar patch in it using two 3-D modeling softwares, namely, Blender [22] and Yafray [23]. We map different natural test images (see Fig. 8) onto this patch. The specular and diffuse lightings are allowed for realistic measurements. A virtual parabolic omnidirectional camera then captures the test plane from different positions and orientations. The resolution of the synthetic images is 512×512. Fig. 9 shows some of the resulting images.

We compare the proposed feature detection algorithm to planar SIFT and spherical SIFT algorithms. All tests are repeated for each image patch, and the average performance are

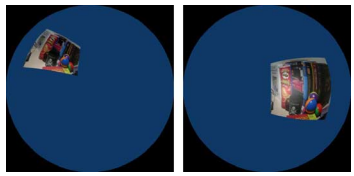


Fig. 9. Two of the synthetic omnidirectional test images. The first image is selected as the reference image for the rotation tests. The second image is selected as the reference image for the translation tests.

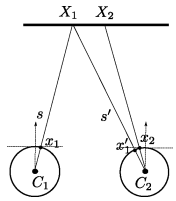


Fig. 10. Mapping of the feature point in the first image to the second image. Distances s and s' are used for the update of the scale in the mapped feature point. Note that x_1, x_2, and x_1' are points mapped on the sphere.

reported. We use the SIFT implementation in [24] for planar SIFT. We modify it to implement our proposed method. We use the implementation in [9] for the comparison with the spherical SIFT method. We select the value of the smoothing interval $k = 2^{1/S}$, where S is the number of scale levels for each octave. We select $S = 3$ and use four successive octaves for all methods. These values are same as the original SIFT method. We use two descriptors for the proposed feature detector. The first descriptor is the novel polar descriptor. We use three radial divisions, eight orientation divisions, and gradient histograms with eight orientation bins. This corresponds to a descriptor vector of length, i.e., $8 * (1 + 2 * 8) = 136$. The second descriptor that is denoted as the virtual camera plane (VCP) is based on SIFT descriptors that are computed on the plane tangent to the sphere at the location of the feature point [10].

We measure the repeatability of the features for different transformations such as rotation and translation. The measure of repeatability is the ratio of repeating features to the number of detected features in a reference image. To find the repeating features, we first map the detected features in the reference image to the destination image and update the scale of the mapped features. Fig. 10 illustrates this mapping. The update of the scale of the mapped feature is performed by multiplying the initial scale σ_1 by the ratio of the distances of the corresponding 3-D point to both cameras. The updated scale $\hat{\sigma_1}$ is then

$$\hat{\sigma_1} = \sigma_1 \frac{s'}{s} \quad (14)$$

where s is the distance to the reference camera and s' is the distance to the destination camera. In the case of no translation, this ratio is equal to $s'/s = 1$. Note that we know the 3-D position and orientation of the cameras, as well as the position of the 3-D planar patch. Thus, we are able to map each feature in the reference image to the destination image.

For omnidirectional images, there is a scale change due to the nonuniform sampling density in addition to the scale changes due to the camera movement. The proposed feature detection algorithm takes this scale change into account, unlike the planar

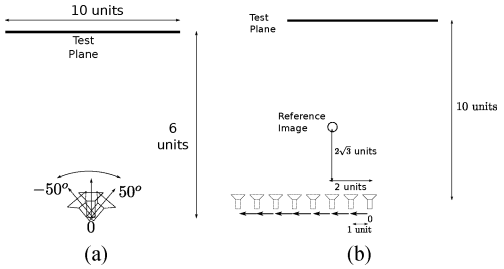


Fig. 11. (a) Rotation around one axis. (b) Translation parallel to the test plane.

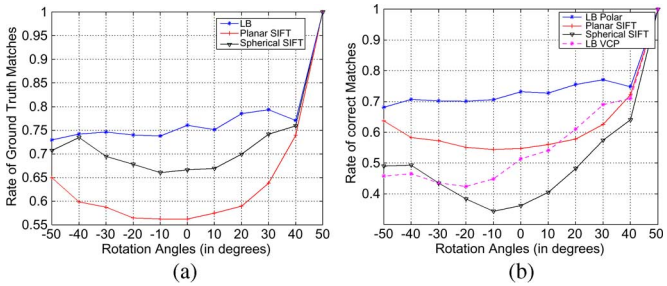


Fig. 12. (a) Ground -truth and (b) correct match rates for the rotation tests.

