This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

Short Papers

sRD-SIFT: Keypoint Detection and Matching in Images With Radial Distortion

Miguel Lourenço, Joao P. Barreto, and Francisco Vasconcelos

***Abstract*—Keypoint detection and matching is of fundamental importance for many applications in computer and robot vision. The association of points across different views is problematic because image features can undergo significant changes in appearance. Unfortunately, state-of-the-art methods, like the scale-invariant feature transform (SIFT), are not resilient to the radial distortion that often arises in images acquired by cameras with microlenses and/or wide field-of-view. This paper proposes modifications to the SIFT algorithm that substantially improve the repeatability of detection and effectiveness of matching under radial distortion, while preserving the original invariance to scale and rotation. The scale-space representation of the image is obtained using adaptive filtering that compensates the local distortion, and the keypoint description is carried after implicit image gradient correction. Unlike competing methods, our approach avoids image resampling (the processing is carried out in the original image plane), it does not require accurate camera calibration (an approximate modeling of the distortion is sufficient), and it adds minimal computational overhead. Extensive experiments show the advantages of our method in establishing point correspondence across images with radial distortion.**

***Index Terms*—Image keypoints, radial distortion (RD), scale-invariant feature transform (SIFT) features.**

I. Introduction

Finding point correspondences between two images of the same scene is a key step of many computer and robot vision algorithms, such as structure-from-motion (SfM), visual recognition, and image content retrieval. Current methods for associating points across different views typically comprise three steps: 1) the *detection* of keypoints, e.g., corners and blobs, at distinctive locations that can be repeatedly found under different viewing conditions; 2) the *description* of a keypoint neighborhood patch, usually represented through a feature vector that must be distinctive and robust to geometric and photometric transformations; and, finally, 3) the *matching* of descriptor vectors which is typically carried using a distance defined in the feature space, e.g., Euclidean distance [1]. The literature reports several approaches for finding image correspondences that differ in one or more of the steps enumerated previously [1], [2]. The scale-invariant feature transform (SIFT) [3] is arguably one of the most popular matching algorithms, being broadly used in robotics because of its invariance to common image transformations such as scale, rotation, and moderate viewpoint change [4], [5].

Manuscript received May 18, 2011; revised October 27, 2011; accepted January 16, 2012. This paper was recommended for publication by Associate Editor D. Kragic and Editor D. Fox upon evaluation of the reviewers' comments. This work was supported by the Portuguese Science Foundation under Grant PTDC/EEA-ACR/68887/2006 and Grant SFRH/BD/63118/2009.

The authors are with the Institute for Systems and Robotics Coimbra, University of Coimbra, 3030-290 Coimbra, Portugal (e-mail: miguel@isr.uc.pt; jpbar@deec.uc.pt, fpv@isr.uc.pt).

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

Digital Object Identifier 10.1109/TRO.2012.2184952

Many robotic systems employ cameras with unconventional optical arrangements that introduce radial distortion (RD): fish-eye lenses provide a wide field-of-view (FOV), which is advantageous for many robotic tasks like egomotion estimation [6] and visual place recognition [7]; minilenses are often used in flying robots because of their small size that enables complying with payload requirements [8]; boroscopes are employed in medical endoscopy and industrial inspection for visualizing small cavities with difficult or limited access [9]. Unfortunately, the SIFT algorithm, as well as the majority of competing methods, is meant for perspective images and cannot handle the strong distortion introduced by these optics [6]–[10]. At the image level, the RD causes a nonuniform displacement of the pixel positions along radial directions and toward the center. This leads to a compression of the image structures that affects the scale-invariant detection in multiple manners, with some keypoints, that are previously found at fine scales, being missed; other keypoints being assigned to incorrect scales; and false keypoints being detected because of spurious image artifacts (e.g., straight lines that become curves) [11]. In addition, and since RD also changes the image gradients, the SIFT description varies with the position where the feature is projected, which has a pernicious effect in terms of matching results [11].

This paper presents a set of modifications to the SIFT algorithm that improve the detection repeatability and matching performance under RD, while preserving the original invariances to scale and rotation. Keypoints are detected by looking for extrema in a scale-space representation obtained using a kernel that adapts the distortion at each image pixel position. It is shown that this adaptive filtering can be well approximated by a horizontal and vertical 1-D correlation using a Gaussian kernel with standard deviation that varies with the pixel image radius. Such approximation enables a computational efficiency that is comparable with the original SIFT algorithm. Additionally, we propose to achieve description invariance to RD by performing implicit gradient correction using the Jacobian of the distortion function. The main virtue of our algorithm, i.e., dubbed sRD-SIFT, is that all the operations are carried in the original image plane, avoiding the introduction of spectral artifacts, while implicitly reconstructing the image signal before resampling [10]. Extensive experiments show that sRD-SIFT has important advantages with respect to alternative approaches such as explicit distortion correction [4] and the pSIFT algorithm [7]. The paper extends a previous conference publication [12], providing a more thorough analysis and validation of the framework.

A. Related Work

SIFT has been applied, in the past, to images with significant distortion. While some works simply ignore the pernicious effects of RD and directly apply the original algorithm over distorted images [9], others perform a preliminary correction of distortion through image rectification and, then, apply SIFT [4]. The latter approach is quite straightforward, but it has two major drawbacks: the explicit distortion correction can be computationally expensive for the case of large frames and, more importantly, the interpolation required by the image rectification introduces artifacts that affect the detection repeatability.

Daniilidis *et al*. [10] were the first ones arguing that the warping of wide FOV images should be avoided, and that optical flow in catadioptric views should be computed assuming the sphere $\mathbb{S}^2$ as the underlying domain of the image function. In [13], Bulow proposes a scale-space representation for functions defined in $\mathbb{S}^2$ by solving the spherical heat

1552-3098/$31.00 © 2012 IEEE

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

diffusion equation. Inspired by [13], Hansen *et al*. investigated the generalization of the SIFT algorithm for images with domain on the sphere [7]. The advantages of such generalization are twofold: First, the SIFT on the sphere can be indistinguishably applied to any type of central projection image. The only requirement is to know in advance the intrinsic camera calibration in order to map the image plane into $\mathbb{S}^2$; second, the formulation of SIFT on the sphere enables us to achieve full invariance to pure camera rotation motion. The original SIFT algorithm that is proposed by Lowe [3], despite being invariant to rotations on the plane $\mathbb{P}^2$, is unable to handle the projective transformations due to camera rotation [14].

The main difficulty in extending the SIFT algorithm to the sphere is the computation of a suitable scale-space representation that passes, in an implicitly or explicitly manner, by backprojecting the image $\mathtt{I}$ into $\mathbb{S}^2$ and convolving the result with a spherical Gaussian function G_S [13]. Ideally, this operation must avoid the resampling of the original image signal [10] and must be computationally efficient. So far, the proposed approaches are the following.

