Spherical Diffusion for Scale-Invariant Keypoint Detection in Wide-Angle Images

Peter Hansen
Queensland Univ. of Technology
CSIRO ICT Centre
Brisbane, Australia
peter.hansen@csiro.au

Wageeh Boles
Queensland Univ. of Technology
Brisbane, Australia
w.boles@qut.edu.au

Peter Corke
CSIRO ICT Centre
Brisbane, Australia
peter.corke@csiro.au

Abstract

Two variants of the SIFT algorithm are presented which operate on calibrated central projection wide-angle images characterised as having extreme radial distortion. Both define the scale-space kernel, termed the spherical Gaussian, as the solution of the heat diffusion equation on the unit sphere. Scale-space images are obtained as the convolution of the image mapped to the sphere with the spherical Gaussian which is shift invariant to pure rotation and the radial distortion in the original image. The first method termed sSIFT implements convolution in the spherical Fourier domain, and the second termed pSIFT approximates this process more efficiently in the spatial domain using stereographic projection. Results using real fisheye and equiangular catadioptric image sequences show improvements in the overall matching performance (recall vs 1-precision) of these methods versus SIFT, which treats the image as planar perspective.

1. Introduction

Robust vision-based localisation of mobile robots may be considered as the primary goal of many computer vision researchers. Recent advances in methods of visual SLAM and loop closure detection highlight the current interest in the use of vision for such applications [7, 9, 16, 24].

A considerable amount of research has focused on the problems of camera modelling, calibration and geometry of wide-angle images [1, 12, 27]. This work has made it possible to calibrate wide-angle cameras with sufficient accuracy for applications relating to mosaicing and vision-based localisation [6]. Their ability to capture a wide field of view not only removes ambiguities in motion estimation [11], but has potential advantages in loop closure scenarios where certain camera configurations allow the camera to obtain a similar view of the operating environment irrespective of orientation [23]. The ability to find corresponding keypoints in wide-angle images across wide baselines has potential advantages for localisation and is the interest of this work.

Although numerous advances have been made relating to the problem of wide-baseline feature matching [2, 3, 18, 21, 22, 26], with the most prominent algorithm used in the literature being the Scale-Invariant Feature Transform (SIFT) [20], these methods have typically been designed for use with perspective images. With some exceptions [8, 14], these algorithms are frequently applied *blindly* to wide-angle images without compensation for the extreme radial distortion present in the images. While one could argue that any wide-angle image could simply be converted to perspective, this is not suitable since perspective projection is limited to representing images with less than a hemispherical field of view where the transformation produces artifacts during the interpolation. Furthermore, we argue that image processing with wide-angle images should be shift-invariant for a camera subject to pure rotation. Both these conditions may be met by considering the image as a function defined on the unit sphere, with image processing considered in the same domain.

The contribution of this work is a suitable means for scale-invariant keypoint detection and matching with wide-angle images which is invariant to radial distortion and shift-invariant to pure rotation. Given any central calibrated image, it may be mapped to the unit sphere. Scale-invariant keypoints are then found using a variant of the SIFT algorithm which defines the scale-space kernel as the solution of the spherical heat-diffusion equation on the sphere which was solved by Bülow [5] who considered its use for scale-space analysis with wide-angle images [4]. Scale-space images are obtained as the convolution of the image mapped to the sphere with the spherical Gaussian. In this work we implement this convolution in two ways; the first in the spherical Fourier domain which we have previously con-

sidered [14], and the second using a more efficient approximation for the image and spherical Gaussian mapped to the stereographic image plane. We consider the second approach since we found issues related to aliasing for the former [15]. Furthermore, keypoint descriptors are found by sampling the keypoint support region on the sphere which is again invariant to the radial distortion in the original image.

2. Scale-Space with Wide-Angle Images

We assume that any wide-angle image may be considered as a function f on the unit sphere $\mathbb{S}^2$, where each pixel maps to a unique point $\eta(\theta,\phi) = (\sin\theta\,\cos\phi, \sin\theta\,\sin\phi, \cos\theta)^T$ with $\theta \in [0,\pi)$ an angle of colatitude and $\phi \in [0, 2\pi)$ an angle of longitude. Bülow [5] proposed scale-space for functions defined on the sphere as the solution of the spherical heat diffusion equation

$$\Delta_{\mathbb{S}^2} f(\theta,\phi;t) = \frac{1}{k}\partial_t f(\theta,\phi;t), \tag{1}$$

where $\Delta_{\mathbb{S}^2}$ is the spherical Laplace operator which restricted to the unit sphere ($r = 1, \frac{\partial f}{\partial r} = 0$) is defined as [17]:

$$\Delta_{\mathbb{S}^2} f = \frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial f}{\partial\theta}\right) + \frac{1}{\sin^2\theta}\frac{\partial^2 f}{\partial\phi^2}. \tag{2}$$