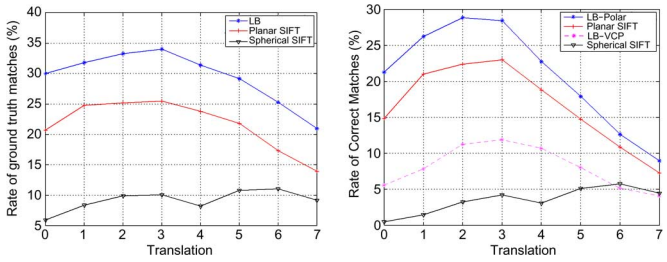


Fig. 13. (Left) Ground-truth and (right) correct match rates for the translation tests.

SIFT method. For a fair comparison, however, the mapped scale $\hat{\sigma_1}$ for features computed with planar SIFT is adapted as

$$\hat{\sigma_1} = \sigma_1 \frac{s'}{s} \frac{\gamma}{\gamma'} \tag{15}$$

where $\gamma = \sqrt{(4 + x^2 + y^2)^2/16}$. The gamma scaling factor is obtained from (8), and γ' is computed on the destination image.

After mapping, we decide on the correct matches by checking the scale ratios and the distance between the positions of the mapped feature point and the feature points on the other image. The condition on the feature position distance is given by

$$\|\mathbf{x_1}' - \mathbf{x}_2\| \leq \delta_d \tag{16}$$

where $\mathbf{x}'_1$ denotes the mapped point and the distance metric is the L_2 norm. The distance threshold δ_d is adapted to capture the matches with different scale levels and is computed as

$$\delta_d = \delta_0 \hat{\sigma_1} \tag{17}$$

where δ_0 is the distance threshold parameter.

The matching condition on the scale ratio is

$$\left| \frac{\min(\hat{\sigma_1}, \sigma_2)}{\max(\hat{\sigma_1}, \sigma_2)} \right| \leq \delta_s \tag{18}$$

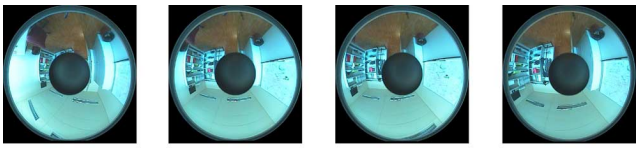
where δ_s is the scale ratio threshold.

Equipped with the correct match conditions from (16) and (18), we measure the repeatability for rotation and translations, as illustrated in Fig. 11. We first consider rotations in SO(3) around the camera center. We apply rotations around the Y-axis to the camera looking toward the Z-axis. The camera captures images for rotation values changing from -50 to $50°$ with $10°$ intervals. The image capture is performed from two different positions, namely, six and eight units away from the test plane in the virtual synthetic environment. The image corresponding to the rotation of $-50°$ is selected as the reference image. For each rotation value, we compare the features in the reference image to the features in the image corresponding to the rotation value. We apply the distance and the scale ratio conditions for each feature in the reference image. The feature pairs satisfying these two conditions are considered as ground-truth matches. Note that we do not apply any descriptor matching method for the ground-truth match computation. The ground-truth matches give the performance of the feature detection algorithm independently of the descriptor quality. Fig. 12(a) shows the performance for rotation tests using the ground-truth matches with our proposed feature detection method (LB), the spherical SIFT, and the planar SIFT. The rate of the ground-truth matches is the number of ground-truth matches divided by the number of features in the reference image. It can be seen that with the proposed method most of the features in the reference image are detected again in the other images independently of the rotation value. The proposed method outperforms both the planar SIFT and spherical SIFT algorithms. The spherical SIFT, similar to the proposed method, preserves the scale when a rotation in SO(3) is applied. The planar SIFT is however highly affected by the rotation changes.

