2010 IEEE International Conference on Robotics and Automation
Anchorage Convention District
May 3-8, 2010, Anchorage, Alaska, USA

Feature Detection and Matching in Images with Radial Distortion

Miguel Lourenço, João P. Barreto and Abed Malti
Institute of Systems and Robotics,
Faculty of Science and Technology,
University of Coimbra.
3030 Coimbra, Portugal
{miguel,jpbar,amalti}@isr.uc.pt

***Abstract*— Image keypoints are broadly used in robotics for different purposes, ranging from recognition to 3D reconstruction, passing by SLAM and visual servoing. Robust keypoint matching across different views is problematic because of the relative motion between camera and scene that causes significant changes in feature appearance. The problem can be partially overcome by using *state-of-the-art* methods for keypoint detection and matching, that are resilient to common affine transformations such as changes in scale and rotation. Unfortunately, these approaches are not invariant to the radial distortion present in images acquired by cameras with wide field-of-view. This article proposes modifications to the Scale Invariant Feature Transform (SIFT), that improve the repeatability of detection and effectiveness of matching in the presence of distortion, while preserving the characteristics of invariance to scale and rotation. These modifications require an approximate modeling of the image distortion, and consist in using adaptative gaussian filtering for detection and implicit gradient correction for description. Extensive experiments, with both synthetic and real images, show that our method outperforms explicit distortion correction using image rectification.**

I. Introduction

The Scale-Invariant Feature Transform (SIFT) [1] enables keypoint detection and description in conventional perspective images, providing invariance to common image transformation such as scale, rotation, illumination, and minimal viewpoint changes [2]. In the past SIFT has been successfully applied in robotics for performing different tasks such as visual servoing and SLAM [3], [4]. In addition, robotic systems can benefit from the usage of wide field-of-view images. Panoramic cameras enable a more thorough visual coverage of the environments, and are highly advantageous in egomotion estimation by avoiding ambiguities between translation and rotation whenever the translation direction lies outside the field of view [5], [6]. However, the projection in cameras with wide angle lens presents strong radial distortion caused by the bending of the light rays when crossing the optics. The distortion increases as we go far a way from the center, and it is typically described by non-linear terms that are function of the image radius. Since the original SIFT algorithm was not designed to handle this type of image deformation, keypoint detection and matching in wide-angle imagery can be highly problematic [7].

The authors acknowledge the Portuguese Science Foundation, that generously funded this work through grant PTDC/EEA-ACR/68887/2006

The SIFT algorithm performs keypoint detection in a scale-space representation of the image [8], [9] applying an approximating of Laplacian-of-Gaussian (LoG) by the Difference-of-Gaussian (DoG). The detection is carried in the DoG pyramid by looking for extrema simultaneously in scale and space, with the extrema being illustrative of the correlation between the characteristic length of the signal feature and the standard deviation of the filter σ. After the detection of the keypoints, the processing is carried at the level of the gaussian pyramid where the extrema occurred, and a main orientation, based on the spatial gradients, is assigned to each keypoint. The final descriptor is computed using a patch of 16×16, after rotation according to the previously assigned orientation, providing invariance to image rotation.

Radial distortion (RD) is a non-linear geometric deformation that moves the pixel position along the radial direction and towards the center of distortion. In broad terms, the compression induced by the RD diminishes the characteristic length of the signal features and, as a consequence, the corresponding extrema tend to occur at lower levels of scale than they would occur in the absence of distortion. In addition, the image gradients are also affected by the pixel shifting induced by RD. The SIFT descriptor, despite of being robust to small changes in the gradient contributions, suffers a considerable deterioration for significant amounts of distortion, which has a negative impact in the recognition performance.

Despite of the fact that the SIFT algorithm is not invariant to RD, it has been applied in the past to images with significant distortion. While ones ignore the pernicious effects of RD and directly apply the original SIFT algorithm over distorted images [10], others perform a preliminary correction of distortion through image rectification and then apply SIFT [11]. This last approach has two major drawbacks: (i) the rectification is computationally expensive, specially when dealing with large sized images; (ii) the image re-sampling requires interpolation that, depending on the choice of reconstruction filter, can adulterate the spectrum of the image signal and affect the response of the DoG [12]. Recently, Hansen et al. [13] proposed an approach to extend SIFT for wide angle images. The method assumes that camera calibration is known and they suggest to back-project the image onto an unitary sphere and build a scale-space representation that is the solution of the diffusion

equation over the sphere. Such representation minors the problems inherent to planar perspective projection, enabling RD invariance and extra invariance to rotation. However, the approach requires perfect camera calibration and tends to be highly complex and computationally expensive.