1) *Mapping G_S into $\mathbb{P}^2$ [10]*: Instead of backprojecting $\mathtt{I}$ into $\mathbb{S}^2$, the kernel G_S is projected into $\mathbb{P}^2$ and the convolution is carried directly in the image plane. This avoids image resampling but leads to an adaptive filtering, with the mapped Gaussian kernel changing at every image pixel location, and the filtering not being separable in *X* and *Y* [15]. Such complexity makes the solution unsuitable for generating the multiple levels of the difference-of-Gaussian (DoG) pyramid.
2) *Diffusion in the Spectral Domain [7]*: The Gaussian smoothing is performed in the spectral domain. Let $\mathtt{I}_S$ be the result of backprojecting the original image $\mathtt{I}$ into the sphere. The spectrum of $\mathtt{I}_S$ can be found via a discrete spherical Fourier transform (DSFT), and the filtering result is achieved by applying the inverse DSFT to the product of the image spectrum with the transform of G_S. This operation can be efficiently implemented as long as it is imposed an upper limit in bandwidth to keep computation tractable. The problem is that such limit can lead to aliasing issues [7].
3) *Approximated Diffusion (pSIFT) [7]*: The diffusion on the sphere can be efficiently approximated by mapping the image $\mathtt{I}$ via the sphere into the stereographic plane and by convolve the result with the stereographic projection of G_S. The projected Gaussian kernel, despite changing at every image pixel location, is always a symmetric function that can be approximated by successive 1-D convolutions along *X*- and *Y*-directions (separation property). This enables us to achieve a computational efficiency similar to the original SIFT, while avoiding the aliasing problems of the spectral approach. However, the mapping of $\mathtt{I}$ requires image resampling that introduces pernicious artifacts [7].
4) *Laplace–Beltrami Operator [16], [17]*: Recently, some authors have applied Riemannian geometry concepts to compute the scale-space representation of central catadioptric images. The Gaussian smoothing on the sphere is achieved through a suitable Laplace– Beltrami (LB) operator that preserves the geometry of the visual contents and adapts to the nonuniform resolution, while using the original image pixel values. Unfortunately, the derived LB operators are specific for catadioptric images and cannot be applied to cameras with lens distortion.

Comparing with the previously described methods, our framework is less general in the sense that it requires the distortion to be described by the division model [18] (this excludes catadioptric images) and is not invariant to the effects of pure camera rotation motion. However, in sRD-SIFT, every processing step is carried on the plane using original pixel values and, in a similar manner to the pSIFT algorithm, the computational efficiency of the adaptive filtering is improved by considering an approximate kernel function that is separable in X- and Y-directions. Another advantage is that, unlike the aforementioned approaches, sRD-SIFT does not require accurate intrinsic camera calibration (an approximate modeling of the distortion suffices).

B. Article Structure and Notation

The structure of this paper is as follows. Section II is a background section that briefly reviews the SIFT algorithm, the assumed camera model, and evaluation metrics that will be used throughout the paper. The modifications to SIFT detection and description leading to sRD-SIFT are, respectively, discussed in Sections III and IV. The design of the algorithm is guided by tests on a representative set of perspective images to which RD is artificially added. This enables fully controlled experiments with accurate ground truth and assurance that observations are only due to the distortion effect. Finally, Section V conducts several tests with real distorted images undergoing changes in scale, rotation, and viewpoint.

Convolution kernels are represented by symbols in sans serif font, e.g., G, and image signals are denoted by symbols in typewriter font, e.g., $\mathtt{I}$. Vectors and vector functions are typically represented by bold symbols, and scalars are indicated by plain letters, e.g., $\mathbf{x} = (x, y)^\mathsf{T}$ and $\mathbf{f}(\mathbf{x}) = (f_x(\mathbf{x}), f_y(\mathbf{x}))^\mathsf{T}$.

II. Background

A. Scale-Invariant Feature Transform

The keypoint detection uses a scale-space representation of the image [19] where the Laplacian-of-Gaussian is approximated by LoG [20]. Let $\mathtt{I}(x, y)$ and $\mathsf{G}(x, y; \sigma)$ be, respectively, an image signal and a 2-D Gaussian function with standard deviation σ. The blurred version of $\mathtt{I}(x, y)$ is obtained by its convolution with the Gaussian kernel

$$\mathtt{L}(x, y; \sigma) = \mathtt{I}(x, y) * \mathsf{G}(x, y; \sigma) \tag{1}$$

and the DoG pyramid is computed as the difference of consecutive filtered images with the standard deviation differing by a constant multiplicative factor:

$$\mathtt{DoG}(x, y, k^{n+1}\sigma) = \mathtt{L}(x, y; k^{n+1}\sigma) - \mathtt{L}(x, y; k^n\sigma)\,. \tag{2}$$

Each pixel in the DoG pyramid is compared with its neighbors in order to find local extrema in scale and space dimensions. These extrema are, subsequently, filtered and refined to obtain keypoints. The next step is the computation of the descriptor vectors using the image gradients of a local patch around each detected keypoint. Scale invariance is achieved by performing all the computations at the scale of selection in the Gaussian pyramid. The method starts by finding the dominant orientation of the local gradients and uses it for rotating the image patch toward a normalized position. Finally, the SIFT descriptor is computed by performing a Gaussian weighting of gradient contributions, quantizing the orientations, and building histograms that accumulate magnitudes. For further details, see [3].

B. Division Model for Radial Distortion

We will assume that the image distortion follows the first-order division model [18], with the amount of distortion being quantified by a single parameter ξ (typically, $\xi < 0$), and the distortion center being approximated by the image center. Let $\mathbf{x} = (x, y)^\mathsf{T}$ and $\mathbf{u} = (u, v)^\mathsf{T}$ be the coordinates of corresponding points in the distorted and undistorted images expressed with respect to a reference frame with origin in the center. $\mathbf{f}$ is a vector function that maps points in the undistorted image

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

plane $\mathtt{I}^u$ into points in the distorted image $\mathtt{I}$ [18]:

$$\mathbf{x} = \mathbf{f}(\mathbf{u}) = \begin{pmatrix} f_x(\mathbf{u}) \\ f_y(\mathbf{u}) \end{pmatrix} = \begin{pmatrix} \dfrac{2u}{1+\sqrt{1-4\xi(u^2+v^2)}} \\ \dfrac{2v}{1+\sqrt{1-4\xi(u^2+v^2)}} \end{pmatrix}. \quad (3)$$

The function is bijective, and the inverse mapping from $\mathtt{I}$ to $\mathtt{I}^u$ is given by

$$\mathbf{u} = \mathbf{f}^{-1}(\mathbf{x}) = \begin{pmatrix} f_u^{-1}(\mathbf{x}) \\ f_v^{-1}(\mathbf{x}) \end{pmatrix} = \begin{pmatrix} \dfrac{x}{1+\xi(x^2+y^2)} \\ \dfrac{y}{1+\xi(x^2+y^2)} \end{pmatrix}. \quad (4)$$

Given a particular value for the radius $r = \sqrt{x^2+y^2}$ in the distorted image, the corresponding undistorted radius is

$$r^u = \frac{r}{1+\xi r^2}. \quad (5)$$

The RD is quantified in the division model by the parameter ξ. Henceforth, and in order to make the compression undergone by a particular image more intuitive, the amount of distortion will be quantified by

$$\%_{\text{distortion}} = \frac{r_M^u - r_M}{r_M^u} \times 100 = -\xi\, r_M \times 100 \quad (6)$$

with r_M denoting the distance from the center to the image corner (the maximum distorted radius).