The Green's function of (1) may be considered as the spherical Gaussian $G_{\mathbb{S}^2}$ and was solved by Bülow with initial condition $G_{\mathbb{S}^2}(\theta,\phi;0) = \delta_{Ss}$ being the spherical Dirac function at the north pole $n = (0,0,1)^T$ defined as

$$f(n) = \int_{\eta\in\mathbb{S}^2} f(\eta)\delta_{\mathbf{S}^2}(\theta,\phi)d\eta, \quad f \in L^2(\mathbb{S}^2), \tag{3}$$

to obtain the solution

$$G_{\mathbb{S}^2}(\theta,\phi;kt) = \sum_{l\in\mathbb{N}} \sqrt{\frac{2l+1}{4\pi}} Y_l^0(\theta,\phi) e^{-l(l+1)kt}, \tag{4}$$

where Y_l^0 are the zonal spherical harmonic functions ($G_{\mathbb{S}^2}$ is rotationally symmetrical about the pole). The reader is referred to standard texts [13] for a description of spherical harmonic functions and [5] for a derivation of the solution. The solution in (4) describes how a unit heat source at the north pole of a thin spherical vessel of constant thermal conductivity k evolves over time t. In the context of image processing and scale-space analysis, the variable kt will simply be referred to as the scale.

Given the spherical Gaussian function, it is possible to find the scale-space representation $L_{\mathbb{S}^2}$ of an image I at scale kt as the convolution of the image mapped to the sphere with the spherical Gaussian function $G_{\mathbb{S}^2}$. Using the definition of spherical convolution in [10], for all rotations $R \in SO(3)$ define the operator $\Lambda(R)$ which rotates a point on the sphere to new position $\Lambda(R)f(\eta) = f(R^{-1}\eta)$. These may, for example, be parameterised by Euler rotations $R = R_z(\gamma)R_y(\beta)R_z(\alpha)$. Convolution of two square integratable functions f and h on the sphere is defined as

$$(f * h)(\eta) = \int_R f(Rn)\, h(R^{-1}\eta)\, dR, \quad \eta \in \mathbb{S}^2. \tag{5}$$

The convolution defined in (5) may be implemented in either the spatial or spherical Fourier domain.

2.1. Diffusion in Spatial Domain

Considering that the spherical Gaussian function is rotationally symmetrical, using the analogy in [8], integration in (5) may be restricted to the subgroup of rotations $R = R_z(\gamma)R_y(\beta)$. Then, the scale-space representation $L_{\mathbb{S}^2}$ of an image at scale kt may be obtained as

$$L_{\mathbb{S}^2}(\beta,\gamma;kt) = \int_{\eta\in\mathbb{S}^2} I(\eta)\, G_{\mathbb{S}^2}(R^{-1}\eta;kt)\, d\eta, \tag{6}$$

where $d\eta = \sin(\theta)d\theta d\phi$. Rather then mapping the image to the sphere, the convolution defined in (6) may be performed by mapping the spherical Gaussian to the image plane as discussed in [8]. This has the advantage that the original image values are used without any interpolation when mapping them to the sphere. Note that the scale-space image $L_{\mathbb{S}^2}$ will be represented on the original image plane.

Unfortunately, implementing convolution in such a way requires a unique non-symmetrical kernel at all positions on the image plane. Assuming that for a single scale kt all kernels at all locations are precomputed offline with size $n \times n$ pixels on the $m \times m$ sized image, convolution requires a total of $2n^2m^2$ computions ($\mathcal{O}(n^2)$). In contrast, convolution with a standard Gaussian function may exploit the symmetry of the kernel and use the separability criteria to rewrite convolution as $I * G_{x,y} = (I * G_x) * G_y$ which requires a total of $4nm^2$ computions ($\mathcal{O}(n)$). It is for this reason that convolution in the spherical Fourier domain is considered.

2.2. Diffusion in Spherical Fourier Domain

Given that the image I may be mapped to the sphere, the spectrum of the image $\hat{I}$ may be found via a discrete spherical Fourier transform $\hat{I} = SFT(I)$. This spectrum includes the set of coefficients $\hat{I}_l^m$ for degree l and order m, where $l \in \{0,1,\ldots,b-1\}$ and $|m| \leq l$ for the image sampled on a $2b \times 2b$ equiangular polar grid. Then setting initial condition as the original image data $\hat{L}(0) = \hat{I}$, the convolution theorem in [10] for a symmetrical filter may be used to obtain the scale-space representation of the image $\hat{L}(kt)$ as

a response in the spherical Fourier domain via convolution with the spherical Gaussian [5]:

$$\hat{L}_l^m(kt) = \hat{L}_l^m(0)e^{-l(l+1)kt}. \quad (7)$$

This scale-space representation in the spatial domain $L_{\mathbb{S}^2}(kt)$ may then be found via an inverse spherical Fourier transform $L_{\mathbb{S}^2}(kt) = ISFT(\hat{L}(kt))$ and mapped to the original image plane. Using the discrete forward and inverse spherical Fourier transform algorithm of Driscoll and Healy [10], for bandwidth (sample rate) b the computational expense is of order $\mathcal{O}(b^2(\log b)^2)$.

