SCHOOL OF ENGINEERING - STI
SIGNAL PROCESSING LABORATORY LTS4
Zafer Arıcan and Pascal Frossard

CH-1015 LAUSANNE

Phone: +4121 6934329
Fax: +4121 6937600
e-mail: zafer.arican@epfl.ch

SCALE INVARIANT FEATURES AND POLAR DESCRIPTORS IN OMNIDIRECTIONAL IMAGING

Zafer Arıcan and Pascal Frossard

Ecole Polytechnique Fédérale de Lausanne (EPFL)

Signal Processing Laboratory LTS4 Technical Report

TR-LTS-2010-002

August 24th, 2010

Part of this work has been submitted to the IEEE Transactions on Image Processing.

Scale Invariant Features and Polar Descriptors in Omnidirectional Imaging

Zafer Arıcan and Pascal Frossard
Ecole Polytechnique Fédérale de Lausanne (EPFL)
Signal Processing Laboratory (LTS4), Lausanne, 1015 - Switzerland.
{zafer.arican, pascal.frossard}@epfl.ch
Fax: +41 21 693 7600, Phone: +41 21 693 4329

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 describe the mirror and lens structure 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 any omnidirectional image that can be mapped on the sphere. We also present a new descriptor and feature matching solution for omnidirectional images. The descriptor builds on the log-polar planar descriptors and adapts the descriptor computation to the specific geometry and the non-uniform sampling density of spherical and omnidirectional images. We further propose a rotation-invariant matching method that eliminates the orientation computation during the feature detection phase and thus decreases the computational complexity. Finally, we show that the proposed framework also permits to match features in images with different geometries. Experimental results demonstrate that the new feature detection method combined with the proposed descriptors offers promising performance and improves on the common SIFT features computed on the planar omnidirectional images as well as other state-of-the art methods for omnidirectional images.

Index Terms

scale-invariant features, omnidirectional imaging, Riemannian geometry, polar descriptors

I. Introduction

APPLICATIONS such as camera calibration, object detection, recognition or tracking generally rely on the localization and matching of visual features in multiple images. These features need to be invariant to different transformations and illumination changes so that they can be matched successfully in different images with similar content. Scale invariance is an important characteristic of visual features that permits to be less sensitive to imperfect camera settings and transformations. 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 descriptors and feature detection approaches [2]–[5] that apply to images from classical cameras.

Meanwhile, omnidirectional vision has been an active research field with applications in robotics and surveillance where sensors with large fields of view present several advantages in terms of deployment and maintenance. The omnidirectional cameras typically consist of either a fisheye lens or a catadioptric system with a lens and a mirror with a smooth surface (e.g., parabolic, hyperbolic mirrors). The characteristics of the resulting images is highly dependent on the geometry of the system. This particular geometry also causes partial scale changes in different regions of the same image. 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 should be taken into account for an appropriate processing of the light information. 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].

Recent works such as [9]–[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 an equiangular grid. An 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 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 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.

After 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. Descriptors based on histogram computation provide the best robustness to these transformations. The

SIFT descriptor 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 recently been 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 take the geometry of omnidirectional images into consideration. Furthermore, they assume that the sampling is uniform along the image. For omnidirectional images, however, the sampling density 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; however, an extra interpolation step is performed that may change the true scale of the feature and the gradient values. In another approach [9], the SIFT descriptors are directly computed on the spherical surface. This approach is however affected by different sampling densities as it does not take into account the different number of samples in the spatial bins.

We propose in this paper a novel framework for the computation of scale invariant features in omnidirectional images captured by sensors with specific 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 propose 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 then propose a polar descriptor for scale invariant features on the sphere that builds on log-polar descriptors used for planar images. The new descriptors exploit the geometry of the sphere and take 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 relax the computational complexity in the construction of the descriptors. Tests 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 and the state-of-the-art methods performed on spherical images. Finally, we also show that it is possible to efficiently match features detected in different types of images. Our framework therefore provides a promising solution for feature detection applications in both omnidirectional and hybrid camera networks where it outperforms the state-of-the art methods.

The rest of the 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 non-oriented descriptors is explained in Section IV. Section V discusses the experimental results for feature matching between images from omnidirectional and heterogenous cameras.

II. Scale-Space Analysis on Non-Euclidian Manifolds

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 performed by convolution with Gaussian kernels. For non-euclidian manifolds, however, the Gaussian kernels cannot be applied directly as they violate the shift invariance requirement listed in the scale-space axioms [17] in this case. The heat diffusion equation can however be used to smooth these images and to form the scale-space representation. In this section, we give an overview of the heat diffusion equation and the Riemannian geometry for the computation of differential operators on manifolds. We then give an example of omnidirectional images, namely parabolic omnidirectional images and explain their scale-space representation in a spherical framework. The spherical representation is obtained by using the property that any central catadioptric omnidirectional image can be uniquely mapped onto the sphere [12]. Two other examples namely spherical images and planar images are also discussed briefly.

A. Riemannian Geometry Framework

The 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 as these kernels are not shift invariant on these surfaces. However, one can still compute the scale-space representation $I(x, y, t)$ on non-Euclidian manifolds with help of the heat diffusion equation and differential operators that can be computed on the non-euclidean manifolds. It reads

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

where Δ is the Laplacian operator and t is the scale level. The initial condition is given as $I(x, y, t_0) = I(x, y)$ where the original images is denoted by $I(x, y)$. It can be noted that the Gaussian function with standard deviation $\sqrt{t}$ is the Green's function for the heat diffusion equation (1) on planar images [17].

The heat diffusion 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 [18]. In order to give a brief definition of

these operators, let $\mathcal{M}$ be a parametric surface on $\mathcal{R}^3$ 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}. Note that for sake of simplicity, Einstein notation is adopted for the formulation. Furthermore, the divergence of a vector field $\mathbf{V}$ on $\mathcal{M}$ is given as $div\mathbf{V} = \frac{1}{\sqrt{g}}\partial_i(V^i\sqrt{g})$, 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

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

This operator that corresponds to the Laplace operator on the plane can be used to solve the heat diffusion equation (1) on non-Euclidian manifolds and eventually for scale-space analysis. It has to be noted that the specific form of the Laplace-Beltrami operator depends on the particular geometry of the manifold $\mathcal{M}$ through g.

B. Parabolic Mirror Systems

We now consider the 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 by inverse stereographic projection [12], similarly to images from most simple mirrors and catadioptric systems. This enables easier processing of the parabolic images and enables the capture of the light rays from a central point which is the center of the sphere. Thus, we use the mapping from the parabolic omnidirectional images to the sphere for the computation of the differential operators explained in the previous section. In this case, the manifold $\mathcal{M}$ becomes the sphere. We then derive the metric necessary to the construction of differential operators on the sphere similarly to [18], in order to perform the 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 $\mathcal{S}^2$ in terms of the variables r, θ and ϕ that represent the spherical coordinates. The line element satisfies

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

The corresponding angles for θ and ϕ are given in Figure 1(a). The stereographic projection maps each point on the sphere to a plane $\mathcal{R}^2$ of coordinate (x, y). A point in polar coordinates (R, ϕ) on the stereographic plane is related to the point (θ, ϕ) on the sphere by $R = 2r\tan(\frac{\theta}{2})$ and $\phi = \phi$. Then the terms in the line element of (3) read

$$\begin{aligned} d\theta^2 &= \frac{16r^2}{(r^2+4R^2)^2}dR^2 \\ \sin^2(\theta) &= \frac{16r^2R^2}{(4r^2+R^2)^2} \end{aligned} \tag{4}$$

and the line element, dl^2 is

$$dl^2 = \frac{16r^4}{(4+R^2)^2}(dR^2 + R^2d\phi^2) \tag{5}$$

.

