Increasing the Precision of Junction Shaped Features

Kai Cordes and Jörn Ostermann
Institut für Informationsverarbeitung (TNT), Appelstr. 9A, 30167 Hannover, Germany
{cordes, ostermann}@tnt.uni-hannover.de

Abstract

The scale invariant feature operator (SFOP) detects circular features from an image using a spiral shape model. Special cases of the spiral model are junctions and circular symmetric shapes. The spatial localization is determined with subpixel accuracy which is obtained by an interpolation of the structure tensor in the scale space. For the interpolation, SFOP uses a 3D quadratic function. This leads to suboptimal solutions since the structure tensor surrounding a feature does not show the shape of a 3D quadratic. The aim of this paper is to improve the localization of the features detected by SFOP. A Difference of Gaussians function is proposed for the signal approximation which leads to improved precision values and to more accurate features. The proposed method improves the localization such that 72.5% of the features increase their precision. Hence, more features are extracted while increasing their repeatability by up to 9% on standard benchmarks.

1 Introduction

Scale invariant features play an important role in many computer vision applications, such as object recognition or scene reconstruction. In object recognition, the *scale invariant feature transform* (SIFT) technique [11] provides very good results for localization and matching. For scene reconstruction, the SIFT detector becomes more and more important, especially when a wide baseline correspondence analysis is required [3, 14]. The detected features are robust to changes in illumination, rotation, scale, and to surprisingly large viewpoint changes.

Recently, several scale invariant feature detectors have been proposed to focus on different shapes that occur in an image. The goal is to complete the description of the observed scene using complimentary feature types. The *scale invariant feature operator* (SFOP) junctions (cf. Fig. 1) proposed in [7] are of special interest because of its complementary property in conjunction with SIFT features. The combination of SIFT features and SFOP junctions maximizes the completeness with respect to image content [6, 8]. Hence, a very good distribution of features in the image can be achieved, which is essential for image coding as well as for scene reconstruction.

The SFOP junctions are localized by maximizing the *precision* in the Laplacian of Gaussians (LoG) scale space [7, 10]. For subpixel and subscale feature localization, the precision is approximated by a 3D quadratic. This approach commonly used by scale invariant feature detectors [1, 2, 9, 11] due to its computational efficiency. However, the interpolated gradient signal does not show the shape of a 3D quadratic which leads to a systematic error, as shown for SFOP in [4].

Figure 1: The detected junctions marked with cyan color are better explained with the proposed localization technique than with the reference [7] (yellow). They are selected because of their higher precision.

We improve the SFOP subpixel localization by a Difference of Gaussians approximation of the LoG scale space. This is well-justified, because scale invariant features and junctions are modeled by using a signal model with Gaussian shape [10]. For the evaluation, the repeatability criterion on two image benchmark data sets is used [5, 12]. The repeatability measure thresholds the overlap error to decide if detected feature pairs in two images are correct and counts valid feature pairs.

In the following Sect. 2, the reference localization technique using the SFOP approach is briefly explained. In Sect. 3, the proposed signal adapted localization technique is presented. Sect. 4 shows the experimental results on natural image data. In Sect. 5, the paper is concluded.

2 SFOP Feature Localization

The detection of SFOP features is done by localizing image positions $\mathbf{x}_0 = (x_0, y_0)$ with locally maximal *precision* [7]. The *precision* $\tilde{d}(\tau)$ introduced by Lindeberg [10] is calculated from the gradient vector $\nabla_\tau g(\mathbf{x}')$ and a Gaussian $G_\sigma(\mathbf{x}')$ in the scale space [7]:

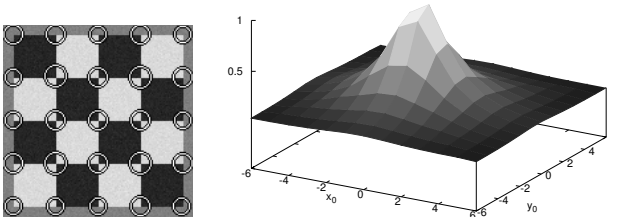
$$\tilde{d}(\tau) = \frac{\int |(\mathbf{x}_0 - \mathbf{x}')^\top \nabla_\tau g(\mathbf{x}')|^2 G_\sigma(\mathbf{x}') d\mathbf{x}'}{\int |\nabla_\tau g(\mathbf{x}')|^2 G_\sigma(\mathbf{x}') d\mathbf{x}'} \qquad (1)$$

Here, τ depicts the differentiation scale and σ is the integration scale.