Unfortunately, although convolution in the spherical Fourier domain over multiple scales kt provides an efficient means for obtaining multiple scale-space representations of the image, there is an upper limit on the bandwidth b which needs to be enforced due to computational requirements. As a result, there may be aliasing issues when finding the discrete SFT due to the variable spatial resolution of the image, as discussed in detail in [15] where a suitable anti-aliasing filter implementation is described.

2.3. Approximate Spherical Diffusion

An efficient approximation to spherical diffusion is proposed using stereographic projection. Importantly, the method is not limited by bandwidth related issues. Given any central projection wide angle image, it is first mapped via the sphere to the stereographic image plane. Then for some scale kt, the spherical Gaussian function $G_{\mathbb{S}^2}$ centred at the pole is mapped to the stereographic image plane. The approximate spherical diffusion is then found simply as the convolution of the stereographic image with this spherical Gaussian function at all locations on the image. Since stereographic projection is conformal and locally preserves angles, the inverse stereographic projection of this spherical Gaussian at any location on the stereographic image plane back to the sphere is a near isotropic function on the sphere. However, the resulting convolution operation results in a non-uniform diffusion scale on the sphere due to the variable spatial resolution of the stereographic image.

To illustrate, for simplicity we assume the image plane and sphere to be smooth manifolds. Then, an infinitesimal small change in angle $d\psi$ along any great circle on the sphere from a point $\eta(\theta, \phi) = (x, y, z)^T$ may be written as $d\psi^2 = dx^2 + dy^2 + dz^2$. Substituting for the angle of colatitude θ and longitude ϕ yields the expression

$$d\psi^2 = d\theta^2 + \sin^2\theta \, d\phi^2. \quad (8)$$

Since the polar coordinates of a point on the image plane $\mathbf{x}(R, \zeta)$ project by inverse stereographic projection to a point $\eta(\theta, \phi)$ on the sphere by

$$\phi = \zeta, \quad \theta = 2\arctan(R \,/\, m_p) \quad (9)$$

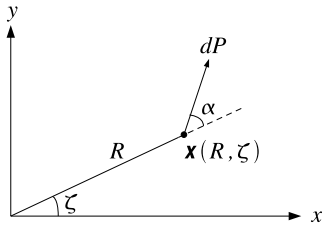


Figure 1. The vector dP represents a small shift at angle α on the image at radius R from the principal point (centre of distortion).

where m_p is the distance of the stereographic image plane from the north pole, then substituting into (8) gives the following expression for $d\psi^2$ parameterised by polar coordinates R, ζ on the image plane:

$$d\psi^2 = \frac{4m_p^2}{\left(m_p^2 + R^2\right)^2}\left(dR^2 + R^2 d\zeta^2\right). \quad (10)$$

Referring to figure 1, the change in polar coordinates on the wide-angle image plane may be rewritten as a shift dP at angle α as

$$dR^2 = dP^2 \cos^2\alpha \quad (11)$$

$$d\zeta^2 = \begin{cases} 0 & \text{if } R = 0 \\ \frac{dP^2 \sin^2\alpha}{R^2} & \text{if } R > 0 \end{cases}, \quad (12)$$

where dP is measured in pixels. Substituting into (10) yields a simplified expression for $d\psi^2$ dependent only on the radius from the principal point R and magnitude of the shift dP at any angle α:

$$d\psi^2 = \frac{4m_p^2}{\left(m_p^2 + R^2\right)^2} dP^2. \quad (13)$$

It may be seen from this result that using the approximate diffusion operation, the ratio $d\psi^2/dP^2$, and hence the spherical diffusion scale, reduces with radius R from the principal point. Although this makes the method unsuitable for uniform diffusion, for scale-space analysis this is not a limiting factor as the image is analysed across a wide range of scales during keypoint detection. Additionally, if the scale of the spherical Gaussian mapped to the image at the principal point kt_0 is known, the *corrected* scale on the sphere kt_R at a given radius R on the image plane may be found. This may be achieved simply by comparing the local sample rates ψ_0^2 and ψ_R^2 on the sphere at the image centre and at a given radius R respectively. From (13), the ratio of the local sample rates $d\psi_R^2$ and $d\psi_0^2$ for radius from the image centre R and 0 respectively is

$$\frac{d\psi_R^2}{d\psi_0^2} = \frac{m_p^4}{\left(m_p^2 + R^2\right)^2}. \quad (14)$$

Then, the corrected scale at radius R relates to the scale at radius 0 by

$$kt_R = \frac{kt_0\, m_p^4}{\left(m_p^2 + R^2\right)^2}, \tag{15}$$

where the ratio kt_R/kt_0 is independent of the magnitude of the scale kt_0 at the image's principal point. Assuming the kernel values are precomputed a priori, the same separability criteria used for standard Gaussian convolution may be used resulting in the same computational cost.