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

$$dl^2 = \frac{16}{(4+x^2+y^2)^2}(dx^2 + dy^2) \tag{6}$$

giving the induced Riemannian metric

$$g_{ij} = \begin{pmatrix} \frac{16}{(4+x^2+y^2)^2} & 0 \\ 0 & \frac{16}{(4+x^2+y^2)^2} \end{pmatrix} \tag{7}$$

and the inverse metric

$$g^{ij} = \begin{pmatrix} \frac{(4+x^2+y^2)^2}{16} & 0 \\ 0 & \frac{(4+x^2+y^2)^2}{16} \end{pmatrix} \tag{8}$$

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

$$\left|\nabla_{\mathcal{S}^2} I\right|^2 = \frac{(4+x^2+y^2)^2}{16}\left|\nabla_{\mathcal{R}^2} I\right|^2 \tag{9}$$

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

$$\left|\Delta_{\mathcal{S}^2} I\right|^2 = \frac{(4+x^2+y^2)^2}{16}\left|\Delta_{\mathcal{R}^2} I\right|^2 \tag{10}$$

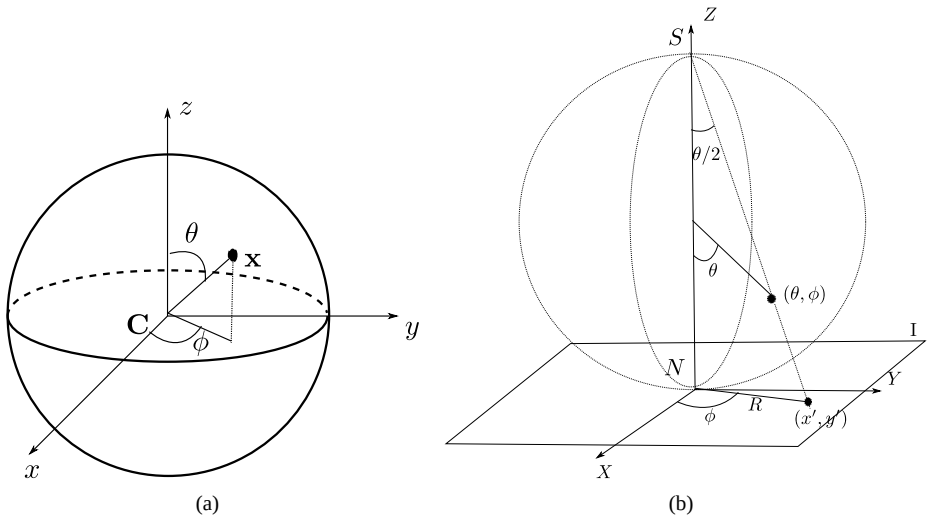


Fig. 1. Geometry of the 2-sphere (a) Spherical polar coordinates and (b) stereographic projection: The plane corresponds to the image plane and the θ is the corresponding colatitude angle on the sphere.

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 Laplace-Beltrami (LB) operator by first computing the Laplacian of the image as if it is a planar image with fast methods and then weighting by the metric to compute the LB operator.

C. Other Imaging Systems

We discuss the induced Riemannian metric for two other imaging manifolds for the sake of completeness. The first example is the planar image. In such a case, the metric is the identity matrix

$$g_{ij} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \tag{11}$$

and thus the LB operator is equal to the Laplacian as expected.

The second example is the spherical manifold with equiangular grid where the image plane becomes the unwarped spherical image. The coordinates of the image plane are in discrete longitude, ϕ and co-latitude angles, θ. In such a scenario, the induced Riemannian metric is

$$g_{ij} = \begin{pmatrix} 1 & 0 \\ 0 & sin^2\theta \end{pmatrix} \tag{12}$$

and the Laplace-Beltrami operator is

$$\Delta I = \frac{cos(\theta)}{sin(\theta)}\partial_\theta I + \partial_{\theta\theta} I + \frac{1}{\sin^2(\theta)}\partial_{\phi\phi} I \tag{13}$$

where ∂_θ is the gradient with respect to θ and $\partial_{\theta\theta}$ is the Laplacian with respect to θ. Note that, the Laplacian might have some numerical instability around the poles. An alternative method based on spherical Fourier transform and gaussian kernels on the sphere has been proposed to perform the smoothing on the sphere [11], [19]. That method is also based on the same heat diffusion equation but aims to provide better numerical stability particularly around the poles.

III. Feature Detection

Equipped with the scale-space representation framework, 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 scale-normalized Laplacian of Gaussian images provide the most stable scale-invariant features [20] and differences 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 base smoothing level, σ_o. We form the scale levels such that scale-normalized difference images are obtained after 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 and we use discrete differential operators on the plane for the computation of the gradient and Laplacian. We use $[-1 \quad 1]/d_s$ as 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. Figure 2 illustrates some smoothed images for as a parabolic omnidirectional images. Note that the central regions is smoothed less than the outer regions. This is due to the non-uniform sampling density on the surface of the sphere caused by the inverse stereographic projection.

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

Fig. 2. (a) Original parabolic omnidirectional image, (b) smoothing $t = 2.25$, (c) smoothing $t = 5.76$, (d) the smoothed image with a Gaussian kernel on the planar image. The resulting smoothing on the sphere is non-uniform in this case

Finally, the images are down-sampled 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 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: smoothing level
d_t: time interval
n_i: 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 the 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: $I(x, y, t) = \bar{I}(x, y)$

Once the scale-space images are formed, the scale-normalized Laplacian of Gaussian images are approximated by the difference of Gaussian images similarly to the SIFT framework [1]. This permits to detect the most important visual features by computing the difference of images of successive smoothing levels. Note however 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

$$L(x, y, t_i) = I(x, y, t_i) - I(x, y, t_{i-1}) \tag{14}$$

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 x 3 pixels in the current difference image $L(x, y, t_i)$ and 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 [20]. The detected feature point is assigned 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 is performed to increase stability and accuracy. First the unstable points are removed. 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 pre-defined threshold are removed. This is a result of heat diffusion based smoothing where high textured regions will have higher difference between smoothed images compared to the low textured regions. Difference images are also used to decide on the edge response. Again, as in [1], edge responses are removed by checking the ratio

between maximum and minimum principle curvatures of the difference image at the feature position and features with a ratio greater than a pre-defined threshold are deleted. Finally, the position of the feature point is refined to a sub-pixel level by fitting a 3D quadratic function to the pixel position and scale. The extrema of the fitted function specifies the refined position of the feature. Figure 3 illustrates the detected features in two omnidirectional images with rotation.

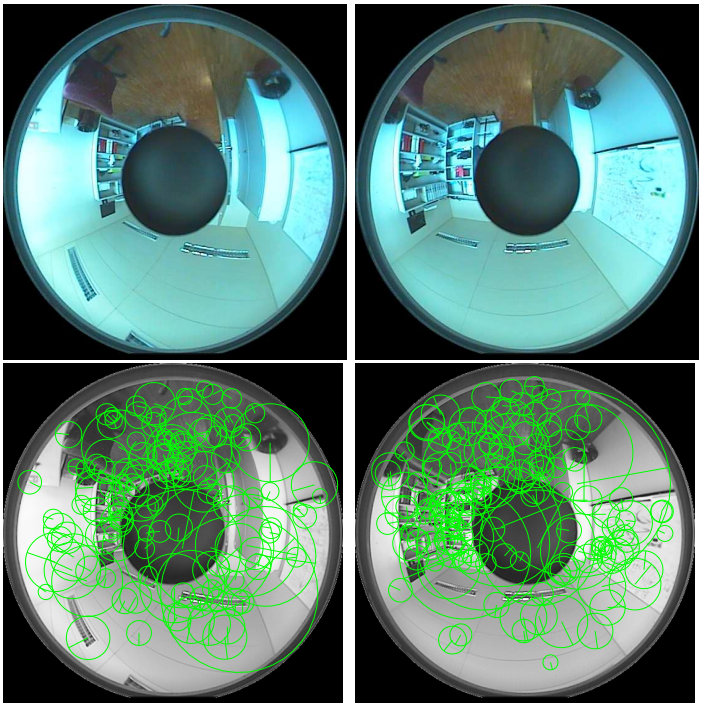


Fig. 3. Original omnidirectional images (Top) and the detected features in these images (Bottom). The diameter of the circles denote the scale level.

IV. Feature Descriptors and Matching

Once scale-invariant features are computed by 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 non-oriented version of this polar descriptor. We also describe a feature matching criteria based on Kullback-Leibler (KL) divergence and propose two matching criteria for non-oriented 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 pixel densities on the surface of the sphere are taken into account. Note that the processing on the sphere enables the computation of descriptors for both spherical and omnidirectional images.

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, σ and the radius of this region is proportional to $\sin(\sigma)$, similarly to the method in [10]. As illustrated in Figure 4, 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. For spherical images with equiangular grid, pixels inside the circular support region on the sphere are used directly.

