Proc. International Conference on Scale Space and Variational Methods in Computer Vision (SSVM 2013),
Schloss Seggau, Graz region, Austria, Springer Lecture Notes in Computer Science, vol 7893, pages 355-367.

Image Matching using Generalized Scale-Space Interest Points

Tony Lindeberg*

School of Computer Science and Communication
KTH Royal Institute of Technology, Stockholm, Sweden

Abstract. The performance of matching and object recognition methods based on interest points depends on both the properties of the underlying interest points and the associated image descriptors. This paper demonstrates the advantages of using generalized scale-space interest point detectors when computing image descriptors for image-based matching. These generalized scale-space interest points are based on linking of image features over scale and scale selection by weighted averaging along feature trajectories over scale and allow for a higher ratio of correct matches and a lower ratio of false matches compared to previously known interest point detectors within the same class. Specifically, it is shown how a significant increase in matching performance can be obtained in relation to the underlying interest point detectors in the SIFT and the SURF operators. We propose that these generalized scale-space interest points when accompanied by associated scale-invariant image descriptors should allow for better performance of interest point based methods for image-based matching, object recognition and related vision tasks.

Key words: interest points, scale selection, scale linking, matching, object recognition, feature detection, scale invariance, scale space

1 Introduction

A common approach to image-based matching consists of detecting interest points with associated image descriptors from image data and then establishing a correspondence between the image descriptors. Specifically, the SIFT operator [1] and the SURF operator [2] have been demonstrated to be highly useful for this purpose with many successful applications, including object recognition, 3-D object and scene modelling, video tracking, gesture recognition, panorama stitching as well as robot localization and mapping.

In the SIFT operator, the intitial detection of interest points is based on differences-of-Gaussians from which local extrema over space and scale are computed. Such points are referred to as scale-space extrema. The difference of Gaussian operator can be seen as an approximation of the Laplacian operator,

* The support from the Swedish Research Council (contract 2010-4766), the Royal Swedish Academy of Sciences and the Knut and Alice Wallenberg Foundation is gratefully acknowledged.

and it follows from general results in [3] that the scale-space extrema of the Laplacian have scale-invariant properties that can be used for normalizing local image patches or image descriptors with respect to scaling transformations. The SURF operator is on the other hand based on initial detection of image features that can be seen as approximations of the determinant of the Hessian operator with the underlying Gaussian derivatives replaced by an approximation in terms of Haar wavelets. From the general results in [3] it follows that scale-space extrema of the determinant of the Hessian do also lead to scale-invariant behaviour, which can be used for explaining the good performance of the SIFT and SURF operators under scaling transformations.

The subject of this article is to show how the performance of image matching can be improved by using a generalized framework for detecting interest points from scale-space features involving (i) new Hessian feature strength measures at a fixed scale, (ii) linking of image features over scale into feature trajectories to allow for a better selection of significant image features and (iii) scale selection by weighted averaging along feature trajectories to allow for more robust scale estimates. By replacing the interest points in the regular SIFT and SURF operators by generalized scale-space interest points to be described below, it is possible to define new scale-invariant image descriptors that lead to better matching performance compared to the performance obtained by corresponding interest point detection mechanisms as used in the SIFT and SURF operators.

2 Generalized Scale-Space Interest Points

Basic requirements on the interest points on which image matching is to be performed are that they should (i) have a clear, preferably mathematically well-founded, *definition*, (ii) have a well-defined *position* in image space, (iii) have local image structures around the interest point that are *rich in information content* such that the interest points carry important information to later stages and (iv) be stable under local and global deformations of the image domain, including perspective image deformations and illumination variations such that the interest points can be reliably computed with a high degree of *repeatability*.

2.1 Differential Entities for Detecting Scale-Space Interest Points

As basis for performing local image measurements on a two-dimensional image f, we will consider a *scale-space representation* [4–10]

$$L(x, y;\ t) = \int_{(u,v)\in\mathbb{R}^2} f(x-u, y-v)\, g(u, v;\ t)\, du\, dv \tag{1}$$

generated by convolution with Gaussian kernels $g(x, y;\ t) = \frac{1}{2\pi t}\, e^{-(x^2+y^2)/2t}$ of increasing width, where the variance t is referred to as the scale parameter, and with *scale-normalized derivatives* with $\gamma = 1$ defined according to a $\partial_\xi = t^{\gamma/2}\, \partial_x$ and $\partial_\eta = t^{\gamma/2}\, \partial_y$ [3]. To detect interest points within this scale-space framework, we will consider:

(i) either the *scale-normalized Laplacian operator* [3]

$$\nabla^2_{norm} L = t\,(L_{xx} + L_{yy}) \tag{2}$$

or the *scale-normalized determinant of the Hessian* [3]

$$\det \mathcal{H}_{norm} L = t^2\,(L_{xx} L_{yy} - L^2_{xy}), \tag{3}$$

(ii) either of the following differential analogues/extensions of the Harris operator [11] proposed in [12, 13]; the *unsigned Hessian feature strength measure I*

$$\mathcal{D}_{1,norm} L = \begin{cases} t^2\,(\det \mathcal{H}L - k\ \mathrm{trace}^2\,\mathcal{H}L) & \text{if } \det \mathcal{H}L - k\ \mathrm{trace}^2\,\mathcal{H}L > 0 \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

or the *signed Hessian feature strength measure I*

$$\tilde{\mathcal{D}}_{1,norm} L = \begin{cases} t^2\,(\det \mathcal{H}L - k\ \mathrm{trace}^2\,\mathcal{H}L) & \text{if } \det \mathcal{H}L - k\ \mathrm{trace}^2\,\mathcal{H}L > 0 \\ t^2\,(\det \mathcal{H}L + k\ \mathrm{trace}^2\,\mathcal{H}L) & \text{if } \det \mathcal{H}L + k\ \mathrm{trace}^2\,\mathcal{H}L < 0 \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

where $k \in [0, \frac{1}{4}[$ as derived in [12] with the preferred choice $k \approx 0.06$, or

(iii) either of the following differential analogues and extensions of the Shi and Tomasi operator [14] proposed in [12, 13]; the *unsigned Hessian feature strength measure II*

$$\mathcal{D}_{2,norm} L = t\ \min(|\lambda_1|, |\lambda_2|) = t\ \min(|L_{pp}|, |L_{qq}|) \tag{6}$$

or the *signed Hessian feature strength measure II*

$$\tilde{\mathcal{D}}_{2,norm} L = \begin{cases} t\,L_{pp} & \text{if } |L_{pp}| < |L_{qq}| \\ t\,L_{qq} & \text{if } |L_{qq}| < |L_{pp}| \\ t\,(L_{pp} + L_{qq})/2 & \text{otherwise} \end{cases} \tag{7}$$

with L_{pp} and L_{qq} denoting the eigenvalues of the Hessian matrix ordered such that $L_{pp} \leq L_{qq}$ [10].

2.2 Scale Selection Mechanisms

To perform scale selection for the abovementioned differential feature detectors, we will consider two different approaches:

- Detection of *scale-space extrema* $(\hat{x}, \hat{y}, \hat{t})$ where the scale normalized differential entities assume local extrema with respect to space and scale [3], and with image features ranked by the magnitude of the scale-normalized response $|\mathcal{D}_{norm} L|$ at the scale-space extremum.
- Linking image features at different scales into feature trajectories over scale and performing scale selection by weighted averaging of scale values along each feature trajectory T delimited by bifurcation events [12, 13]

$$\hat{\tau}_T = \frac{\int_{\tau \in T} \tau\, \psi((\mathcal{D}_{\gamma-norm} L)(p(\tau);\ \tau))\, d\tau}{\int_{\tau \in T} \psi((\mathcal{D}_{\gamma-norm} L)(p(\tau);\ \tau))\, d\tau} \tag{8}$$

with the integral expressed in terms of effective scale $\tau = \log t$ to give a scale covariant construction of the corresponding scale estimates $\hat{t}_T = \exp \hat{\tau}_T$, and with significance measure taken as the integral of the scale-normalized feature responses along the feature trajectory [12, 13]

$$W_T = \int_{\tau \in T} \psi(|(\mathcal{D}_{norm} L)(p(\tau);\ \tau)|)\, d\tau \tag{9}$$

where $\psi(|\mathcal{D}_{norm} L|) = w_{\mathcal{D}L}\, |\mathcal{D}_{norm} L|^a$ represents a monotonically increasing self-similar transformation and $w_{\mathcal{D}L} = (L_{\xi\xi}^2 + 2L_{\xi\eta}^2 + L_{\eta\eta}^2)/(A(L_\xi^2 + L_\eta^2) + L_{\xi\xi}^2 + 2L_{\xi\eta}^2 + L_{\eta\eta}^2 + \varepsilon^2)$ with $A = 4/e$ representing the relative weighting between first- and second-order derivatives [15] and with $\varepsilon \approx 0.1$ representing an estimated noise level for image data in the range $[0, 255]$.

