

# Bivariate Feature Localization for SIFT Assuming a Gaussian Feature Shape

Kai Cordes, Oliver Müller, Bodo Rosenhahn, and Jörn Ostermann

Institut für Informationsverarbeitung (TNT)  
Leibniz Universität Hannover

{cordes,omueller,rosenhahn,ostermann}@tnt.uni-hannover.de

**Abstract.** In this paper, the well-known SIFT detector is extended with a bivariate feature localization. This is done by using function models that assume a Gaussian feature shape for the detected features. As function models we propose (a) a bivariate Gaussian and (b) a Difference of Gaussians. The proposed detector has all properties of SIFT, but provides invariance to affine transformations and blurring. It shows superior performance for strong viewpoint changes compared to the original SIFT. Compared to the most accurate affine invariant detectors, it provides competitive results for the standard test scenarios while performing superior in case of motion blur in video sequences.

## 1 Introduction

The precise detection of feature points in an image is a requirement for many applications in image processing and computer vision. In most cases, feature points are used to establish correspondences between different images containing the same scene captured by a camera. These corresponding points can be used to perform measurements to reconstruct the geometry of the observed scene [1] or to detect objects. To obtain reasonable and comparable feature points, the image signal surrounding the feature position is analyzed and distinctive characteristics of this region are extracted. These characteristics are usually assembled to a vector which describes the feature. This vector, called descriptor [2], is used to establish correspondences by calculating a similarity measure ( $L_2$  distance) between the current descriptor and the feature descriptors of a second image. Feature detection and correspondence analysis is more challenging if the baseline between the cameras capturing the images is large. The region description of corresponding features before and after the viewpoint change should be the same. Several methods have been proposed in the literature to model invariants, such as brightness change and rotation [3, 4], scale change [2, 5, 6], affine transformation [7–12], and perspective transformation [13].

For viewpoint changes, the mapping between the local regions of two images can be modeled by an affine transformation. This is motivated by the assumption that the scene surface is locally planar and perspective effects are small on local

regions [11]. Based on affine normalization, the Harris-Affine [8] and Hessian-Affine [9] detectors determine the elliptical shape with the second moment matrix of the intensity gradient. In [14], maximally stable extremal regions (MSER) are constructed using a segmentation process. Then, an ellipse is fit to each of these regions. Although these detectors initially do not select features in an affine scale space, their evaluation shows excellent performance for significant viewpoint changes [11, 15].

Extensive work has been done on evaluating feature point detectors and descriptors [9, 11, 16–18]. Commonly, the most important criterion for affine invariant detectors is the repeatability criterion [9, 18]. For evaluation, the reference test set of images offered in [11] is widely used. The data set contains sequences of still images with changes in illumination, rotation, perspective, and scale. Using this data set, the most accurate localization of affine invariant features is performed by the Harris-Affine, Hessian-Affine, and MSER detector [15]. However, their performances depend on the image transformation emphasized throughout the specific test sequence and the type of texture present in the images.

Since the results of different detectors lead to differently structured features, the detectors are categorized by the type of features they detect. In [15], these categories for the affine invariant feature detectors are *corner detectors* (Harris-Affine), *blob detectors* (Hessian-Affine), and *region detectors* (MSER). For computer vision challenges like object detection, the combination of complementary features can be used to increase the performance [11, 19, 20].



**Fig. 1.** Examples of motion blurred image content. From left to right: input image, Hessian-Affine, MSER, SIFT, and our approach DoG-Affine. Harris-Affine provides no feature for this image (part of the sequence shown in Figure 5).

