Detecting Interpretable and Accurate Scale-Invariant Keypoints

Wolfgang Förstner, Timo Dickscheid, Falko Schindler
Department of Photogrammetry
Institute of Geodesy and Geoinformation, University of Bonn
Nussallee 15, 53115 Bonn, Germany
wf@ipb.uni-bonn.de,dickscheid|falko.schindler@uni-bonn.de

Abstract

This paper presents a novel method for detecting scale invariant keypoints. It fills a gap in the set of available methods, as it proposes a scale-selection mechanism for junction-type features. The method is a scale-space extension of the detector proposed by Förstner (1994) and uses the general spiral feature model of Bigün (1990) to unify different types of features within the same framework. By locally optimising the consistency of image regions with respect to the spiral model, we are able to detect and classify image structures with complementary properties over scale-space, especially star and circular shapes as interpretable and identifiable subclasses. Our motivation comes from calibrating images of structured scenes with poor texture, where blob detectors alone cannot find sufficiently many keypoints, while existing corner detectors fail due to the lack of scale invariance. The procedure can be controlled by semantically clear parameters. One obtains a set of keypoints with position, scale, type and consistency measure. We characterise the detector and show results on common benchmarks. It competes in repeatability with the Lowe detector, but finds more stable keypoints in poorly textured areas, and shows comparable or higher accuracy than other recent detectors. This makes it useful for both object recognition and camera calibration.

1. Introduction

Local image features are an important aspect of computer vision research. The idea is to represent the image content by a set of small, possibly overlapping representative parts, which are invariant to distortions arising from the acquisition process, from illumination or viewpoint, and can reliably be found in other images of the same object. Corresponding features in different views may then be determined by nearest neighbour search in the space of descriptions of the surrounding image region, providing both sparseness and robustness compared to a search over the whole image domain.

Keypoints play a central role as they are anchored to a specific position in the image which is useful for both matching and recognition. We distinguish two types: *Point-like* keypoints refer to a specific *point* in the image, as a junction or the centre of a round spot, whereas *blob-like* keypoints refer to small *regions*, not necessarily round, where no specific point needs to be identifiable in the image within the region. Procedures for using keypoints consist of two parts: a keypoint detector and a keypoint descriptor. Here we are only concerned with the detection, and rely on the power of Lowe's SIFT descriptor [10].

"There is no such thing as generic keypoints" [21]. The choice of a particular detector must reflect the task at hand. The motivation for the new detector proposed in this paper arose in the context of automatic image orientation in poorly textured, structured scenes, as shown in Figure 1. We found that in such cases, state of the art keypoint detectors often yield too few features or poor geometric configurations. Sometimes multiple features are computed in very nearby locations, which have to be eliminated during matching to fulfil the uniqueness constraint. The Harris affine detector does not reliably extract the corner features, as one would expect. However, a combined set of features from two or three complementary detectors may well give stable correspondences for camera calibration and orientation. Thus we require the following properties from a keypoint detector:

Completeness & Complementarity: The detector should as much as possible exploit structural elements visible in the image to yield a maximally complete set of keypoints. This implies that different types of complementary keypoints are extracted at the same time.

Invariance and repeatability: The detected keypoints should be scale and rotation invariant and provide high repeatability in order to support image matching.

Accuracy: The keypoints should have high localisation accuracy to support camera calibration.

Interpretability: Basic interpretable elements in the

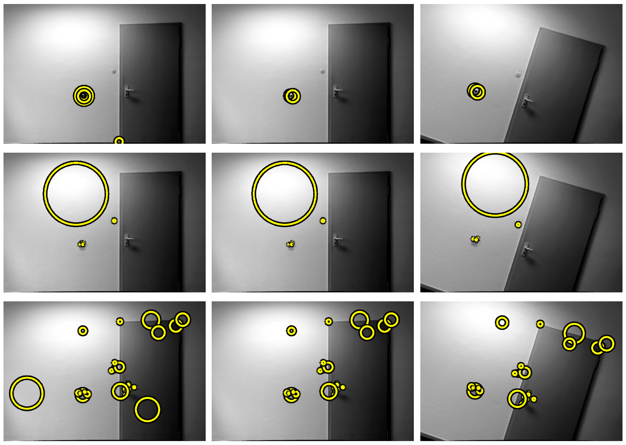


Figure 1: A challenging image pair of a poorly textured scene with features extracted by three popular keypoint detectors (Top: Harris affine, center: Lowe, bottom: our detector). The left column shows the set of extracted features in one image, while the middle and right column show the reduced set of matched features of the pair using SIFT descriptors. Note that the Harris affine detector does not extract corners, while our detector delivers corners as a subset.

scene, especially corners and circles, should be part of the detected keypoints.

Control parameters: The procedure should have as few control parameters as possible, with clear semantics.

To meet these requirements, we developed a scale-space extension of the keypoint detector in [3] and generalised it from junctions to all types of spiral features according to Bigün [1]. The idea is to find points where the consistency of image regions with respect to a spiral model is locally optimal. This allows for the simultaneous detection and classification of image structures with complementary properties over scale-space, especially star and circular shapes as interpretable and identifiable subclasses. One obtains a set of keypoints with position, scale, type and consistency measure.

We start by giving an overview of the most related existing feature detectors as a basis for describing and characterising our detector in detail in section 3. In section 4 we compare the new method to state of the art detectors in well known benchmarks. It turns out that it competes in repeatability with the Lowe detector. However, in poorly textured areas it finds more stable keypoints, which is useful for object recognition, and its accuracy is comparable or higher than that of well known keypoint detectors, which is favourable for camera calibration. We conclude with some critical remarks and possible future work.