In [13] it is shown that when applied to a rotationally symmetric Gaussian blob model $f(x, y) = g(x, y;\ t_0)$, both scale-space extrema detection and weighed scale selection lead to similar scale estimates $\hat{t} = t_0$ for all the above interest point detectors. When, subjected to non-uniform affine image deformations outside the similarity group, the determinant of the Hessian $\det \mathcal{H}_{norm} L$ and the Hessian feature strength measures $\mathcal{D}_{1,norm} L$ and $\tilde{\mathcal{D}}_{1,norm} L$ do, however, have theoretical advantages in terms of affine covariance or approximations thereof [12, 13].

3 Scale-Invariant Image Descriptors for Matching

For each interest point, we will compute a complementary image descriptor in analogous ways as done in the SIFT and SURF operators, with the difference that the feature vectors will be computed from Gaussian derivative responses in a scale-space representation instead of using a pyramid as done in the original SIFT operator [1] or a Haar wavelet basis as used in the SURF operator [2].

For our SIFT-like image descriptor, we compute image gradients ∇L at the detection scale $\hat{t}$ of the interest point. An orientation estimate is computed in a similar way as by Lowe [1], by accumulating a histogram of gradient directions $\arg \nabla L$ quantized into 36 bins with the area of the accumulation window proportional to the detection scale $\hat{t}$, and then detecting peaks in the smoothed orientation histograms. Multiple peaks are accepted if the height of the secondary peak(s) are above 80 % of the highest peak. Then, for each point on a 4×4 grid with the grid spacing proportional to the detection scale measured in units of $\hat{\sigma} = \sqrt{\hat{t}}$, a weighed local histogram of gradient directions $\arg \nabla L$ quantized into 8 bins is accumulated around each grid point, with the weights proportional to the gradient magnitude $|\nabla L|$ and a Gaussian window function with its area proportional to the detection scale $\hat{t}$ with trilinear interpolation for distributing the weighted increments for the sampled image measurements into adjacent histogram bins. The resulting 128-dimensional descriptor is normalized to unit sum to achieve contrast invariance, with the relative contibution of a single bin limited to a maximum value of 0.20.

For our SURF-like image descriptor, we compute the following sums of derivative responses $\sum L_x$, $\sum |L_x|$, $\sum L_y$, $\sum |L_y|$ at the scale $\hat{t}$ of the interest point, for each one of 4×4 subwindows around the interest point as Bay et al [2] and with similar orientation normalization as for the SIFT operator. The resulting 64-D descriptor is then normalized to unit length for contrast invariance.

4 Matching Properties under Perspective Transformations

To evaluate the quality of the interest points with their associated local image descriptors, we will apply bi-directional nearest-neighbour matching of the image descriptors in Euclidean norm. In other words, given a pair of images f_A and f_B with corresponding sets of interest points $A = \{A_i\}$ and $B = \{B_j\}$, a match between the pair of interest points (A_i, B_j) is accepted only if (i) A_i is the best match for B_j in relation to all the other points in A and, in addition, (ii) B_j is the best match for A_i in relation to all the other points in B.

To suppress matching candidates for which the correspondence may be regarded as ambiguous, we will furthermore require the ratio between the distances to the nearest and the next nearest image descriptor to be less than $r = 0.9$.

Next, we will evaluate the matching performance of such interest points with local image descriptors over a dataset of poster images with calibrated homographies over different amounts of perspective scaling and foreshortening.

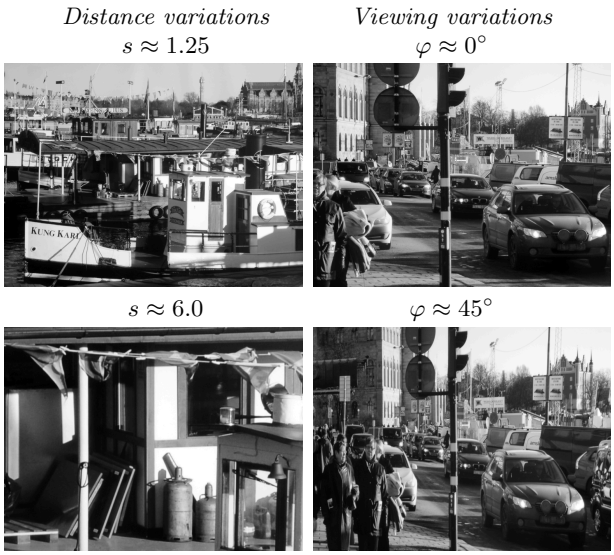

Fig. 1. Illustration of images of posters from multiple views (left) by varying the distance between the camera and the object for different frontal views, and (right) by varying the viewing direction relative to the direction of the surface normal. (Image size: 768×576 pixels.)