C. Performance Metrics

1) Measuring Detection Performance: The repeatability of keypoint detection in different views of a scene is an important metric to characterize the performance of a particular detection scheme. Let S_l and S_r be the sets of keypoints that are independently detected in images $\mathtt{I}_l$ and $\mathtt{I}_r$. The repeatability is given by

$$\%_{\text{Repeatability}} = \frac{\#S^{\text{true}}}{\min(\#S_l, \#S_r)} \times 100 \quad (7)$$

with S^{true} being the keypoints that are simultaneously detected in the two views ($S^{\text{true}} = S_l \cap S_r$), and $\#$ denoting set cardinality. A keypoint belongs to S^{true} *iff* it is a common detection that satisfies consistency criteria in space and scale [1].[1] The space consistency concerns the keypoint pixel location in the two images and can be verified using the multiview geometry between $\mathtt{I}_l$ and $\mathtt{I}_r$ (e.g., a plane homography H mapping one image into the other, the epipolar constraint, etc.). The scale consistency refers to the fact that the scales of detection in the two images must agree. Note that if a keypoint in a distorted image has scale σ_d, then in the absence of distortion, the corresponding scale would be

$$\sigma_0 = \frac{\sigma_d}{1+\xi r^2} \quad (8)$$

with r denoting the original keypoint radius (5). Since the distortion causes a nonlinear compression that diminishes the size of the image structures, the evaluation must take into account this effect and perform an adaptive correction of scale using a linear approximation of the distortion function.

[1]We follow the criteria that are proposed in [1] where the consistency in space and scale implies an overlap between keypoint regions of more than 70%.

2) Measuring Matching Performance: Two keypoints are considered to be a match *iff* the Euclidean distance between their SIFT descriptors is below a certain threshold λ [3], [7]. Let M be the set of keypoints in the image $\mathtt{I}_l$ for which the matching algorithm finds a correspondence in $\mathtt{I}_r$. The set M can be divided into the correct matches M^{true} and incorrect matches M^{false}. Thus, the ability of the matching algorithm in finding correct matches can be quantified using the recall. This metric must be complemented by the *precision* that measures how well the algorithm discards keypoints that have no correspondence:

$$\text{recall}(\lambda) = \frac{\#M^{\text{true}}}{\#S^{\text{true}}}, \qquad \text{precision}(\lambda) = \frac{\#M^{\text{true}}}{\#M}. \quad (9)$$

In general, a good matching performance is achieved whenever there is a choice for λ that makes both the *precision* and the *recall* close to 1. Thus, the matching performance can be evaluated by verifying if the curve *1-precision versus recall* for varying λ passes at a short distance of the ideal operation point $(0, 1)$ [1].

III. Keypoint Detection in Images With Radial Distortion

The distortion causes a nonuniform compression of the image structures that affects SIFT detection performance. This can be observed in the synthetic experiment of Fig. 1(a), where the repeatability of keypoint detection decreases with increasing amounts of added distortion. A straightforward strategy to avoid the harmful effects of RD is to explicitly correct the distortion and run the standard SIFT detection on the rectified frame [4]. Fig. 1 also evaluates this approach, with the test frames being first distorted and then restored using successive image resampling. It would be to expect a repeatability close to 100%; however, and despite the significant improvements with respect to standard SIFT, the results are far from this score.

The problem is that the distortion correction by image resampling implicitly requires reconstructing the signal from the initial discrete image. Thus, not only there are high-frequency components that cannot be recovered (e.g., low resolution and aliasing), but also the reconstruction filters are imperfect. The bilinear and bicubic interpolations are, respectively, the first- and second-order approximations of the ideal reconstruction kernel, i.e., the infinite *sinc* function [15]. These approximations introduce spurious frequency components and other signal artifacts that affect the keypoint detection. The skeptical reader can easily confirm this fact by observing the experiment of Fig. 2. The right-most image is the result of a linear rescaling of the left-most image by a factor of 1.5. Note that since the signal resolution is increased, there are neither aliasing effects nor losses of high-frequency components. We would expect for the SIFT detector to find the same keypoints in the original and expanded frames, but this is clearly not the case. This largely explains the repeatability results shown in Fig. 1(a). It is interesting to observe that for RD $< 15\%$ the standard SIFT detection outperforms rectSIFT, meaning that for small amounts of distortion, the harmful effects of image resampling surpass the benefits of the explicit correction.

A. Adaptive Gaussian Filtering (RD-SIFT)

We propose to improve the detection repeatability using a model-based approach for image blurring that compensates for the spectral modifications caused by RD. While in rectSIFT, the $\mathtt{DoG}$ pyramid is computed after warping the image, in this section, the scale-space representation is generated directly from the frame with distortion using adaptive Gaussian filtering. The outcome is a $\mathtt{DoG}$ pyramid equivalent to the one that would be obtained by following the steps.

1) Correct the RD of the image $\mathtt{I}$.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

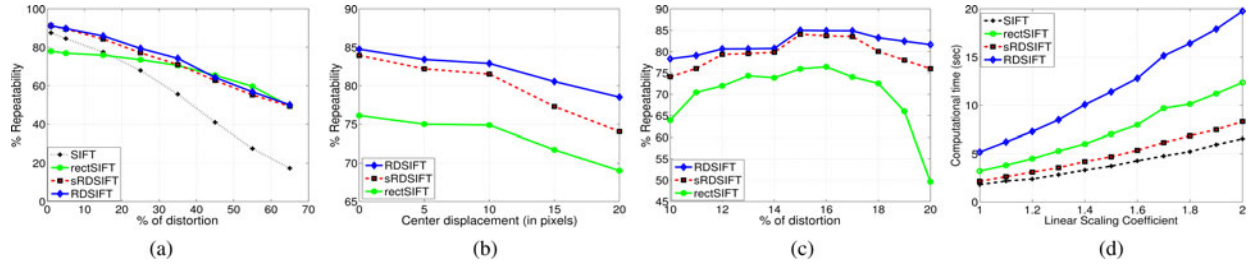


Fig. 1. Keypoint detection in images to which increasing amounts of distortion is artificially added. The curves are obtained by averaging results over 15 images with size 640×480 and different visual contents. The detection in the reference image (RD $=$ 0%) is always performed using the standard SIFT. We compare SIFT directly applied to distorted images (SIFT), SIFT applied over frames where the distortion is corrected using explicit image warping (rectSIFT), the accurate adaptive filtering derived in Section III-A (RD-SIFT), and its approximated counterpart proposed in Section III-B to diminish the computational overhead (sRD-SIFT). (a) Detection repeatability. (b) and (c) Robustness to errors in the calibration parameters. (d) Comparison of the computation times for building the DoG pyramid. The robustness to calibration errors was tested assuming as ground truth the detection results for RD $= 15\%$. The computational time was evaluated for a constant distortion of 25% and increasing image sizes. (a) Detection repeatability. (b) Disturbance in the center. (c) Disturbance in % of RD. (d) Time profiling.

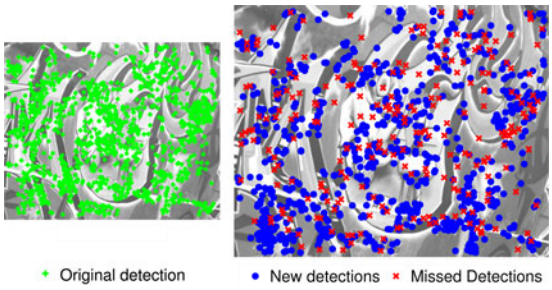


Fig. 2. SIFT detection in resampled images. The size of the left-most image is increased by 50% using bilinear interpolation. SIFT keypoint detection is independently run in each frame, the results are compared, and the differences overlaid. The reasons for new detections are explained in [3]. More surprisingly is the fact that there are keypoints in the original image that are not detected in the scaled version. Replacing bilinear for bicubic interpolation leads to similar observations (not shown).

2) Blur the undistorted image $\mathtt{I}^u$ through successive convolutions with a Gaussian function.
3) Apply RD to the blurred images $\mathtt{L}^u$.
4) Subtract the distorted blurred images $\mathtt{L}$ for obtaining the final DoG pyramid.