We then perform the matching using the computed descriptors at the feature points. Fig. 12(b) shows the performance of the proposed method for polar descriptor (LB-Polar) and for the tangent plane descriptor (LB-VCP) [10]. The proposed method with the proposed polar descriptor performs best compared with the planar SIFT and spherical SIFT algorithms. It also performs better than the SIFT descriptor computed on the tangent plane due to its geometry adaptiveness and its ability to compute the descriptor without any extra interpolation.

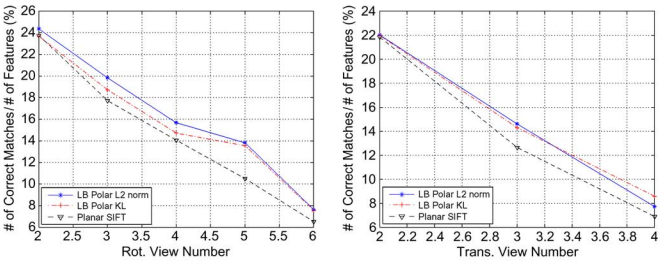
Next, we apply translation on the cameras. The displacement of the camera consists of a translation along the X-axis when the camera is facing the test plane. Seven different displacements with one unit intervals are performed. Similar to the rotation case, the proposed descriptor captures better the characteristics of the feature with its sampling awareness and geometry adaptation. Fig. 13 shows the ground-truth rate and the correct match rate for the translation test. Note that the spherical SIFT poorly performs particularly for the translation test due to the change of resolution introduced during mapping to the equiangular grid.

Finally, we test the performance of the proposed feature detection method on natural omnidirectional images. We capture two sets of omnidirectional images. The first set is composed of images captured by rotating cameras and includes six images with the resolution of 1024×1024. The second set has four



(a) (b) (c) (d)

Fig. 14. Some of the input natural test images. (a) and (b) are for the rotation tests, and (c) and (d) are for the translation tests.



(a) (b)

Fig. 15. Correct match performance on the real images for (a) different rotation values and (b) translation values.

omnidirectional images with the resolution of 1024×1024 captured by a camera moving in the direction of its optical axis (see Fig. 14). Furthermore, each camera pair is calibrated so that the essential matrix, the relative orientation, and the translation between the cameras are known.

In order to measure the matching performance, we check the conditions for the correct matches given in (16) and (18). We use the view illustrated in Fig. 14(a) as the reference image and perform the matching with respect to the features in this image for the rotation tests. Similarly, the image shown in Fig. 14(c) is used as the reference for the translation tests. We compare the proposed method to the planar SIFT method. We also test the matching performance for both L_2-norm and KL-divergence metrics. For a fair comparison, we arrange the scaling levels so that the average number of detected features are close for both the proposed method and the planar SIFT. We set the ambiguity factor [10] to 1.5 for all three methods and compare the correct matching performances.

Fig. 15(a) shows the performance of the two methods for the rotating cameras. It illustrates the ratio of the correct matches for different rotation values with respect to the number of features in the reference image. The view numbers correspond to the different images under test. The proposed method detects more correct matches than the planar SIFT method. The results suggest that the proposed feature detection and matching method is robust to rotation. The performance gain becomes more significant as the angular difference between the camera orientations increases.

Finally, Fig. 15(b) illustrates the correct match performance for the translating cameras. Similar to the rotating cameras, it illustrates the correct match ratio. The results suggest that the proposed method provides more robustness to translation than the planar SIFT method. The KL divergence, however, does not improve the matching performance for the natural omnidirectional images because the number of pixels for each spatial bin in the support regions is not large enough to form a precise histogram at the position where the KL divergence is computed.

Fig. 16. Some synthetic omnidirectional test images for the matching performance test.

B. Matching Performance

We now study the performance of the proposed oriented and nonoriented polar descriptors for synthetic omnidirectional images. The depth information for each image is available so that a homography can be computed in order to define the ground-truth information. For omnidirectional images given by parabolic mirrors, two different types of transformations are applied, and the performance of the descriptors is measured by computing recall and precision. The first transformation is rotation around the Y-axis (i.e., the axis that points up and is aligned with the image coordinate system). The camera is rotated from 0° to 50° with 10° intervals. The second transformation is the translation along the optical axis with one unit intervals reaching up to the translation of four units. We create 512×512 parabolic omnidirectional images for both transformation types. Fig. 16 shows some of the resulting test images. Interested readers can refer to [25] for tests on some other types of transformations and tests on spherical images.