The support regions are first used for computing the orientation of the feature. For rotation invariance, an orientation is assigned to the visual feature. The orientation of the feature is determined by forming first an 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 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

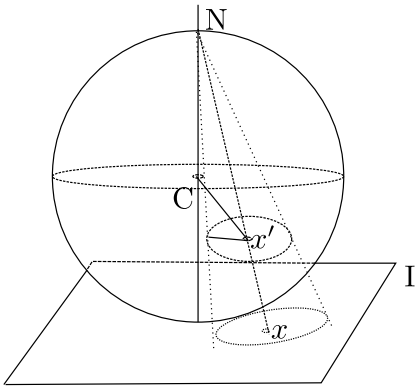


Fig. 4. The circular region around the mapped feature point x' on the sphere is mapped to the omnidirectional image plane I by stereographic projection from the north pole, N

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, however, is 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 depending on this distance. Recall that for omnidirectional images, the gradients are computed using the induced Riemannian metric explained in Section II

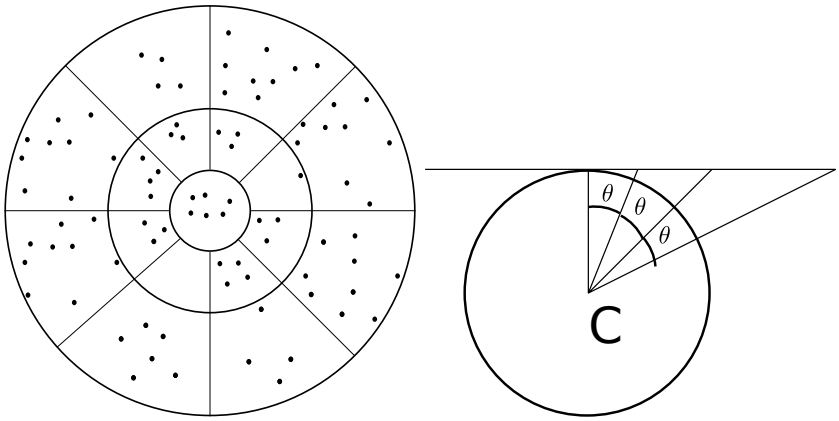


Fig. 5. The log-polar descriptor and non-uniform samples inside spatial bins (left). Radial bin divisions correspond to equal division of θ on the sphere (right).

Once an orientation is assigned to a visual feature, a support region around the same feature point but with a bigger size is formed again proportionally 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]. Figure 5 (left) shows the descriptor structure with 8 orientation bins and 3 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 and correspond to the geometry of the sphere. In addition, the selection of increasing radius values for the radial bins is achieved by dividing the latitude angles uniformly. Figure 5 (right) illustrates this relation. Similarly to the orientation computation, the support region and the subdivisions for the descriptor computation are mapped onto the omnidirectional plane using stereographic projection. For spherical images with equiangular grid the pixels in the support regions are used directly.

For rotation invariance, the support region and the subdivisions are rotated on the sphere around the feature point using the assigned orientation of the feature. For each spatial bin, the 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 rotation of the support region. Similarly 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 histogram-comparison based method described later in the paper, the histograms are normalized for each spatial bin and concatenated to the descriptor vector afterwards. The latter approach 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.

B. Feature matching with oriented polar descriptors

The most common method to compare descriptors is the L_2-norm between the two descriptor vectors. It is a simple and fast method but it is shown that it does not necessarily provide the best matching scores [21]. Histogram-comparison based methods such as the Kullback-Leibler (KL) divergence [16] or Earth Mover's distance (EMD) [21], [22] are shown to provide better matching performance at the price of more computation. The KL divergence method still provides a good trade-off between accuracy and computation cost. The symmetric KL divergence between the histograms h_1 and h_2 is computed as

$$KL(h_1, h_2) = \sum_{n}^{M} \left((h_1(n) log \frac{h_1(n)}{h_2(n)}) \right) + \sum_{n}^{M} \left((h_2(n) log \frac{h_2(n)}{h_1(n)}) \right) \tag{15}$$

where M is the number of bins in the histograms.

We propose to use the KL-divergence for matching the polar descriptors proposed in this paper. If we denote by D^{ik} the spatial bin of the polar descriptor D for the i^{th} radial division and the k^{th} orientation division, the KL divergence between descriptors D_p and D_q is $\sum_{ik} KL(D_p^{ik}, D_q^{ik})$.

It has been shown that the ratio of the best matching score to the second best matching score gives better matching performance compared to checking only the best matching score [23]. This ratio is called the ambiguity factor parameter [10] and used also in this paper.

Figure 6 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.

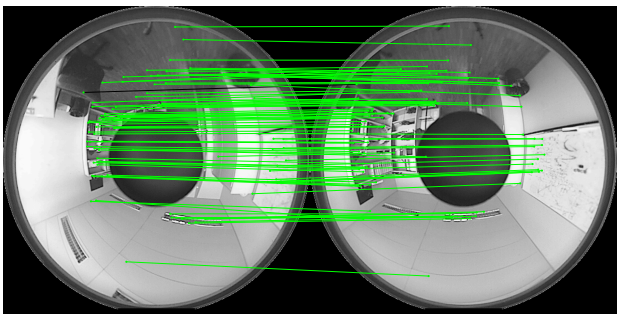


Fig. 6. The matched features using the KL divergence metric and polar descriptors.

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 corresponds simply to a shift of the spatial orientation bin indexes and gradient histogram bin indexes. Hence, the descriptor can capture the distinctive characteristics of the feature even without rotating the descriptor. A non-oriented 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 non-oriented 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 histograms and the L_1 distance between the histograms is computed. The L_1 distance is particularly efficient for finding a peak orientation. The amount of shift with the smaller distance is chosen as the relative orientation. Formally, the shift α is expressed as

$$\alpha = \underset{\alpha}{\operatorname{argmin}} \left[\sum_{n} (\left\| D_p^{11}(n) - D_q^{11}(n+\alpha) \right\|_1) \right] \tag{16}$$

where D^{11} denotes the central spatial bin.

After the shift α has been computed, any matching criteria for polar descriptors can be modified accordingly. 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

$$\|D_p - D_q\|_2 = \sqrt{\sum_{ik}\sum_{n}^{M}\left(D_p^{ik}(n) - D_p^{i(k-\alpha)}(n+\alpha)\right)} \tag{17}$$

The KL divergence used for polar descriptors is computed as $\sum_{ik} KL(D_p^{ik}, \hat{D}_q^{i(k-\alpha)})$ where $\hat{D}_q(n) = D_q(n+\alpha)$ and the shift is circular.

V. Experimental results

A. Feature Detection Performance

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 10x10 unit planar patch in it. We map different natural test images given in Figure 7 onto this patch. A virtual parabolic omnidirectional camera then captures the test plane from different positions and orientations. The resolution of the synthetic images is 512×512. Figure 8 shows some of the resulting images.

We compare the proposed feature detection algorithm to planar SIFT and spherical SIFT algorithms. We select the value of the smoothing interval $k = 2^{1/3}$ and use 4 successive octaves for all methods. We use two descriptors for the proposed feature detector. The first descriptor is the novel polar descriptor. We use 3 radial divisions, 8 orientation divisions and gradient histograms with 8 orientation bins. This corresponds to a descriptor vector of length, $8 * (1 + 2 * 8) = 136$. The second descriptor which is denoteds as VCP is based on SIFT descriptors that are computed on the plane tangent to the sphere at the location of the feature point [10].

Fig. 7. Illustration of the natural test images that are mapped on the synthetic planar patch.

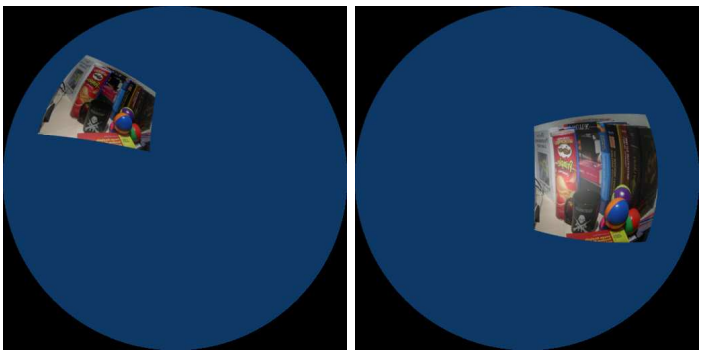


Fig. 8. Two of the synthetic parabolic omnidirectional test images. The first image is selected as the reference image for the rotation tests. The second image is selected as the reference image for the translation tests.