3. Scale-Invariant Keypoint Detection

We present here two variants of the SIFT algorithm which define scale-space as the convolution of the image with the spherical Gaussian. The first termed spherical SIFT (sSIFT) implements diffusion in the spherical Fourier domain using the method outlined in section 2.2, the second termed parabolic SIFT (pSIFT) uses the approximate diffusion operation described in section 2.3. The latter is termed parabolic since it operates on the same image plane that would be obtained by using a parabolic catadioptric camera.

3.1. Scale Selection

Unlike SIFT which defines scale σ in pixels relative to the original image resolution, spherical scale-space defines scale kt relative to a function on the sphere. It is necessary to consider how the scales kt may be selected with respect to the original image resolution. In this work, the scales selected in SIFT are used as a guide. For a one-dimensional Gaussian the following is true:

$$\frac{G(x=\sigma;\sigma)}{G(x=0;\sigma)} = e^{-0.5}, \quad \forall \sigma > 0. \tag{16}$$

We consider then if a similar relationship for the spherical Gaussian holds for which

$$\frac{G_{\mathbb{S}^2}(\theta = f(kt); kt)}{G_{\mathbb{S}^2}(\theta = 0; kt)} = e^{-0.5}, \quad \forall kt > 0 \tag{17}$$

where $f(kt)$ is some function of the scale kt. Figure 2 shows the plot of the scale $\sqrt{kt}$ versus angle of colatitude θ for which the condition in (17) is satisfied using the definition of the spherical Gaussian function in (4). Note that the results have been found using non-linear optimisation since a closed form solution to (17) has not been found. From figure 2 it may be observed that there is a linear relationship where $kt = \theta^2/2$.

Define a one pixel sample measurement on the image plane from the principal point as x_s, and θ_s the corresponding angle of colatitude for the sample mapped to the sphere

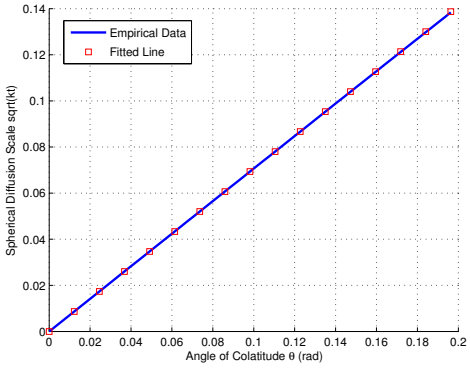


Figure 2. Fitted line obtained via least squares minimisation

for the given wide-angle camera model. It is proposed that a suitable initial scale kt_0 may be selected based on the original image resolution and SIFT starting scale σ_0 as

$$\sqrt{kt_0} = \sigma_0(\theta_s^2/2), \tag{18}$$

where the remaining scales are found using the same scale multiplicative factor used in SIFT for a total of N scales as

$$kt_i = \left(\sqrt{kt_0}\, 2^{i/3}\right)^2, \quad i \in \{0, 1, \ldots, N-1\}. \tag{19}$$

3.2. Keypoint Detection

Given the set of scale-space images $L_{\mathbb{S}^2}$ represented on the image plane, the difference of neighbouring Gaussian images are found $D_{\mathbb{S}^2}$. Keypoints are then found using the same method as SIFT [20] as being a local extrema in the current and adjacent difference of Gaussian images $D_{\mathbb{S}^2}$ whose absolute difference of Gaussian value is above some threshold. Edge responses are removed by enforcing a threshold r on the ratio of maximum to minimum curvature of the eigenvalues of the Hessian matrix $H(D_{\mathbb{S}^2})$. Finally, keypoint position and scale are interpolated using 3D quadratic fit. It is important to note that for edge removal and interpolation, the difference of Gaussian images represented on the wide-angle image plane are assumed locally perspective in the 3×3 neighbourhood surrounding any pixel; this assumption is particularly valid for pSIFT since the image is represented on the stereographic image plane. For pSIFT, the scale correction factor in (15) must be applied to all keypoints.

3.3. Keypoint Descriptors

Referring to figure 3, for a keypoint of scale kt both sSIFT and pSIFT define the support region used to find the

descriptor as a circle on the sphere parameterised by angle $\psi_s = n\sqrt{2kt}$ from the axis passing through the keypoint and the centre of the sphere. Note also that we define β as an angle about this same axis. Here, n is an overall scaling factor applied to all keypoints. The greyscale intensity values on the image which lie within this region are mapped to a fixed sized patch, then the SIFT descriptor is evaluated.

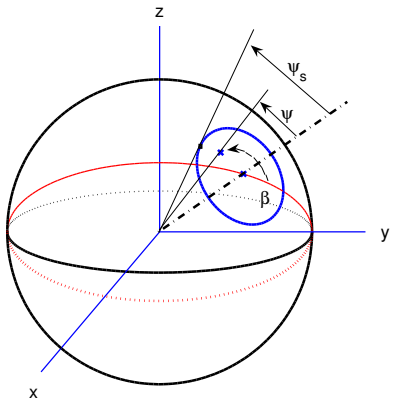


Figure 3. For a keypoint of scale kt, the support region is parameterised by angle $\psi_s = n\sqrt{2kt}$ from the line passing through the keypoint position and the centre of the sphere.