As we will see later, the detection repeatability improves dramatically by avoiding the image resampling required by the warping operation. The adaptive Gaussian function is derived below. Consider the convolution of the undistorted image $\mathtt{I}^u$ with a Gaussian kernel with standard deviation σ. By writing the convolution operation of (1) explicitly, it comes that the blurred image is

$$\mathtt{L}^u(s,t;\sigma) = \sum_{u=-\infty}^{+\infty} \sum_{v=-\infty}^{+\infty} \mathtt{I}^u(u,v)\, \mathsf{G}(s-u, t-v; \sigma)\,. \qquad (10)$$

If $\mathtt{I}$ is the original image, then it follows from Section II-B that $\mathtt{I}^u(\mathbf{u}) = \mathtt{I}(\mathbf{x})$ with $\mathbf{x} = \mathbf{f}(\mathbf{u})$ (3). Replacing $\mathtt{I}^u$ by $\mathtt{I}$ and switching the variables (u, v) by (x, y) using the mapping relation (4), we obtain the result of (11). This equation computes the undistorted blurred image $\mathtt{L}^u$ directly from the original distorted frame $\mathtt{I}$. However, it is no longer a strict convolution because the filter function varies with the image location that is being filtered. Henceforth, we will refer to this operation as being an *adaptive convolution* that is denoted by $\star$ whenever convenient, as shown (11), at the bottom of this page. Let us now apply RD to the blurred image $\mathtt{L}^u$ in order to obtain $\mathtt{L}$. This can be achieved in an implicit manner using again the mapping relations of Section II-B. After replacing the undistorted image coordinates (s, t) by their distorted counterpart (h, k) and performing some algebraic simplifications, we obtain the adaptive filtering of (12), as shown at the bottom of this page, with r being the distance between the center and the image location where the filter is applied, and δ being the ratio between the radius d of each pixel contribution and r

$$\delta = \frac{d}{r} = \frac{\sqrt{x^2+y^2}}{\sqrt{h^2+k^2}}\,. \qquad (13)$$

The keypoints are detected by looking for extrema in the DoG pyramid that is computed by subtracting the images $\mathtt{L}$ of (12) for increasing values of σ (2). The new detection algorithm, henceforth dubbed *RD-SIFT*, is evaluated against SIFT and rectSIFT using the set of images with synthetically added distortion. Fig. 1(a) shows that for increasing RD values, the detection repeatability suffers a much smoother degradation than the one observed for the original SIFT. More importantly, RD-SIFT outperforms rectSIFT for amounts of distortion up to 45%. Beyond this point, the compressive effect is so strong that many image structures disappear and can no longer be filtered out. Since rectSIFT tries to restore the original signal, it tends to provide slightly better repeatability under very extreme amounts of RD that are unlikely to arise in real camera systems.

B. Improving Computational Efficiency (sRD-SIFT)

Unfortunately, the adaptive convolution used in RD-SIFT is computationally demanding both in terms of memory and operations [10].

$$\mathtt{L}^u(s,t;\sigma) = \sum_{x=-\alpha}^{\alpha} \sum_{y=-\alpha}^{\alpha} \mathtt{I}(x,y)\, \mathsf{G}\Big(s - f_u^{-1}(x,y),\, t - f_v^{-1}(x,y);\, \sigma\Big) \qquad \text{with}\, \alpha = \frac{1}{\sqrt{-\xi}} \qquad (11)$$

$$\mathtt{L}(h,k;\sigma) = \sum_{x=-\alpha}^{\alpha} \sum_{y=-\alpha}^{\alpha} \mathtt{I}(x,y)\, \mathsf{G}\Big(\frac{h - x + \xi r^2(h\delta^2 - x)}{1+\xi r^2(1+\delta^2+\xi r^2\delta^2)}, \frac{k - y + \xi r^2(k\delta^2 - y)}{1+\xi r^2(1+\delta^2+\xi r^2\delta^2)}; \sigma\Big) \qquad (12)$$

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

This section derives a filter approximation that enables high detection repeatability while keeping computation tractable. Let us return to (12) and analyze how the filter adapts to the RD present in the image. Consider that the point with coordinates (h, k) is near the image center. In this case, the term ξr^2 is very close to zero and the filtering operation converges to the standard convolution by a Gaussian kernel. This makes sense because, since the effect of RD is usually unnoticeable in the center, there is no need for the filter to make any type of compensation. Consider, now, that the point (h, k) is in the image periphery. Since the filtering kernel dismisses pixel contributions far away from the convolution center, it is reasonable to assume that (x, y) is close to (h, k), and that the ratio δ is approximately unitary. Making $\delta = 1$ in (12) yields

$$\widehat{\mathtt{L}}(h,k;\sigma) = \sum_x \sum_y \mathtt{I}(x,y)\, \mathsf{G}\Big(\frac{h-x}{1+\xi r^2}, \frac{k-y}{1+\xi r^2}; \sigma\Big) \qquad (14)$$

with $\widehat{\mathtt{L}}$ being an approximation of $\mathtt{L}$. The expression can be rewritten using the adaptive convolution operator $\widehat{\mathtt{L}} = \mathtt{I} \star \widehat{\mathsf{G}}$, where $\widehat{\mathsf{G}}$ is given by

$$\widehat{\mathsf{G}} = \mathsf{G}(x, y; (1+\xi r^2)\sigma). \qquad (15)$$

From (15), it follows that $\mathtt{I}$ is filtered by a Gaussian kernel with a standard deviation that varies with the image radius r. As we move far from the center, the filter adapts to the distortion by giving increasing emphasis to the pixel contributions closer to the convolution point. While the exact filtering of (12) uses a different kernel at every image pixel location, the approximation of (14) employs the same filter function for image locations equidistant to the center.

It is well known that the regular 2-D Gaussian function G can be generated by cascading two 1-D Gaussian kernels [15], [19]. The separability property of the regular 2-D Gaussian function [15], [19] is used in standard scale-space implementations for speeding up decreasing the computational complexity of image blurring. The filtering is typically achieved by successively convolving the image with a 1-D Gaussian kernel with horizontal and vertical orientations. Unfortunately, neither the exact filter of (12), nor $\widehat{\mathsf{G}}$, is separable. Despite this, let us consider the adaptive kernel $\mathring{\mathsf{G}}$, which is defined as

$$\mathring{\mathsf{G}} = \mathbf{g}_h(x, y; (1+\xi r^2)\sigma) \star \mathbf{g}_v(x, y; (1+\xi r^2)\sigma) \qquad (16)$$

with $\mathbf{g}_h$ and $\mathbf{g}_v$ being horizontal and vertical 1-D Gaussian functions with standard deviations varying with the radius of the convolution center. Although not discussed in here due to space limitations, it can be shown that $\mathring{\mathsf{G}}$ and $\widehat{\mathsf{G}}$ are equally good approximations of the exact filter function of (12) [11]. Thus, the blurred images $\mathtt{L}$, which are necessary to build the $\mathtt{DoG}$ pyramid, can be approximated by $\mathring{\mathtt{L}}$ obtained by convolving the original distorted image $\mathtt{I}$ with the 1-D filters $\mathbf{g}_h$ and $\mathbf{g}_v$. Fig. 1(a) shows the repeatability of the sRD-SIFT detector that uses separable adaptive filtering for the image blurring. The 1-D kernels are precomputed and stored in a lookup table enabling an implementation with an overall computation performance very close to standard SIFT (see Algorithm 1). The marginal deterioration in repeatability caused by the approximated filtering is largely compensated by the improvements in computational efficiency [see Fig. 1(d)].