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. Figure 9 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 3D point to both cameras. The updated scale $\hat{\sigma_1}$ is then

$$\hat{\sigma_1} = \sigma_1 \frac{s'}{s} \tag{18}$$

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 $\frac{s'}{s} = 1$. Note that we know the 3D position and orientation of the cameras as well as the position of the 3D 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 non-uniform sampling density in addition to the scale changes due to camera movement. The proposed feature detection algorithm takes this scale change into account in contrary to the planar SIFT method. For a fair comparison, however, the mapped scale $\hat{\sigma_1}$ for features computed with planar SIFT is adapted as

$$\hat{\sigma_1} = \sigma_1 \frac{s'}{s} \frac{\gamma}{\gamma'} \tag{19}$$

where $\gamma = \sqrt{\frac{(4+x^2+y^2)^2}{16}}$. The gamma scaling factor is obtained from (10). 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 for the feature position distance is given by

$$||\mathbf{x_1}' - \mathbf{x_2}|| \leq \delta_d \tag{20}$$

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 computed as

$$\delta_d = \delta_0 \hat{\sigma_1} \tag{21}$$

where δ_0 is the distance threshold parameter.

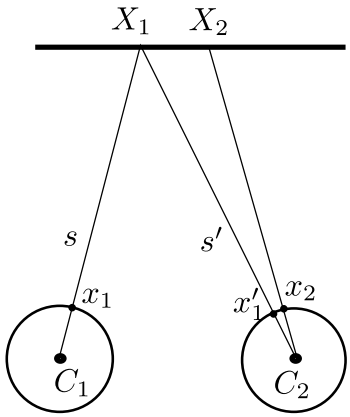


Fig. 9. The mapping of the feature point in the first image to the second image. The distances s and s' are used for the update of the scale in the mapped feature point.

The matching condition on the scale ratio is

$$\left| \frac{\min(\hat{\sigma_1}, \sigma_2)}{\max(\hat{\sigma_1}, \sigma_2)} \right| \leq \delta_s \tag{22}$$

where δ_s is the scale ratio threshold.

Equipped with the correct match conditions (20) and (22), we measure the repeatability first for rotations in SO(3) around the camera center. We apply rotations around the Y axis to the camera looking towards the Z axis. The camera captures images for rotation values changing from -50 to 50 degrees with 10 degrees intervals. The image capture is performed from two different positions, namely 6 and 8 units away from the test plane in the virtual synthetic environment. The image corresponding to the rotation of -50 degrees 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. Figure 10(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, similarly 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. Figure 10(b) shows the performance of the proposed method for polar descriptor (LB-Polar) and with tangent plane descriptor (LB-VCP) [10]. The proposed method with the proposed polar descriptor performs best compared to planar SIFT and spherical SIFT algorithms. It also performs better than the SIFT descriptor computed on the tangent plane by its geometry adaptiveness and its ability to compute the descriptor without any extra interpolation.

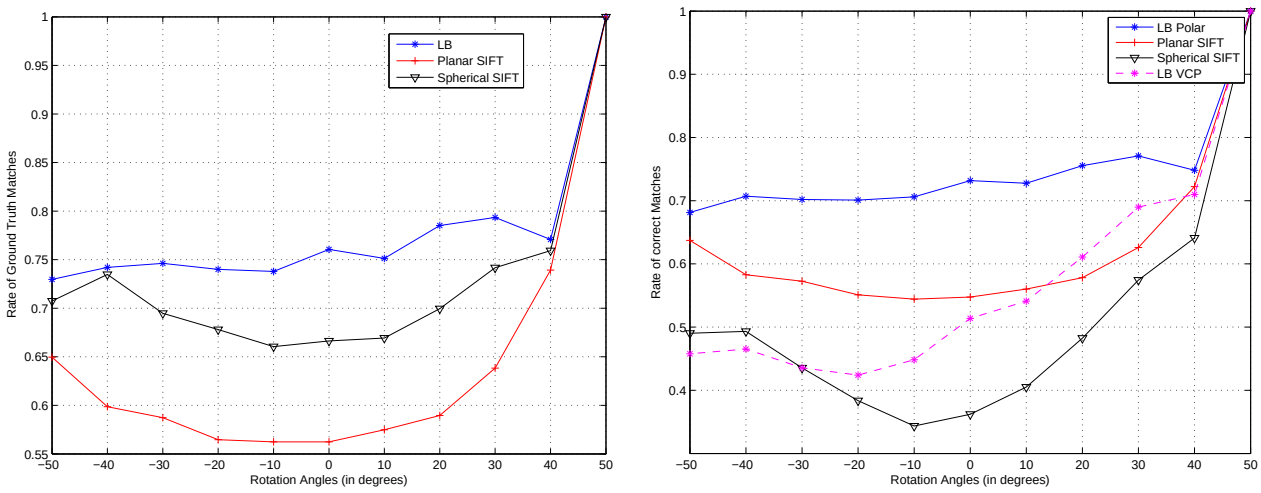


Fig. 10. (a) Ground truth and (b) correct match rates for the rotation tests.

Second, we apply two different types of translation on the cameras. The first translation 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. The second translation type includes both rotation and translation and the translation is performed towards the test plane in order to test scaling effects. The rotation is from 0 to 30 degrees with 10 degrees intervals. The translation reaches up to 4 units with 1 unit intervals. The image corresponding to 30 degrees rotation and 4 units translation is selected as reference. Figure 11 illustrates the two translation schemes.

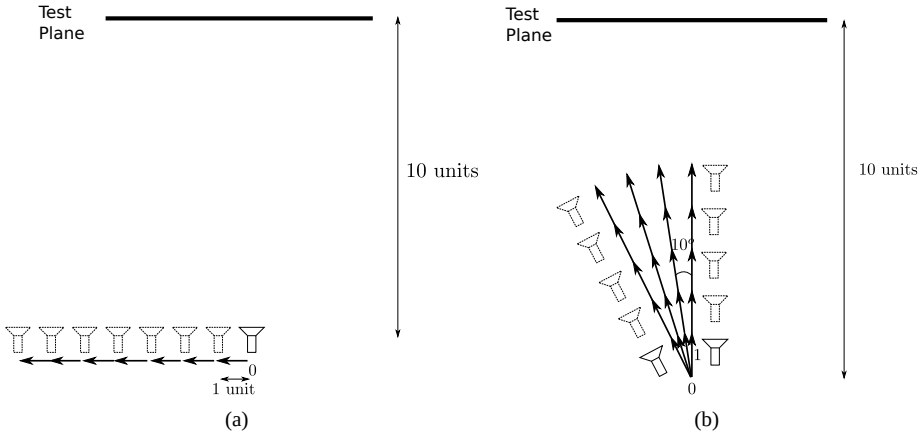


Fig. 11. (a) Translation parallel to the test plane, (b) Translation towards the test plane with rotation

For both types of translation, the proposed feature detection method and the descriptor outperforms the planar SIFT and spherical SIFT. Similarly to the rotation case, the proposed descriptor captures better the characteristics of the feature with its sampling-awareness and geometry adaptation. Figure 12 shows the ground truth rate and the correct match rate for the first type of translation. Similarly, Figure 13 illustrates the repeatability performance for the second type of translation. Note that the spherical SIFT performs poorly particularly for the translation tests 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 set of omnidirectional images. The first set is composed of images captured by rotating cameras and includes 6 images with the resolution of 1024×1024 (See Figure 14). The second set has 4 omnidirectional images with the resolution of 1024×1024 captured by a camera moving in the direction of its optical axis(See Figure 15). Furthermore, each camera pair is calibrated so the essential matrix, the relative orientation and the translation between the cameras are known. Figure 16 illustrates the epipolar great circles computed with the essential matrix for some of the features in the reference image of the translation tests.