4.1 Poster Image Dataset

High-resolution photographs of approximately 4900×3200 pixels were taken of 12 outdoor and indoor scenes in natural city and office environments, from which poster printouts of size 100×70 cm were produced by a professional laboratory. Each such poster was then photographed from 14 different positions:

(i) 11 normal views leading to approximate scaling transformations with relative scale factors s approximately equal to 1.25, 1.5, 1.75, 2.0, 2.5, 3.0, 3.5, 4.0, 5.0 and 6.0, and

(ii) 3 additional oblique views leading to foreshortening transformations with slant angles 22.5°, 30° and 45° relative to the frontal view with $s \approx 2.0$.

For the 11 normal views of each objects, homographies were computed between each pair using the ESM method [16] with initial estimates of the relative scaling factors obtained from manual measurements of the distance between the poster surface and the camera. For the oblique views, for which the ESM method did not produce sufficiently accurate results, homographies were computed by first manually marking correspondences between the four images of each poster, computing an initial estimate of the homography using the linear method in [17, algorithm 3.2, page 92] and then computing a refined estimate by minimizing the Sampson approximation of the geometric error [17, algorithm 3.3, page 98].

The motivation for using such poster image for evaluation is to reflect natural image structures while allowing for easy calibration without 3-D reconstruction.

4.2 Matching Criteria and Performance Measures

Figure 2 shows an illustration of point matches obtained between two pairs of images corresponding to a scaling transformation and a foreshortening transformation based on interest points detected using the $\tilde{\mathcal{D}}_{1,norm}L$ operator.

To judge whether two image features A_i and B_j matched in this way should be regarded as belonging to the same feature or not, we associate a scale dependent circle C_A and C_B to each feature, with the radius of each circle equal to the detection scale of the corresponding feature measured in units of the standard deviation $\sigma = \sqrt{t}$. Then, each such feature is transformed to the other image domain, using the homography and with the scale value transformed by a scale factor of the homography. The relative amount of overlap between any pair of circles is defined by forming the ratio between the intersection and the union of the two circles in a similar way as Mikolajczyk et al [18] define a corresponding ratio for ellipses

$$m(C_A, C_B) = \frac{|\bigcap(C_A, C_B)|}{|\bigcup(C_A, C_B)|}. \tag{10}$$

Then, we measure the performance of the interest point detector by:

$$\text{efficiency} = \frac{\#(\text{interest points that lead to accepted matches})}{\#(\text{interest points})}$$

$$\text{1 - precision} = \frac{\#(\text{rejected matches})}{\#(\text{accepted matches}) + \#(\text{rejected matches})}$$

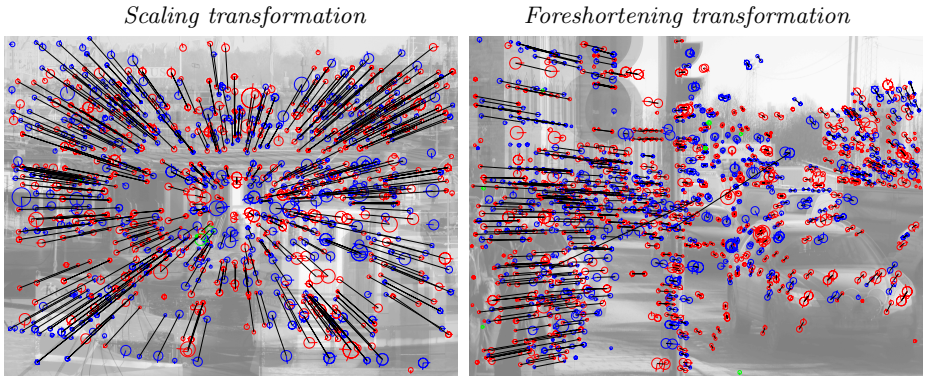

Fig. 2. Illustration of matching relations obtained by bidirectional matching of SIFT-like image descriptors computed at interest points of the signed Hessian feature strength measure $\tilde{\mathcal{D}}_{1,norm}L$ for (left) a scaling transformation and (right) a foreshortening transformation between pairs of poster images of the harbour and city scenes shown in Figure 1. These illustrations have been generated by first superimposing bright copies of the two images to be matched by adding them. Then, the interest points detected in the two domains have been overlayed on the image data, and a black line has been drawn between each pair of image points that has been matched. Red circles indicate that the Hessian matrix is negative definite (bright features), blue circles that the Hessian matrix is positive definite (dark features), whereas green circles indicate that the Hessian matrix is indefinite (saddle-like features).