Algorithm 1: sRD-SIFT: Adaptive (horizontal) convolution

Input: $\mathtt{I}(h,k)$, ξ, σ
Output: $\mathring{\mathtt{L}}(h,k;\sigma)$
/* Compute the filter bank $\mathcal{F}$ */
foreach $r = 0 \rightarrow r_M$ **do**
  $\mathcal{F}[r][\cdot] = \mathbf{g}(\cdot; (1+\xi r^2)\sigma)$
end
/* Perform horizontal convolution */
foreach $h = 0 \rightarrow h_{max}$ **do**
  foreach $k = 0 \rightarrow k_{max}$ **do**
    Compute the radius $r = \sqrt{h^2 + k^2}$
    $\mathring{\mathtt{L}}(h,k;\sigma) = \sum_{t=-w}^{w} \mathtt{I}(h, k+t)\mathcal{F}[r][w+t]$
  end
end

C. Additional Evaluations

1) Robustness to Calibration Errors: The algorithms RD-SIFT, sRD-SIFT, and rectSIFT require prior knowledge about the center and amount of distortion. In this experiment, we evaluate the robustness of the detection to deviations in these parameters eventually due to calibration errors. Fig. 1(b) shows the repeatability behavior when the position error in the distortion center ranges from 0 to 20 pixels (the shift direction is random). As expected, all the methods are affected by inaccurate center calibration, but the break in performance is smooth and proportional to the disturbance. The behavior of the three algorithms is very similar, with RD-SIFT being slightly more robust than the competitors. Fig. 1(c) shows the repeatability when the error is in the quantification of the RD. Both RD-SIFT and sRD-SIFT present a reasonable robustness to the disturbance (the former more than the latter). rectSIFT seems to be more sensitive, especially when the RD is overestimated. We believe that this is due to a poorer image signal reconstruction because of the wider interpolation intervals. From the tests, we can say that the two algorithms that are herein proposed lead to significant improvements in detection repeatability, even when the RD calibration is performed in a coarse manner.

2) Run Time: This experiment compares the execution time of the different detectors with respect to increasing image resolution. Fig. 1(d) shows the average run time on the images of the synthetic dataset after proper scaling and addition of RD $= 25\%$. The measured *detection time* is the sum of the time intervals spent in preprocessing, generating the scale-space representation, and looking for local extrema. In RD-SIFT and sRD-SIFT, the preprocessing consists in computing the adaptive filter masks and storing them into memory, while in rectSIFT, it refers to correcting RD through image re-sampling. Note that for the case of a monocular image sequence, the explicit RD correction must be repeated for each frame, while the adaptive masks are computed only once. From Fig. 1(d), it follows that sRD-SIFT has a computational efficiency close to standard SIFT. We verified, experimentally, that the overhead introduced by the adaptive filtering is usually negligible, and that the time difference is caused by the preprocessing step. The graphic also shows that rectSIFT is substantially less efficient, presenting an execution time that grows exponentially with the image resolution.

IV. Keypoint Description in Images With Radial Distortion

The SIFT description is not invariant to RD because the nonlinear deformation changes the image gradients in the neighborhood of the keypoint. Thus, the SIFT vector is displaced in the description space with respect to the position that it would have in the absence of distortion. Since the RD deformation is nonuniform across the image, this displacement depends on the location where the keypoint is detected. The current section shows how to keep the descriptor vector stationary in order to achieve RD invariance.

A. Implicit Gradient Correction

Since we have a model for the distortion, the RD invariance can be achieved by correcting the deformation before generating the descrip-

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

tors. This can be done explicitly by warping the image and computing the gradients in the undistorted signal, or implicitly, by measuring the gradients in the original image and correcting the result using the derivative chain rule. The implicit approach avoids the propagation of interpolation artifacts inherent to the image resampling and is, computationally, more efficient because the gradient correction is only performed in the description regions around the keypoints. Let $\mathtt{I}$ be the original image and $\mathtt{I}^u$ be its undistorted counterpart. From Section II-B, it follows that

$$\mathtt{I}^u(\mathbf{u}) \;=\; \mathtt{I}(\mathbf{f}(\mathbf{u}))\,.$$

Applying the derivative chain rule, it yields

$$\nabla\mathtt{I}^u \;=\; \mathsf{J}_\mathbf{f}\,.\,\nabla\mathtt{I} \tag{17}$$

with $\nabla\mathtt{I}^u$ and $\nabla\mathtt{I}$ being, respectively, the gradient vectors in $\mathtt{I}^u$ and $\mathtt{I}$, and $\mathsf{J}_\mathbf{f}$ being the 2×2 Jacobian matrix of the mapping function $\mathbf{f}$ given in (3). The Jacobian matrix can be written in terms of distorted image coordinate $\mathbf{x} = (x,y)^\mathsf{T}$ by replacing $\mathbf{u}$ using the inverse mapping of (4). It follows that

$$\mathsf{J}_\mathbf{f} \;=\; \frac{1+\xi r^2}{1-\xi r^2}\begin{pmatrix} 1-\xi(r^2-8x^2) & 8\xi xy \\ 8\xi xy & 1-\xi(r^2-8y^2)\end{pmatrix}$$

with r denoting the radius of $\mathbf{x}$.

In summary, we propose to measure the gradients directly in the original distorted image $\mathtt{I}$, evaluate the Jacobian matrix $\mathsf{J}_\mathbf{f}$ at every relevant pixel location, and correct the gradient vectors $\nabla\mathtt{I}$ using (17). The descriptor is generated from the undistorted gradients $\nabla\mathtt{I}^u$ following the standard procedure described in Section II-A. The only modification is the replacement of the weighting gaussian function $\mathsf{G}(x,y;\sigma)$ by $\widehat{\mathsf{G}} = \mathsf{G}(x,y;(1+\xi r^2)\sigma)$, in order to account for the changes in pixel contributions due to RD.

B. Evaluation in Keypoint Matching

Fig. 3 depicts the curves of precision–recall when the descriptors for the matching are computed before and after performing distortion compensation. The comparison with standard SIFT description shows a dramatic improvement in the retrieval performance. Thus, the first conclusion is that the correction of image gradients enables achieving RD invariance during description, which boosts the overall matching performance. By comparing implicit gradient correction against explicit image warping, it comes that the former outperforms to the latter for amounts of distortion of $\approx 25\%$. This is explained by the fact that the interpolation employed in the resampling process introduces spurious frequency components that propagate for the first-order derivatives that are used in the descriptor vector. For very strong distortions, the explicit image rectification outperforms the implicit gradient correction. As discussed previously, beyond a certain amount of RD, the compressive effect becomes so strong that local variations, that would be observed in the undistorted image, are no longer detectable in the distorted signal. In other words, it is impossible to recover the gradient vector $\nabla\mathtt{I}^u$ using (17) because the corresponding vector $\nabla\mathtt{I}$ cannot be measured. In this case, the interpolation used in the explicit image correction is advantageous because it enables inferring missing information.