To measure the performance, we check the conditions for the correct matches given in (20) and (22) to detect the correct matches similarly to the synthetic scenes. We use the view illustrated in Figure 14(a) as the reference image and perform the

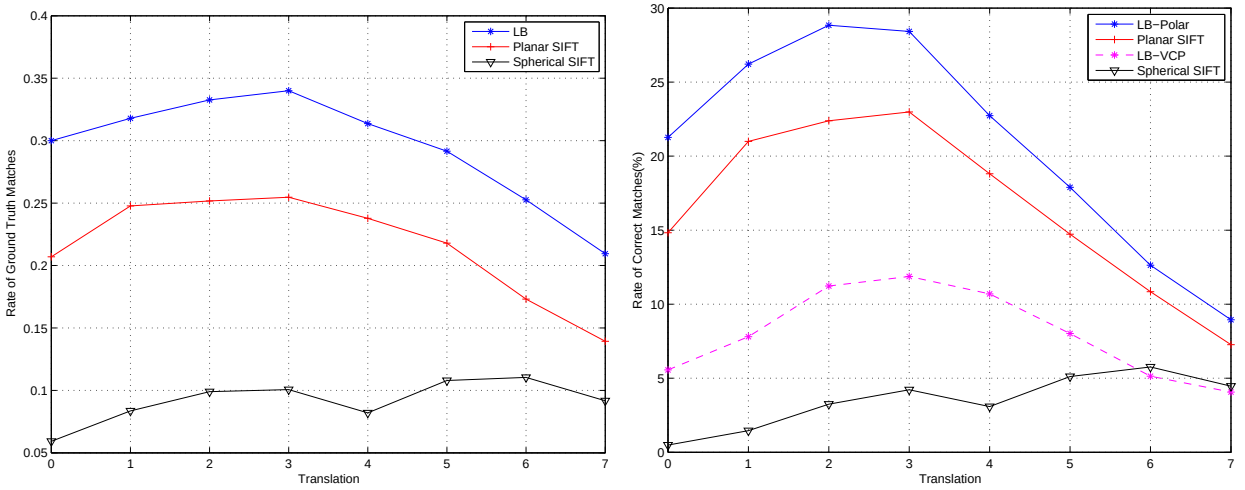


Fig. 12. Ground truth (Left) and correct match (Right) rates for the first type of translation.

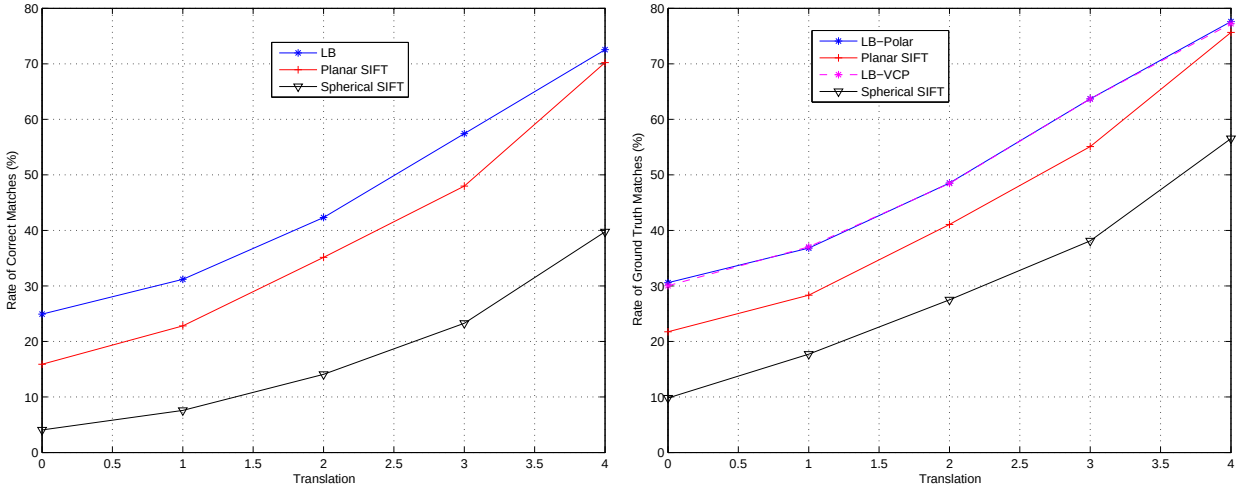


Fig. 13. Ground truth (Left) and correct match (Right) rates for the second type of translation.

matching with respect to the features in this image for the rotation tests. Similarly, the image shown in Figure 15(a) is used as the reference for the translation tests. We compare the proposed method to 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. Table I shows the average number of detected features for both the translation and rotation tests. We set the ambiguity factor to 1.5 for all three methods and compare the correct matching performances.

	LB Polar	Planar SIFT
Rotation	1948	1964
Translation	1980	2079

TABLE I
AVERAGE NUMBER OF DETECTED FEATURES FOR THE ROTATION AND THE TRANSLATION TESTS OF THE NATURAL OMNIDIRECTIONAL IMAGES.

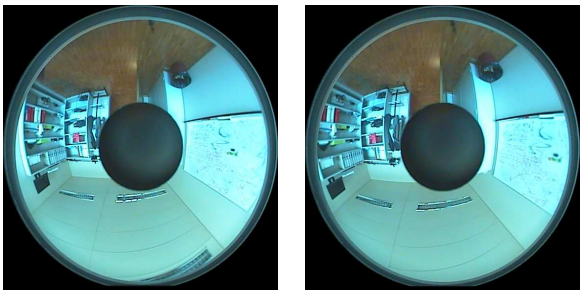
Figure 17 shows the performance of the two methods for the rotating cameras. Figure 17(a) illustrates the ratio of the correct matches for different rotation values with respect to the number of features in the reference image. The view numbers correspond to the different images under test. The proposed method detects more correct matches than the planar SIFT method. Also the proposed descriptor method provides more accurate matching than the planar SIFT method as shown in Figure 17(b). 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, Figure 18 illustrates the correct match performance for the translating cameras. Similarly to the rotating cameras, Figure 18(a) illustrates the correct match ratio. The results suggest that the proposed method provides more robustness to

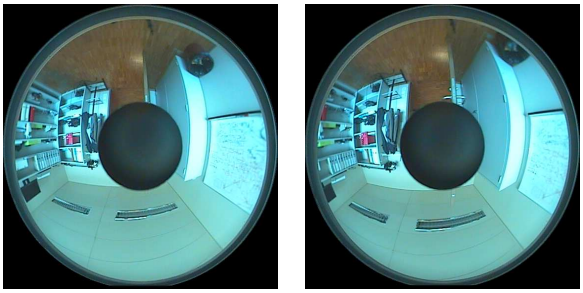
(a) Rot. View 1 (b) Rot. View 2 (c) Rot. View 3

(d) Rot. View 4 (e) Rot. View 5 (f) Rot. View 6

Fig. 14. Natural test images for the feature detection tests for rotating cameras.



(a) Trans. View 1 (b) Trans. View 2



(c) Trans. View 3 (d) Trans. View 4

Fig. 15. Natural test images for the feature detection tests for translating cameras.

translation than the planar SIFT method. As illustrated in the Figure 18(b), the proposed descriptor also improves the matching performance. The KL-divergence, however, does not improve the matching performance for the natural omnidirectional images because the number of pixels for each spatial bins in the support regions is not large enough to form a precise histogram where the KL-divergence is based.

B. Matching Performance

We now study the performance of the proposed oriented and non-oriented polar descriptors for synthetic omnidirectional images and spherical images on an equiangular grid. The depth information for each image is available so that an homography can be computed in order to define the groundtruth information.