In contrast with [13], we propose a set of well engineered modifications to the original SIFT algorithm to achieve RD invariance. Every processing step is carried directly in the distorted image plane and the additional computational cost is marginal. The gaussian pyramid is obtained by convolution with a gaussian filter, whose shape is modified in terms of the image radius. The objective is to take into account the distortion effect, such that the final DoG representation is equivalent to the one that would be obtained by filtering in the absence of distortion and subsequently applying the RD. In a similar manner, the SIFT descriptors are computed directly over the distorted image after correcting the image gradients using the derivative chain rule. Comparative studies show that the modified SIFT algorithm outperforms the approach of correcting the distortion through image rectification in terms of detection repeatability, precision-recall of matching, and computational efficiency, preserving scale and rotation invariance.

The structure is as follows: Section II briefly reviews the SIFT algorithm and the division model [14] that is assumed for describing the image distortion. Section III studies the effect of the radial distortion in keypoint detection, and derives the gaussian adaptative filtering for overcoming the problems caused by image deformation. Section IV evaluates the impact of the distortion in the keypoint description, and proposes implicit gradient correction to account for the RD effect. Finally, section V conducts tests using real distorted images taken from different viewpoints.

Notation: 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 $\mathbf{f}(\mathbf{x}) = (f_x(\mathbf{x}), f_y(\mathbf{x}))^\mathsf{T}$. We will also often use RD to refer to radial distortion.

II. Theoretical Background

A. Scale Invariant Features Transform

Lowe adopts a strategy that approximates the Laplacian-of-Gaussian (LoG), used for the scale-space representation [8], [9], by the DoG operator [1]. Let $\mathtt{I}(x, y)$ be an image signal and $\mathsf{G}_\sigma(x, y)$ a 2D gaussian function with standard deviation σ. The blurred version of $\mathtt{I}(x, y)$ is obtained by its convolution with the gaussian

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

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

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

In the pyramid of DoG images each pixel is compared with its neighborhood pixels in order to find local extrema in scale and space. These extrema are subsequently filtered and refined to obtain the detected keypoints. After the detection of the keypoint, the next steps concern the computation of the final descriptor using the image gradients of a local patch around the point. In order to achieve scale invariance, all the computations are performed at the scale of selection of the keypoint in the gaussian pyramid. The method starts by finding the dominant orientation of the local gradients, and uses it for rotating the image patch towards a normalized position in order to achieve invariance to rotation transformations. For the main orientation assignment, an histogram for 36 bins (10 degrees per bin). Each sample is weighted by a gaussian of 1.5σ to give less emphasis to contributions far from the keypoint. The normalizing rotation is performed and the final SIFT descriptor is computed from a patch of 16×16 pixels divided into subregions of 4×4 pixels, each one providing 8 main orientations [1].

B. The Division Model for Radial Distortion

The effect of lens distortion in image acquisition can be often described using the first order division model [14]. Let $\mathbf{x} = (x, y)$ be a point in the distorted image $\mathtt{I}$, and $\hat{\mathbf{x}} = (u, v)$ the corresponding point in the undistorted image $\hat{\mathtt{I}}$. The origin of coordinate system is assumed to be coincident with the distortion center, which is approximated by the image center [15]. The amount of distortion is quantified by a parameter ξ (typically $\xi < 0$), and undistorted image points $\hat{\mathbf{x}}$ are mapped into distorted points $\mathbf{x}$ by function $\mathbf{f}$:

$$\mathbf{x} = \mathbf{f}(\hat{\mathbf{x}}) = \begin{pmatrix} f_x(\hat{\mathbf{x}}) \\ f_y(\hat{\mathbf{x}}) \end{pmatrix} = \begin{pmatrix} \frac{2u}{1+\sqrt{1-4\xi(u^2+v^2)}} \\ \frac{2v}{1+\sqrt{1-4\xi(u^2+v^2)}} \end{pmatrix}, \tag{3}$$

The distorted image can be rectified using the inverse of distortion function :

$$\widehat{\mathbf{x}} = \mathbf{f}^{-1}(\mathbf{x}) = \begin{pmatrix} f_u^{-1}(\mathbf{x}) \\ f_v^{-1}(\mathbf{x}) \end{pmatrix} = \begin{pmatrix} \frac{x}{1+\xi(x^2+y^2)} \\ \frac{y}{1+\xi(x^2+y^2)} \end{pmatrix} \tag{4}$$

The function $\mathbf{f}$ is radially symmetric around the image center, and its action can be understood as a shift of image points towards the center along the radial direction. The relationship between undistorted and distorted radius is given by :

$$\hat{r} = \frac{r}{1 + \xi r^2} \tag{5}$$

Radial distortion causes a space compression of the image information, which substantially changes the signal spectrum and introduces new high frequency components. To provide the notion of how much the image is compressed, we will often express the amount of distortion through the normalized decrease in the maximum image radius:

$$\%_{distortion} = \frac{\hat{r}_M - r_M}{\hat{r}_M} * 100 \tag{6}$$

with $\hat{r}_M$ and r_M denoting respectively the maximum values for the undistorted and distorted image radius. Through this work we will always assume that image distortion follows the division model.

