

SAMPLING-AWARE POLAR DESCRIPTORS ON THE SPHERE

Zafer Arican and Pascal Frossard

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

ABSTRACT

We present a new descriptor and feature matching solution for omnidirectional images. The descriptor builds on the log-polar planar descriptors, but adapts to the specific geometry and non-uniform sampling density of spherical images. We further propose a rotation-invariant matching method for the proposed descriptor that is particularly interesting for mobile devices. It permits to reduce the computational complexity in the detection phase by eliminating the orientation assignment and moving it to the feature matching step. We then use a criteria based on the Kullback-Leibler divergence in order to improve the feature matching performance. Experimental results with spherical images show that the new descriptors offer promising performance and improve on SIFT descriptors computed on the sphere or on tangent planes.

Index Terms— Omnidirectional imaging, scale-invariant features, polar descriptors, Kullback-Leibler divergence

1. INTRODUCTION

Omnidirectional vision sensors with their large fields of view present many advantages for applications such as scene analysis, representation and detection and thus omnidirectional vision has been an active research field in the related research areas. The omnidirectional cameras typically consist of either a fisheye lens or a lens and a mirror (catadioptric) system such as parabolic or hyperbolic mirrors. The structure of the resulting images is highly dependent on the geometry of the mirror, which should be taken into account for appropriate processing of the light information.

Applications such as camera calibration, object detection, recognition or tracking generally rely on the localization and matching of salient visual features in multiple images. To provide scale-invariance, these features are generally computed by a scale-space analysis framework. This is a very important property of these features which provides robustness to resolution changes and camera translations. 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 feature detection methods and descriptors [2, 3, 4] for classical cameras.

Descriptor computation stays an important step after definition and localization of the visual features. The distinctiveness or character of the feature is defined by the descriptor and plays an important role in matching of the features. The conventional image patch correlation approach is often replaced by more sophisticated descriptors that can deal with scale, rotation and affine transformations as well as 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. The histograms are computed for each spatial bin in the support region. Descriptors with different histograms such as SURF[5] and CHoG[6] have recently been proposed to improve the performance. In addition, the GLOH framework [2] provides a log-polar descriptor that forms spatial bins by radial division of the support region. All these algorithms are however designed for planar images and they assume that the sampling is uniform along the image. For omnidirectional images, the sampling density differs from region to region and it should be taken into account in the computation of the descriptors.

Recent works such as [7, 8, 9] have proposed to process omnidirectional images on the sphere after an inverse stereographic projection that preserves the geometry of the light information [10, 11]. In these works, the scale-space representation is computed with Gaussian kernels on the sphere, while the convolution is performed using the spherical Fourier transform on a equiangular grid. Two types of descriptors are proposed on the sphere. One approach maps the image around the feature to the tangent plane to form a planar image patch and it computes SIFT descriptors on the mapped image patch. This simple approach implicitly deals with different sampling density, however, an extra interpolation is performed which may change the true scale of the feature and change the gradient values. In the second approach, the SIFT descriptor is directly computed on the spherical surface. This approach is affected by different sampling densities as it does not take into account different number of samples per spatial bin.

We propose in this paper a polar descriptor for scale invariant features on sphere that builds on log-polar descriptors for planar images. The shape of the log-polar descriptor interestingly resembles the longitude and colatitude lines around the poles of a sphere. We describe new log-polar descriptors that exploit the geometry of the sphere and take into account the different sampling densities caused by an equiangular grid on the sphere. We use scale invariant feature detection based on spherical Fourier Transform in [7, 8, 9]. We further extend the new polar 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. This is particularly interesting for mobile devices with limited computation power by sending features to a central server, which eventually performs matching operations. The matching process is finally implemented with a criteria based on Kullback-Leibler divergence. This permits to improve the matching performance for both oriented and non-oriented polar descriptors. Experimental results show that both descriptors offer better performance than SIFT based descriptors computed on the sphere or on virtual tangent planes.