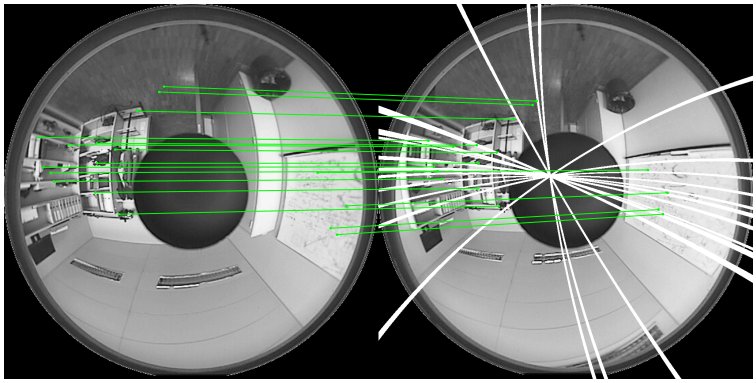
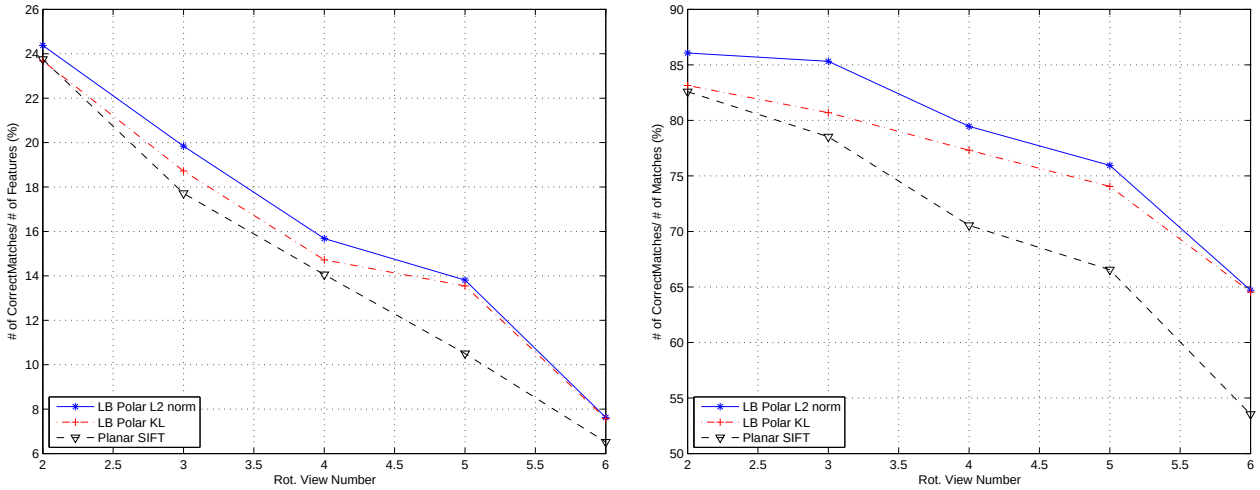


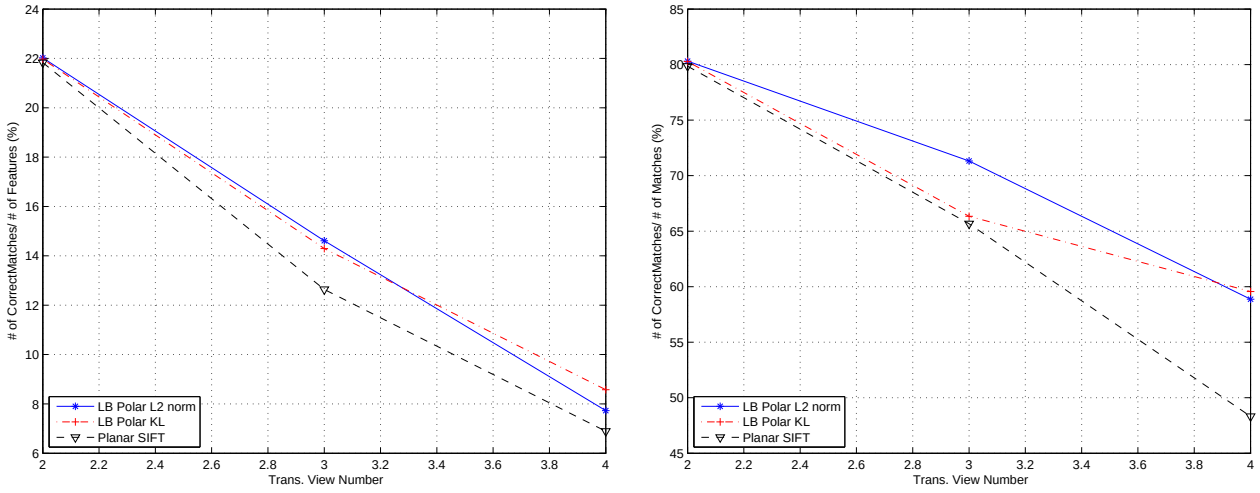
Fig. 16. The great circles corresponding to some features on the reference image for translation. Note that the epipole is close to the optical axis as expected.



(a) Ratio of the correct matches to the number of features in the reference view

(b) Ratio of the correct matches to the number of matches

Fig. 17. Correct match performance on the real images for different rotation values.



(a) Ratio of the correct matches to the number of features in the reference view

(b) Ratio of the correct matches to the number of matches

Fig. 18. Correct match performance on the real images for different translation values.

For omnidirectional images given by parabolic mirrors, 4 different types of transformations are applied and the performance of the descriptors is measured by computing recall and precision. Two of the transformations represent translations on different axes and the other transformations are rotations on different axes. For the first translation test, the camera is moved in one

direction with 1 unit intervals reaching up to the translation of 4 units. For the second translation test, the camera is rotated so that the optical axis looks downwards and the camera is moved in one direction orthogonal to the optical axis with 1 unit intervals reaching up to the translation of 4 units. For rotation tests, the camera is rotated from 0 to 50 degrees with 10 degrees intervals in both X and Y axes. We create 512×512 parabolic omnidirectional images for each translation and rotation. Figure 19 shows some of the resulting test images.

Fig. 19. Some synthetic omnidirectional test images for the matching performance test.

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 virtual camera plane approach. For the SIFT descriptor, the size of the descriptor window is adapted using the computed metric to compensate for the non-uniform sampling density in order to 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 vs 1-precision graphs by sweeping the ambiguity factor from 1 to 8. Note that the recall is computed as

$$recall = \frac{\text{correct matches}}{\text{ground truth matches}}. \tag{23}$$

The 1-precision is computed as

$$1 - precision = 1 - \frac{\text{correct matches}}{\text{all matches}}. \tag{24}$$

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

Figure 20 shows the recall vs 1-precision graphs for the two 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 planar SIFT feature detection method (denoted by SIFTF). The tests show that the feature matching based on Kullback-Leibler divergence based feature matching does not provide an extra performance gain compared to matching based on Euclidean distance based matching for omnidirectional images. As expected the non-oriented version of the descriptor with proposed rotation-invariant feature matching method is slightly worse than the oriented descriptor but still better than both the SIFT descriptor and the virtual camera plane descriptor (VCP) [10] and performs similarly to the SIFTF method.

Similarly, the two translation tests show that the proposed descriptor is more capable to capture the feature characteristics and the proposed detection algorithm together with the polar descriptor outperform the planar SIFT method (See Figure 21). For translation tests, the non-oriented descriptor performs better than the full descriptor. This is due to an implicit extra constraint imposed by the non-oriented descriptor and for the type of transformation under consideration which introduces almost no rotation.

Next, we test the proposed oriented and non-oriented polar descriptors on synthetic spherical images. Figure 22 shows some of the generated synthetic spherical images on equiangular grid. We create 1024×1024 spherical images for 3 positions namely 0, 2 and -4 in one direction to test the matching performance under translation. We choose 5 rotation values, namely 0, 30, 45, 60 and 90 degrees along the Y axis in order to test the matching performance under rotation. We compute the scale-space representation on the sphere and detect scale invariant features using spherical SIFT method with the implementation from [9]. Note that this detection method is also based on the heat diffusion equation but uses spherical Fourier transform and Gaussian kernel on sphere. This method provides an optimized way for smoothing on spherical signals. For a fair comparison between descriptors, we have applied all the matching methods on the same set of features. Similarly to the tests for omnidirectional images, we consider multiple matches with the same position but different orientations as one single 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. We set the distance threshold to 5 pixels and the scale