Fig. 1. Some of the images used for the synthetic experiments and its correspondent distorted views. The data set comprises a broad variety of scenes and visual contexts.

III. SIFT DETECTION UNDER RADIAL DISTORTION

If we apply SIFT directly over a distorted image, the corresponding multi-scale representation is different from the one that would be obtained from the equivalent perspective image in the absence of RD. The distortion compresses the intensity spectrum of the image, introducing new high frequency components. This leads to the detection of some unstable points, which would not be detected in the undistorted image, as well as the non-detection of others.

A. Evaluation using Images with Artificially added RD

To study SIFT detection under RD, we used a set of images from the internet, and we artificially injected radial distortion (Fig. 1). We decide to perform such synthetic experiment in order to control the amount of distortion, to know the positions where keypoint detection should occur (ground truth), and because it would not be practically feasible to acquire multiple images with different distortions from the same viewpoint. Let's consider an image of the data set and one of its distorted versions, and assume S_0 and S as being the set of keypoints detected in the original and distorted images, respectively. The elements of S can either be points already detected in the original image, or new keypoints that appear due to the high frequency components introduced by radial distortion. Henceforth, we will denote the former by S^d and the latter by S^{new} such that:

$$S = S^d \cup S^{new} \tag{7}$$

$$S^d = S_0 \cap S \tag{8}$$

The set S^d contains keypoints in the distorted image detected at a correct spatial location. However, the correct assignment of scale is fundamental for achieving reliable matching across different views. Therefore set S^d is split in two subsets: S^c containing the points detected at correct scale and location, and S^{ws} being the set of points close in space but not in scale (detections at wrong scale).

$$S^c = S_0 \cap (S - (S^{new} \cup S^{ws})) \tag{9}$$

From the subset introduced, the repeatability in keypoint detection is evaluated using the following metric:

$$\%_{\text{Repeatability}} = \frac{\#S^c}{\#S_0} * 100 \tag{10}$$

with $\#$ denoting the number of keypoints in each set. The occurrence of new spurious detections due to radial distortion is quantified as follows:

$$\%_{\text{New detections}} = \frac{\#S^{new}}{\#S} * 100 \tag{11}$$

And finally the detection at wrong scale is characterized by the percentage of points detected at incorrect scale with respect to the points detected at a correct image location [1]:

$$\%_{\text{Keypoints at wrong scale}} = \frac{\#S^{ws}}{\#S^d} * 100 \tag{12}$$

B. How does RD affect Keypoint Detection?

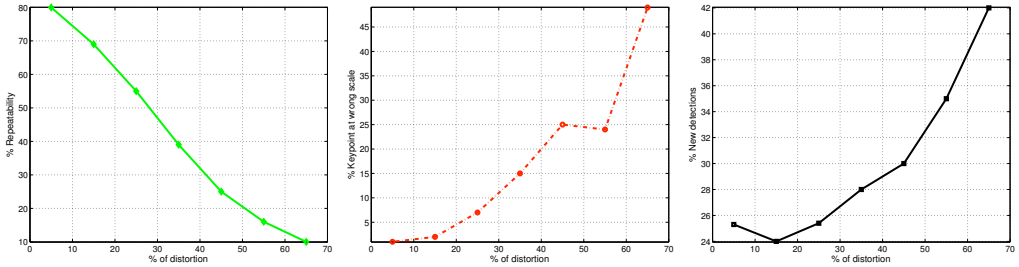
The compressing effect induced by radial distortion is responsible for several problems during keypoint detection. Since the level σ of the DoG pyramid at which detection occurs reflects the characteristic length of a certain feature in the image, the compressive effect of RD pushes the extrema detection towards lower values of scales. Since SIFT starts filtering at $\sigma_0 = 1.6$, some keypoints will no longer be picked as an extrema because the value of their *scale* will no longer be considered in SIFT band-pass filtering. In addition there will be keypoints detected at different scales and, the high frequency components introduced by RD, can even lead to new detections. Fig. 2 shows experimental evidence of the degradation of SIFT detection in images with increasing RD. The observed behavior is in accordance with the stated theoretical interpretation: (i) the loss of repeatability is more pronounced at lower levels of the DoG pyramid, and detections at wrong scales arise at coarser levels of scale, which reflects the fact that RD makes the keypoints smaller; (ii) the compression induced by RD in the image spectrum creates new unstable keypoints that were not detected in the original image.

C. Adaptative gaussian filtering

We introduce a new approach for image adaptative blurring that accounts for the RD effect. The objective is to generate a scale-space representation equivalent to the one that would be obtained by filtering the image in the absence of distortion, followed by applying the distortion over all the levels of the DoG pyramid. Remark that this is different from the DoG obtained by simply convolving the distorted image with the standard isotropic gaussian kernel, in the sense that in this case the action of the distortion is before (and not after) the gaussian filtering. To achieve such goal we will perform the distortion correction in an implicit manner, by adapting the convolution kernel that is used directly over the distorted image.