2. Related work

The *basic local feature detectors* by Förstner [5] and Harris and Stephens [7] detect windows of given size having locally maximal localisation accuracy with respect to least-squares matching onto a shifted, noisy copy of themselves. This is achieved by evaluating the structure tensor, or second moment matrix

$$M_{\tau,\sigma} = \overline{\nabla_\tau g \nabla_\tau g^\mathsf{T}} = G_\sigma * \left[\begin{array}{cc} g_{x,\tau}^2 & g_{x,\tau} g_{y,\tau} \\ g_{x,\tau} g_{y,\tau} & g_{y,\tau}^2 \end{array} \right] \quad (1)$$

which is computed by first determining the gradient $\nabla_\tau g = [g_{x,\tau}, g_{y,\tau}]^\mathsf{T}$ using the *differentiation scale* τ, and then taking the average with a kernel having *integration scale* σ, which is usually chosen as a Gaussian. [19] have shown that selecting windows is most effective when using local maxima of the smallest eigenvalue $\lambda_2(M)$ above a threshold, thus requiring a sufficiently large gradient content within the window to be locally optimal. This is both in contrast to the proposals in [5], where a minimum ratio λ_2/λ_1 is required for excluding points on edges, and the heuristic measure $\det M_{\sigma,\tau} - k \operatorname{tr}^2 M_{\sigma,\tau}$ in [7], where k is determined empirically. As shown in [5], the detected windows are not only highly suitable for matching, but also optimal for locating centres of junctions and circular symmetric features. The window centres at corners are biased, as will be discussed further below. Triggs [21] generalised these properties, and developed a rigorous theoretical framework for designing detectors which explicitly optimise localisation accuracy in the least-squares self-matching approach under selectable combinations of translation, orientation, scaling and illumination parameters.

Scale invariance of keypoint detectors has been addressed thoroughly by Lindeberg [9] for all types of features. For finding *blob-like* keypoints, he proposed to use the Hessian

$$H_\tau = \left[\begin{array}{cc} g_{xx,\tau} & g_{xy,\tau} \\ g_{xy,\tau} & g_{yy,\tau} \end{array} \right] \quad (2)$$

of the image function and to search for maxima of the determinant $\det H_\tau$ and the scale-normalised trace $\tau^2 \operatorname{tr} H_\tau$ over increasing values of τ. As the trace $\operatorname{tr} H_\tau$ is identical to the Laplacian of Gaussians (LoG) of the image function $\nabla_\tau^2 g = \nabla_\tau^2 * g$, the proposed maximum search is equivalent to template matching with the Mexican hat form of ∇_τ^2 and can hence be interpreted as finding dark and bright blobs at *characteristic scales*. The LoG scale-space is used by the popular keypoint detector of Lowe [10].

Mikolajczyk and Schmid [12] also followed Lindeberg's approach and proposed to extract blob-like keypoints where both $\operatorname{tr} H_\tau$ and $\det H_\tau$ attain a local maximum, leading to the so-called Hessian-Laplace detector. Furthermore, they proposed a scale-invariant version of the Harris detector that works iteratively in two steps: First, the positions of candidate keypoints are attained by searching local extrema within each level of the multi-scale Harris function. Then a characteristic scale is selected by searching a local peak over scale at each of these candidate positions in the LoG scale-space. The integration scale σ of the Harris measure is taken from the optimal scale in the Laplacian pyramid and the differentiation scale is related to σ by a constant ratio $\sigma/\tau = k$.

To our knowledge, no scheme for selecting the scale from the structure tensor is proposed.

Junction-type keypoints have also been addressed by Lindeberg [9], who proposed a two-step scale selection procedure: It first detects points and selects the integration scale σ based on local curvature of the contours of the image function, and then finds the differentiation scale τ by maximising the precision

$$\tilde{d}(\tau) = \frac{\int_{x\in\mathbb{R}^2} |(\boldsymbol{x}-\boldsymbol{x}')^\mathsf{T}\nabla_\tau g(\boldsymbol{x}')|^2 G_\sigma(\boldsymbol{x}')\mathrm{d}\boldsymbol{x}'}{\int_{x\in\mathbb{R}^2} |\nabla_\tau g(\boldsymbol{x}')|^2 G_\sigma(\boldsymbol{x}')\mathrm{d}\boldsymbol{x}'} \qquad (3)$$

of the junction [9, eq. (65)]. A final localisation step according to Förstner [5] is carried out on the resulting scale level. As the scale of the corner points measures their roundness, the scale of most corners is quite small.

Interpretability of keypoints is sometimes the main motivation for developing a scheme. This of course refers to all corner detectors based on edge analysis, but also to direct schemes as the one step procedure in [3] or the method of Parida *et al.* [17]. The latter describe how to detect and classify junctions explicitly by measuring coherency of gradients in a local window with a junction model containing a variable number of intersecting edges, equivalent to the numerator in (3). The detector considers different scales, but provides no explicit scale selection mechanism. Other detectors based on the consistency of local image structure with a junction model have been proposed in [15; 18].

The bias induced by window detectors at corners, i. e. the distance between the optimal window's centre and the actual corner, is scale dependent. It is reduced by the second step in the procedures [5] and [9]. Ouellet *et al.* [16] employ this bias to develop an accumulator-based detector: By using uniform instead of Gaussian windows for the integration step, they reduce the effect and select those bias-corrected positions which are supported by the largest image region. However, they only investigate three scales. The procedure described in [3] integrates window selection and bias reduction into one step. It can classify the detected keypoints as junctions and circular symmetric features, but works only on prespecified scales.