V. Experiments With Real Imagery

The evaluations of Figs. 1 and 3, that guided the design process, were carried using a set of artificially distorted images. This section aims to confirm the results so far by running experiments in images

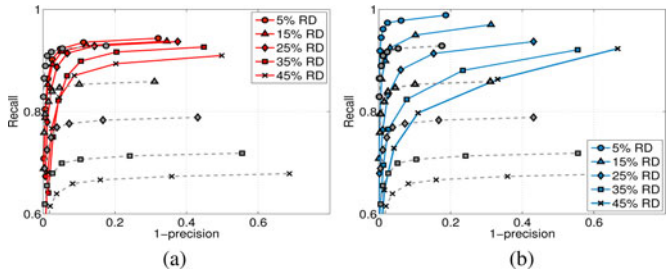


Fig. 3. Keypoint matching in images to which distortion is artificially added. The matching is between the original perspective frames and the corresponding distorted versions by a certain amount of RD (markers depict the amount of distortion). The graphics show the curves of *1-precison versus recall* averaged across the different images of the dataset. (a) Matching of SIFT descriptors computed after explicitly correcting the distortion (red lines) with the original SIFT framework (gray-dashed lines). (b) Matching of SIFT descriptors computed after implicit gradient correction, as proposed in Section IV-A, with the original SIFT framework. (a) rectSIFT vs SIFT. (b) sRD-SIFT vs SIFT.

Fig. 4. Calibration grid and three images (out of 13) of each sequence used in the experiments described in Section V-A. The images were acquired with a camera with low lens distortion (RD$\approx 10\%$), a 4-mm minilens commonly used in flying robotics' applications (RD$\approx 25\%$) [8], and a fish-eye lens with a wide FOV (RD$\approx 45\%$) [6], [7]. The resolution was 640×480 for all cases.

acquired by real cameras with lens distortion that undergo changes in scale, rotation, and viewpoint. The sRD-SIFT keypoint detection and matching is compared against the original SIFT algorithm, the SIFT run after performing explicit RD correction via image warping (rectSIFT), and the pSIFT framework [7]. As discussed in Section I, the pSIFT detection approximates the spherical diffusion using a stereographic projection and computes the descriptor by considering a support region on the sphere, which is resampled to a canonical patch of size 41×41.

A. Planar Textured Surfaces

This experiment uses three images sequences of planar scenes acquired using lenses that introduce different amounts of distortion (see Fig. 4). The results of each sequence are averaged over 78 image pairs obtained from 13 frames. For the case of rectSIFT and sRD-SIFT, the distortion center is assumed to be coincident with the image center, and the distortion parameter ξ is roughly estimated by straightening up lines in the image periphery [21]. For the case of pSIFT, the camera intrinsics are fully calibrated from images of a checkerboard pattern using the method proposed in [22]. Since the scenes are planar, the frames are related by an homography that can be used to verify the correctness of the matches and the repeatability of detection [1], [2]. We apply a robust estimation algorithm that uses hundreds of correspondences to compute these ground truth homographies [2], [14].

The table in Fig. 5 compares the performance of the four studied algorithms. The two left-most columns concern the computational

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

		Time (sec)		Detect. & Match. (Constant Threshold)					Detect.&Match. (500 strongest detections)			
		Detect.	Total	#S	$\#S^{true}$	%Rep.	$\#M^{true}$	%Prec.	$\#S^{true}$	%Rep.	$\#M^{true}$	%Prec.
10%	SIFT	1.57	1.97	1052	596	**57**	405	**62**	179	**36**	122	**68**
	rectSIFT	3.31	3.73	1057	644	**61**	431	**70**	178	**36**	121	**74**
	pSIFT	2.05	2.78	1224	756	**61**	524	**67**	207	**41**	141	**71**
	sRD-SIFT	1.61	2.32	1080	777	**72**	528	**71**	219	**44**	146	**77**
25%	SIFT	1.95	2.79	1332	871	**65**	458	**47**	189	**37**	101	**53**
	rectSIFT	4.85	5.66	1375	1022	**74**	539	**68**	203	**41**	113	**75**
	pSIFT	2.21	3.37	1558	1168	**75**	654	**57**	213	**43**	121	**67**
	sRD-SIFT	1.99	3.02	1412	1110	**78**	641	**65**	228	**46**	127	**74**
45%	SIFT	1.87	2.35	900	295	**27**	78	**30**	102	**20**	32	**40**
	rectSIFT	18.22	20.88	752	419	**56**	165	**67**	181	**36**	72	**74**
	pSIFT	2.33	4.28	1557	795	**51**	286	**63**	231	**46**	83	**67**
	sRD-SIFT	2.01	3.98	1663	809	**49**	295	**65**	179	**36**	78	**71**

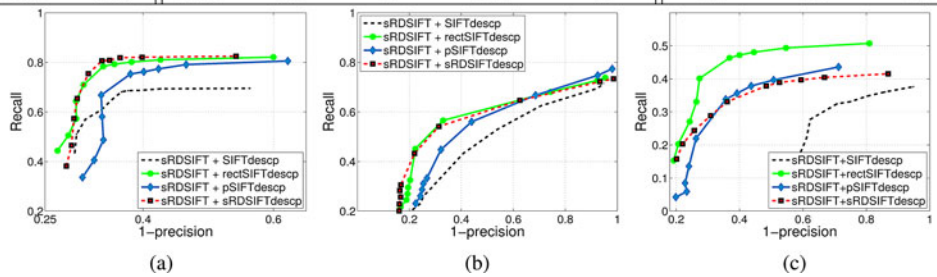


Fig. 5. Table compares the performance of the four algorithms in planar scenes. The left-most group of columns concern the computational overhead, the middle group refers to detection and matching when the threshold value for keypoint selection in the DoG pyramid is the same for all methods, and the right-most group repeats the analysis considering only the 500 strongest keypoint detections. $\#S$, $\#S^{\text{true}}$, and $\#M^{\text{true}}$ are, respectively, the average numbers of independent image detections ($\#S = \min(\#S_l, \#S_r)$), of common detections in the image pair (matching potential), and of correctly established correspondences. We also show the detection repeatability and the matching precision as defined in Section II-C. (a)–(c) *1-precision*versus *recall* curves that characterize the retrieval performance of the four descriptors being tested (in this case, the keypoints were detected using sRD-SIFT). (a) Recall for RD $= 10\%$. (b) Recall for RD $= 25\%$. (c) Recall for RD $= 45\%$.

overhead and show the time for detection[2] and the total runtime. It can be observed that the overhead of pSIFT and sRD-SIFT with respect to the original SIFT is very small, with the former being slightly slower than the latter because of the rendering of the stereographic image. In rectSIFT, the exponential growth of computation time with RD is justified by the increasing size of the corrected warped frames.

The middle columns show the average results for detection and matching when the threshold for selecting keypoints in the DoG pyramid is 1.25×10^{-2}. The relative performance of SIFT, rectSIFT, and sRD-SIFT in terms of repeatability and matching precision is in accordance with the synthetic experiments in Figs. 1 and 3. For RD $= 45\%$, rectSIFT presents the highest repeatability score, but sRD-SIFT achieves substantially more detections thanks to the adaptive filtering that avoids an excessive blurring in the image periphery. Comparing sRD-SIFT with pSIFT, the former tends to achieve better repeatability and precision scores, but in overall terms, the two methods behave quite similarly. Since the test images undergo significant changes in viewpoint (see Fig. 4), the pSIFT invariance to camera rotation is an advantage that seems to compensate the drawbacks of the resampling used for rendering the stereographic image.

For some applications, like robot navigation, it is usually preferable to have few high-quality keypoints than many points that are often unstable and cannot be reliably detected and matched across different frames. Taking this into account, we decided to repeat the evaluation using only the 500 strongest detection responses in each frame. The results are shown in the right side of the table. The relative performance of the four methods is roughly the same as above, with the comparative advantages of sRD-SIFT and pSIFT becoming less pronounced, and pSIFT emerging as top-performer for RD $= 45\%$.