2. SCALE-INVARIANT FEATURES ON THE SPHERE

Prior to computation of descriptors, one has to detect visual features. This is usually done with a scale-space analysis of the images. We briefly summarize here the scale-space analysis on the sphere using spherical Fourier transform, as well as the detection of scale invariant features discussed in [9].

Scale-space representation of planar images is obtained by smoothing the image with a gaussian kernel with variance, σ^2 . However, for a signal I lying on a sphere, the application of such approach is not straightforward due to shift-variance on the equiangular grid. An alternative solution for smoothing is discussed in [9] using heat diffusion equation and the Green's function in terms of spherical harmonics. The heat equation is defined as

$$\frac{\partial I(\theta, \phi, t)}{\partial t} = \Delta I(\theta, \phi, t) \quad (1)$$

where Δ is the laplacian on the sphere and the initial condition is $I(\theta, \phi, 0) = I(\theta, \phi)$. The colatitude angle θ is in the range $[0, \pi]$ and the longitude angle ϕ is in $[0, 2\pi)$. As shown in [12], the solution of the heat diffusion equation on the sphere is

$$\hat{I}_l^m(t) = \hat{I}_l^m(0)e^{-l(l+1)t} \quad (2)$$

where $\hat{I}(t)$ is the spectrum of the smoothed image at smoothing level t and $\hat{I}(0)$ is the spectrum of the original image. The spectrum of a spherical signal is computed by a spherical Fourier transform, which is defined as

$$I = \sum_{l \in \mathbb{N}} \sum_{|m| \leq l} \hat{I}_l^m Y_l^m, \quad \hat{I}_l^m = \int_{\mathbb{S}^2} I(\nu) \bar{Y}_l^m(\nu) d\nu \quad (3)$$

where \hat{I}_l^m are the spherical Fourier coefficients (spectrum) of I , and Y_l^m is the spherical harmonics of degree l and order m [13]. Using (3) and (2), one can compute the smoothed spherical image at the smoothing level t by first applying the forward spherical Fourier transform, and then applying Eq. (2). Finally an inverse spherical Fourier Transform is performed.

After scale-space images are computed for different scale levels t , the differences of gaussian images, $D(\theta, \phi, t)$ are computed by

$$D(\theta, \phi, t) = I(\theta, \phi, kt) - I(\theta, \phi, t) \quad (4)$$

where k is a scale multiplication factor. The SIFT features are computed on the difference images by first finding the local extrema compared to 26 neighboring pixels in the current and adjacent difference images. The low contrast extremum points that are below a pre-defined threshold are discarded. Edge responses that give unstable feature points are removed by checking the maximum ratio between maximum and minimum principal curvature of the difference images at the pixel position. Finally, the position and scale of the feature are refined by fitting a 3D quadratic.

3. SAMPLING-AWARE POLAR DESCRIPTORS ON THE SPHERE

We can now propose a new descriptor inspired from log-polar descriptors [2] on planar images. We tailor this descriptor by considering the specific geometry and sampling structure of spherical images. A circular support region around the feature point is computed based on the scale of the visual features, σ . The radius of this region is proportional to $\sin(\sigma)$. An orientation is first assigned to the visual feature. This orientation is computed by forming an histogram

of gradient orientations weighted by the magnitude of the gradients. The peak of the histogram is selected as the orientation of the feature. If there are more than one peak, a new feature is added with the same position and scale information, but with a different orientation.

For spherical images, the sampling density is not constant over the surface and thus affects the gradient computation. On a unit-sphere with a $N \times N$ equiangular grid, the sampling distance between two neighbor samples differing in colatitude angles θ is constant and equal to π/N . The sampling distance between two neighbor samples with different longitude angle ϕ 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.

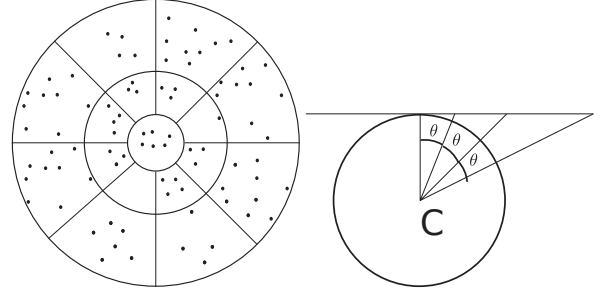


Fig. 1: The log-polar descriptor and non-uniform samples inside spatial bins (left). Radial bin divisions correspond to equal division of θ on sphere (right).