For a fixed sized patch of size $p \times p$ pixels with polar coordinates R_p, ζ_p, an equiangular mapping is used where

$$R_p = 0.5(p-1)(\psi/\psi_s), \quad \zeta_p = \beta. \tag{20}$$

This mapping is selected as the spherical Gaussian function itself will appear identical when mapped to the patch irrespective of its scale kt. This is explained by considering that a spherical Gaussian centred at the keypoint position of scale kt may be written as a function of the angles ψ, β as $G_{\mathbb{S}^2}(\psi, \beta; kt)$. It may then be rewritten in scale-normalised coordinates relative to the keypoint support region scaling n as $G_{\mathbb{S}^2}(\psi', \beta; 1/n^2)$ where $\psi' = \psi/\psi_s$. One may see from (20) that the mapping may be written as a linear function of the scale-normalised coordinates.

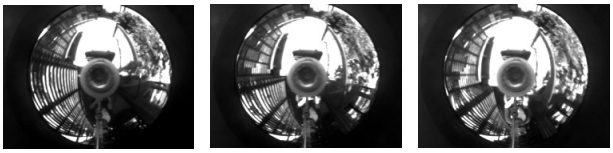
4. Experiments

The experiments presented compare the matching performance of sSIFT and pSIFT to standard SIFT where the image is treated as planar perspective. This comparison is made in the context of visual odometry where the goal is to find keypoint correspondences between successive images.

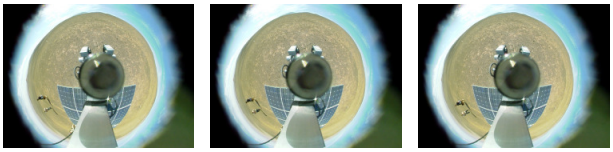
4.1. Input Data

The data used includes three separate image sequences, with an example of three consecutive images for each shown in figure 4. The first (fisheye) includes images taken from a camera moving in an outdoors industrial environment. The second (tractor) uses images from a downwards facing equiangular catadioptric camera on a mobile robotic tractor moving outdoors through a university campus. The third (Hyperion) again uses downwards facing equiangular catadioptric images from a camera on the mobile robot Hyperion operating in the Atacama desert. All images are converted to greyscale for image processing, where a suitable *mask* is used to remove keypoints detected in regions outside the camera's field of view, and in the case of the tractor and Hyperion data sets, keypoints on the mobile robot itself.

(a) Fisheye - 1024×768 pixels.



(b) Tractor (equiangular catadioptric) - 1024×768 pixels.



(c) Hyperion (equiangular catadioptric) - 640×480 pixels.

Figure 4. Example of three consecutive images in each data set.

4.2. Keypoint Detection

For each image sequence, SIFT, sSIFT and pSIFT keypoints are found in each image. When finding SIFT keypoints, the original wide-angle images are treated as perspective with the masks used to remove invalid keypoints.

For sSIFT, the initial scale is selected using the method described in section 3.1 based on the original camera models. The sample bandwidth used to find the forward and inverse SFT's is set to to $b = 512$ which is the maximum computationally feasible. Since there is some degree of aliasing with both the fisheye and tractor data sets, a suitable anti-aliasing filter is used when sampling the image for the forward SFT (see [15] for details). For sSIFT the scale-space representations of the image are always mapped back to the original sized image plane for image processing at

any scale.

When finding pSIFT keypoints, each image is first mapped to the stereographic image plane via the sphere. For all sequences, the stereographic image is the same size as the original images where the scaling is set such that a point on the equator of the sphere maps to the same radius on the stereographic image as the original image. In the following experiments, results for pSIFT are found using two separate initial scales (both applied to the same stereographic image). The first initial scale is found based on the stereographic image model, and the second for the original camera model (ie. fisheye or equiangular). Furthermore, the same octave based approach to image processing used in SIFT (3 scales per octave) is implemented with pSIFT where all kernels are precomputed offline.

For both sSIFT and pSIFT, the difference of Gaussian threshold is set to 0.01 (assuming greyscale values in the range 0 to 1), with the ratio of maximum to minimum curvature used in edge removal set to $r = 10$. The descriptor support regions are set to size $\psi_s = 8\sqrt{2kt}$, where kt is the keypoint scale, and the local region mapped to a fixed sized 41×41 patch from which the SIFT descriptors are found.