For each point (x_0, y_0, σ) those scales σ are selected which have locally highest precision. The local maxima are found by evaluating, if the precision value in each scale σ is bigger or smaller that its 26 neighbors. A subpixel and subscale localization step is applied [2] to increase the accuracy of initially detected features.

Since the precision is calculated from a Laplacian Pyramid, the shape of the precision values nearby a feature is still Laplacian. This is demonstrated for an examples in Fig. 2. The gradients have Laplacian shape, they are not quadratic. Our motivation

is to show that exchanging the obviously suboptimal quadratic approximation with a signal adapted function leads to higher accuracy of the feature. We also show that our proposal leads to larger precision values.



(a) Junction features (b) Precision shape

Figure 2: Input image (from [7]) and the precision shape surrounding one of the detected junctions.

3 Signal Adapted Feature Localization

The localization procedure aims at determining subpixel coordinates for the detected feature. In [7], the precision (1) is interpolated using a 3D quadratic approximation and the $3\times3\times3$ neighborhood. As shown in Fig. 2 for a detected junction, the precision does not show a quadratic shape in the neighborhood of the feature position. To account for the shape of the gradients, a signal adapted localization procedure is introduced.

Inspired by [4], we approximate the image gradients by a circular Difference of Gaussians shape:

$$D_{\mathbf{x}_0,\sigma,l}(\mathbf{x}) = l \cdot \left(exp(-\frac{(\mathbf{x}-\mathbf{x}_0)^2}{2\sigma^2}) - exp(-\frac{(\mathbf{x}-\mathbf{x}_0)^2}{2\sigma_k^2}) \right) \quad (2)$$

This function is determined by the parameters $\mathbf{p} = (\mathbf{x}_0, \sigma, l)$, $\mathbf{x}_0 = (x_0, y_0)$. The second standard deviation σ_k is determined by the known distance k between two scales of the pyramid. An example is shown in Fig. 3. The parameters are obtained using the Levenberg-Marquardt optimization. The initially detected scale provides a reasonable initialization of the parameters.

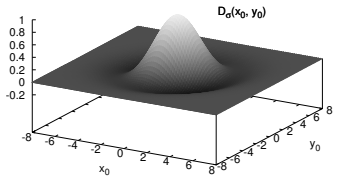


Figure 3: Proposed regression function for the approximation of the gradient image signal of the scale space as demonstrated in Fig. 2.

After reaching the minimal distance between the function (2) and the precision values in the $3\times3\times3$ neighborhood of the detected feature, the precision at the maximum is computed and compared to the precision achieved by the original interpolation using the 3D quadratic. The feature localization with the larger precision value is chosen. As shown in the results section, the DoG precision is larger in most cases. Sometimes the precision of the 3D quadratic is larger due to a diverging optimization procedure of the more challenging DoG optimization. In Sect. 4, a comparison of the resulting precision is shown as well as the accuracy evaluation using the repeatability criterion.

4 Experimental Results

For evaluation, the repeatability on two benchmark data sets with perspective distortion is used (cf. Fig. 4). The sets consist of sequences of six images which show a planar scene captured from a varying viewpoint with increasing angle relative to the first image. Since the data from [5] provide higher accuracy, we show the whole data set [5] (2048×1365) and the most prominent sequence *Graffiti* from [12] (800×640). An example sequence *Posters* is shown in Fig. 5. The results for the precision are presented in Sect. 4.1, the repeatability results are shown in Sect. 4.2. The reference localization is denoted as REF SFOP and the proposed localization as DOG SFOP.

Certainly, the computational expense of DOG SFOP is larger than of REF SFOP. The time for the localization increases by a factor of 3.

Figure 4: First images of the benchmark image sequences [5, 12] presented in the results section. From left to right: Graffiti [12], Grace, Underground, Posters, There, Colors [5].