Once an orientation is assigned to a visual feature, a support region is formed proportionally to the number of radial divisions and $\sin(\sigma)$. The samples inside the support region are grouped into spatial bins. The log-polar descriptors divide the support region by discretizing the polar coordinates of the regions around the feature point. Figure 1 (left) shows such a descriptor structure with 8 orientation bins and 3 radial bins. The central bin is not divided into orientation bins in order to deal with feature localization errors. As one may observe, the structure is similar to longitudes and latitudes around the poles of a sphere for an equiangular grid. In addition, the selection of increasing radius for the radial bins is achieved by dividing the latitude angles uniformly. Figure 1 (right) illustrates this relation.

The samples in the support region are rotated in $SO(3)$ such that the feature point coincides with the north pole of the sphere. The north pole is defined as the point where $\theta = 0, \phi = 0$. Simply, quantizing the ϕ and θ then assigns samples to the corresponding bins. After the spatial bins are assigned, the log-polar descriptors compute the orientation and magnitude of the samples inside the spatial bins and form an orientation histogram weighted by the magnitude of the gradients similar to SIFT descriptor. Again, the sampling structure on the sphere is taken into account while computing the gradients. The gradients are weighted by the inverse of the sampling distance. Another factor to consider is that the number of samples falling into a spatial bin also depends on the position of the feature. A normalization is performed by dividing the histogram values by the number of the samples inside the spatial bin.

Finally, as in the SIFT framework, a descriptor vector is formed by concatenating the histograms and normalizing this vector. For a descriptor with 3 radial divisions, 8 orientation divisions and gradient histograms with 8 orientation bins, the length of the vector is $8 * (1 + 2 * 8) = 136$. For histogram-comparison based method de-

scribed in this paper, the histograms are normalized for each spatial bin and concatenated to the descriptor vector. The latter approach implicitly performs normalization with respect to number of samples.

4. FEATURE MATCHING

4.1. Feature matching on 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 [14] that it does not necessarily provide the best matching scores. Histogram-comparison based methods such as Kullback-Leibler (KL) divergence [6] or Earth Mover's distance (EMD) [15, 14] are shown to provide better matching performance at the price of more computation. Among these methods, KL-divergence has a good trade-off between accuracy and computation cost. The symmetric KL-divergence is computed as

$$KL(h_1, h_2) = \sum_n^M \left(h_1(n) \log \frac{h_1(n)}{h_2(n)} \right) + \sum_n^M \left(h_2(n) \log \frac{h_2(n)}{h_1(n)} \right) \quad (5)$$

where M is the number of bins in the histograms.

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

As in [1], after the distances are computed, the pair with the minimum distance that is one factor smaller than the second minimum distance is selected as a matching pair. The comparison factor that is called the ambiguity factor is a parameter typically set to 1.5.

4.2. Rotation invariant matching criteria for polar descriptors

We now exploit an interesting characteristic of our new descriptors 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 captures the orientation of the feature. In addition, for a polar descriptor, the rotation of the descriptor is just the 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 orienting the descriptor. 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 propose a matching algorithm to match these non-oriented descriptors.

First the approximate relative orientation is computed from the central spatial bins. Circular shift is applied to one of the central bin histograms, and the L_1 distance between the histograms is computed. The amount of shift giving the smaller distance is assigned as the relative orientation. Formally, the shift α is expressed as

$$\alpha = \underset{\alpha}{\operatorname{argmin}} \left[\sum_n (|D_p^{11}(n) - D_q^{11}(n + \alpha)|_1) \right] \quad (6)$$

where D^{11} denotes the central spatial bin. After the shift α has been computed, the L_2 distance is computed between the descriptors by shifting the indexes of the bins. In other words, we have

$$\|D_p - D_q\|_2 = \sqrt{\sum_{ik} \sum_n (D_p^{ik}(n) - D_p^{i(k-\alpha)}(n + \alpha))^2} \quad (7)$$

The KL-divergence is computed similarly.