4.3. Performance metric

The performance of each keypoint detection method is measured using recall vs 1-precision which has previously been used in similar studies related to keypoint detection and descriptors [19, 22] and suitable for use when the exact number of false matches between images is unknown. For a given data set all frame to frame keypoint correspondences are found, combined into a single global set and then ordered based on their similarity score. The Recall vs 1-precision results are then found which are defined as:

$$\text{recall} = \#_{\text{correct matches}} \;/\; \#_{\text{total correct matches}} \qquad (21)$$

$$1 - \text{precision} = \#_{\text{false matches}} \;/\; \#_{\text{all matches}}. \qquad (22)$$

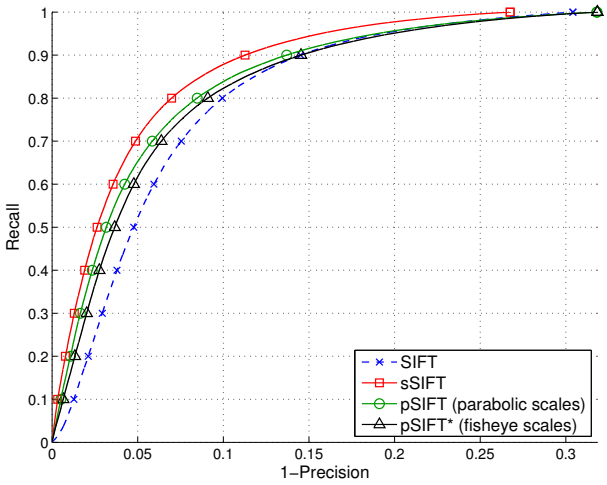
where the number of correct matches ($\#_{\text{correct matches}}$) and number of false matches ($\#_{\text{false matches}}$) refers to the number in the subset of matches whose score is below the matching threshold in the global set.

Two commonly used similarity scores are considered in these experiments. The first is the Euclidean distance between descriptors, and the second the ambiguity metric which is the ratio of nearest to second nearest Euclidean distance to all keypoints in the other image. For both, a keypoint in one image may only correspond to one keypoint in the other image. If two or more are found for any keypoint, only the one with the highest similarity score is retained.

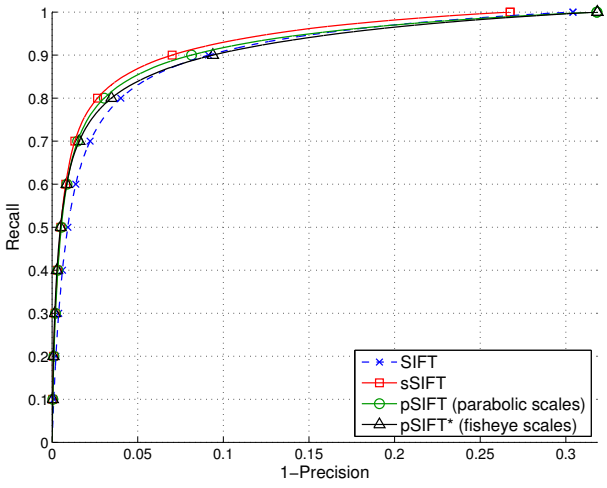
For a set of (calibrated) keypoint matches, between frames, with spherical coordinates η and η' in image 1 and 2 respectively, the essential matrix E is found using a subset of all keypoint matches for all keypoint detection modes with the highest similarity. The five point algorithm [25] is used to solve for E with RANSAC used to remove outliers. Then assuming the operating environment is rigid, a keypoint is considered correct if $|\eta'^T E \eta| <$ threshold.

4.4. Results and Discussion

The results for each keypoint detection method and matching mode are shown in figures 5, 6 and 7 for the fisheye, tractor and Hyperion data sets respectively. The mean number of frame to frame correspondences for each keypoint type and image sequence are given in table 1.