The evaluation of the matching score is only performed for image features that are within the image domain for both images before and after the transformation. Moreover, only features within corresponding scale ranges are evaluated. In other words, if the scale range for the image f_A is $[t_{min}, t_{max}]$, then image features are searched for in the transformed image f_B within the scale range $[t'_{min}, t'_{max}] = [s^2\, t_{min}, s^2\, t_{max}]$, where s denotes an overall scaling factor of the homography. In the experiments below, we used $[t_{min}, t_{max}] = [4, 256]$.

4.3 Experimental Results

Table 1 shows the result of evaluating 2×9 different types of scale-space interest point detectors with respect to the problem of establishing point correspondences between pairs of images on the poster dataset. Each interest point detector is applied in two versions (i) with scale selection from local extrema of scale-normalized derivatives over scale, or (ii) using scale linking with scale selection from weighted averaging of scale-normalized feature responses along feature trajectories.

In addition to the 2×7 differential interest point detectors described in section 2, we have also included 2×2 additional interest point detectors derived from the Harris operator [11]: (i) the Harris-Laplace operator [19] based on

Efficiency: SIFT-like image descriptor

	scaling		foreshortening		average	
Interest points	extr	link	extr	link	extr	link
$\nabla^2_{norm} L \quad (\mathcal{D}_1 L > 0)$	0.7484	0.7994	0.7512	0.7574	0.7498	0.7784
$\det \mathcal{H}_{norm} L \ (\mathcal{D}_1 L > 0)$	0.7721	*0.8225*	0.7635	*0.7932*	0.7678	*0.8079*
$\det \mathcal{H}_{norm} L \ (\tilde{\mathcal{D}}_1 L > 0)$	0.7691	0.8163	0.7602	0.7841	0.7647	0.8002
$\mathcal{D}_{1,norm} L$	0.7719	**0.8280**	0.7596	**0.7977**	0.7658	**0.8128**
$\tilde{\mathcal{D}}_{1,norm} L$	0.7698	0.8241	0.7578	*0.7916*	0.7638	*0.8079*
$\mathcal{D}_{2,norm} L \quad (\mathcal{D}_1 L > 0)$	0.7203	0.8187	0.7111	0.7776	0.7157	0.7981
$\tilde{\mathcal{D}}_{2,norm} L \quad (\mathcal{D}_1 L > 0)$	0.7204	*0.8261*	0.7113	0.7766	0.7159	0.8014
Harris-Laplace	0.7002	0.7855	0.7046	0.7535	0.7024	0.7695
Harris-detHessian	0.7406	0.7608	0.7561	0.7319	0.7406	0.7463

Efficiency: SURF-like image descriptor

	scaling		foreshortening		average	
Interest points	extr	link	extr	link	extr	link
$\nabla^2_{norm} L \quad (\mathcal{D}_1 L > 0)$	0.7424	0.7832	0.7280	0.7140	0.7352	0.7486
$\det \mathcal{H}_{norm} L \ (\mathcal{D}_1 L > 0)$	0.7656	0.8072	0.7402	*0.7504*	0.7529	*0.7788*
$\det \mathcal{H}_{norm} L \ (\tilde{\mathcal{D}}_1 L > 0)$	0.7628	0.8015	0.7372	0.7430	0.7500	0.7723
$\mathcal{D}_{1,norm} L$	0.7661	**0.8126**	0.7354	**0.7537**	0.7507	**0.7831**
$\tilde{\mathcal{D}}_{1,norm} L$	0.7640	*0.8081*	0.7334	*0.7478*	0.7487	*0.7779*
$\mathcal{D}_{2,norm} L \quad (\mathcal{D}_1 L > 0)$	0.7157	0.8014	0.6870	0.7284	0.7013	0.7649
$\tilde{\mathcal{D}}_{2,norm} L \quad (\mathcal{D}_1 L > 0)$	0.7158	*0.8100*	0.6873	0.7328	0.7015	0.7714
Harris-Laplace	0.6948	0.7620	0.6724	0.6944	0.6836	0.7282
Harris-detHessian	0.7345	0.7381	0.7192	0.6705	0.7268	0.7043