5. EXPERIMENTAL RESULTS

We test the proposed oriented and non-oriented polar descriptors on synthetic spherical images. The depth information for each image is available so that an homography can be computed in order to define the groundtruth information. Figure 2 shows some of these images. We create 3 spherical images of resolution 1024×1024 for 3 positions namely, 0, 2 and -4 in the same direction to test the matching performance under translation and 5 spherical images with the same resolution for 5 rotations, namely, 0, 30, 45, 60 and 90 degrees rotated on Y axis to test the matching performance under rotation. We compute the scale space representations on the sphere and detect scale invariant features using the method summarized in section 2, with the implementation from [7]. We detect 833 features in average for translation and 910 features for rotation images. For fair comparisons between descriptors, we apply all the matching methods on the same set of features. In addition, we consider 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 distance and scale of the possible matches.

We set the distance to 5 pixels and the ratio of the scales 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. Note that the recall is computed as

$$\text{recall} = \frac{\text{correct matches}}{\text{ground truth matches}} \quad (8)$$

The 1-precision is computed as

$$1 - \text{precision} = 1 - \frac{\text{correct matches}}{\text{all matches}} \quad (9)$$

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

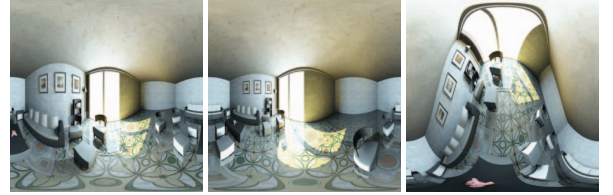


Fig. 2: 3 of the synthetic spherical room images used in experiments

We compare the polar descriptors to two other methods namely, virtual camera planes(VCP) [8], and local spherical SIFT descriptors(LSD) [7]. The VCP method forms image patches by projecting the spherical image to a plane tangent at the feature point and computes SIFT descriptor on the image patch. The LSD scheme computes the SIFT directly on the spherical image by forming a rectangular support region for the descriptor.

We denote the polar descriptor as PSD and non-oriented polar descriptor as NoOrPSD. Figure 3 shows the recall vs 1-precision graph for the rotation. In case of rotation, polar descriptors provide the best performance. Non-oriented descriptors perform as good as VCP and LSD with less computation cost than the oriented descriptor on the detection phase. It takes around 15% less time when the

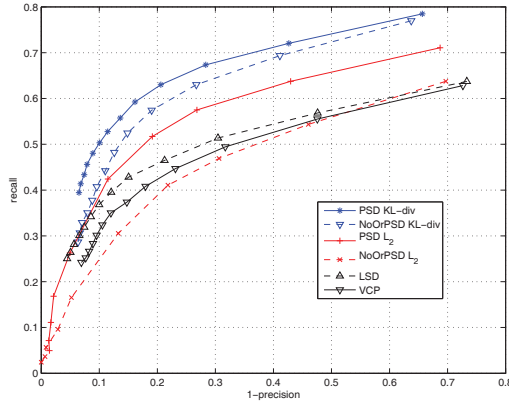


Fig. 3: recall vs 1-precision for rotation tests. Both polar descriptors provide better performance for all matching criteria

orientation step is discarded. KL-divergence together with the polar descriptors gives a clear performance increase.

Figure 4 shows the performance of different descriptors for translation of the cameras. Polar descriptors perform better than VCP and LSD methods. Note that translation causes only a slight change in orientation of the feature. This favors the non-oriented descriptors as there is no orientation computation. On the other hand, extra orientations computed for the oriented descriptors increases ambiguity which causes mismatches.