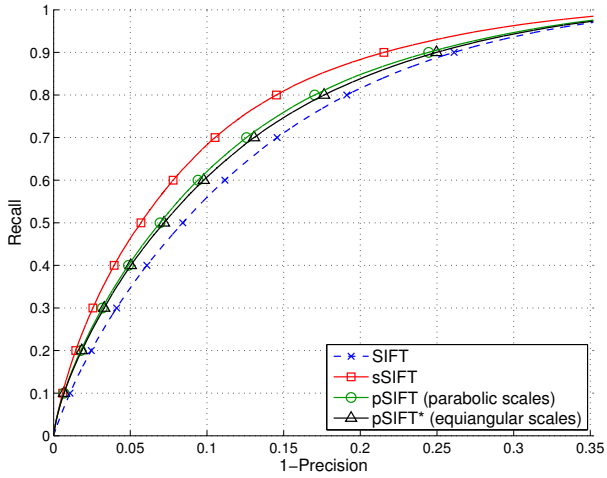
(a) Euclidean Similarity



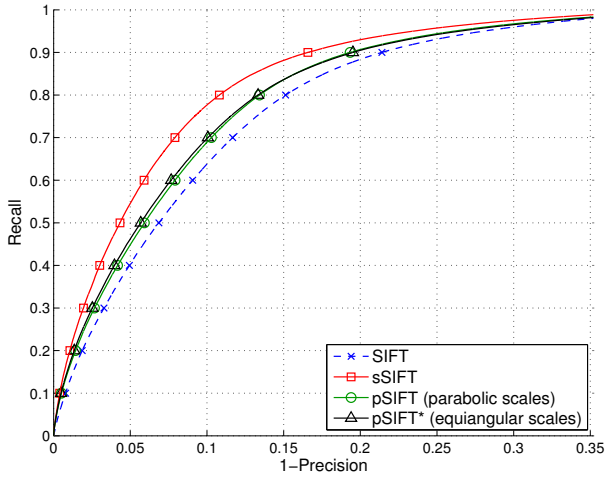
(b) Ambiguity Similarity

Figure 5. Recall versus 1-precision results for the fisheye sequence.

For the fisheye sequence, sSIFT outperforms all other keypoint detection modes for both similarity metrics with

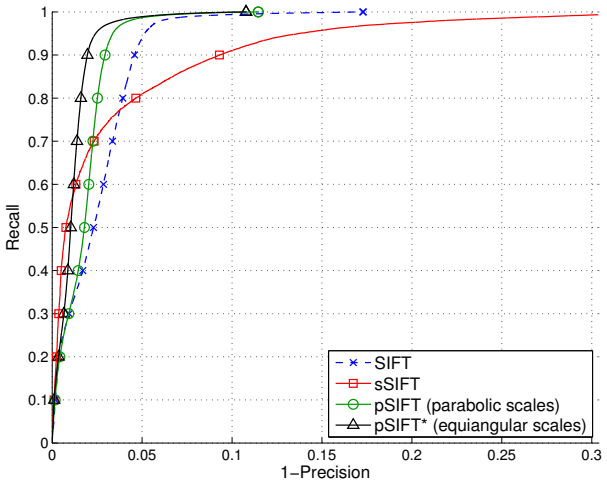


(a) Euclidean Similarity

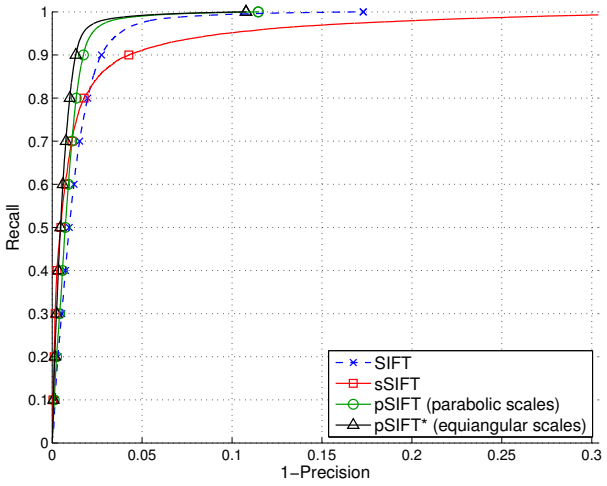


(b) Ambiguity Similarity

Figure 6. Recall versus 1-precision results for the tractor sequence.



(a) Euclidean Similarity



(b) Ambiguity Similarity

Figure 7. Recall versus 1-precision results for the Hyperion sequence.

more average frame to frame matches found. The results using pSIFT are similar irrespective of the initial scale used with both showing improved performance over SIFT.

Similar results are found for the tractor data set with sSIFT outperforming all other keypoint detection modes for recall vs 1-precision, however, with fewer average frame to frame matches. Although the number of matches is fewer than SIFT, pSIFT again gives improved recall vs 1-precision results compared to SIFT for each similarity metric and initial scale used.

Although the the Hyperion and tractor data sets both use an equiangular catadioptric camera, the relative results between keypoint detection modes is different. Again, pSIFT outperforms SIFT in the recall vs 1-precision results with a significant increase in the average number of frame to frame matches. For sSIFT, the recall vs 1-precision results degrade more quickly than the other keypoint detection mode with only a minimal number of average frame to frame matches. We attribute this poor recall vs 1-precision result to the fact that this data set has minimal detail in each image where the mapping to the sphere and back to the image plane required for convolution in the spherical Fourier domain effectively destroys much of the original information contained within the image. This result re-enforces the argument that image processing should be implemented on the original image plane. Although this is not done for pSIFT, there is only one transformation from the original image plane to the stereographic image plane.

Table 1. Mean number of frame to frame keypoint matches for each method of SIFT. pSIFT* refers to the original camera model values used for scale selection.

	Fisheye	Tractor	Hyperion
SIFT	251.9	351.2	48.8
sSIFT	288.5	187.9	17.0
pSIFT	236.5	285.8	116.8
pSIFT*	274.1	269.3	241.3

5. Conclusions

Two variants of the SIFT algorithm were presented suitable for use with wide-angle images. Both define scale-space for wide-angle images as the solution of the spherical heat diffusion equation on the sphere (spherical Gaussian). The first implements convolution in the spherical Fourier domain, and the second approximates this process on the stereographic image plane. Overall, results showed that although pSIFT does not outperform sSIFT on all data sets, it is a suitable approximation which is efficient to implement and gives improved performance over a *blind* application of SIFT where the image is treated as perspective.