4.1 Precision

The proposed localization technique leads to an increase in precision for all tested images. In Tab. 1, the numbers of features with a larger precision are shown. The number of junctions with larger precision value is 24392 for the 3D quadratic and 64396 for DoG, eq. (2). This means that 72.5% of the junction features are better explained by the proposed DoG shape although this approach builds on a much more challenging optimization scheme. Overall, 10.7% more junction features are extracted by DOG SFOP because of their increase in precision. The only sequence, in which DOG SFOP does not provide more features than the reference REF SFOP [7], is *Colors*.

Table 1: Numbers of extracted junction features by REF SFOP or DOG SFOP subsumed for all six images of each sequence. The proposed method provides more features (cf. Sect. 4.1).

Approach	REF SFOP	DOG SFOP	
Localization	Quadr.	Quadr.	DoG
Graffiti	5008	1262	4347
Grace	15563	4337	12410
Underground	10382	3375	7998
Posters	35240	9733	29257
There	6885	3651	5349
Colors	7117	2034	5035
Σ	80195	24392	64396
		88788	

Figure 5: Example benchmark sequence *Posters* with the largest number of junctions in the test set (cf. Tab. 1).

4.2 Repeatability

For the evaluation with natural images, the same benchmark data sets are used (Fig. 4). For the mapping from the first image to the others, ground truth homographies are provided. If the feature which is detected in the first image is also detected in the second image, the feature pair is deemed correct. A threshold value $\varepsilon_O = 0.4$ defines the maximally allowed error of a feature pair. It is calculated by the *overlap error* [12]. The repeatability is the ratio of correctly detected feature pairs to the maximally possible number of correct feature pairs. It is calculated by the *Matlab* script provided by the authors of [12]. The benchmark [12] provides image resolution of 0.5 megapixels while the data set [5] provides higher resolutions. To account for the higher resolution of the images [5], the number of octaves in the SFOP detector is set to 4 while the standard 3 is used for Graffiti. Additionally to the junction (JUNC) evaluation, the circular symmetric (CIRC) features are evaluated with the same localization techniques. To extract the junctions using the SFOP detector, the parameter *type* is set to 0, for the circular symmetric features, *type* is set to 90.

The results are shown in Fig. 6 and Fig. 7. The repeatability increases for every sequence, except for *Colors*. Here, the junction repeatability for REF SFOP is larger than DOG SFOP for the image pairs 1-2 and 1-3. The *Colors* sequence is known as a very challenging scenario [13] and only 20 valid feature pairs are extracted. The performance of SFOP is poor on these images. Thus, the repeatability results are very sensitive to the selection. On the other hand, the circular symmetric features (CIRC) show better results with DOG SFOP on *Colors*.

The maximum gain for DOG SFOP compared to REF SFOP is ~9%, not regarding *Colors*. The number of valid feature pairs increases in every case, which means that more accurate features are extracted due to the proposed localization technique. For *Posters*, ~30% more features are extracted while increasing the repeatability.

5 Conclusions

The proposed approach exchanges the quadratic subpixel and subscale localization of the SFOP detector. This is motivated by the signal shape of the junction gradients, which is not 3D quadratic. The proposed Difference of Gaussians shape increases the precision for 72.5% of the features. The number of extracted valid feature pairs increase in every case. The number of pairs increase by up to 30% while providing an increased repeatability. The results are found on standard benchmark data sets. Additionally, it is shown that the results are valid for the circular symmetric features. The only drawback of the proposed method is the computational expense which increases by a factor of 3.

Nevertheless, the proposed approach provides a significant improvement of the SFOP feature localization and can be applied to other feature detectors as well.

References

[1] H. Bay, T. Tuytelaars, and L. Van Gool. Surf: Speeded up robust features. In *European Conference on Computer Vision*, volume 3951 of *LNCS*, pages 404–417. Springer, 2006.

[2] M. Brown and D. G. Lowe. Invariant features from interest point groups. In *British Machine Vision Conference*, pages 656–665, 2002.

[3] K. Cordes. *Occlusion Handling in Scene Reconstruction from Video*, volume 10. VDI, 2014.