Affine invariant detectors [11; 12; 22] have been motivated by the fact that scale invariance alone is not sufficient for robust matching in case of very strong viewpoint variations. They have been characterised and analysed in detail in [6; 13]. Although the new operator presented here is not affine invariant, the most prominent scale and affine invariant detectors will be included in our experimental comparison in section 4.

The point detectors of Förstner [3; 5] include the junction model and measure consistency of local image structure with respect to both junction-type and circular symmetric features. Bigün [1] generalised this idea in a unified theory for all local feature types that can be derived as special cases of spiral features, including junctions and circular features. This theory has been a major inspiration for the detector proposed in this paper. However, Bigün did not explicitly solve the problem of scale selection, although addressing feature detection at different scales.

3. Theory

The principle. Referring to Figure 2, the task is to find representative keypoints. Besides the white and black regions that a Laplacian blob detector easily finds (region *d*), we want to detect all junction regions including their local scale. The junction detector of Lindeberg [9] finds these points, but with small local scales only, as indicated by points *a*, *b*, and *c*. However, we would like to have regions around the junctions which are as large as possible, but still represent the junction and not the neighbouring image structures. This is indicated by the lower right junctions *e*, *f*, *g*, *h* and *i*, where the regions indicated by the circles are approximately proportional to the area covered by the junction.

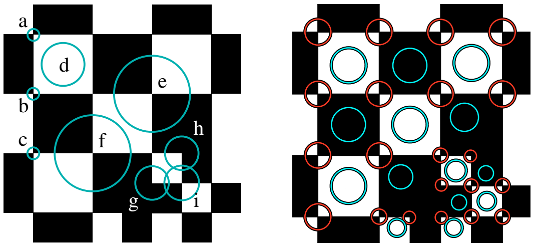
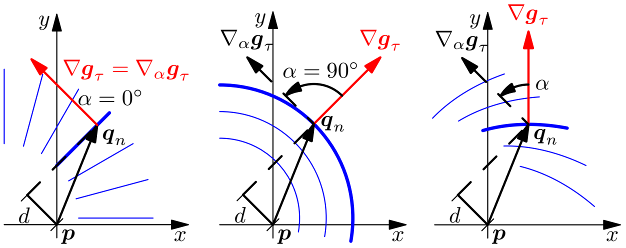


Figure 2: Left: Chequerboard with two different scales and some exemplary keypoints shown with circles approximately having the desired radius. The local scales at keypoints *a*, *b*, and *c* are not desired: They are not representative for the junction. Right: Keypoints computed by the new detector (1-σ-circles): red: corners, white: circular symmetric features.

In contrast to Lindeberg's selection scheme based on the local curvature we directly optimise (3), but using the more general spiral model $\mathcal{M}(\alpha)$ of Bigün [1], with star shaped ($\alpha = 0$) and circular symmetric ($\alpha = 90°$) regions as special cases.

In a nutshell, we search for keypoints $[\boldsymbol{p}, \sigma] = [x, y, \sigma]$ in the smoothed image $g(\boldsymbol{p}, \tau)$ with smoothing parameter τ, where the position $\boldsymbol{p}$ of a spiral structure with parameter α in a window of integration scale σ can be determined with locally highest precision $[\boldsymbol{p}, \sigma]_{\text{opt}} = \arg\max_{\boldsymbol{p},\alpha,\tau,\sigma} w(\boldsymbol{p}, \alpha, \tau, \sigma)$, which may also be interpreted as a weight. The precision w is measured by the inverse of the empirical variance $\tilde{d}$ from (3) of the centre position.

The image model. An image patch around $\boldsymbol{p}$ has an ideal spiral structure in case the edge direction at a point $\boldsymbol{q}$ in the neighbourhood has a constant angle α with the radius vector. This structure has been suggested by [1] for image analysis and is illustrated in Figure 3. In case $\alpha = 0°$ the edge directions point toward the centre of rotation and form an ideal star shaped image function as a special case (Figure 3a). In case $\alpha = 90°$ the gradient directions point toward the centre of rotation and form circular symmetric features (Figure 3b). Hence this model of spirals meets our requirement of interpretability in section 1 and yields a straightforward criterion for a keypoint detector.


y
$\nabla \boldsymbol{g}_\tau = \nabla_\alpha \boldsymbol{g}_\tau$
$\alpha = 0°$
$\boldsymbol{q}_n$
d
$\boldsymbol{p}$
x
y
$\nabla_\alpha \boldsymbol{g}_\tau$
$\nabla \boldsymbol{g}_\tau$
$\alpha = 90°$
$\boldsymbol{q}_n$
d
$\boldsymbol{p}$
x
y
$\nabla \boldsymbol{g}_\tau$
$\nabla_\alpha \boldsymbol{g}_\tau$
α
$\boldsymbol{q}_n$
d
$\boldsymbol{p}$
x

(a) Junction (b) Circular feature (c) Logarithmic spiral

Figure 3: Measuring the distance d of an edge from the reference point $\boldsymbol{p}$ in a spiral type feature. The angle α is measured between the radial and the tangential direction and is constant per figure, thus the spirals are logarithmic spirals. (a) Junction, where $\alpha = 0°$. (b) Circular symmetric feature, $\alpha = 90°$. (c) Logarithmic spiral with arbitrary α.