6. Acknowledgements

The authors thank the Field Robotics Centre at Carnegie Mellon University for the Hyperion data set, and Kane Usher, Ashley Tews and Jon Roberts for the tractor data set.

References

[1] S. Baker and S. Nayar. A theory of single-viewpoint catadioptric image formation. *International Journal of Computer Vision*, 35(2):175–196, 1999.

[2] A. Baumberg. Reliable feature matching across widely separated views. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pages 774–781, 2000.

[3] H. Bay, T. Tuytelaars, and L. V. Gool. SURF: Speeded Up Robust Features. In *Proc. European Conference on Computer Vision*, pages 346–359, 2006.

[4] T. Bülow. Multiscale image processing on the sphere. In L. van Gool, editor, *Proc. 24th Symposium Pattern Recognition of the DAGM*, pages 609–617, 2002.

[5] T. Bülow. Spherical diffusion for 3D surface smoothing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(12):1650–1654, 2004.

[6] P. Corke, D. Strelow, and S. Singh. Omnidirectional visual odometry for a planetary rover. In *Proc. International Conf. on Intelligent Robots and Systems*, pages 4007–4012, 2004.

[7] M. Cummins and P. Newman. Probabilistic appearance based navigation and loop closing. In *Proc. International Conf. on Robotics and Automation*, pages 2042–2048, 2007.

[8] K. Daniilidis, A. Makadia, and T. Bülow. Image processing in catadioptric planes: spaciotemporal derivatives and optical flow computation. In *Proc. Third Workshop on Omnidirectional Vision*, pages 3–12, 2002.

[9] A. Davison, I. Reid, N. Molton, and O. Stasse. MonoSLAM: real-time single camera slam. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(6):1–16, 2007.

[10] J. R. Driscoll and D. M. Healy. Computing fourier transforms and convolutions on the 2-sphere. *Advances in Applied Mathematics*, 15:202–250, 1994.

[11] C. Fermüller and Y. Aloimonos. Ambiguity in structure from motion: Sphere versus plane. *International Journal of Computer Vision*, 28(2):137–154, 1998.

[12] C. Geyer and K. Daniilidis. Catadioptric projective geometry. *Int'l Journal of Computer Vision*, 45(3):223–243, 2001.

[13] H. Groemer. *Geometric Applications of Fourier Series and Spherical Harmonics*. Cambridge University Press, 1996.

[14] P. Hansen, P. Corke, W. Boles, and K. Daniilidis. Scale invariant feature matching with wide angle images. In *Proc. International Conference on Intelligent Robots and Systems*, pages 1689–1694, 2007.

[15] P. Hansen, P. Corke, W. Boles, and K. Daniilidis. Scale-invariant features on the sphere. In *Proc. International Conference on Computer Vision*, pages 1–8, 2007.

[16] V. Ila, J. Andrate-Cett, R. Valencia, and A. Sanfeliu. Vision-based loop closing for delayed state robot mapping. In *Proc. International Conference on Intelligent Robots and Systems*, pages 3892–3897, 2007.

[17] J. D. Jackson. *Classical Electrodynamics*. John Wiley & Sons, 2nd edition, 1975.

[18] T. Kadir and M. Brady. Saliency, scale and image description. *Int'l Journal of Computer Vision*, 45(2):83–105, 2001.

[19] Y. Ke and R. Sukthankar. PCA-SIFT: A more distinctive representation for local image descriptors. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pages 506–513, 2004.

[20] D. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.

[21] J. Matas, O. Chum, M. Urban, and T. Pajdla. Robust wide-baseline stereo from maximally stable extremal regions. *Image and Vision Computing*, 22(10):761–767, 2004.

[22] K. Mikolajczyk and C. Schmid. Scale and affine invariant interest point detectors. *International Journal of Computer Vision*, 60(1):63–86, 2004.

[23] D. Prasser, M. Milford, and G. Wyeth. Outdoor simultaneous localisation and mapping using RatSLAM. *Field and Service Robotics*, 25:143–154, 2006.

[24] J. Sivic and A. Zisserman. Video Google: A text retrieval approach to object matching in videos. In *Proc. International Conf. on Computer Visison*, pages 1470–1477, 2003.

[25] H. Stewénius, C. Engels, and D. Nistér. Recent developments on direct relative orientation. *ISPRS Journal of Photogrammetry and Remote Sensing*, 60(4):284–294, 2006.

[26] T. Tuytelaars and L. V. Gool. Matching widely separated views based on affine invariant regions. *International Journal of Computer Vision*, 59(1):61–85, 2004.

[27] X. Ying and Z. Hu. Can we consider central catiodioptric cameras and fisheye cameras within a unified imaging model. In *Proc. ECCV*, pages 442–455, 2004.