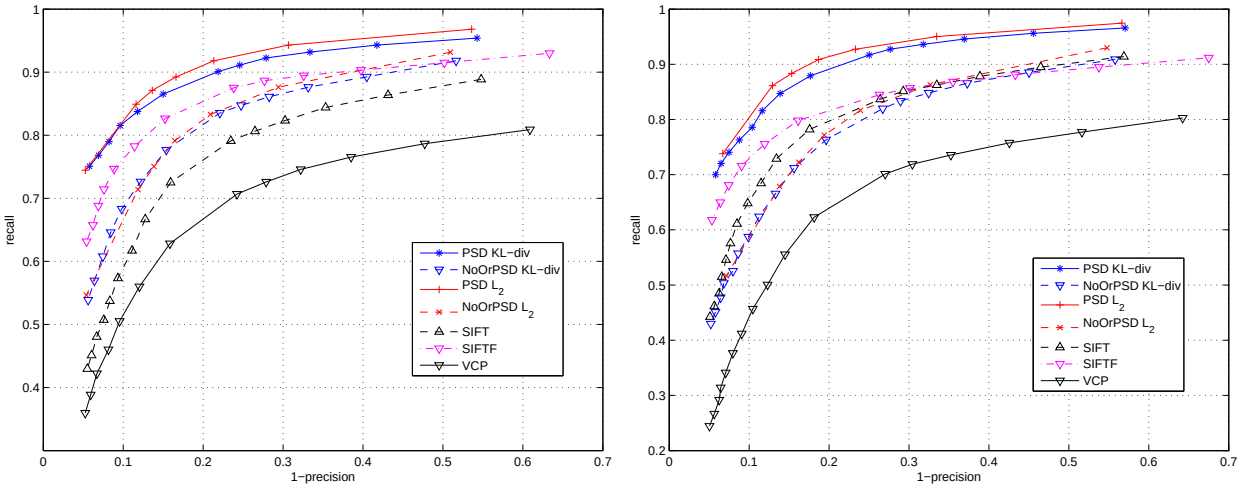


Fig. 20. Recall vs 1-precision for two rotation tests. First test (left) is for sideway rotation and the second test (right) is for the up-down rotation. SIFTF denotes the planar SIFT for which both detection and descriptor computation are based on SIFT method.

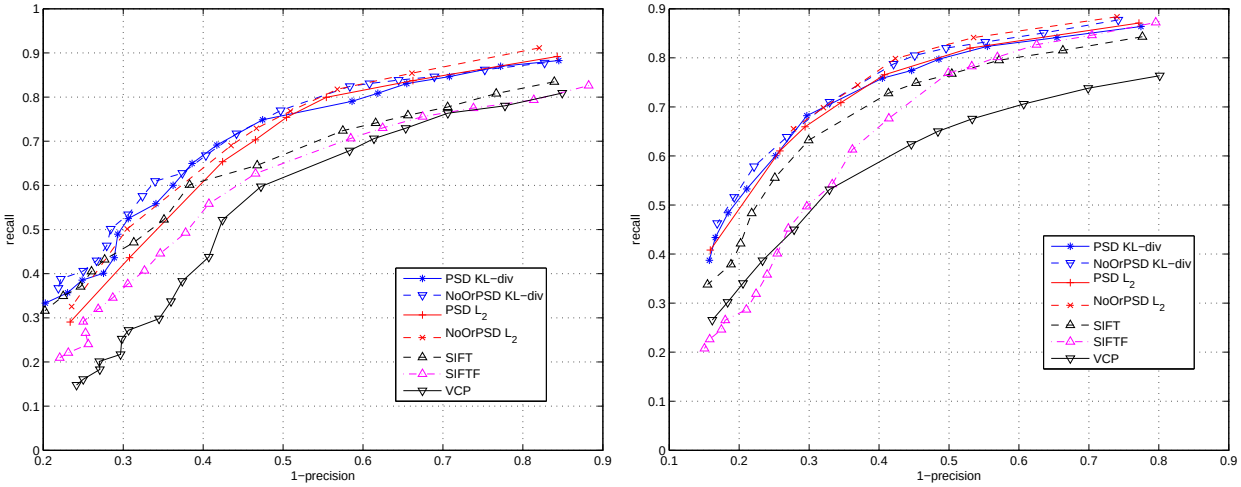


Fig. 21. Recall vs 1-precision for two translation tests. First test (left) is for translation along the optical axis. The second test (right) is for the translation of downward looking camera. SIFTF denotes the planar SIFT for which both detection and descriptor computation are based on SIFT method.

ratio threshold to 0.8. We compute recall and precision for translation and rotation match results and form recall vs 1-precision graphs by sweeping the ambiguity factor from 1 to 4.

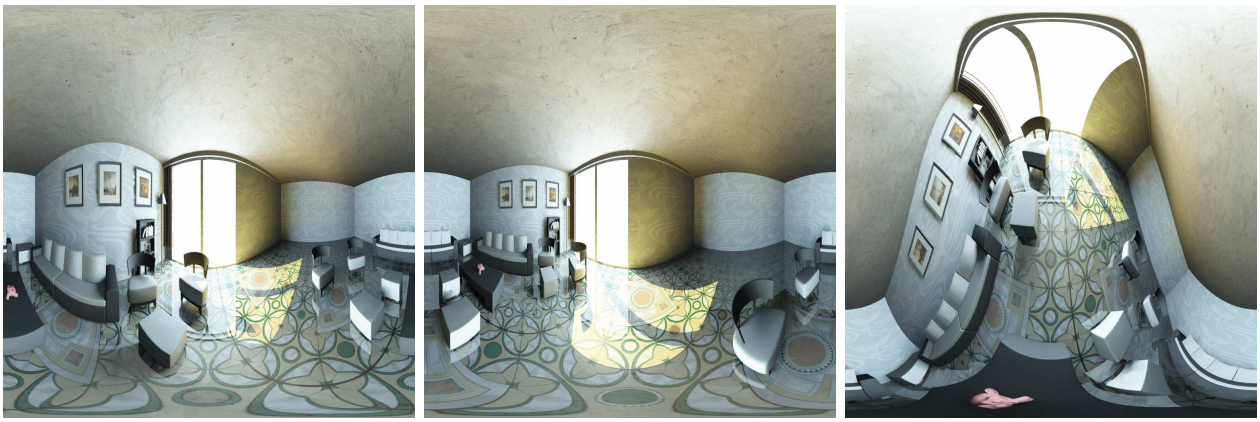
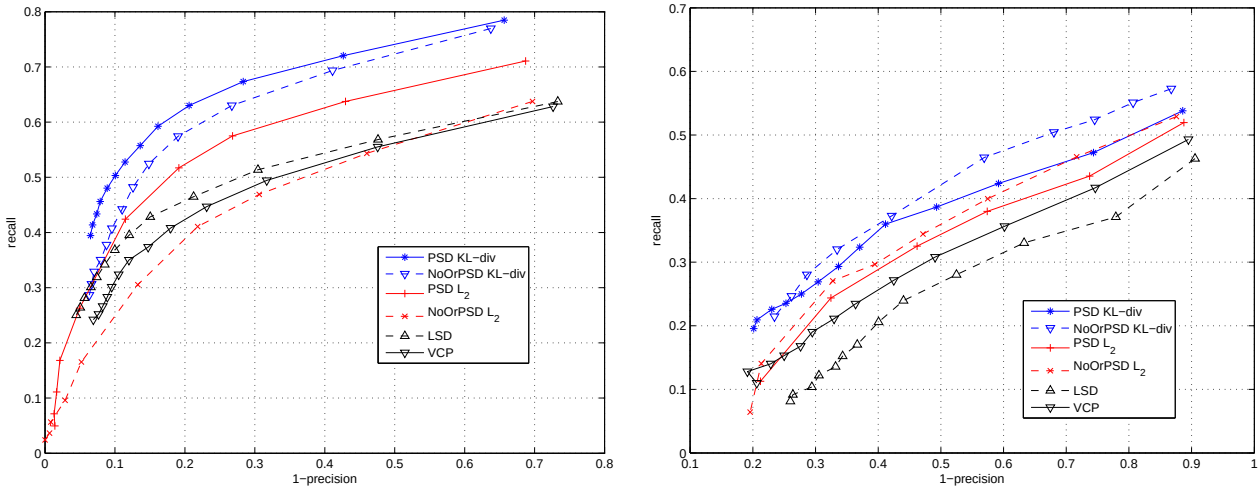


Fig. 22. Three of the synthetic spherical images used in the experiments

We perform the comparison for the polar descriptors and two other methods namely (VCP) [10], and local spherical SIFT

descriptors (LSD) [9]. The VCP method forms image patches by projecting the spherical image to a plane tangent at the feature point. It computes the SIFT descriptor on the image patch. The LSD scheme computes the SIFT descriptor directly on the spherical image by forming a rectangular descriptor regions.