We compute the scale-space representation and detect features using the algorithm proposed in Section III. When features are detected, we compute descriptors on the omnidirectional planar images using the proposed descriptor, the SIFT descriptor, and the VCP approach. For the SIFT descriptor, the size of the descriptor window is adapted in order to compensate for the nonuniform sampling density and provide a fair comparison. The size of the support regions is multiplied by the norm of the metric computed at the feature point. We also compare the proposed descriptor to the planar SIFT descriptor computed at the feature points detected by the planar SIFT feature detection method (denoted by SIFTF). For a fair comparison, we have considered multiple matches with the same position but different orientations as one match. As we know the homography between the images, we compute the ground-truth matches and correct matches by checking the conditions for scale ratio and distance between the feature pairs given in Section V-A. For the matching tests, we set the distance threshold to 0.5 pixels and the scale ratio threshold to 0.9. We compute the recall and precision performance for translation and rotation matches. We form recall versus 1-precision graphs by sweeping the ambiguity factor [10] from 1 to 8. Note that the recall is computed as

$$\text{recall} = \frac{\text{correct matches}}{\text{ground-truth matches}}. \tag{19}$$

The 1-precision is computed as

$$\text{1-precision} = 1 - \frac{\text{correct matches}}{\text{all matches}}. \tag{20}$$

The performance measure of the 1-precision factor is described by the proximity of points to the upper-left corner of the graph.

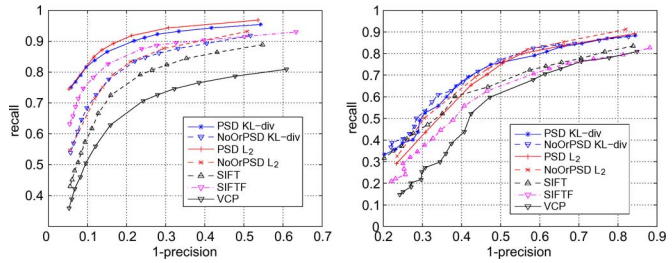


Fig. 17. Recall versus 1-precision for the (left) rotation and (right) translation tests. SIFTF denotes the planar SIFT for which both detection and descriptor computation are based on the SIFT method.

Fig. 17 shows the recall versus 1-precision graphs for the rotation and translation tests. For rotation tests, the proposed descriptor outperforms both SIFT-based descriptors computed at the feature points. The feature detection algorithm, together with the proposed descriptor, also performs better than the planar SIFT method with detection based on the planar SIFTF detection method. The tests show that the feature matching based on KL-divergence-based feature matching does not provide an extra performance gain compared with matching based on the Euclidean distance for omnidirectional images. As expected, the nonoriented version of the descriptor with the proposed rotation-invariant feature matching method is slightly worse than the oriented descriptor but still better than both the SIFT and VCP descriptors [10]. It performs similarly to the SIFTF method. Similarly, the translation tests show that the proposed descriptor is more capable of capturing the feature characteristics and the proposed detection algorithm, together with the polar descriptor, outperform the planar SIFT method. Interestingly, the nonoriented descriptor performs better than the full descriptor in translation tests. This is due to an implicit extra constraint imposed by the nonoriented descriptor and to the type of transformation under consideration, which introduces almost no rotation.