The SIFT detector performs very well under moderate affine distortions. It has been shown [12] that in this case the repeatability rates are better than those from affine invariant detectors. However, the performance of SIFT decreases under substantial viewpoint changes. Also, elliptically shaped features are not localized accurately. Hence, our motivation is to enhance the affine invariant properties of SIFT while preserving its performance for moderate affine distortions. This is done without building an affine scale space pyramid, but by using a modified localization scheme that assumes (a) a bivariate model and (b) a Gaussian shape for the features located in the scale space. These two characteristics can also increase the stability of the detection in images with motion blur as shown in Figure 1. In this example, an elliptical description of the feature is preferable to represent the region. The detection of Harris-Affine, Hessian-Affine,

and MSER lead to only a few features. In video sequences taken by a moving camera, motion blur is prevalent even if it is not visible to the human eye [21].

This work shows that our approach for determining the feature localization increases the performance of the SIFT detector in case of strong viewpoint changes as well as motion blur. The **contributions** of this paper are

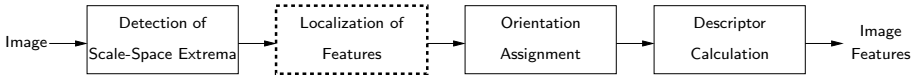
- the extension of the SIFT detector with a bivariate localization procedure,
- the comparison of the extended detector to the most accurate state of the art affine invariant feature detectors using synthetic and natural images, and
- the evaluation of the detectors using a structure from motion approach with naturally motion blurred images.

In the following Section, the SIFT detector is briefly presented. In Section 3, the proposed affine invariant extension of SIFT is derived. Section 4 shows experimental results using synthetically constructed and real image data. In Section 5, the paper is concluded.

## 2 The SIFT Detector

The **Scale Invariant Feature Transform** provides distinctive feature points which can be used to establish correspondences between images with the same scene content. The resulting features are invariant to illumination, rotation, and scale changes between the analyzed images. The SIFT detector demonstrates superior performance in this field. However, affine invariance that is commonly used to approximate the perspective distortion resulting from a viewpoint change is not modeled.

The workflow of SIFT using an image as input data is shown in Figure 2. First, feature points are characterized as extrema in the Difference of Gaussians pyramid of the image. The Difference of Gaussians pyramid is used as an approximation of the Laplacian pyramid, which has been proven to provide stable scale invariant features. In the next step, the localization is refined by an interpolation of the 26 surrounding grid points with  $x$ -,  $y$ -, and scale - coordinates. The interpolation is done by fitting a second order function to the Difference of Gaussians output. It provides subpel and subscale accuracy of the localization of a feature point. After the localization, features are rejected that do not fulfill the criteria of (a) minimal contrast, and (b) maximal ratio of principal curvatures. This guaranties the repeatability stability of the accepted features. In order to apply an orientation parameter to a feature, the main orientation of the surrounding image gradients is estimated. If the orientation estimation procedure leads to ambiguous results, multiple features with the same coordinates, but different orientation are constructed. Finally, a 128 dimensional vector is computed using the surrounding gradient orientations. This vector is called descriptor. It is used to establish the correspondence to a feature in the other image. Correspondences between two images are found by associating feature points with a minimal distance between their descriptors.



**Fig. 2.** Workflow diagram of the detection of image features with the SIFT detector. The part modified in our approach is marked with a dotted box border.

### 3 Bivariate Feature Localization

In order to estimate the localization parameters of a feature, a parabolic interpolation of the Difference of Gaussians signal is used by SIFT based feature detectors [2, 6, 10, 16]. If the neighborhood of a feature in the input image has Gaussian shape, the output of the Difference of Gaussians filter has Difference of Gaussians shape [22]. Therefore, the approach of interpolating with a parabolic curve is suboptimal and can lead to localization errors. This motivates the approximation of a scale space extremum using a Gaussian (Section 3.1) or a Difference of Gaussians (Section 3.2). Thereby, we can expect a better approximation for features with a Gaussian shape. An example result compared to the other affine invariant detectors for a Gaussian feature is shown in Figure 3.