Table 1. Performance measures obtained by matching different types of scale-space interest points with associated SIFT- and SURF-like image descriptors for the poster image dataset. The columns show from left to right: (i) the average efficiency over all pairs of scaling transformations, (ii) the average efficiency over all pairs of foreshortening transformations and (iii) the average total computed as the mean of the scaling and foreshortening scores. The columns labelled "extr" and "link" indicate whether the features have been detected with scale selection from extrema over scale or by scale linking.

spatial extrema of the Harris measure and scale selection from local extrema over scale of the scale-normalized Laplacian, (ii) a scale-linked version of the Harris-Laplace operator with scale selection by weighted averaging over feature trajectories of Harris features [12], and (iii-iv) two Harris-detHessian operators analogous to the Harris-Laplace operators, with the difference that scale selection is performed based on the scale-normalized determinant of the Hessian instead of the scale-normalized Laplacian [12].

The experiments are based on detecting the $N = 800$ strongest interest points extracted from the first image, regarded as reference image for the homography. To obtain an approximate uniform density of interest points under scaling transformations, an adapted number $N' = N/s^2$ of interest points is searched for (i) within the subwindow of the reference image that is mapped to

Interest points and image descriptors ranked on matching efficiency

Interest points	Scale selection	Descriptor	Efficiency
$\mathcal{D}_{1,norm}L$	link	SIFT	0.8128
$\tilde{\mathcal{D}}_{1,norm}L$	link	SIFT	0.8079
$\det \mathcal{H}_{norm}L$ $(\mathcal{D}_1 L > 0)$	link	SIFT	0.8079
$\tilde{\mathcal{D}}_{2,norm}L$ $(\mathcal{D}_1 L > 0)$	link	SIFT	0.8014
$\det \mathcal{H}_{norm}L$ $(\tilde{\mathcal{D}}_1 L > 0)$	link	SIFT	0.8002
$\vdots$			$\vdots$
$\det \mathcal{H}_{norm}L$ $(\mathcal{D}_1 L > 0)$	extr	SIFT	0.7721
$\det \mathcal{H}_{norm}L$ $(\mathcal{D}_1 L > 0)$	extr	SURF	0.7656
$\nabla^2_{norm}L$ $(\mathcal{D}_1 L > 0)$	extr	SIFT	0.7484
Harris-Laplace	extr	SIFT	0.7002

Table 2. The five best combinations of interest points and image descriptors among the $2 \times 2 \times 9 = 36$ combinations considered in this experimental evaluation as ranked on the ratio of interest points that lead to correct matches. For comparison, results are also shown for the SIFT descriptor based on scale-space extrema of the Laplacian, the SIFT or SURF descriptors based on scale-space extrema of the determinant of the Hessian and the SIFT descriptor based on Harris-Laplace interest points.

the interior of the transformed image and (ii) in the transformed image, with s denoting relative scaling factor between the two images.

This procedure is repeated for all pairs of images within the groups of distance variations or viewing variations respectively, implying up to 55 image pairs for the scaling transformations and 6 image pairs for the foreshortening transformations, *i.e.* up to 61 matching experiments for each one of the 12 posters, thus up to 732 experiments for each one of 2×9 interest point detectors.

As can be seen from the results of matching SIFT- or SURF-like image descriptors in Table 1, the interest point detectors based on scale linking and with scale selection by weighted averaging along feature trajectories generally lead to significantly higher efficiency rates compared to the corresponding interest point detectors based on scale selection from local extrema over scale. Specifically, the highest efficiency rates are obtained with the scale linked version of the unsigned Hessian feature strength measure $\mathcal{D}_{1,norm}L$, followed by scale-linked versions of the unsigned signed Hessian feature strength measure $\tilde{\mathcal{D}}_{1,norm}L$ and the determinant of the Hessian operator $\det \mathcal{H}_{norm}L$ with complementary thresholding on $\mathcal{D}_{1,norm}L > 0$.

Corresponding experimental results that cannot be included here because of lack of space show that the lowest and thus the best 1-precision score is obtained with the determinant of the Hessian operator $\det \mathcal{H}_{norm}L$ with complementary thresholding on $\tilde{\mathcal{D}}_{1,norm}L > 0$, followed by the determinant of the Hessian operator $\det \mathcal{H}_{norm}L$ with complementary thresholding on $\mathcal{D}_{1,norm}L > 0$.