Let G_σ be a bi-dimensional gaussian function with standard deviation σ, $\hat{\mathtt{I}}$ the undistorted image, and $\mathtt{I}$ the distorted image. The value of the blurred undistorted image $\hat{\mathsf{L}}_\sigma$ at pixel (s,t) is given by

$$\hat{\mathsf{L}}_\sigma(s,t) = \sum_{u=-\infty}^{+\infty} \sum_{v=-\infty}^{+\infty} \hat{\mathtt{I}}(u,v)\, \mathsf{G}_\sigma(s-u, t-v) \tag{13}$$



(a) Repeatability of the detection (b) Detections at wrong scale (c) New detections

Fig. 2. Experimental evaluation of the SIFT detector under RD images. As it can be seen the SIFT detector is severely affected by RD, being the repeatability clearly affected even for lower levels of distortion.

This is the convolution that SIFT performs for the case of the image being rectified for correcting the distortion. However, and since we want to work directly the with distorted image $\mathtt{I}$, the undistorted image $\hat{\mathtt{I}}$ can be replaced by its distorted counterpart, taking into account the inverse of the mapping function $\mathbf{f}()$ (4). Considering that

$$\hat{\mathtt{I}}(u,v)) = \mathtt{I}(f_u^{-1}(x,y), f_v^{-1}(x,y)), \tag{14}$$

and changing the variables (u,v) by (x,y) in (13), it arises:

$$\hat{\mathtt{L}}_\sigma(s,t) = \sum_{x=-\frac{1}{\sqrt{-\xi}}}^{\frac{1}{\sqrt{-\xi}}} \sum_{y=-\frac{1}{\sqrt{-\xi}}}^{\frac{1}{\sqrt{-\xi}}} \mathtt{I}(x,y)\, \mathtt{G}_\sigma(s - f_u^{-1}(x,y), t - f_v^{-1}(x,y)) \tag{15}$$

Since $\mathtt{L}_\sigma$ is the distorted version of the smoothed image $\hat{\mathtt{L}}_\sigma$, we can repeat the reasoning and change the undistorted coordinates (s,t) by their distorted counterparts (h,k). It follows that

$$\mathtt{L}_\sigma(h,k) = \sum_{x=-\frac{1}{\sqrt{-\xi}}}^{\frac{1}{\sqrt{-\xi}}} \sum_{y=-\frac{1}{\sqrt{-\xi}}}^{\frac{1}{\sqrt{-\xi}}} \mathtt{I}(x,y)\, \mathtt{G}_\sigma(f_u^{-1}(h,k) - f_u^{-1}(x,y), f_v^{-1}(h,k) - f_v^{-1}(x,y)), \tag{16}$$

which after some algebraic manipulations leads to

$$\mathtt{L}_\sigma(h,k) = \sum_{x=-\frac{1}{\sqrt{-\xi}}}^{\frac{1}{\sqrt{-\xi}}} \sum_{y=-\frac{1}{\sqrt{-\xi}}}^{\frac{1}{\sqrt{-\xi}}} \mathtt{I}(x,y)\, \mathtt{G}_\sigma\Big(\frac{h - x + \xi r^2(h\delta^2 - x)}{1 + \xi r^2(1 + \delta^2 + \xi r^2 \delta^2)}, \frac{k - y + \xi r^2(k\delta^2 - y)}{1 + \xi r^2(1 + \delta^2 + \xi r^2 \delta^2)}\Big). \tag{17}$$

with

$$\begin{cases} r &= \sqrt{h^2 + k^2} \\ \delta &= \sqrt{\frac{x^2+y^2}{h^2+k^2}} \end{cases} \tag{18}$$

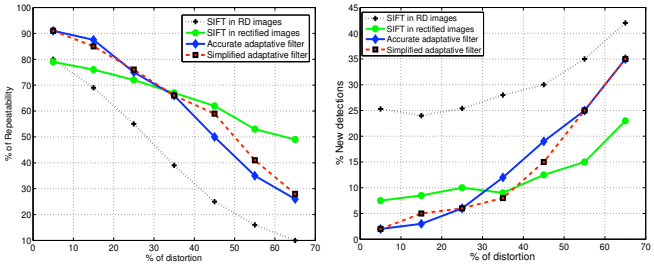
Remark that now the smoothing kernel depends on (x,y) and (h,k) and (17) is no longer a straightforward convolution. However, if the pixel coordinates (h,k) is very close to the center, then $\xi r^2 \approx 0$ and the expression becomes a standard convolution. This makes sense because the distortion in the central region is negligible and there is no need for the filter to make any compensation. On the other hand, if the pixel (h,k) is far from the center, then the filtering kernel only takes significant values for (x,y) close to the location (h,k) (the center of convolution), for which the ratio δ is approximately unitary ($\delta \approx 1$). In this particular case (17) can be simplified to