#### 3.1 Gaussian Function Model

The Gaussian function model we propose with the covariance matrix  $\Sigma = \begin{pmatrix} a^2 & b \\ b & c^2 \end{pmatrix}$  and its determinant  $|\Sigma| = \det(\Sigma)$  is the following

$$G_{\mathbf{x}_0, \Sigma}(\mathbf{x}) = \frac{r_G}{\sqrt{|\Sigma|}} \cdot e^{-\frac{1}{2}((\mathbf{x}-\mathbf{x}_0)^\top \Sigma^{-1}(\mathbf{x}-\mathbf{x}_0))} \tag{1}$$

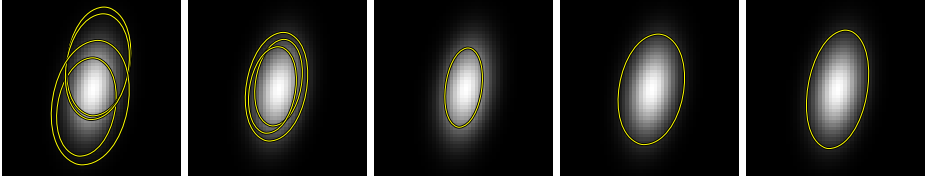
Here,  $\mathbf{x}_0 = (x_0, y_0) \in [-1; 1]$  is the subpel position of the feature point. The parameters  $a, b, c$  define the surrounding elliptical region and  $r_G$  is a peak value parameter. The parameter vector  $\mathbf{p}_G = (x_0, y_0, a, b, c, r_G)$  determines a member of the function model.

#### 3.2 Difference of Gaussians Function Model

If we assume a Gaussian image signal  $G_{\mathbf{x}_0, \Sigma}$  surrounding a feature point as in Equation (1), the correct response of the Difference of Gaussians filter is a Difference of Gaussians function with modified standard deviations  $\sigma$ . This motivates the following Difference of Gaussians regression function

$$\begin{aligned} D_{\mathbf{x}_0, \sigma} &= r_D(G_{\mathbf{x}_0, \Sigma_\sigma} - G_{\mathbf{x}_0, \Sigma_{k\sigma}}) * G_{\mathbf{x}_0, \Sigma} \\ &= r_D(G_{\mathbf{x}_0, \Sigma_\sigma + \Sigma} - G_{\mathbf{x}_0, \Sigma_{k\sigma} + \Sigma}) \end{aligned} \tag{2}$$

with  $\Sigma_\sigma = \begin{pmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{pmatrix}$  and the standard deviation  $\sigma$  of the detected scale of the feature. Like in equation (1),  $r_D$  is a peak value parameter.



**Fig. 3.** Different results of feature localization of an elliptical Gaussian blob obtained by the Harris-Affine, Hessian-Affine, MSER, and our approaches Gaussian-Affine and DoG-Affine (from left to right). The right three images show desirable results.

Like the Gaussian function (Section 3.1), the Difference of Gaussians function can be described by a six-dimensional parameter vector  $\mathbf{p}_D = (x_0, y_0, a, b, c, r_D)$ .

### 3.3 Optimization and Computational Complexity of the Approaches

The bivariate Gaussian (Section 3.1) and the Difference of Gaussians function models (Section 3.2) evaluate 9 and 27 pixel values, respectively. A member of the function model is determined by a parameter vector  $\mathbf{p} = (x_0, y_0, a, b, c, r)$ . The parameter vector  $\mathbf{p}$  is identified through a regression analysis which is initialized with the fullpel position, a circular shape  $\Sigma = \mathbf{E}$ , and a peak level parameter equal to 1. To obtain the optimal parameter vector, the distance between the pixel neighborhood and the model function is minimized using the Levenberg-Marquardt algorithm.

**Convergence.** As only the feature localization scheme of SIFT is modified, the original SIFT and the proposed localization approaches use the same fullpel features selected in the scale space pyramid as input. Nevertheless, some features are rejected after the localization by SIFT that are considered valid using our approaches and vice versa. The overall numbers of features per image  $N_f$  are approximately equal as shown in Table 1 for the *Graffiti* sequence. The number of features that are localized with our methods but discarded by SIFT, are 173 and 210 for Gaussian-Affine and DoG-Affine, respectively.