Fig. 5(a)– (c) aims to compare the four descriptors being tested. The precision–recall of each method is measured over the same set of keypoints detected using sRD-SIFT. The results are consistent with the observations made in Fig. 3, with the implicit gradient correction of Section IV-A being the top-performer for RD amounts up to 25%. For very high distortion, the explicit correction by interpolation provides the best keypoint description and can be used as an alternative to the implicit gradient correction technique to further improve the matching results of our framework. Surprisingly, the pSIFT descriptor presents a break in terms of descriptor distinctiveness for all levels of distortion. This fact is due to the additional resampling step for mapping the sphere support regions into a canonical patch of 41×41 pixels [7]. The pernicious effects of the operation might be negligible for coarse scale features, but for fine structures, the interpolation intervals are often too large and induce gross errors in the rendered patch.

[2]The detection time does not include the offline computation of the filter masks used by pSIFT and sRD-SIFT. For pSIFT, the MATLAB implementation supplied by the authors took around 5 min to compute the octave filters for each sequence. For sRD-SIFT, the MATLAB and C implementations took, respectively, 1.25 and 0.35 s to accomplish the task.

B. Structure-From-Motion

Accurate point correspondence across frames is of key importance in multiple-view geometry [8], [14]. In this section, we consider 21 image

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

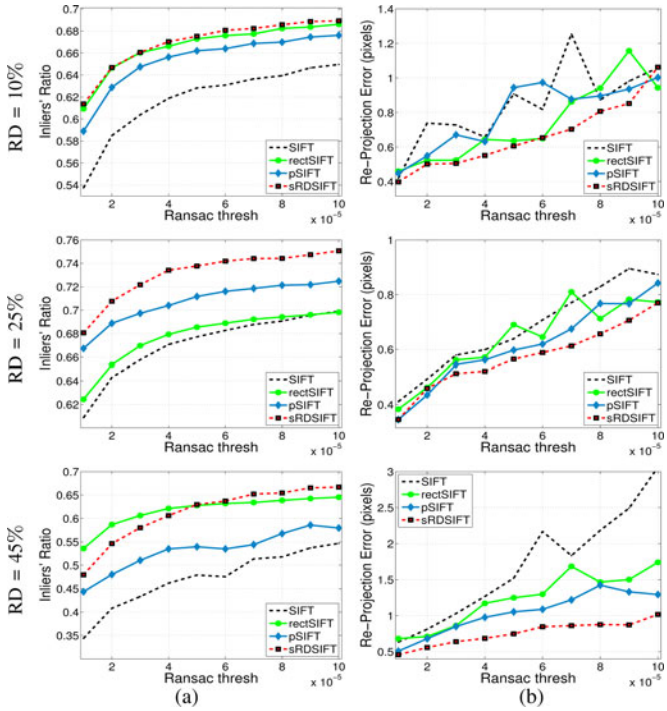


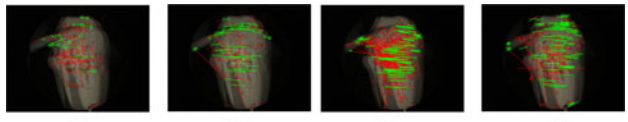
Fig. 6. Estimation of epipolar geometry from image correspondences obtained using the four keypoint methods. The graphics show average results for the ratio of inliers and the reprojection error when the value of the RANSAC threshold increases. We consider 21 image pairs for each camera and run the estimation algorithm 50 times for each RANSAC threshold value. (a) Inliers' ratio. (b) Reprojection error.

pairs of a scene with depth variation and estimate the relative camera motion using epipolar geometry. The experiment is performed for each one of the cameras used in the previous section, which are accurately calibrated employing the method described in [22]. The rigid camera motion is estimated by the well-known five-point algorithm [23], which is run in a robust RANSAC procedure [24]. Fig. 6 compares the SfM results for each camera when the input correspondences are obtained using SIFT, rectSIFT, pSIFT, and sRD-SIFT.

The RANSAC algorithm is an iterative scheme that computes the essential matrix from five randomly chosen correspondences and counts the number of point matches that agree with the achieved estimation. A point match is considered to be an inlier *iff* the Sampson distance to the corresponding epipolar line is below a certain threshold value [14]. We decided to vary this threshold and analyze the inlier correspondences and the reprojection error after estimating the relative displacement between cameras. Although not shown in the graphics, sRD-SIFT provides the largest number of inliers, being closely followed by pSIFT. rectSIFT and SIFT achieve roughly 20% and 30% less correct correspondences than our framework. Fig. 6(a) depicts the percentage of point matches that are classified as inliers during the SfM estimation. It can be observed that sRD-SIFT is consistently the top-performer, significantly ahead of pSIFT for all levels of distortion. The same happens for the reprojection error meaning that our approach is more accurate in localizing the keypoints. This advantage is explained by the fact that we completely avoid image resampling that introduces subpixel position errors during the interpolation.

C. Structure-From-Motion in Medical Endoscopy

We are currently engaged in a project that aims to implement SfM from arthroscopic images for the purpose of computer-aided navigation



	#M	#M^{true}	% Prec	Reproj. Error
SIFT	110	13	11%	—
rectSIFT	130	52	40%	0.987
pSIFT	300	94	31%	0.927
sRD-SIFT	306	134	42%	0.457

Fig. 7. SfM in endoscopic stereo images with low texture. (a)–(d) Two images are overlaid and the point correspondences for each method are marked in red. The table shows the number of input matches for the RANSAC algorithm, the number of inliers (green matches), and the final reprojection error. The relative motion between views is refined by minimizing the reprojection error using iterative bundle adjustment [14]. For the case of SIFT correspondences, the RANSAC algorithm [24] was unable to converge to a plausible solution. (a) SIFT. (b) rectSIFT. (c) pSIFT. (d) sRD-SIFT.

in orthopedic surgery. In this context, finding accurate image correspondences is specially difficult because not only the endoscopic lens introduces severe RD (RD $\approx 35\%$), but also the surfaces in the joint cavity tend to be textureless (e.g., bones). Fig. 7 compares the estimation of the rigid motion between two views of a knee model, when the point correspondences are obtained using SIFT, rectSIFT, pSIFT, and sRD-SIFT. The scene is very poorly textured, but there are small image structures that can be potentially matched to accomplish the task. The original SIFT provides unreliable matches because it cannot handle the joint effect of RD and lack of texture. The results improve when the keypoint detection and matching is carried out after distortion correction via image resampling. However, and given the reprojection error, the accuracy of the motion estimation is not the best. The problem when using rectSIFT is that the interpolation tends to smooth the fine image structures, and the number of keypoint detections is relatively small. pSIFT improves the number of input matches, but the interpolation errors during resampling propagate to the camera motion estimation and the final reprojection error is similar to the one achieved with rectSIFT. sRD-SIFT seems to handle well the situation, providing the largest number of inliers and reducing the reprojection error in 50%. This is an example where using sRD-SIFT makes all the difference in terms of the achieved results.