$$\mathtt{L}_\sigma(h,k) \approx \sum_{x=-\frac{1}{\sqrt{-\xi}}}^{\frac{1}{\sqrt{-\xi}}} \sum_{y=-\frac{1}{\sqrt{-\xi}}}^{\frac{1}{\sqrt{-\xi}}} \mathtt{I}(x,y)\, \mathtt{G}_\sigma\Big(\frac{1}{1+\xi r^2}(h-x), \frac{1}{1+\xi r^2}(k-y)\Big) \tag{19}$$

The result above is an approximation of (17), and henceforth we will call it the *simplified adaptative filter*. While in the original SIFT detection the image is blurred using a standard isotropic gaussian kernel with standard deviation σ, in our case the standard deviation of the filter decreases as a function of the images radius $(1+\xi r^2)\sigma$. The convolution kernel follows the deformation caused by RD, and emphasizes the contribution of pixels increasingly closer to the convolution point while the filter moves far from the center of distortion. The blurring using a standard gaussian filter uses the same kernel mask over the entire image. Moreover the computational efficiency of the convolution can be largely improved by taking advantage of the decoupling properties in X and Y of the gaussian [9]. Unfortunately, the dependence of the adaptative filter with respect to the radius requires using different kernels for different concentric image circle locations. However, while the accurate adaptative filtering of equation (17) cannot be decoupled, the convolution with its simplified version in (19) can be separately done in X and Y dimensions, adding a minimal computational overhead when compared to a spatial invariant gaussian filter.

D. Detection Results

In terms of detection evaluation, the repeatability of keypoint detection is unarguably the most important property of a reliable detector [16]. Figure 3 compares the repeatability of detection at the correct location and scale by running different approaches over the synthetically distorted imagery. The properties of the derived adaptative filters allow to overcome the main limitations of SIFT under RD (Fig. 3). For the initial octaves of the scale pyramid, the adaptative gaussian filters allow to detect points that original SIFT does



(a) Repeatability of the detection (b) New detections

Fig. 3. We compare the proposed adaptative filters against the original SIFT algorithm ran in both distorted and rectified images (Fig.1). The repeatability of detection for different amounts of distortion is shown in (a). The adaptative filtering provides the highest repeatability rates for amounts of RD up to $\approx 35\%$. The performance of accurate and simplified adaptation is very similar, and henceforth we will consider the latter because of computational efficiency reasons. The graphic (b) concerns the percentage of new spurious detection showing that the improved repeatability of the adaptative filtering is not achieved at the expenses of an increase in detection specificity.

not consider anymore. They also allow to model the structure size at higher levels of the pyramid in order to avoid detections at wrong scale (Fig. 2(b)). It is somewhat surprising the fact that adaptive filtering outperforms image rectification for medium-small amounts of RD. The image re-sampling for distortion compensation implicitly requires reconstructing the discrete signal. Depending on the type of low-band pass filtering (in our case we use first order interpolation), the reconstruction can either remove high frequency components and/or introduce new spurious frequencies [12]. Thus, the rectification causes changes in the image spectrum that have consequences in terms of the detection repeatability. The skeptical reader can easily verify this by performing a linear image rescaling (expansion to avoid aliasing effects) and compare the SIFT detections. Contrary to the expected, not every keypoint in the original images is detected in the scaled version.

For high amounts of RD (above 35%) the image rectification outperforms the adaptive filtering. When the compressive effect of RD is too high there are image structures that vanish and become impossible to detect without performing some kind of image reconstruction. This partially explains this experimental observation.

IV. Matching in Radial Distorted Space

By applying a certain amounts of distortion to an image, the pixels are shifted towards the center along the radial direction. This will deform the image gradients and consequently corrupt the SIFT descriptors (see Fig. 4 (c)). However, if we consider that the distortion can be reversed using the inverse mapping of equation (14), we can compute the distorted image gradients and correct them by applying a chain-rule derivation. The correction of image gradients using the chain rule can be carried only on the neighborhood of detected keypoints at the scale of selection in the distorted scale-space, which avoids a significant computational overhead.