In order to measure the consistency of the neighbourhood of a point $\boldsymbol{p}$ with the model we determine its distances d_n to the edge line through a neighbouring point $\boldsymbol{q}_n$ having angle α with respect to the gradient direction at $\boldsymbol{q}_n$ (Figure 3). The gradient $\nabla_\tau g = \nabla G_\tau * g$ depends on the differentiation scale τ. Transformed with the rotation matrix R_α, the distance is given by $d(\boldsymbol{p}, \boldsymbol{q}_n, \alpha, \tau) = (\boldsymbol{q}_n - \boldsymbol{p})^{\mathsf{T}} \mathsf{R}_\alpha \nabla_\tau g(\boldsymbol{q}_n) / |\nabla_\tau g(\boldsymbol{q}_n)|$. The variance of the centre point is determined as the Cramer Rao bound derived from a Maximum Likelihood Estimation of all pixels in the neighbourhood of the reference point $\boldsymbol{p}$ based on the model $\mathcal{M}$: The stochastic variable $d(\boldsymbol{q}|\mathcal{M})$ for the distance d is assumed to be normally distributed with mean 0 and variance

$$\text{Var}\left(d(\boldsymbol{q}|\mathcal{M})\right) = s^2 / \left(|\nabla_\tau g(\boldsymbol{q})|^2 G_\sigma(\boldsymbol{q} - \boldsymbol{p})\right). \qquad (4)$$

The weight, or inverse variance $1/\text{Var}(d(\boldsymbol{q}))$ of each d has three important properties: It increases with the squared magnitude of the gradient, depends on the distance $|\boldsymbol{q} - \boldsymbol{p}|$ of $\boldsymbol{q}$ from the reference point $\boldsymbol{p}$ and is scaled by some variance factor s^2.

Discrete optimisation function. We first assume fixed parameters α, τ and σ and want to estimate the optimal position of the spiral. The negative log-likelihood function to be minimised is

$$\Omega = \sum_{n=1}^{N(\sigma)} \left[(\boldsymbol{q}_n - \boldsymbol{p})^{\mathsf{T}} \boldsymbol{R}_\alpha \nabla_\tau g(\boldsymbol{q}_n) \right]^2 G_\sigma(\boldsymbol{q}_n - \boldsymbol{p}) \qquad (5)$$

with respect to $\boldsymbol{p}$, the sum ranging over all $N(\sigma)$ pixels $\boldsymbol{q}_n$ in the neighbourhood of $\boldsymbol{p}$ defined by the Gaussian G_σ. The Cramer Rao bound for the covariance matrix $\Sigma_{\widehat{p}\widehat{p}}$ of the estimated position $\widehat{\boldsymbol{p}}$ is given by

$$\Sigma_{\widehat{p}\widehat{p}} = \widehat{s^2} \mathsf{M}^{-1} \quad \text{with} \quad \mathsf{M} = \textstyle\sum_n \mathsf{J}_n \mathsf{W} \mathsf{J}_n^{\mathsf{T}} \qquad (6)$$

using the Jacobian $\mathsf{J}_n = \partial \boldsymbol{d}_n / \partial \boldsymbol{p}$ and the diagonal weight matrix $\mathsf{W} = \text{Diag}(s^2 / \text{Var}(\underline{d}_n))$, with the variance $\text{Var}(\underline{d}_n)$ from (4). The unbiased estimated variance factor $\widehat{s^2}$ is given by $\widehat{s^2} = \Omega(\widehat{\boldsymbol{p}}) \,/\, (N(\sigma) - 2)$. In order to come to a scalar measure, we take the maximum eigenvalue $\lambda_1(\Sigma_{\widehat{p}\widehat{p}})$ of the covariance matrix, indicating the maximum variance of $\widehat{\boldsymbol{p}}$ in some direction in the flavour of Shi *et al.* [19]. For finding keypoints with maximum localisation precision we take the inverse variance. Making all dependencies explicit, we obtain the formal definition of the precision

$$w(\boldsymbol{p}, \alpha, \tau, \sigma) = \frac{1}{\lambda_1(\Sigma_{\widehat{p}\widehat{p}})} = \frac{(N(\sigma) - 2)\ \lambda_2(\mathsf{M}(\boldsymbol{p}, \alpha, \tau, \sigma))}{\Omega(\boldsymbol{p}, \alpha, \tau, \sigma)} \qquad (7)$$

with the matrix $\mathsf{M}(\boldsymbol{p}, \alpha, \tau, \sigma)$ from (6) and using the fact that $1/\lambda_1(\Sigma_{\widehat{p}\widehat{p}}) = \lambda_2(\Sigma_{\widehat{p}\widehat{p}}^{-1})$.

Continuous optimisation function. For numerical reasons we now replace the sums by adequate integrals and exploit the resulting convolutions. The matrix M turns out to be the moment matrix of the gradients or the structure tensor $\mathsf{M}(\boldsymbol{p}, \alpha, \tau, \sigma) = N(\sigma) G_\sigma(\boldsymbol{p}) * \left(\mathsf{R}_\alpha \nabla_\tau \nabla_\tau^{\mathsf{T}} \mathsf{R}_\alpha^{\mathsf{T}} \right)$ referring to the spiral model and specialising to the classical structure tensor for $\alpha = 0$. The average squared distance can be written as $\Omega(\boldsymbol{p}, \alpha, \sigma, \tau) = N(\sigma)\, \text{tr} \left\{ \mathsf{R}_\alpha \nabla_\tau \nabla_\tau^{\mathsf{T}} \mathsf{R}_\alpha^{\mathsf{T}} * \boldsymbol{p}\boldsymbol{p}^{\mathsf{T}} G_\sigma(\boldsymbol{p}) \right\}$ due to $\text{tr}(\mathsf{AB}) = \text{tr}(\mathsf{BA})$ and identifying the integral as a convolution. Obviously, the factor $N(\sigma)$ cancels when determining the precision in (7). The smaller eigenvalue $\lambda_2(\mathsf{M})$ of the structure tensor measures isotropy of the local texture. The threshold for this quantity is discussed at the end of this section. As the smallest eigenvalue λ_2 of $\mathsf{M}(\boldsymbol{p}, \alpha, \sigma, \tau)$ is independent on α, it only needs to be calculated once.