C. Feature Matching With Hybrid Cameras

We finally test the performance of matching with hybrid cameras on synthetic planar, omnidirectional, and spherical images captured at the same camera positions. Fig. 18 shows the test images from different types of cameras. Gaussian kernels for planar images and Gaussian functions computed in terms of spherical harmonics for spherical images are optimized solutions for the heat diffusion equation to form scale-space representations for planar and spherical images, respectively. The method in this paper rather proposes a direct solution to heat diffusion equation using an iterative numerical solution. As all these three methods solve the same heat diffusion equation but with different methods, these three scale-invariant feature detection methods can be combined in order to perform hybrid matching by computing a common type of descriptor. The proposed descriptor is defined on the sphere and can be used to perform hybrid matching between omnidirectional and spherical images by using their optimized scale-invariant feature detection method. That is, the features on the spherical images are detected using the spherical SIFT method, and the features on the omnidirectional images are computed using the proposed

Fig. 18. Planar, omnidirectional, and spherical test images for hybrid matching tests.

TABLE I
MATCHING PERFORMANCE FOR HYBRID MATCHING BETWEEN OMNIDIRECTIONAL AND SPHERICAL IMAGES

	Polar KL	Polar L_2	SIFT
Correct Matches	127	123	104
Ground Truth Matches	274	274	245
Detected Features (Omni. Image)	437	437	462
Detected Features (Spherical Image)	1288	1288	1552

TABLE II
MATCHING PERFORMANCE FOR HYBRID MATCHING BETWEEN OMNIDIRECTIONAL AND PLANAR IMAGES

	Polar KL	Polar L_2	SIFT
Correct Matches	27	28	10
Ground Truth Matches	106	106	24
Detected Features (Omni. Image)	437	437	676
Detected Features (Planar Image)	541	541	544

method. The descriptors for each type of methods are then computed using the method proposed in this paper, and the matching is performed. The polar descriptor can be used for planar images as well. For planar images, the descriptor is modified using the small angle approximation. That is, the planar image is assumed to be tangent to a sphere with a very large radius so that the small angle approximation holds. In this case, the radial bins have equal ranges as $\sin(\alpha) \approx \alpha$.

We compare the performance to the planar SIFT method applied to the three types of images. We compute ground-truth and correct matches by checking the distance condition explained in Section V-A. As the images are synthetically obtained, the depth and homography information are known. In addition, there is no scale change among the images because all three images are captured from the same camera position. The features on the omnidirectional images are mapped onto the planar images to measure the distance between the positions of the features. Similarly, the features on the omnidirectional images are mapped onto the sphere to check the distance condition between features in the omnidirectional image and features in the spherical image. We compare the performance of the L_2 and KL-divergence matching criteria during the tests. Table I shows the matching performance between the omnidirectional and spherical images. The ambiguity factor is set to 1 so that the matches with the best distance score are selected. The results show that geometry-adaptive feature detection with the proposed descriptor performs better than applying SIFT to all types of images. Table II shows the results for the hybrid matching test between omnidirectional and planar images. A similar performance is observed for this test that favors the geometry-adaptive feature detection and the proposed descriptor.

D. Discussion

The experiments on different transformations and manifolds have shown that the proposed approach can capture the scale space better than the planar SIFT and the spherical SIFT. As the images are directly processed on the planar image plane but with a formulation on the sphere, the true geometry of the images are preserved. Detection performance tests have proven that our approach based on the Riemannian geometry has better repeatability than the other two methods. Similarly, the new log-polar descriptor has better discrimination than VCP, planar SIFT, and local spherical descriptor due to the gradient computations that adapted the geometry of the mirror. Finally, since the gradient estimation in the proposed descriptor is adaptable to different types of manifolds, it is possible to match features from different imaging systems and to perform matching in hybrid camera systems.

We have also proposed simplified descriptors where the central region of the polar descriptor is not divided into spatial radial bins and the orientation of the descriptor is computed from the gradients in this region. If the number of regions is large enough, this results into the same descriptor orientation as the computation from the whole descriptor region. The descriptor uses index shifting based on a eight-bin histogram for computing the gradient orientation. This gives only a 45° resolution, but the gradient-histogram-based descriptors are known to be robust to small orientation changes. Nevertheless, both the central region assumption and the eight-bin index shifting used in the proposed simplified descriptor cause some degradation in the performance for rotation tests, but the degradation is only marginal.