We denote the polar descriptor as PSD and non-oriented polar descriptor as NoOrPSD. Figure 23(a) shows the recall vs 1-precision graph for the rotation tests. In this case, the polar descriptors provide the best performance. Non-oriented descriptors perform as good as VCP and LSD with less computation cost on the detection phase compared to the oriented descriptor. The KL-divergence metric together with the polar descriptors gives an additional performance gain.



(a) (b)

Fig. 23. Recall vs 1-precision for (a) the rotation and (b) the translation tests.

Figure 23(b) shows the performance of different descriptors for the translation of the cameras. Again, the polar descriptors perform better than VCP and LSD methods. Note that the translation causes only slight changes in the orientation of the features. This favors the non-oriented descriptors which do not compute any orientation. On the other hand, the extra orientations computed for the oriented descriptors increases ambiguity and causes mismatches.

Both Figures 23(a) and 23(b) show the potential of the non-oriented descriptors together with the rotation-invariant matching method for feature-matching applications on low-end sensors. The increase in performance for the polar descriptors is due to better handling of the sampling without the extra interpolation introduced by VCP, as well as better handling of position dependent gradients. It is also shown that histogram-based matching is more precise for spherical images, similarly to what has been reported in the planar case [21].

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. Figure 24 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 their particular type of images. 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 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 too. 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 compute the descriptors for all features using the polar descriptors. 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 obtained synthetically, 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

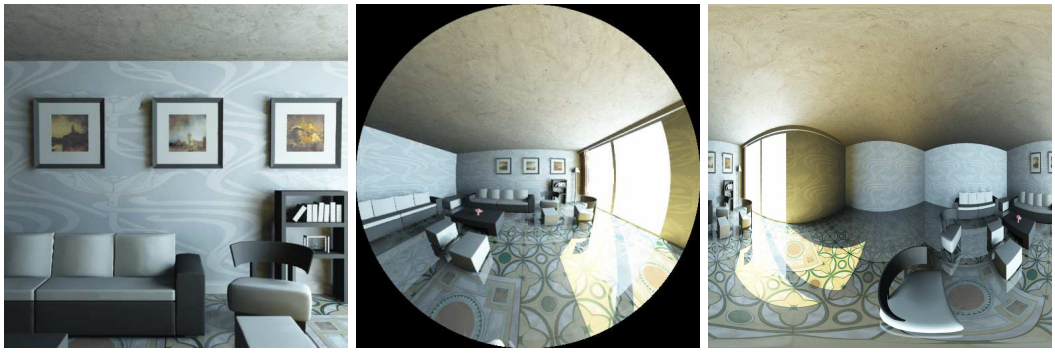


Fig. 24. Planar, omnidirectional and spherical test images for hybrid matching tests

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 in the spherical image. We compare the performance of L_2 and KL-divergence during the tests. Table II lists the number of detected features in omnidirectional, spherical and planar images. Table III shows the matching performance between omnidirectional, spherical and planar images. The ambiguity factor is set to 1 so that the matches with the best distance score are selected. The results suggest that geometry-adaptive feature detection with the proposed descriptor performs better than applying SIFT to all types of images.

TABLE II
NUMBER OF DETECTED FEATURES IN OMNIDIRECTIONAL, SPHERICAL AND PLANAR IMAGES.

	Polar	SIFT
Detected Features (Omni. Image)	437	462
Detected Features (Spherical Image)	1288	1552
Detected Features (Planar Image)	541	544

TABLE III
MATCHING PERFORMANCE FOR HYBRID MATCHING BETWEEN OMNIDIRECTIONAL (OMNI.) AND SPHERICAL IMAGES AND PLANAR IMAGES

	Polar KL	Polar L_2	SIFT
Correct Matches (Omni.- Spherical)	127	123	104
Ground Truth Matches (Omni. - Spherical)	274	274	245
Correct Matches (Omni. - Planar)	27	28	10
Ground Truth Matches (Omni. - Planar)	106	106	24

We also test the matching criteria by changing the ambiguity factor. Table IV and Table V show the performance for different ambiguity factors. The KL divergence better fits the proposed descriptor for the hybrid matching test. The performance decrease with the L_2 norm based matching criteria is due to the different discretizations of the signals on images of different camera types.

	Polar KL			Polar L_2			SIFT		
Ambiguity Fact.	1	2	3	1	2	3	1	2	3
Correct Match	27	21	13	28	6	0	10	10	9
Matches	437	22	14	437	6	0	437	18	12

TABLE IV
MATCHING PERFORMANCE FOR HYBRID MATCHING BETWEEN OMNIDIRECTIONAL AND PLANAR IMAGES UNDER DIFFERENT AMBIGUITY FACTORS.

	Polar KL			Polar L_2			SIFT		
Ambiguity Fact.	1	2	3	1	2	3	1	2	3
Correct Match	127	60	34	123	10	1	104	55	23
Matches	437	84	34	437	10	1	462	66	23

TABLE V
MATCHING PERFORMANCE FOR HYBRID MATCHING BETWEEN OMNIDIRECTIONAL AND SPHERICAL IMAGES UNDER DIFFERENT AMBIGUITY FACTORS.

VI. Conclusions

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 works directly 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 invariance of the features to rotation in SO(3), and to competitive performance in the case of translation between the compared images.

We have also proposed a polar descriptor that takes the specific geometry and non-uniform sampling grids into consideration. We have implemented a matching method that can successfully match non-oriented 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 the proposed feature detection and descriptor can be used to match features detected in images captured with different types of cameras. Our framework provides a promising solution for feature detection applications in both omnidirectional and hybrid cameras and 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," *International Journal of Computer Vision*, vol. 60, 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, 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," *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 IEEE Conference onComputer Vision and Pattern Recognition, CVPR*, pp. 1–8, 2008.

[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, pp. 283–318, Jan 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. of Workshop on Omnidirectional Vision, Camera Networks and Non-classical Cameras. OMNIVIS*, 2005.

[7] Y. Bastanlar, L. Puig, P. Sturm, J. Guerrero, and J. Barreto, "Dlt-like calibration of central catadioptric cameras," in *Proc. of the Eighth Workshop on Omnidirectional Vision, OMNIVIS*, 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, CVWW*, 2006.

[9] J. Cruz-Mota, I. Bogdanova, B. Paquier, M. Bierlaire, and J. Thiran, "Scale invariant feature transform on the sphere: Theory and applications," *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 IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS*, pp. 1689–1694, 2007.

[11] P. Hansen, P. Corke, W. Boles, and K. Daniilidis, "Scale-invariant features on the sphere," in *Proc. of IEEE International Conference on Computer Vision, CVPR*, pp. 1 – 8, Oct 2007.

[12] C. Geyer and K. Daniilidis, "Catadioptric Projective Geometry," *International Journal of Computer Vision*, vol. 45, no. 3, pp. 223–243, 2001.

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

[14] P. Hansen, W. Boles, and P. Corke, "Spherical diffusion for scale-invariant keypoint detection in wide-angle images," in *Proc. of Digital Image Computing: Techniques and Applications, DICTA*, (Los Alamitos, CA, USA), pp. 525–532, 2008.

[15] H. Bay, A. Ess, T. Tuytelaars, and L. Gool, "Speeded-up robust features (surf)," *Computer Vision and Image Understanding*, vol. 110, pp. 346–359, Jan 2008.

[16] V. Chandrasekhar, G. Takacs, D. Chen, S. Tsai, R. Grzeszczuk, and B. Girod, "CHoG: Compressed Histogram of Gradients," in *Proc. of International Conference on Computer Vision and Pattern Recognition, CVPR*, pp. 2504–2511, 2009.

[17] T. Lindeberg, *Scale-space theory in computer vision*. Springer, 1994.

[18] I. Bogdanova, X. Bresson, J. 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.

[19] T. Bulow, "Spherical diffusion for 3d surface smoothing," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1650–1654, 2004.

[20] K. Mikolajczyk and C. Schmid, "An affine invariant interest point detector," in *Proc. of European Conference on Computer Vision, ECCV*, pp. 128–142, Springer, 2002.

[21] H. Ling and K. Okada, "An efficient earth movers distance algorithm for robust histogram comparison," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 5, pp. 840–853, 2007.

[22] Y. Rubner, C. Tomasi, and L. Guibas, "The earth mover's distance as a metric for image retrieval," *International Journal of Computer Vision*, vol. 40, no. 2, pp. 99–121, 2000.

[23] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1615–1630, 2005.