Applying the chain rule derivation on (14), we obtain

$$\begin{pmatrix} \frac{\partial \hat{\mathtt{I}}}{\partial u} \\ \\ \frac{\partial \hat{\mathtt{I}}}{\partial v} \end{pmatrix} = \begin{pmatrix} \frac{\partial \mathtt{I}}{\partial x}\frac{\partial f_x}{\partial u} + \frac{\partial \mathtt{I}}{\partial y}\frac{\partial f_y}{\partial u} \\ \\ \frac{\partial \mathtt{I}}{\partial x}\frac{\partial f_x}{\partial v} + \frac{\partial \mathtt{I}}{\partial y}\frac{\partial f_y}{\partial v} \end{pmatrix} = \mathsf{J} \begin{pmatrix} \frac{\partial \mathtt{I}}{\partial x} \\ \\ \frac{\partial \mathtt{I}}{\partial y} \end{pmatrix} \quad (20)$$

J denotes the jacobian of function $\mathbf{f}$, which can be expressed in terms of point coordinates in the distorted image

$$\mathsf{J} = \frac{1+\xi r^2}{1-\xi r^2} \begin{pmatrix} 1-\xi(r^2-8x^2) & 8\xi xy \\ 8\xi xy & 1-\xi(r^2-8y^2) \end{pmatrix} \quad (21)$$

with r denoting the radius of the distorted pixel (x, y).

Thus, instead of correcting the distortion in the entire image using rectification, we propose to apply the gradient compensation around the keypoints detected in the distorted image. This provides gains in computational efficiency and avoids interpolation artifacts that can change the local appearance of the features.

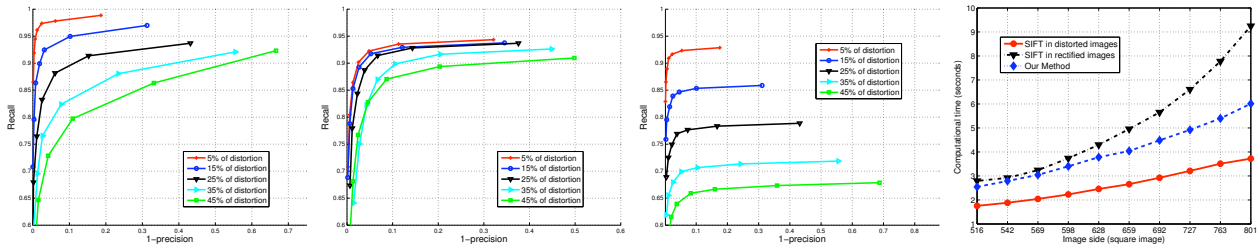
A. Evaluating Matching Performance

In order to evaluate the effectiveness of image gradient correction, we will match features extracted in the original (undistorted) image of the data set shown in Fig. 1, with features detected in the corresponding artificially distorted images. The gradient correction is compared against matching results obtained by applying standard SIFT and by applying SIFT over corrected images after interpolation. The performance evaluation is described using Recall *vs* 1-Precision curves [2].

We can observe that, even for low amounts of distortion, the SIFT descriptor starts to be affected by radial distortion (Fig. 4(c)). It is easy to understand that when the image is compressed the local patch around each keypoint receives contributions that do not occur for the original undistorted image. As mentioned early, the SIFT descriptor is prepared to deal with small shifts inside each subregion histogram. However, for high amounts of distortion this effect becomes too noticeable and, as a consequence, the descriptor drifts in the feature space precluding a successful match.

Another relevant constraint for the SIFT descriptor usage in RD images is that the gaussian weighting, used to give more emphasis to contributions close to the keypoint, starts to loose its effectiveness. As we increase the distortion, some pixels, that were initially far from the keypoint, are shifted inside its neighborhood and actively contribute for the descriptor building. We reduce this pernicious effect by considered a weighting gaussian function with standard deviation $(1+\xi r^2)\sigma$, instead of a standard deviation of σ as used in the original SIFT approach. This allows to have similar contributions in the distorted and in the original undistorted image, and then improve the descriptor resilience.

From experimental evidence, it is clear that for low levels of distortion the method of implicit gradient correction outperforms the classic approaches, Fig. 4. It is also observed that the image rectification is the most valid approach for higher levels of distortion. Nevertheless, our method always provides better matching results when comparing with the



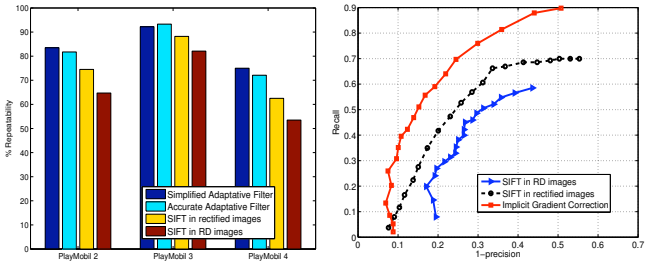
(a) Implicit gradient correction (b) Image rectification (c) SIFT in radial distorted image

Fig. 4. The curves (a)-(c) show the recall against 1-precision for increasing amounts of radial distortion. The recall indicates the percentage of correct matches obtained (true positives) over the entire set of possible correct matches (S^c subset). The 1-precision is a measure of specificity corresponding to the percentage of false positives over the total number of matches obtained. We can observe that the rectification from distortion allows high percentages of successful matches for all levels of distortion. However, until $\approx 25\%$ of RD the implicit gradient correction outperforms the rectification, being the most suitable approach for moderate levels of distortion. In (d) can be seen the comparison of computational time varying image size at constant distortion of 25%. Our method (simplified adaptative filter with implicit gradient correction) adds minimal computational complexity to the original method when compared with the explicit distortion correction.