**Computation Time.** The computation time of the two approaches using the same SIFT implementation basis<sup>1</sup> is shown in Table 2. Here, The *Graffiti* (Section 4) sequence is processed. The overall time needed is divided by the number of detected features. Compared to the original SIFT (*o-sift*), the computational overhead of Gaussian-Affine (*gauaff*) and DoG-Affine (*dogaff*) is a factor of 2.7 and 7.1, respectively. While the Difference of Gaussians function model appears to be mathematically correct, the Gaussian-Affine provides a good approximation.

<sup>1</sup> <http://web.engr.oregonstate.edu/~hess/>

**Table 1.** Detected number of features of the original SIFT (*o-sift*) and our approaches Gaussian-Affine (*gauaff*) and DoG-Affine (*dogaff*). About 200 of the features localized by our methods are rejected by the original SIFT - detector as shown in the columns *gauaff \ o-sift* and *dogaff \ o-sift*, respectively.

	o-sift	gauaff	gauaff \ o-sift	dogaff	dogaff \ o-sift
$N_f$	1148	1234	173	1270	210

**Table 2.** The computation time used per feature for the original SIFT (*o-sift*) and our approaches Gaussian-Affine (*gauaff*) and DoG-Affine (*dogaff*)

o-sift	gauaff	dogaff
0.97ms	2.67ms	6.91ms

## 4 Experimental Results

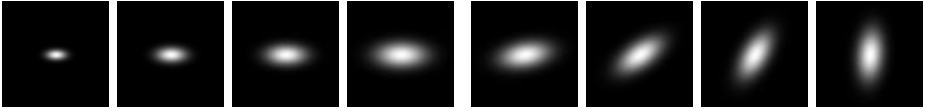
For the evaluation of our approaches, we use the following input data:

- synthetic test images to prove the accuracy of our approaches for Gaussian shaped features (Figure 4)
- natural image set with viewpoint changes to evaluate the competitiveness compared to the most accurate affine invariant detectors [11]
- natural motion blur images to show the improvement in a structure from motion approach (Figure 5)

As pointed out in [15], the most accurate affine invariant feature point detectors are the Hessian-Affine [10] (*hesaff*), the Harris-Affine [9] (*haraff*), and Maximally Stable Extremal Regions (MSER) [14] (*mseraf*). Hence, these detectors will serve as benchmarks for the comparison with our detectors Gaussian-Affine (Section 3.1) and DoG-Affine (Section 3.2), denoted as *gauaff* and *dogaff*. The input data description used for the evaluation follows in Section 4.1. The results are presented in Section 4.2.

### 4.1 Data Construction

**Synthetic Image Data.** The synthetic test images are constructed using elliptical Gaussian blobs with varying size and orientation determining the covariance matrix  $\Sigma_0$ . The image size is  $64 \times 64$ . Some examples are shown in Figure 4. The elliptical region is defined by the center position  $(x_0, y_0)$  and the ground truth covariance matrix  $\Sigma_0 = \begin{pmatrix} a_0^2 & b_0 \\ b_0 & c_0^2 \end{pmatrix}$ . Gaussian noise is added to the images. All resulting covariance matrices estimated in the results section will be invariant to a global scale factor. Therefore, a normalization is applied to the covariance matrix results.



**Fig. 4.** Some examples of the synthetic test images with varying size (left) and angle (right) determined by the covariance matrix  $\Sigma_0$

**Reference Image Data Set.** For the evaluation with natural images, the well-known Mikolajczyk data set<sup>2</sup> is used. The data set provides image pairs with a camera performing a rotational movement, or observing a planar scene. Thus, all detected feature ellipses in one image can be mapped to the corresponding feature ellipses in the other image by a homography. The estimated homography matrices are also provided within the data set. These images are widely used by the computer vision community, especially for evaluating feature descriptors.