[4] K. Cordes, O. Müller, B. Rosenhahn, and J. Ostermann. HALF-SIFT: High-accurate localized features for SIFT. In *IEEE Conference on Computer Vision and Pattern Recognition Workshop*, pages 31–38, 2009.

[5] K. Cordes, B. Rosenhahn, and J. Ostermann. High-resolution feature evaluation benchmark. In *International Conference on Computer Analysis of Images and Patterns*, volume 8047 of *LNCS*, pages 327–334. Springer, 2013.

[6] T. Dickscheid, F. Schindler, and W. Förstner. Coding images with local features. *International Journal of Computer Vision*, pages 1–21, 2010.

[7] W. Förstner, T. Dickscheid, and F. Schindler. Detecting interpretable and accurate scale-invariant keypoints. In *IEEE International Conference on Computer Vision*, pages 2256–2263, 2009.

[8] W. Förstner, T. Dickscheid, and F. Schindler. On the Completeness of Coding with Image Features. In *British Machine Vision Conference*, 2009.

[9] S. Leutenegger, M. Chli, and R.Y. Siegwart. BRISK: Binary robust invariant scalable keypoints. In *IEEE International Conference on Computer Vision*, pages 2548–2555, 2011.

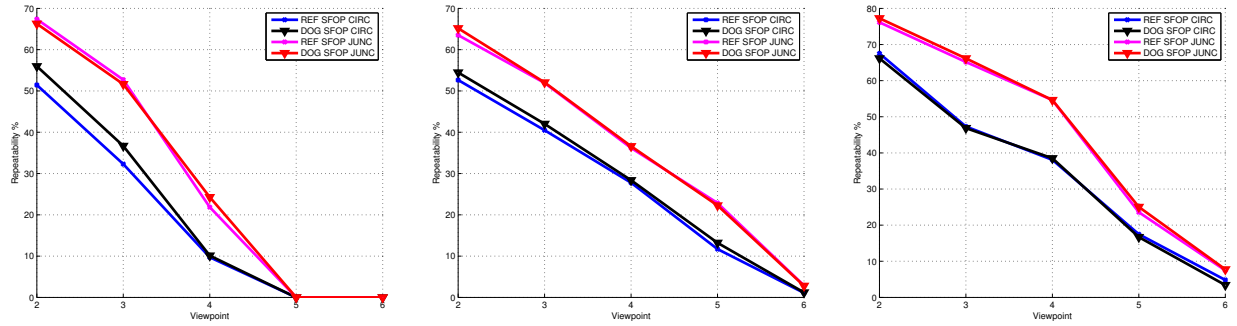
[10] T. Lindeberg. Feature detection with automatic scale selection. *International Journal of Computer Vision*, 30:79–116, 1998.

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

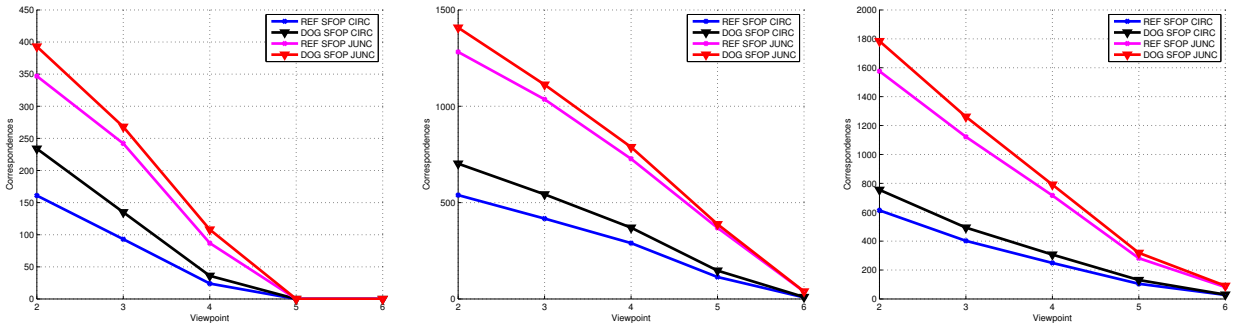
[12] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. Van Gool. A comparison of affine region detectors. *International Journal of Computer Vision*, 65(1-2):43–72, 2005.