Finally, since both the scale-space computation and the gradient computations are based on simple operations, the complexity of our proposed method is comparable with the planar SIFT method. As introduced in Algorithm 1, the scale-space smoothing is composed of multiple simple metric computations, subtraction, and image addition operations. Other feature detection steps are similar to the planar SIFT method. To be fully complete, some additional complexity is added in the descriptor computation with the scaling factor driven by the geometry metric; compared with the other steps such as histogram computation, this scaling process, however, has a marginal influence in terms of computational complexity.

VI. Conclusion

We have proposed a scale-invariant feature computation framework for omnidirectional images. We have exploited the foundations of the Riemannian geometry to formulate the scale-space analysis and a feature detection framework that directly works on the original image plane without the 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 the invariance of the features with respect to rotation in SO(3) and to competitive performance in the case of translation between images.

We have also proposed a polar descriptor that takes the specific geometry and nonuniform sampling grids into consideration. We have implemented a matching method that can successfully match nonoriented polar descriptors. The complexity on the descriptor computation phase is reduced in this case, which makes it particularly interesting for mobile applications. Finally, we have shown that our novel framework can be used to match features in images captured with different types of cameras. Our framework provides a promising solution for feature detection applications in both omnidirectional and hybrid cameras where it outperforms the state-of-the art methods in terms of both detection and matching performance.

References

[1] D. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, Nov. 2004.

[2] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," *IEEE Trans. Pattern Anal. Mach. Intell.*, 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. Gool, "A comparison of affine region detectors," *Int. J. Comput. Vis.*, vol. 65, no. 1/2, pp. 43–72, Nov. 2005.

[4] E. Tola, V. Lepetit, and P. Fua, "A fast local descriptor for dense matching," in *Proc. IEEE Conf. 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-ofattention," *Int. J. Comput. Vis.*, vol. 11, no. 3, pp. 283–318, Dec. 1993.

[6] T. Geodeme, T. Tuytelaars, G. Vanacker, M. Nuttin, and L. V. Gool, "Omnidirectional sparse visual path following with occlusion-robust feature tracking," in *Proc. Workshop OMNIVIS*, 2005.

[7] Y. Bastanlar, L. Puig, P. Sturm, J. Guerrero, and J. Barreto, "DLT-like calibration of central catadioptric cameras," in *Proc. 8th Workshop OMNIVIS*, 2008.

[8] T. Mauthner, F. Fraundorfer, and H. Bischof, "Region matching for omnidirectional images using virtual camera planes," in *Proc. CVWW*, 2006, pp. 93–98.

[9] J. Cruz-Mota, I. Bogdanova, B. Paquier, M. Bierlaire, and J. Thiran, Scale invariant feature transform on the sphere: Theory and applications Transport and Mobility Laboratory, Ecole Polytechnique Fıacute;àale de Lausanne, Lausanne, Switzerland, Tech. Rep., May 2009.

[10] P. Hansen, P. Corke, W. Boles, and K. Daniilidis, "Scale invariant feature matching with wide angle images," in *Proc. IEEE/RSJ Int. Conf. IROS*, 2007, pp. 1689–1694.

[11] P. Hansen, P. Corke, W. Boles, and K. Daniilidis, "Scale-invariant features on the sphere," in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2007, pp. 1–8.

[12] C. Geyer and K. Daniilidis, "Catadioptric projective geometry," *Int. J. Comput. Vis.*, vol. 45, no. 3, pp. 223–243, Dec. 2001.

[13] X. Ying and Z. Hu, "Can we consider central catadioptric cameras and fisheye cameras within a unified imaging model," in *Proc. ECCV*, 2004, vol. 3021, Lecture Notes Computer Science, pp. 442–455.

[14] P. Hansen, W. Boles, and P. Corke, "Spherical diffusion for scale invariant keypoint detection in wide-angle images," in *Proc. DICTA*, Los Alamitos, CA, 2008, pp. 525–532.

[15] H. Bay, A. Ess, T. Tuytelaars, and L. Gool, "Speeded-up robust features (SURF)," *Comput. Vis. Image Understand.*, vol. 110, no. 3, pp. 346–359, Jun. 2008.

[16] V. Chandrasekhar, G. Takacs, D. Chen, S. Tsai, R. Grzeszczuk, and B. Girod, "CHoG: Compressed histogram of gradients," in *Proc. Int. Conf. CVPR*, 2009, pp. 2504–2511.