Finally, we need to give an expression for $N(\sigma)$. This is not straight forward as the Gaussian has unlimited support, i. e. one in principle would need to count all pixels of the image. Practically only pixels in a small range around the reference position $\boldsymbol{p}$ contribute to the estimation process. We take the number of pixels in a box filter having the same energy as the Gaussian, $N(\sigma) = 12\sigma^2 + 1$, being aware that this is an approximation, as the used gradients are not stochastically independent.

Estimating α. Before being able to detect local maxima in the precision cube $w(\boldsymbol{p}, \tau, \sigma)$, we need to find the optimal angle α that maximises w for each point in concern. Whereas the number $N(\sigma)$ and the eigenvalue λ_2 of the structure tensor $\boldsymbol{M}$ are independent on the angle α, the sum $\Omega(\alpha)$ is a periodic function in α represented by $\Omega(\alpha) = a - b\cos(2\alpha - 2\alpha_0)$ with a minimum $\Omega_{\text{min}} = a - b$ at $\alpha = \alpha_0$. We can determine a, b and α_0 from three particular values for Ω, e. g. using $\alpha = \{0°, 60°, 120°\}$. The parameter $\alpha = \alpha_0$ maximises the precision p for a certain location and scales $(\boldsymbol{p}, \sigma, \tau)$.

Non-maximum suppression over scale and space. The local maxima still may be too close to each other in position or scale. Moreover, only points with similar spiral characteristic, thus similar angle α should be compared. Therefore we suppress all non-maxima in the scale-angle-space. The neighbourhood is defined by a weighted distance (Mahalanobis distance):

$$r^2 = \frac{|\boldsymbol{p} - \boldsymbol{p}'|^2}{((f_p\sigma)^2 + (f_p\sigma')^2)} + \frac{\sin^2(\alpha - \alpha')}{\sigma_a^2} + \frac{(\log\sigma/\sigma')^2}{(f_s \log 2)^2} \tag{8}$$

Keypoints $\boldsymbol{p}'$ closer to another keypoint $\boldsymbol{p}$ are eliminated, in case $r < T_r$. For $f_p = 1$, $f_s = 1/2$, $T_r = 1$ and $\sigma_a = \sqrt{2}$, for example, points should have a distance larger than σ in the image plane, a factor $\sqrt{2}$ (half an octave) in scale or $45°$ with respect to α. Non-maximum suppression with these default parameters is optional in our algorithm.

The algorithm. The optimisation of the five parameters, position $\boldsymbol{p}$, model parameter α, the differentiation and the integration scales τ and σ is realised by locally finding the optimal α, and imposing a constraint on the two scales by fixing their ratio, hence $\tau = \sigma/k$ with $k = 3$, for example (Algorithm 1). We evaluate the precision of each point $\boldsymbol{p}$ and each scale σ in the scale-space image $g(\boldsymbol{p}, \sigma, \tau(\sigma))$ and search for local maxima using a spline approximation of the function $w(\boldsymbol{p}, \sigma)$ [8, pp. 78–82]. Interpolating w allows us to localise precise keypoints with subpixel/subscale accuracy.

Threshold for the isotropy $\lambda_2(\boldsymbol{M})$. By putting a threshold on $\lambda_2(\boldsymbol{M})$ we can eliminate spurious keypoints, i. e. keypoints caused by noise. In case of a white noise image with standard deviation s the eigenvalues of $\boldsymbol{M}$ are equal, hence $\lambda_2 = 1/2\ \text{tr}(\boldsymbol{M})$, where tr$(\boldsymbol{M})$ is the sum of $N(\sigma)$ absolute squared gradients $a = g^2_{\tau,x} + g^2_{\tau,y}$. As the expectation of the derivative square $E(g^2_{\tau,x}) = s^2 \int G^2_{\tau,x}(\boldsymbol{x})\mathrm{d}\boldsymbol{x} = s^2/(8\pi\tau^4)$ and $2a/E(a)$ is χ^2_2-distributed we obtain the threshold $T_\lambda(s^2, \tau, \sigma, S) = \frac{N(\sigma)}{16\pi\tau^4} s^2 \chi^2_{2,S}$ with a specified significance level S, depending on the noise variance s^2 and both scales, but not on α. Note that s^2 can be estimated from the image data. This is an approximation, as the pixels in the window are not statistically independent. However we found that with a threshold $1.5\,T_\lambda$ we do not obtain spurious features on a pure noise image, indicating the dependency on the scales and the noise variance holds empirically.

```
Algorithm 1: Proposed keypoint detection scheme
$g, k, T_\lambda, T_w, S, f_p, f_s, T_r \rightarrow \{\boldsymbol{p}, \sigma, \alpha\}$

for all scales $\sigma$ do
    determine gradient: $\nabla_\tau(g(\boldsymbol{p}, \sigma))$;
    determine $\lambda_2$ of moment matrix: $\boldsymbol{M}(\boldsymbol{p}, \alpha = 0, \sigma)$;
    for three angles $\alpha = 0°, 60°, 120°$ do
        determine $\Omega(\boldsymbol{p}, \alpha, \sigma)$;
    end
    determine best angle $\alpha_0(\boldsymbol{p}, \sigma)$ yielding $\Omega_{\text{min}}$;
    compute precision $w(\boldsymbol{p}, \sigma)$;
end
detect local maxima in 26-neighbourhood of $w$;
keep keypoints with $\lambda_2(\boldsymbol{M}) > T_\lambda$;
perform non-maximum suppression;
optimise keypoint locations by interpolating $w$;
```