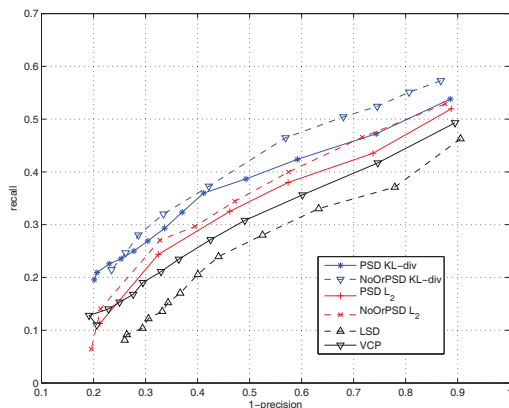


Fig. 4: recall vs 1-precision for translation tests.

Both graphs show the potential of the non-oriented descriptors together with rotation-invariant matching for accurate matching with reduced computational complexity on the detection phase. The increase in performance for the proposed polar descriptors are due to better handling of the sampling without any extra interpolation and geometry adaptiveness. It is also shown that histogram based matching is more precise, similarly to what has been reported in the planar case.

6. CONCLUSION

We have proposed two scale invariant polar descriptors inspired from log-polar descriptors on the plane. We have built the descriptor so that it considers non-uniform sampling density on the sphere. The

proposed descriptor is not limited to equiangular grid and can be applied to any sampling scheme on the sphere. We have also proposed a matching method that can successfully match non-oriented polar descriptors through the KL-divergence criteria. The complexity on the descriptor computation phase is reduced in this case, which enable its use for mobile applications.

7. REFERENCES

- [1] D.G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, Jan 2004.
- [2] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 10, pp. 1615 – 1630, Oct 2005.
- [3] E. Tola, V. Lepetit, and P. Fua, "A fast local descriptor for dense matching," in *Proc. of CVPR*, 2008, pp. 1–8.
- [4] T. Lindeberg, "Detecting salient blob-like image structures and their scales with a scale-space primal sketch - a method for focus-of-attention," *International Journal of Computer Vision*, vol. 11, no. 3, pp. 283–318, Jan 1993.
- [5] H. Bay, A. Ess, T. Tuytelaars, and L.V. Gool, "Speeded-up robust features (surf)," *Computer Vision and Image Understanding*, vol. 110, no. 3, pp. 346–359, Jan 2008.
- [6] V. Chandrasekhar, G. Takacs, D. Chen, S.S.Tsai, R. Grzeszczuk, and B. Girod, "CHoG: Compressed Histogram of Gradients," in *Proc. of CVPR*, 2009, pp. 2504–2511.
- [7] J. Cruz-Mota, I. Bogdanova, B. Paquier, M. Bierlaire, and J.P. Thiran, "Scale invariant feature transform on the sphere: Theory and applications," *EPFL Technical Report*, pp. 1–44, May 2009.
- [8] P. Hansen, P. Corke, W. Boles, and K. Daniilidis, "Scale invariant feature matching with wide angle images," in *Proc. of IROS*, 2007, pp. 1689–1694.
- [9] P. Hansen, P. Corke, W. Boles, and K. Daniilidis, "Scale-invariant features on the sphere," in *Proc. of ICCV*, Oct 2007, pp. 1 – 8.
- [10] C. Geyer and K. Daniilidis, "Catadioptric projective geometry," *International Journal of Computer Vision*, vol. 45, no. 3, pp. 223–243, Jan 2001.
- [11] X.G. Ying and Z.Y. Hu, "Can we consider central catadioptric cameras and fisheye cameras within a unified imaging model," *Lecture Notes in Computer Science*, vol. 3021, pp. 442–455, Jan 2004.
- [12] T. Bulow, "Spherical diffusion for 3d surface smoothing," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1650–1654, 2004.
- [13] J.R. Driscoll and D.M Healy, "Computing Fourier transforms and convolutions on the 2-sphere," *Advances in Applied Mathematics*, vol. 15, no. 2, pp. 202–250, 1994.
- [14] 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.
- [15] Y. Rubner, C. Tomasi, and L.J. 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.