**Natural Motion Blur Images.** To obtain natural motion blurred images, a sequence is captured from a moving vehicle. Radial distortion is compensated. The scene content consists of rigid geometry. The level of motion blur as well as the viewpoint angle increases with a decreasing distance of camera and scene structure. Some images of the sequence (94 frames) are shown in Figure 5.



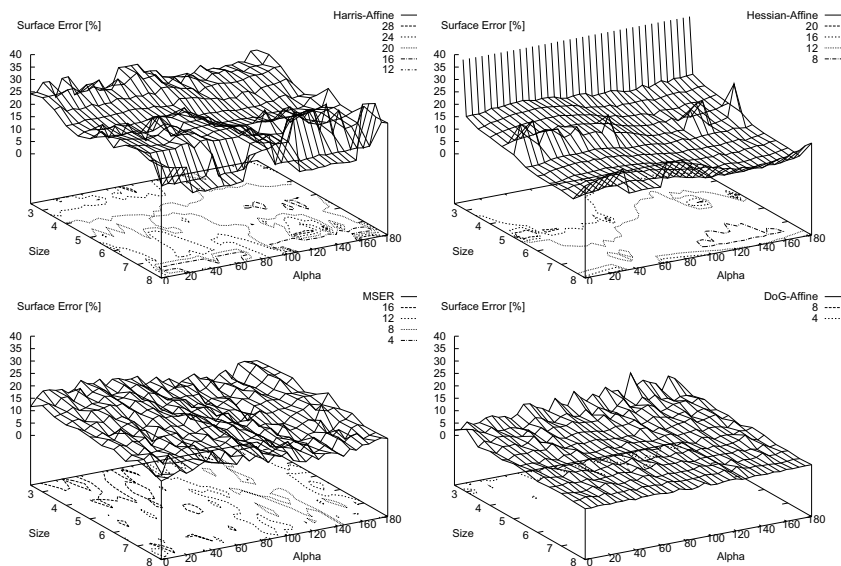
**Fig. 5.** Some example frames  $f_0, f_{30}, f_{60}, f_{90}$  of the natural motion blur sequence (94 frames). The motion blur level on the sidewalk is higher than on the wall.

## 4.2 Results

**Synthetic Images.** The methods lead to different results regarding the localization and the amount of detected features as shown in Figure 3. The subjectively best results are obtained by MSER and our approaches Gaussian-Affine and DoG-Affine, which are very similar (see Figure 3). For an objective evaluation of the synthetic images, the Surface Error [9] is calculated. The surface error is a percentage value that is minimal if a detected ellipse area is exactly matching the ellipse determined by the ground truth values. For the evaluation, the detected covariance matrices are normalized. This avoids the results being dependent on a global scale of the features. If multiple features are detected, the best of them

<sup>2</sup> [www.robots.ox.ac.uk/~vgg/research/affine/index.html](http://www.robots.ox.ac.uk/~vgg/research/affine/index.html)

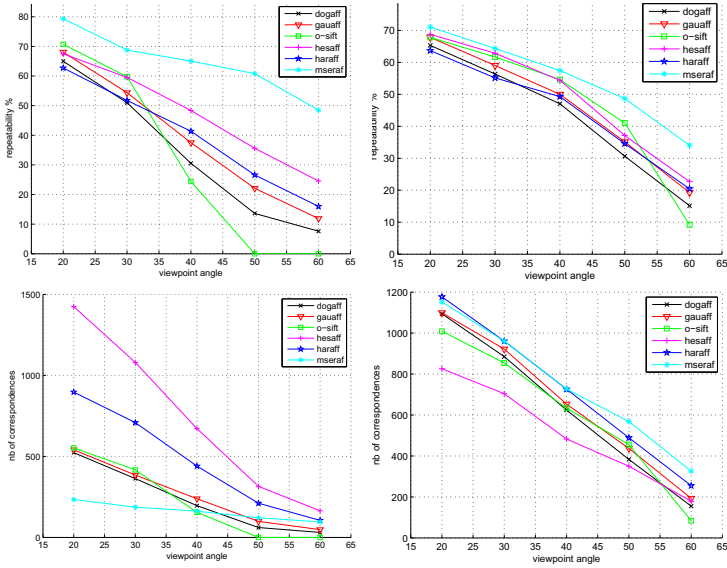
is chosen. For the evaluation, Gaussian noise of 30 dB is added to the images. The results are shown in Figure 6. Our detector DoG-Affine provides the best results for this type of features with added noise. The Gaussian-Affine detector yields nearly the same results with slightly increased error. Harris-Affine, a *corner detector*, results in the highest Surface Errors of the four detectors. Ellipses with small sizes of  $\leq 2.5$  are not detected by Hessian-Affine.