4. Experiments

Empirical proof of theory. We will now demonstrate the interpretability and complementarity of the points extracted on an image of a Siemens star with regular beams. The results for several detectors are depicted in Figure 4. The thick blobs, found by the Lowe-detector, perfectly match to the dark ends of the beams with intuitive local scale. The smaller blobs, anchored to the outer corners of each ray's end, are not easy to interpret, and not symmetric either. MSER clearly detects one region for each beam in perfect symmetry, and no non-interpretable features at all. Results for Harris-Laplace, Hessian-Laplace and IBR are neither symmetric nor easy to interpret. Especially the first two yield many redundant points at different scales and slightly differing positions. Here a non-maximum suppression certainly would help. The comments on IBR apply for EBR, which is not shown here, as well. After all, while at least the Lowe detector and MSER provide stable and inter-

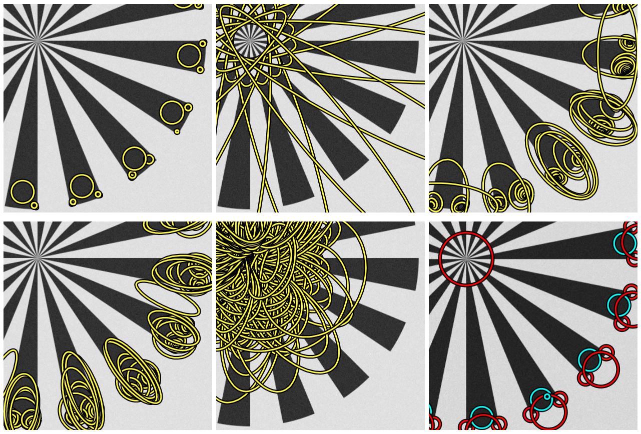


Figure 4: Features detected by different methods on the beams of a Siemens star. The image contains a small amount of Gaussian noise (2 %). Top left to bottom right: Lowe, MSER, Harris affine, Hessian affine, IBR, SFOP. Lowe, MSER and SFOP clearly show rotation invariance. SFOP detects the same larger blobs in the ends of the beams of the star as LOWE. Only SFOP explicitly detects the centre junction.

pretable results, none of the detectors actually gives a complete structural representation, because the corner points are not detected by any of the methods. Our detector (bottom right) finds both blob-like structures comparable to the thick blobs extracted by the Lowe detector, shown in cyan, and the complete set of junctions, shown in red. Note the junction points detected at the end of each fan as special corners. The feature type is explicitly returned by the new detector via the angle α, which we mapped directly to the 180°-colormap of the HSV color space for plotting. This illustrates well that the feature type is robustly distinguished.

Repeatability on planar structures. We want to follow the well known repeatability benchmark scheme proposed in [13], using the software supplied by the authors. However, we agree upon Haja *et al*. [6] to modify the scheme in two ways: First, we only consider the best matches in case of multiple possible assignments in order to forbid repeatability values greater than 100 %. Second, for two keypoints (i, j) in an image pair (m, n), we not only want to measure the area overlap $e_o^{i,j}(m, n)$ of their shapes for determining valid correspondences, but also the position error $e_p^{i,j}(m, n)$ of their window centres, cf. [6; 13]. As precision is an important factor, we choose to plot the repeatability for one specific image pair over varying thresholds on both e_o and e_p as in [13, Figure 21 (a)].

The results of the Lowe detector are computed with the original implementation kindly provided by the author, but using the original, not double image resolution. All other implementations are taken from the website maintained by the authors of [13], with parameters set to the same values as proposed there.

Figure 5: Example images from our experiments. From left to right: *Boat*, *Graffiti*, *Leuven City Hall*.

We investigate the repeatability on three sequences with six or seven images each, depicted in Figure 5 and 1, using carefully estimated 2D homographies between image pairs as a ground truth for the point transfer. It has been found by several authors that under moderate affine distortions (i. e. less than 40° viewpoint angle for a planar surface) affine-invariant detectors show lower repeatability than scale-invariant ones on average. Hence we expect different results for the three datasets.

The first two sequences are popular examples from the authors of [13].The *Boat sequence* shows a textured scene with natural and man-made objects from a constant viewpoint under rotation and zoom variations. We expect our detector to show good repeatability here. The *Graffiti sequence* shows a structured pattern with strong increasing affine distortions, so that in this case we expect better results from affine-invariant detectors between non-neighbouring images. Our own *Door sequence* (Figure 1) shows an indoor environment under rotation and scale variations with small affine distortions. The poor texturedness makes it a challenge for most state of the art detectors.

The results are shown in Figure 6. For the two standard datasets *Boat* and *Graffiti*, we have chosen to set the significance level of our detector so that the number of correspondences is almost equal to that of the Lowe detector. We see from the upper row of Figure 6 that, together with Lowe, our detector shows highest repeatability w.r.t. overlap error, and significantly better repeatability w.r.t. position error than all other detectors. Considering the *Graffiti* sequence, the results for our detector are below average for all scores. Note that both our detector and the Lowe detector did not return keypoints with less than around 20 % overlap error here, which may be explained by the windows shapes being restricted to circles. However, the Lowe detector behaves otherwise better than our detector, as the blobs are more robust to affine distortions than the spiral family. Although we may still improve the behaviour by computing the affine parameters of the local patch from the second moment matrix, we emphasise that the result is already satisfying for many applications.