(a) Playmobil data set ($\approx 15\%$ of distortion)

(b) Smiles data set ($\approx 30\%$ of distortion)

Fig. 5. The Playmobil data set was acquired with a lens of $\approx 15\%$ of distortion and the Smiles data set with one of $\approx 30\%$. Both data sets englobe a set of images taken form different viewpoint angles of a planar surface. As it can be seen the image appearance considerably changes due to the distortion effect allied to the viewpoint in which is taken.



(a) Detection repeatability (b) Curves of precision-recall

Fig. 6. Evaluation of the playmobil data set. In (a) is compared the repeatability of the detection for the 4 methods in evaluation in Fig. 3. The images undergo significant scale and viewpoint changes (Fig. 5), while the RD invariance is preserved. In (b) it can be seen that the implicit gradient correction overcomes the main limitations of the SIFT descriptor for moderate amounts of distortion. Since the images suffer scale and rotation changes, we can conclude that the invariances of the original SIFT descriptor are preserved.

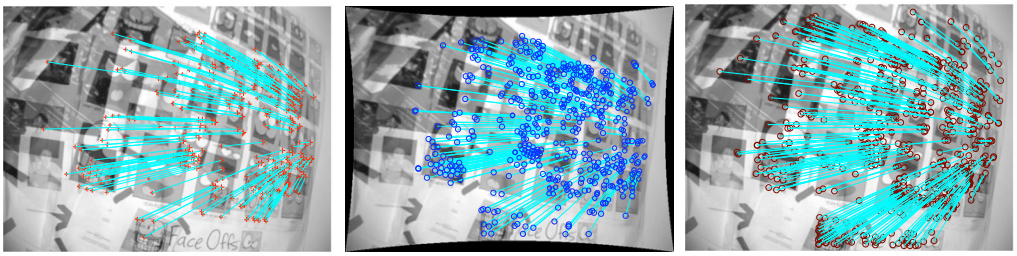
use of SIFT directly in distorted images. The implicit gradient correction technique allows to minimize the effect of the pixels shifting for moderate amounts of radial distortion.

V. Validation with Real Images

The tests performed so far with synthetic imagery provide reliable ground truth and enable to test the uniquely RD invariance. However, we aim to match images with different acquisition conditions, like scale, rotation and viewpoint changes. In this section we will carry on tests with real images with radial distortion undergoing significant viewpoint changes. This enables to evaluate the resilience to RD and also the invariance to rotation and scale that the original SIFT provides. The data set is composed by a set of images of a textured planar surface. This means that every two images are related by an homography that enables to generate the ground truth between images. In order to do this, a estimation of the distortion parameter [17] is performed. Then, the homography between the different image views was generated by hand using 10 correspondences. Then, this homography is used to select hundreds of automatically detected and matched keypoints between the two views and a new estimation based on these points is performed. We considered two data sets one with 15% of distortion and the other with RD of $\approx 30\%$ (Fig. 5).

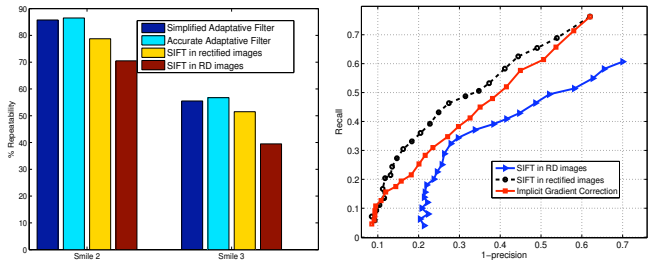
The playmobil data set is composed by set of images with moderate distortion, undergoing scale, rotation and viewpoint changes. We observe in Fig.6(a) that the proposed filters allow an improvement in detection under RD, outperforming the explicit distortion correction. We also confirmed that, for moderate amount of distortion, the implicit gradient correction performs better than the two classic approaches (Fig. 6(b)).

The smiles data set presents a set of images undergoing considerable viewpoint changes, with the estimated value of distortion being $\approx 30\%$. In terms of detection (Fig. 7(a)), our method is the one with higher score of successful detections. The derived filters preserve the scale invariant in feature detection, as can be observed for the real data sets repeatability. As proved under simulation, the implicit gradient correction starts to be affected by high values of distortion in the same manner as the original SIFT descriptor computed over RD images. In here, the rectification provides better performance than our method (Fig. 7(b)). However, since the recall measure depends on the S^c set, we can



(a) RD images (201 matches) (b) Rectified images (310 matches) (c) Our method (401 matches)

Fig. 8. Matches between smile2 and smile3. Our method provides better matching results in the image periphery, where the RD makes the others methods fail. To obtain the matches we use the ambiguity distance [13] and the threshold of 0.8 proposed by Lowe [1] to compute the descriptors distance. The outliers were discarded recurring to the homography between the two views.