**Fig. 6.** Surface Error for the test signals shown in Figure 4 with varying angle alpha and varying size. Gaussian noise of 30 dB is added to the images. From left top to bottom right: Harris-Affine, Hessian-Affine, MSER, and our approach DoG-Affine. Contour lines are displayed in the ground plane. Together with the legend at the top right of each diagram, they indicate the error level.

**Natural Image Pairs with Viewpoint Change.** The image data sets provide image pairs with associated homography matrices that map all features of one image to the corresponding features in the other image. For the evaluation, the repeatability criterion with default parameters is used [9]. A correspondence is deemed correct, if the surface error between the ellipses of the features is below 40 %. The repeatability value is the ratio of correct correspondences and the minimal number of feature points detected in the first and the second image. The detectors SIFT, Gaussian-Affine, and DoG-Affine use the same default SIFT threshold values [6]. The evaluation results are shown in Figure 7. On top, the repeatability values between the first and the other images are displayed. The bottom row shows the number of correct matches. It can be seen that our

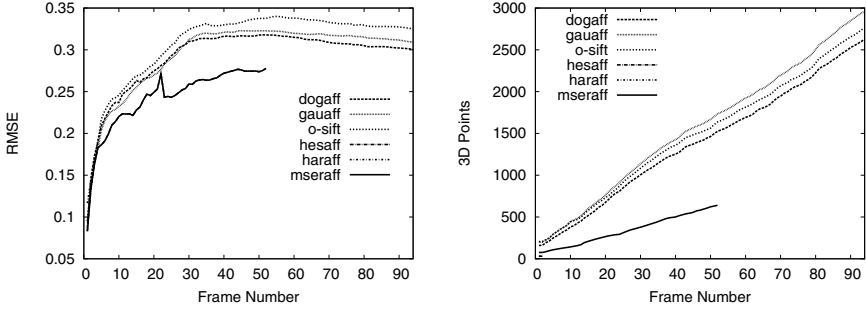




**Fig. 7.** Results of the *Graffiti* sequence (left) and the *Wall* sequence (right) demonstrating perspective change. On top, the repeatability curves are shown, the bottom diagrams show the number of correctly detected feature pairs.

detectors Gaussian-Affine and DoG-Affine provide better results for strong viewpoint change than the original SIFT approach. MSER performs best for these still images with well-bounded textures. For most natural images in this setup, the Gaussian-Affine provides slightly better repeatability results than the DoG-Affine approach. Overall, our approaches provide competitive accuracy, MSER and Hessian-Affine provide slightly better repeatability rates. This result is also valid for the other test scenarios [11], which are not shown here.