[13] T.-N. Nguyen and K. Miyata. Multi-scale region perpendicular local binary pattern: an effective feature for interest region description. *The Visual Computer*, pages 1–16, 2014.

[14] N. Snavely, S. M. Seitz, and R. Szeliski. Modeling the world from internet photo collections. *International Journal of Computer Vision*, 80:189–210, 2008.

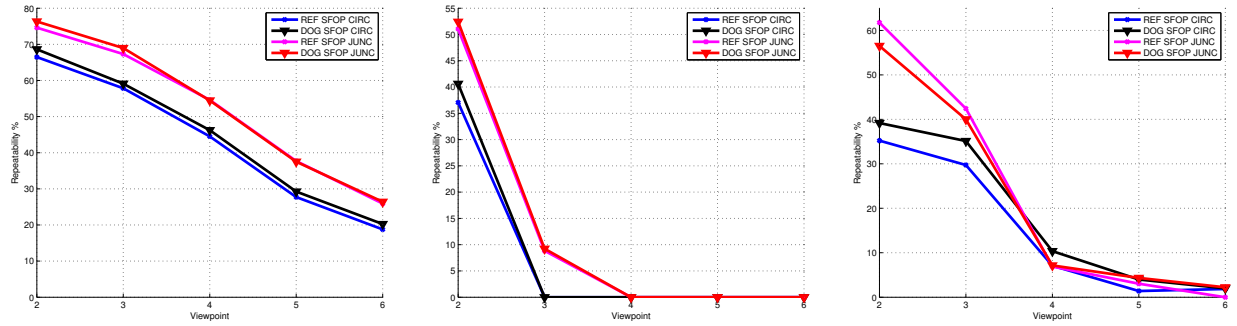


(a) *Graffiti*: Repeatability (b) *Grace*: Repeatability (c) *Underground*: Repeatability

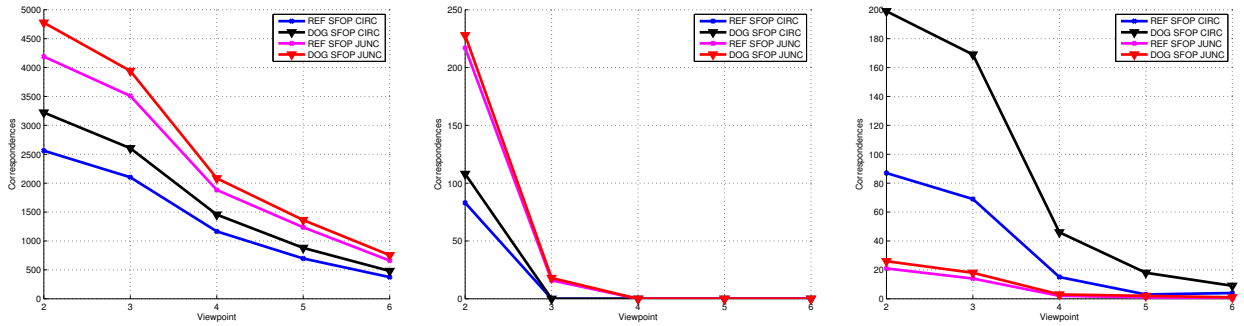


(d) *Graffiti*: Correspondences (e) *Grace*: Correspondences (f) *Underground*: Correspondences

Figure 6: Repeatability (top) and absolute number of correct feature pairs (bottom row) for *Graffiti* (800×640) [12], *Grace*, and *Underground* (2048×1365) [5]. The triangles ▼, ▼ denote the results of the proposed localization for junctions and circular symmetric features, respectively.



(a) *Posters*: Repeatability (b) *There*: Repeatability (c) *Colors*: Repeatability



(d) *Posters*: Correspondences (e) *There*: Correspondences (f) *Colors*: Correspondences

Figure 7: Repeatability (top) and absolute number of correct feature pairs (bottom row) for *Posters*, *There*, and *Colors* (2048×1365) [5]. The triangles ▼, ▼ denote the results of the proposed localization for junctions and circular symmetric features, respectively.