Among the more traditional feature detectors based on scale selection from local extrema over scale, we can also note that the determinant of the Hessian operator $\det \mathcal{H}_{norm}L$ performs significantly better than both the Laplacian operator $\nabla^2_{norm}L$ and the Harris-Laplace operator. We can also note that the

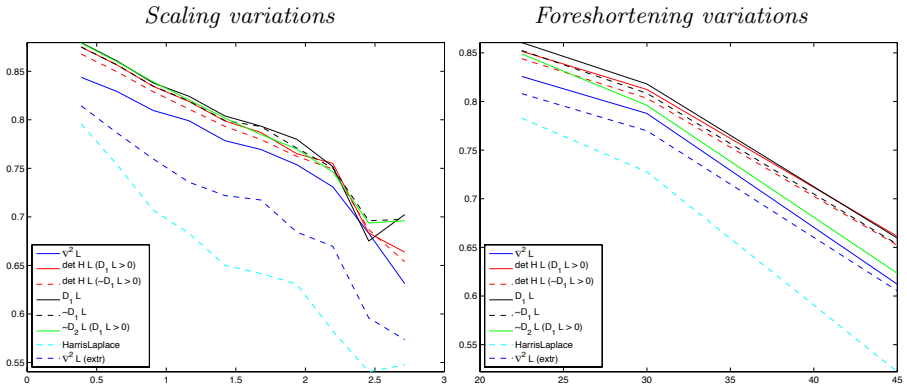

Fig. 3. Graphs showing how the matching efficiency depends upon (left) the amount of scaling $s \in [1.25, 6.0]$ for scaling transformations (with $\log_2 s$ on the horizontal axis) and (right) the difference in viewing angle $\varphi \in [22.5°, 45°]$ for the foreshortening transformations for interest point matching based on SIFT-like image descriptors.

Harris-Laplace operator can be improved by either scale linking or by replacing scale selection based on the scale-normalized Laplacian by scale selection based on the scale-normalized determinant of the Hessian.

When comparing the results obtained for SIFT-like and SURF-like image descriptors, we can see that the SIFT-like image descriptors lead to both higher efficiency rates and lower 1-precision scores than the SURF-like image descriptors. This qualitative relationship holds over all types of interest point detectors. In this respect, the pure image descriptor in the SIFT operator is clearly better than the pure image descriptor in the SURF operator. Specifically, more reliable image matches can be obtained by replacing pure image descriptor in the SURF operator by the pure image descriptor in the SIFT operator.

Table 2 lists the five best combinations of interest point detectors and image descriptors in this evaluation as ranked on their efficiency values. For comparison, the results of our corresponding analogues of the SIFT operator with interest point detection from scale-space extrema of the Laplacian and our analogue of the SURF operator based on scale-space extrema of the determinant of the Hessian are also shown. As can be seen from this ranking, the best combinations of generalized points with SIFT-like image descriptors perform significantly better than the corresponding analogues of regular SIFT or regular SIFT based on scale-space extrema of the Laplacian or the determinant of the Hessian.

Figure 3 shows graphs of how the efficiency rate depends upon the amount of scaling for the scaling transformations and the difference in viewing angle for the foreshortening transformations. As can be seen from these graphs, the interest point detectors $\det \mathcal{H}_{norm} L$, $\mathcal{D}_{1,norm} L$ and $\tilde{\mathcal{D}}_{1,norm} L$ that possess affine covariance properties or approximations thereof [12, 13] do also have the best matching properties under foreshortening transformations. Specifically, the generalized interest point detectors based on scale linking perform significantly better than

scale-space extrema of the Laplacian or the determinant of the Hessian as well as better than the Harris-Laplace operator.

5 Summary and Conclusions

We have presented a set of extensions of the SIFT and SURF operators, by replacing the underlying interest point detectors used for computing the SIFT or SURF descriptors by a family of generalized scale-space interest points.

These generalized scale-space interest points are based on (i) new differential entities for interest point detection at a fixed scale in terms of new Hessian feature strength measures, (ii) linking of image structures into feature trajectories over scale and (ii) performing scale selection by weighted averaging of scale-normalized feature responses along these feature trajectories [12].

The generalized scale-space interest points are all *scale-invariant* in the sense that (i) the interest points are preserved under scaling transformation and that (ii) the detection scales obtained from the scale selection step are transformed in a scale covariant way. Thereby, the detection scale can be used for defining a local scale normalized reference frame around the interest point, which means that image descriptors that are defined relative to such a scale-normalized reference frame will also be scale invariant.

By complementing these generalized scale-space interest points with local image descriptors defined in a conceptually similar way as the pure image descriptor parts in SIFT or SURF, while being based on image measurements in terms of Gaussian derivatives instead of image pyramids or Haar wavelets, we have shown that the generalized interest points with their associated scale-invariant image descriptors lead to a higher ratio of correct matches and a lower ratio of false matches compared to corresponding results obtained with interest point detectors based on more traditional scale-space extrema of the Laplacian, scale-space extrema of the determinant of the Hessian or the Harris-Laplace operator.