**Motion Blur Sequence.** The test scenario for the motion blur images evaluates a structure from motion technique like in [1] to demonstrate the usability of the feature detectors in the field of multiview scene reconstruction. Resulting from the camera motion, blurring occurs as well as a moderate viewpoint angle between consecutive frames. The motion blur level as well as the viewpoint angle increase with a decreasing distance between scene content and camera position. For the comparison, feature point sets are generated for each of the tested detectors. For the reconstruction of the scene, feature correspondences are established using the SIFT descriptor [6]. Then, the outliers are removed using the fundamental matrix and the RANSAC algorithm [23]. Finally, the cameras  $A_k$  and 3D object points  $\mathbf{P}_j$  are estimated and refined using incremental bundle adjustment [1, 24]. The incremental bundle adjustment minimizes the reprojection error, which is defined by the distances between an estimated 3D object point  $\mathbf{P}_j$



**Fig. 8.** Structure from motion results using different detectors for the natural motion blur sequence for each frame  $k$ . Harris-Affine, Hessian-Affine, MSER, Gaussian-Affine, DoG-Affine. Left: mean reprojection error [pel], right: Number of valid 3D object points until frame  $k$ . haraff and hesaff fail in frame  $k_s = 2$ .

projected by the estimated camera matrix  $A_k$  and the detected positions of corresponding 2D feature point  $\mathbf{p}_{j,k}$  in the images  $k$  (see e.g. [24]).

$$RMSE = \sqrt{\frac{1}{JK} \sum_{j=1}^J \sum_{k=1}^K d(\mathbf{p}_{j,k}, A_k \mathbf{P}_j)^2} \tag{3}$$

The  $RMSE$  is used for the evaluation which is shown for each frame  $k$  in Figure 8, left. The number of valid reconstructed object points  $J$  for each  $k$  is shown in Figure 8, right. It can be seen that the proposed detectors Gaussian-Affine and DoG-Affine provide a more accurate reconstruction of the scene than the original SIFT detector. The reconstruction with DoG-Affine provides a smaller RMSE than with Gaussian-Affine, but about 11% less object points. The estimation using Harris-Affine, Hessian-Affine, or MSER fails after an early frame  $k_s$  in the bundle adjustment step. In Table 3 the failure frame number  $k_s$ , the overall RMSE and the number of object points in the sequence for each detector are shown. The main reason for the failure is that the motion blurred sideways region is not represented by sufficient accurate features as shown in Figure 1.

**Table 3.** Detection results of the natural motion blur sequence (see Figure 5, 94 frames). After frame number  $k_s$  the bundle adjustment diverges if  $k_s < 94$ . The overall resulting number of valid object points is denoted with  $J$ .

	haraff	hesaff	mseraff	o-sift	gauaff	dogaff
$k_s$	2	2	52	94	94	94
$J$	30	31	641	2760	2962	2627
$RMSE$	—	—	0.278	0.325	0.309	0.301

## 5 Conclusion

In this work, an extension to the SIFT detector is proposed which enables the detection of elliptical structures in the Difference of Gaussians gradient signal. Therefore, the subpel estimation technique of the original SIFT detector is exchanged by a regression analysis using two proposed function models, a Gaussian and a Difference of Gaussians. These function models provide features that are rejected by SIFT. The resulting feature detector has all properties of the widely respected SIFT detector, but provides invariance to affine transformation and motion blur while preserving the stable convergence behavior.

The accuracy of the approach is proved using synthetically constructed image data with added Gaussian noise. For the viewpoint change scenario, our approach shows superior performance compared to SIFT, while being competitive to state of the art affine invariant feature detectors. The natural motion blur images show superior structure from motion reconstruction results for our approach compared to all other detectors regarding accuracy and stability.