(a) Detection repeatability (b) Smile

Fig. 7. Evaluation of the smiles data set. In (a) is compared the repeatability of the detection for the 4 method in evaluation in Fig. (3). We can observe that, although the repeatability diminishes as we augment the viewpoint, our method is the more resilient to distortion, showing the highest rates of repeatability. The final matching performance of our method is slightly poorer than image rectification in terms of precision-recall. However, the adaptive filtering still provides in absolute terms the highest number of correct matches (see Fig. 8)

argue that if a method for image descriptor presents lower performance in terms of recall but if the integrated detector is really efficient, the algorithm can provide better retrieval performance. Our method is advantageous when the images are acquired with lens that induce radial distortion since they allow an improvement of keypoints tracking across different views of the same scene, Fig. 8. The methods herein proposed allies the invariance of the original SIFT to a more resilient detection and description under radial distortion. From the experimental results (simulation and real cases), we can conclude that our method is a suitable approach for use in cameras where the lens induce radial distortion.

VI. Conclusions

In this paper we presented modifications to the broadly used SIFT algorithm that enhance it with invariance to image radial distortion. Extensive experiments prove that our method outperforms explicit image rectification for considerable amounts of distortion, preserving all the original invariance of SIFT with respect to scale and rotation. The proposed modifications add a minimal computational overhead to the original method, being potentially applicable to several robotic tasks. As future work we aim to improve the resilience of the descriptor built using the derivative chain rule. The proposed detection using adaptative filtering is extremely effective under considerable amounts of distortion, however increasing the distinctiveness of the descriptor is a priority in order to improve global performance.

References

[1] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vision*, 2004.

[2] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. V. Gool, "A comparison of affine region detectors," *Int. J. Comput. Vision*, vol. 65, no. 1/2, 2005.

[3] S. Se, D. Lowe, and J. Little, "Mobile robot localization and mapping with uncertainty using scale-invariant visual landmarks," *Int. J. of Robotics Research*, vol. 21, 2002.

[4] T. Nierobisch, J. Krettek, U. Khan, and F. Hoffmann, "Optimal large view visual servoing with sets of sift features," in *IEEE Int. Conf. on Robotics and Automation*, 2007.

[5] P. Baker, C. Fermuller, Y. Aloimonos, and R. Pless, "A spherical eye from multiple cameras (makes better models of the world," in *IEEE Int. Conf. on Comput. Vision and Pattern Recognition*, 2001.

[6] J. Gluckman and S. Nayar, "Egomotion and omnidirectional cameras," in *IEEE Int. Conf. on Comput. Vision*, 1998.

[7] M. Lourenço, "Techniques for keypoint detection and matching in endoscopic images," Coimbra, July 2009. [Online]. Available: http://sites.google.com/site/miguelrlourenco/research-interest

[8] 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.*, 1984.

[9] T. Lindeberg, "Feature detection with automatic scale selection," *Int. J. Comput. Vision*, vol. 30, no. 2, 1998.

[10] D. Burschka, M. Li, R. H. Taylor, and G. D. Hager, "Scale-invariant registration of monocular endoscopic images to ct-scans for sinus surgery," in *MICCAI (2)*, 2004.

[11] R. Castle, D. Gawley, G. Klein, and D. Murray, "Towards simultaneous recognition, localization and mapping for hand-held and wearable cameras," in *IEEE Int. Conf. on Robotics and Automation*, April 2007.

[12] L. Velho, A. Frery, and J. Gomes, *Image Processing for Computer Graphics and Vision*. Springer London, 2008.

[13] P. Hansen, P. Corke, W. Boles, and K. Daniilidis, "Scale-invariant features on the sphere," in *Int. Conf. on Comput. Vision*, Oct. 2007.

[14] A. Fitzgibbon, "Simultaneous linear estimation of multiple view geometry and lens distortion," in *Int. Conf. on Comput. Vision and Pattern Recognition*, 2001.

[15] R. Willson and S. Shaffer, "What is the center of the image," *Int. Conf. on Comput. Vision and Pattern Recognition*, 1993.

[16] T. Tuytelaars and K. Mikolajczyk, "Local invariant feature detectors: a survey," *Found. Trends. Comput. Graph. Vis.*, vol. 3, no. 3, 2008.

[17] J. P. Barreto and H. Araujo, "Geometric properties of central catadioptric line images and their application in calibration," *IEEE Trans. Pattern Anal. Mach. Intell.*, 2005.