D. Visual Odometry for Recovering Robot Trajectory

In this indoor experiment, a mobile robot with a fish-eye camera describes a loop around a table. The objective is to recover the motion from a sparse sequence of 19 images with RD $\approx 45\%$. The motion estimation is carried by a sequential SfM pipeline that uses as input the point matches obtained by the four competing keypoint methods. This pipeline iteratively adds new consecutive frames with a five-point RANSAC initialization (using two views) [23], a scale factor adjustment (using three views) [14], and a final refinement with a sliding window bundle adjustment. This experiment is, particularly, challenging due to the wide-baseline displacements between frames, with the usage of a fish-eye lens being crucial to have a sufficient overlay between views and, hence, obtain enough feature matches for the motion estimation. The trajectory is a closed loop; however, we do not perform matching between the last and the first view. In this manner, the motion estimation tends to accumulate drift error, with the corresponding magnitude being an indicator of the quality of input features. Fig. 8 shows that the motion estimation using sRD-SIFT keypoints is substantially

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

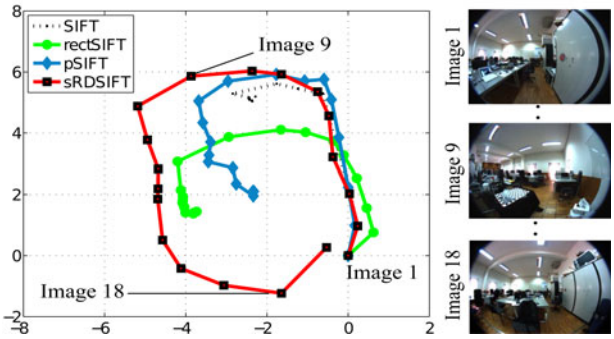


Fig. 8. Robot motion estimation. The trajectory is a closed loop, with the last image being captured in the same position as the first one.

superior to the others. The reason behind this result is that pSIFT and rectSIFT establish a lower number of geometrically consistent matches than sRDSIFT. The absence of good features, conjugated with the lower subpixel precision and with the stochastic nature of the RANSAC initialization, leads to high variability of motion estimations in repeated runs of the SfM pipeline, making rectSIFT and pSIFT unreliable.

VI. Conclusion

This paper proposes modifications to the broadly used SIFT framework that make it resilient to image RD, while preserving the original invariance to scale, rotation, and moderate viewpoint change. The only assumptions are that the camera follows the division model [18], and that the amount of distortion is coarsely known. We ran several experiments, both in synthetic and real frames, that prove the advantages of sRD-SIFT whenever there is significant image distortion. Our method provides significantly more correct point correspondences than the SIFT algorithm run after correcting the distortion via image warping. Comparing with the pSIFT algorithm [7], the gains in terms of number of matches are marginal, but sRD-SIFT has a higher accuracy in keypoint localization as proven by the experiments described in Sections V-B–V-D. These benefits are achieved at the expense of a small computational overhead when compared with the standard SIFT implementation. sRD-SIFT can be advantageous in several robot vision tasks, ranging from SfM to visual recognition, as well as in medical applications that rely in endoscopic imagery.

The main virtue of our approach is that it completely avoids image signal resampling. The interpolation used in previous works, which require image warping operations [4], [7], severely affects the keypoint detection performance. With sRD-SIFT, we show that the RD can be locally compensated using an adaptive kernel, and that this adaptive filtering can be implemented in a computationally affordable manner. The latest version of sRD-SIFT binaries are available for download at *http://arthronav.isr.uc.pt/ mlourenco/srdsift*.

Acknowledgment

The authors would like to thank P. Hansen and P. Corke for making available the implementation of the pSIFT algorithm that was used to run the comparative tests, and A. Malti for useful discussion at the beginning of this project.

References

[1] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 10, pp. 1615–1630, Oct. 2005.
[2] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. Van Gool, "A comparison of affine region detectors," *Int. J. Comput. Vision*, vol. 65, pp. 43–72, 2005.
[3] D. G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," *Int. J. Comput. Vision*, vol. 60, pp. 91–110, 2004.
[4] R. Castle, D. Gawley, G. Klein, and D. Murray, "Towards simultaneous recognition, localization and mapping for hand-held and wearable cameras," in *Proc. IEEE Int. Conf. Robot. Autom*, 2007, pp. 4102–4107.
[5] S. Se, D. Lowe, and J. Little, "Mobile robot localization and mapping with uncertainty using scale-invariant visual landmarks," *Int. J. Robot. Res.*, vol. 21, pp. 735–758, 2002.
[6] P. Baker, C. Fermuller, Y. Aloimonos, and R. Pless, "A spherical eye from multiple cameras (Makes better models of the world," in *Proc. IEEE Conf. Comp. Vis. Pattern Recognit.*, 2001, pp. I-576–I-583.
[7] P. Hansen, P. Corke, and W. Boles, "Wide-angle visual feature matching for outdoor localization," *Int. J. Robot. Res.*, vol. 29, pp. 267–297, 2010.
[8] B. Steder, G. Grisetti, C. Stachniss, and W. Burgard, "Visual slam for flying vehicles," *IEEE Trans Robot.*, vol. 24, no. 5, pp. 1088–1093, Oct. 2008.
[9] D. Burschka, M. Li, R. H. Taylor, and G. D. Hager, "Scale-invariant registration of monocular endoscopic images to CT-Scans for sinus surgery," *Med. Image Anal.*, vol. 5, pp. 413–426, 2004.
[10] K. Daniilidis, A. Makadia, and T. Bulow, "Image Processing in Catadioptric Planes: Spaciotemporal Derivatives and Optical Flow Computation," in *Proc. Int. Workshop on Omndirectional Vision*, 2002, pp. 3–12.
[11] M. Lourenco, J. Barreto, and F. Vasconcelos, "Model-based keypoint detection in images with radial distortion," Inst. Syst. Robot., Univ. Coimbra, Coimbra, Portugal, Tech. Rep., 2011.
[12] M. Lourenco, J. P. Barreto, and A. Malti, "Feature detection and matching in images with radial distortion," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2010, pp. 1028–1034.
[13] T. Bulow, "Spherical diffusion for 3D surface smoothing," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 12, pp. 1650–1654, Dec. 2004.
[14] Y. Ma, S. Soatto, J. Kosecka, and S. Sastry, *An Invitation to 3D Vision: From Images to Geometric Models*. New York: Springer-Verlag, 2003.
[15] L. Velho, A. Frery, and J. Gomes, *Image Processing for Computer Graphics and Vision*. London, U.K.: Springer, 2008.
[16] Z. Arican and P. Frossard, "OmniSIFT: Scale invariant features in omnidirectional images," in *Proc. IEEE Int. Conf. Image Process.*, 2010, pp. 3505–3508.
[17] L. Puig and J. J. Guerrero, "Scale space for central catadioptric systems. Towards a generic camera feature extractor," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2011, pp. 1599–1606.
[18] A. Fitzgibbon, "Simultaneous linear estimation of multiple view geometry and lens distortion," in *Proc. IEEE Conf. Comp. Vis. Pattern Recognit.*, 2001, pp. I-125–I-132.
[19] T. Lindeberg, "Feature detection with automatic scale selection," *Int. J. Comput. Vision*, vol. 30, pp. 77–116, 1998.
[20] J. L. Crowley and A. C. Parker, "A representation for shape based on peaks and ridges in the difference of low-pass transform," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-6, no. 2, pp. 156–170, Mar. 1984.
[21] J. P. Barreto, "A unifying geometric representation for central projection systems," *Comput. Vis. Image Underst.*, vol. 103, pp. 208–217, 2006.
[22] J. P. Barreto, J. Roquette, P. Sturm, and F. Fonseca, "Automatic camera calibration applied to medical endoscopy," in *Proc. Brit. Mach. Vis. Conf.*, 2009, pp. 1–10.
[23] D. Nistér, "An efficient solution to the five-point relative pose problem," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 6, pp. 756–770, Jun. 2004.
[24] M. A. Fischler and R. C. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, vol. 24, pp. 381–395, 1981.