[17] T. Lindeberg, *Scale-Space Theory in Computer Vision*. New York: Springer-Verlag, 1994.

[18] I. Bogdanova, X. Bresson, J. Thiran, and P. Vandergheynst, "Scale space analysis and active contours for omnidirectional images," *IEEE Trans. Image Process.*, vol. 16, no. 7, pp. 1888–1901, Jul. 2007.

[19] K. Mikolajczyk and C. Schmid, "An affine invariant interest point detector," in *Proc. ECCV*, 2002, pp. 128–142.

[20] H. Ling and K. Okada, "An efficient earth movers distance algorithm for robust histogram comparison," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 5, pp. 840–853, May 2007.

[21] Y. Rubner, C. Tomasi, and L. Guibas, "The earth mover's distance as a metric for image retrieval," *Int. J. Comput. Vis.*, vol. 40, no. 2, pp. 99–121, Nov. 2000.

[22] "Blender," Blender 2.43 [Online]. Available: http://www.blender.org 2007

[23] Yafray, Yafray 0.09 [Online]. Available: http://www.yafray.org 2007

[24] A. Vedaldi, Sift Implementation [Online]. Available: http://www.vlfeat.org/~vedaldi/code/sift.html Sep. 2007
[25] Z. Arican, "Analysis of multiview omnidirectional images in a spherical framework," Ph.D. dissertation, EPFL, Lausanne, Switzerland, 2010.

Zafer Arıcan (S'06–M'10) received the B.Sc. and M.Sc. degrees in electrical and electronics engineering from Middle East Technical University, Ankara, Turkey, in 2002 and 2004, respectively, and the Ph.D. degree in computer and communication sciences from the Swiss Federal Institute of Technology (EPFL), Lausanne, Switzerland, in 2010.

Since 2011, he has been a Senior Researcher with Turk Telekom Research and Development, Orta Doğu Teknik Üniversitesi Teknokent, Ankara, Turkey. His research interests include computer vision, image and video processing, and mobile applications.

Dr. Arıcan was a recipient of the Best Student Paper Award at the IEEE Advanced Video and Signal-Based Surveillance 2007 conference and the IBM Best Student Paper Award at the International Conference on Pattern Recognition 2008.

Pascal Frossard (S'96–M'01–SM'04) received the M.S. and Ph.D. degrees in electrical engineering from the Swiss Federal Institute of Technology (EPFL), Lausanne, Switzerland, in 1997 and 2000, respectively.

Between 2001 and 2003, he was a Member of the research staff with the IBM T. J. Watson Research Center, Yorktown Heights, NY, where he worked on media coding and streaming technologies. Since 2003, he has been with the faculty of EPFL, where he heads the Signal Processing Laboratory (LTS4). His research interests include image representation and coding, visual information analysis, distributed image processing and communications, and media streaming systems.

Dr. Frossard has been the General Chair of the IEEE International Conference on Multimedia and Expo 2002 and Packet Video 2007. He has been the Technical Program Chair of the European Signal Processing Conference 2008 and a member of the organizing or technical program committees of numerous conferences. He has been an Associate Editor of the IEEE TRANSACTIONS ON MULTIMEDIA (2004), the IEEE TRANSACTIONS ON IMAGE PROCESSING (2010), and the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY (2006–2011). He is the Vice-Chair of the IEEE Image, Video, and Multidimensional Signal Processing Technical Committee (2007). He is an elected member of the IEEE Visual Signal Processing and Communications Technical Committee (2006) and the IEEE Multimedia Systems and Applications Technical Committee (2005). He has served as the Vice-Chair of the IEEE Multimedia Communications Technical Committee (2004–2006) and as a member of the IEEE Multimedia Signal Processing Technical Committee (2004–2007). He was a recipient of the Swiss National Science Foundation Professorship Award in 2003, the IBM Faculty Award in 2005, the IBM Exploratory Stream Analytics Innovation Award in 2008, and the IEEE Transactions on Multimedia Best Paper Award in 2011.