We have also computed results when automatically esti-

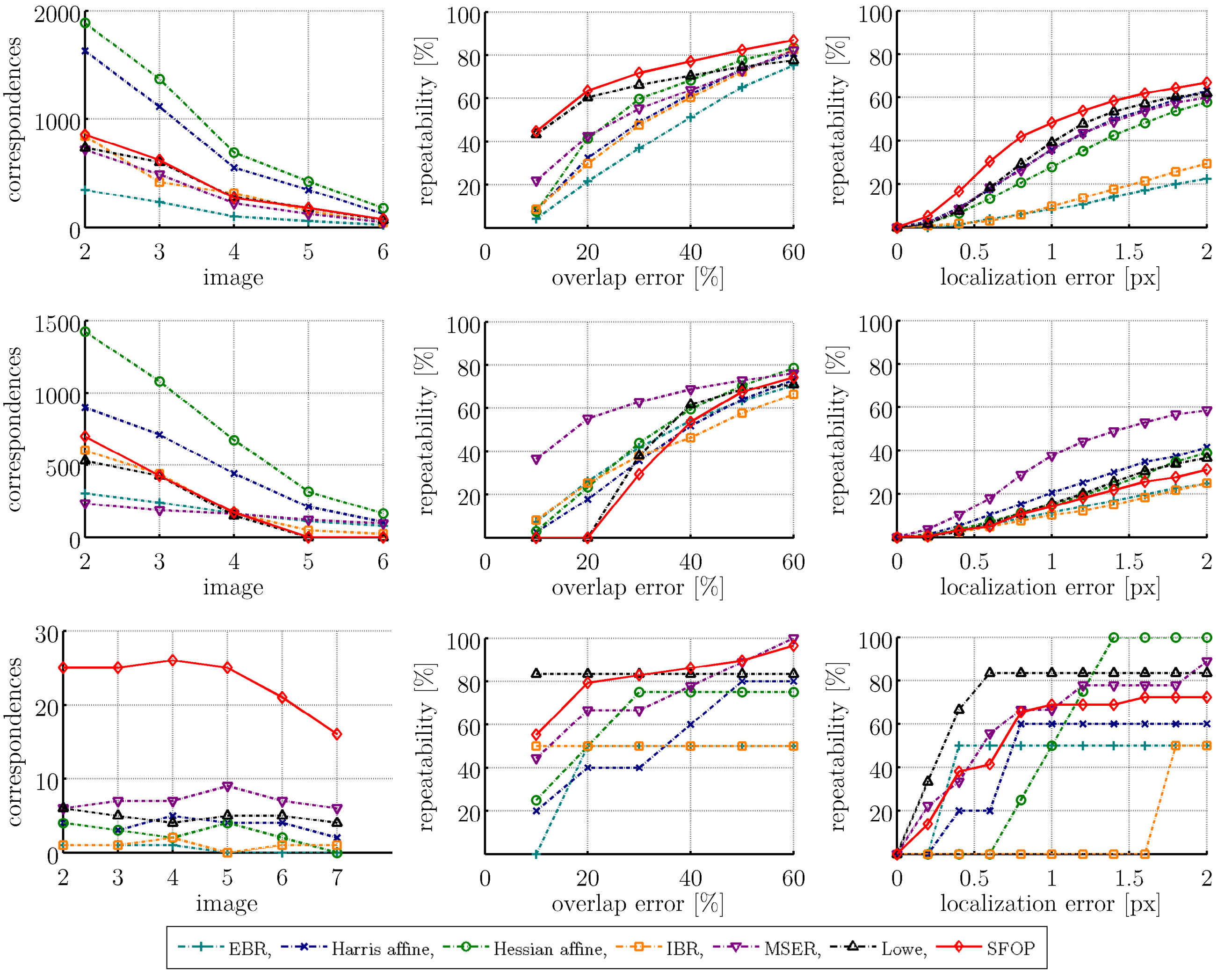


Figure 6: Results of the repeatability benchmark for three sequences. Top row: *Boat*, middle row: *Graffiti*, bottom row: *Door*. The left column shows the number of valid correspondences for each image w.r.t. the first image with less than 40 % area overlap error e_o. The centre column shows the percentage of valid correspondences for varying patch overlap error $e_o^{i,j}(1,3)$ for images 1 and 3. The right column shows the percentage of valid correspondences w.r.t. varying localisation error $e_l^{i,j}(1,3)$ for images 1 and 3. Our new Scale-invariant Feature OPerator is denoted with SFOP.

mating the noise variance of the images and setting T_λ as derived in section 3. The number of valid correspondences increased by a factor of 1.5 on average, while yielding very similar repeatability results. For objectiveness however, we used the simple thresholds in the benchmark.

Considering the results for the *Door* sequence in the bottom row of Figure 6, we see that all state of the art detectors yield less than ten matches, making automatic camera calibration almost impossible. Our detector provides about three times more correspondences, while still showing average repeatability both concerning area overlap and position error, thus building a robust input for camera calibration for this difficult sequence.