In the literature, there has been some debate concerning which one of the SIFT or SURF descriptors leads to the best performance. In our experimental evaluations, we have throughtout found that our SIFT-like image descriptor based on Gaussian derivatives generally performs much better than our SURF-like image descriptor, also expressed in terms of Gaussian derivatives. In this respect, the pure image descriptor in the SIFT operator can be seen as significantly better than the pure image descriptor in the SURF operator.

Concerning the underlying interest points, we have on the other hand found that the determinant of the Hessian operator to generally perform significantly better than the Laplacian operator, both for scale selection based on scale-space extrema and scale selection based on weighted averaging of feature responses along feature trajectories obtained by scale linking. Since the difference-of-Gaussians interest point detector in the regular SIFT operator can be seen as an approximation of the scale-normalized Laplacian, we can therefore regard the underlying interest point detector in the SURF operator as significantly better than the interest point detector in the SIFT operator. Specifically, we

could expect a significant increase in the performance of SIFT by just replacing the scale-space extrema of the difference-of-Gaussians operator by scale-space extrema of the determinant of the Hessian.

In addition, our experimental evaluations show that further improvements are possible by replacing the interest points obtained from scale-space extrema in the SIFT and SURF operators by generalized scale-space interest points obtained by scale linking, with the best results obtained with the Hessian feature strength measures $\mathcal{D}_{1,norm}L$ and $\tilde{\mathcal{D}}_{1,norm}L$ followed by the determinant of the Hessian $\det \mathcal{H}_{norm}L$ and the Hessian feature strength measure $\tilde{\mathcal{D}}_{2,norm}L$.

References

1. Lowe, D.: Distinctive image features from scale-invariant keypoints. Int. J. Comp. Vis. **60** (2004) 91–110
2. Bay, H., Ess, A., Tuytelaars, T., van Gool: Speeded up robust features (SURF). CVIU **110** (2008) 346–359
3. Lindeberg, T.: Feature detection with automatic scale selection. Int. J. Comp. Vis. **30** (1998) 77–116
4. Witkin, A.P.: Scale-space filtering. In: 8th IJCAI. (1983) 1019–1022
5. Koenderink, J.J.: The structure of images. Biol. Cyb. **50** (1984) 363–370
6. Koenderink, J.J., van Doorn, A.J.: Generic neighborhood operators. IEEE-PAMI **14** (1992) 597–605
7. Lindeberg, T.: Scale-Space Theory in Computer Vision. Springer (1994)
8. Florack, L.M.J.: Image Structure. Springer (1997)
9. ter Haar Romeny, B.: Front-End Vision and Multi-Scale Image Analysis. Springer (2003)
10. Lindeberg, T.: Scale-space. In Wah, B., ed.: Encyclopedia of Computer Science and Engineering. Wiley (2008) 2495–2504
11. Harris, C., Stephens, M.: A combined corner and edge detector. In: Alvey Vision Conference. (1988) 147–152
12. Lindeberg, T.: Generalized scale-space interest points: Scale-space primal sketch for differential descriptors. (2010) (under revision for International Journal of Computer Vision, original version submitted in June 2010).
13. Lindeberg, T.: Scale selection properties of generalized scale-space interest point detectors. J. Math. Im. Vis. (2012) Digitally published with DOI:10.1007/s10851-012-0378-3 in September 2012.
14. Shi, J., Tomasi, C.: Good features to track. In: CVPR. (1994) 593–600
15. Lindeberg, T.: On automatic selection of temporal scales in time-casual scale-space. In: Proc. AFPAC'97. Volume 1315 of LNCS., Springer (1997) 94–113
16. Benhimane, S., Malis, E.: Real-time image-based tracking of planes using efficient second-order minimization. In: Intelligent Robots and Systems IROS'2004. (2004) 943–948
17. Hartley, R., Zisserman, A.: Multiple View Geometry in Computer Vision. Cambridge University Press (2000) First Edition.
18. Mikolajczyk, K., Tuytelaars, T., Schmid, C., Zisserman, A., Matas, J., Schaffalitzky, F., Kadir, T., van Gool, L.: A comparison of affine region detectors. Int. J. Comp. Vis. **65** (2005) 43–72
19. Mikolajczyk, K., Schmid, C.: Scale and affine invariant interest point detectors. Int. J. Comp. Vis. **60** (2004) 63–86