## References

1. Pollefeys, M., Gool, L.V.V., Vergauwen, M., Verbiest, F., Cornelis, K., Tops, J., Koch, R.: Visual modeling with a hand-held camera. *International Journal of Computer Vision (IJCV)* 59, 207–232 (2004)
2. Brown, M., Lowe, D.G.: Invariant features from interest point groups. In: *British Machine Vision Conference (BMVC)*, pp. 656–665 (2002)
3. Canny, J.: A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)* 8, 679–698 (1986)
4. Harris, C., Stephens, M.: A combined corner and edge detector. In: *Alvey Vision Conference*, pp. 147–151 (1988)
5. Lindeberg, T.: Feature detection with automatic scale selection. *International Journal of Computer Vision (IJCV)* 30, 79–116 (1998)
6. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision (IJCV)* 60, 91–110 (2004)
7. Lindeberg, T., Garding, J.: Shape-adapted smoothing in estimation of 3-d shape cues from affine deformations of local 2-d brightness structure. *Image and Vision Computing* 15, 415–434 (1997)
8. Mikolajczyk, K., Schmid, C.: An affine invariant interest point detector. In: Heyden, A., Sparr, G., Nielsen, M., Johansen, P. (eds.) *ECCV 2002*. LNCS, vol. 2350, pp. 128–142. Springer, Heidelberg (2002)
9. Mikolajczyk, K., Schmid, C.: Scale & affine invariant interest point detectors. *International Journal of Computer Vision (IJCV)* 60, 63–86 (2004)
10. Mikolajczyk, K., Schmid, C.: A performance evaluation of local descriptors. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)* 27, 1615–1630 (2005)
11. Mikolajczyk, K., Tuytelaars, T., Schmid, C., Zisserman, A., Matas, J., Schaffalitzky, F., Kadir, T., Gool, L.V.: A comparison of affine region detectors. *International Journal of Computer Vision (IJCV)* 65, 43–72 (2005)

12. Yu, G., Morel, J.M.: A fully affine invariant image comparison method. In: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Washington, DC, USA, pp. 1597–1600. IEEE Computer Society, Los Alamitos (2009)
13. Köser, K., Koch, R.: Perspectively invariant normal features. In: IEEE International Conference on Computer Vision (ICCV), pp. 1–8 (2007)
14. Matas, J., Chum, O., Urban, M., Pajdla, T.: Robust wide baseline stereo from maximally stable extremal regions. In: British Machine Vision Conference (BMVC), vol. 1, pp. 384–393 (2002)
15. Tuytelaars, T., Mikolajczyk, K.: Local invariant feature detectors: a survey. In: Foundations and Trends in Computer Graphics and Vision, vol. 3 (2008)
16. Ke, Y., Sukthankar, R.: Pca-sift: A more distinctive representation for local image descriptors. In: International Conference on Computer Vision and Pattern Recognition (ICCV), pp. 506–513 (2004)
17. Schmid, C., Mohr, R., Bauckhage, C.: Comparing and evaluating interest points. In: IEEE International Conference on Computer Vision and Pattern Recognition (ICCV), pp. 230–235 (1998)
18. Schmid, C., Mohr, R., Bauckhage, C.: Evaluation of interest point detectors. *International Journal of Computer Vision (IJCV)* 37, 151–172 (2000)
19. Mikolajczyk, K., Leibe, B., Schiele, B.: Multiple object class detection with a generative model. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), vol. 1, pp. 26–36 (2006)
20. Sivic, J., Russell, B.C., Efros, A.A., Zisserman, A., Freeman, W.T.: Discovering objects and their location in images. In: IEEE International Conference on Computer Vision, vol. 1, pp. 370–377 (2005)
21. Vatis, Y., Ostermann, J.: Adaptive interpolation filter for h.264/avc. *IEEE Transactions on Circuits and Systems for Video Technology* 19, 179–192 (2009)
22. Cordes, K., Müller, O., Rosenhahn, B., Ostermann, J.: Half-sift: High-accurate localized features for sift. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshop, Miami Beach, USA, pp. 31–38 (2009)
23. Fischler, R.M.A., Bolles, C.: Random sample consensus: A paradigm for model fitting with application to image analysis and automated cartography. *Communications of the ACM* 24, 381–395 (1981)
24. Thormählen, T., Hasler, H., Wand, M., Seidel, H.P.: Merging of feature tracks for camera motion estimation from video. In: 5th European Conference on Visual Media Production (CVMP), pp. 1–8 (2008)