Applicability on scenes with 3D structure. The comparison carried out so far restricts to planar surfaces or far away objects. For also getting insight into the behaviour on surfaces with strong 3D structure, we used the same detectors as input for an automatic image orientation procedure on the *Leuven City Hall* sequence [20]. This procedure uses a RANSAC scheme for determining epipolar geometries, so we ran the estimation 10 times with each detector, respectively. We were interested in the average estimated accuracy of the observed features as reported by the bundle adjustment, and found that it was comparable for all detectors, being about 0.5 pel with variations of about 1/10 pel over repeated estimates. Furthermore we examined the average

Euclidean distance of the estimated projection centers from the ground truth, which however was slightly different: We got similar values of about 30 mm for Lowe and our detector, but better results – about 20 mm – for MSER, Hessian affine and the "corner subset" of our detector, i. e. when restricting to $\alpha = 0$. This indicates that the proposed detector is at least comparable to the others on images showing complex 3D structure. For more results on 3D data, we refer to a recent work on the suitability of detectors for image orientation [2], where our new detector performed well under a variety of empirical and statistical indicators. Besides this, we intend to run a recently proposed benchmark test using 3D objects [14], which we expect to verify the current results.

5. Conclusion

We have proposed a new keypoint detector which extends the one of Parida *et al.* [17] by scale-space properties and incorporates the general spiral feature model of Bigün [1]. We have shown that the detector has repeatability comparable to state of the art detectors. It gave the best overall results for a standard zoom-and-rotation sequence, while ranging below average for a sequence with strong affine distortions. The key advantages of this detector over existing ones are complementarity, and hence improved completeness on structured images, and interpretability, while showing the same high localisation accuracy. This makes it a very good choice for camera calibration and object recognition, especially in the case of poorly textured, structured objects. The completeness of several popular detectors w.r.t. the information provided by an image has been addressed in a recent study [4], where the proposed detector showed good results.

In the future we want to investigate the effect of making the differentiation scale independent on the integration scale, the effect of the size of the window around a keypoint on the repeatability, and the effect of classifying the keypoints. Furthermore, we want to develop and investigate an affine invariant extension of the detector.

Acknowledgements. We thank Marko Pilger for valuable initial investigations in his diploma thesis.

References

[1] J. Bigün. A Structure Feature for Some Image Processing Applications Based on Spiral Functions. *Computer Vision, Graphics and Image Processing*, 51(2):166–194, 1990.

[2] T. Dickscheid and W. Förstner. Evaluating the Suitability of Feature Detectors for Automatic Image Orientation Systems. In *7th International Conference on Computer Vision Systems*, Liege, Belgium, 2009.

[3] W. Förstner. A Framework for Low Level Feature Extraction. In *Third European Conference on Computer Vision*, volume III, pages 383–394, Stockholm, Sweden, 1994.

[4] W. Förstner, T. Dickscheid, and F. Schindler. On the Completeness of Coding with Image Features. In *20th British Machine Vision Conference*, London, UK, 2009.

[5] W. Förstner and E. Gülch. A Fast Operator for Detection and Precise Location of Distinct Points, Corners and Centres of Circular Features. In *ISPRS Conference on Fast Processing of Photogrammetric Data*, pages 281–305, Interlaken, 1987.

[6] A. Haja, B. Jähne, and S. Abraham. Localization Accuracy of Region Detectors. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2008.

[7] C. Harris and M. J. Stephens. A Combined Corner and Edge Detector. In *Alvey Vision Conference*, pages 147–152, 1988.

[8] U. Koethe. *Reliable Low-Level Image Analysis*. Department Informatik, University of Hamburg, Hamburg, 2008.

[9] T. Lindeberg. Feature Detection with Automatic Scale Selection. *International Journal of Computer Vision*, 30(2):79–116, 1998.

[10] D. G. Lowe. Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.

[11] J. Matas, O. Chum, M. Urban, and T. Pajdla. Robust Wide Baseline Stereo from Maximally Stable Extremal Regions. *Image and Vision Computing*, 22:761–767, 2004.

[12] K. Mikolajczyk and C. Schmid. Scale and Affine Invariant Interest Point Detectors. *International Journal of Computer Vision*, 60(1):63–86, 2004.

[13] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. V. Gool. A Comparison of Affine Region Detectors. *International Journal of Computer Vision*, 65(1/2):43–72, 2005.

[14] P. Moreels and P. Perona. Evaluation of Features Detectors and Descriptors Based on 3D Objects. In *International Journal of Computer Vision*, 2006.

[15] M. Mühlich and T. Aach. A Theory of Multiple Orientation Estimation. In *9th European Conference on Computer Vision*, Graz, 2006.

[16] J.-N. Ouellet and P. Hebert. ASN: Image Keypoint Detection from Adaptive Shape Neighborhood. In *European Conference on Computer Vision*, pages 454–467, 2008.

[17] L. Parida, D. Geiger, and R. Hummel. Junctions: Detection, Classification, and Reconstruction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(7):687–698, 1998.

[18] K. Rohr. Recognizing Corners By Fitting Parametric Models. *International Journal of Computer Vision*, 9(3):213–230, 1992.

[19] J. Shi and C. Tomasi. Good Features to Track. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 593–600, 1994.

[20] C. Strecha, R. Fransens, and L. V. Gool. Combined Depth and Outlier Estimation in Multi-View Stereo. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2006.

[21] B. Triggs. Detecting Keypoints with Stable Position, Orientation, and Scale under Illumination Changes. In *8th European Conference on Computer Vision*, pages 100–113, 2004.

[22] T. Tuytelaars and L. Van Gool. Matching Widely Separated Views Based on Affine Invariant Regions. *International Journal of Computer Vision*, 59(1):61–85, 2004.