Top-Points as Interest Points for Image Matching

B. Platel, E. Balmachnova, L.M.J. Florack⋆, and B.M. ter Haar Romeny

Technische Universiteit Eindhoven, P.O. Box 513,
5600 MB Eindhoven, The Netherlands
{B.Platel, E.Balmachnova, L.M.J.Florack,
B.M.terHaarRomeny}@tue.nl

Abstract. We consider the use of top-points for object retrieval. These points are based on scale-space and catastrophe theory, and are invariant under gray value scaling and offset as well as scale-Euclidean transformations. The differential properties and noise characteristics of these points are mathematically well understood. It is possible to retrieve the exact location of a top-point from any coarse estimation through a closed-form vector equation which only depends on local derivatives in the estimated point. All these properties make top-points highly suitable as anchor points for invariant matching schemes. By means of a set of repeatability experiments and receiver-operator-curves we demonstrate the performance of top-points and differential invariant features as image descriptors.

1 Introduction

Local invariant features are useful for finding corresponding points between images when they are calculated at invariant interest points. The most popular interest points are Harris points [1], extrema in the normalized scale-space of the Laplacian of the image [2] used in the popular SIFT keypoint detector [3] or a combination of both [4]. For an overview of different interest points the reader is referred to [5].

We propose a novel, highly invariant type of interest point, based on scale-space and catastrophe theory. The mathematical properties and behavior of these so-called top-points are well understood. These interest points are invariant under gray value scaling and offset as well as arbitrary scale-Euclidean transformations. The noise behavior of top-points can be described in closed-form, which enables us to accurately predict the stability of the points. For tasks like matching or retrieval it is important to take into account the (in)stability of the descriptive data.

For matching it is important that a set of distinctive local invariant features is available in the interest points. An overview of invariant features is given in [6]. The choice of invariant features taken in the top-points is free. Because of their simple and mathematically nice nature we have chosen to use a complete set of differential invariants up to third order [7, 8] as invariant features. A similarity measure between these invariant feature vectors based on the noise behavior of the differential invariants is proposed. By means of a set of repeatability experiments and receiver-operator-curves we demonstrate the performance of top-points and differential invariant features as image descriptors.

2   Theory

We present an algorithm for finding interest points in Gaussian scale-space. As input we may use the original image, but we may also choose to use its Laplacian, or any other linear differential entity. The input for our algorithm will be referred to as $u(x, y)$.

2.1   Scale-Space Approach

To find interest points that are invariant to scaling we have to observe the input function at all possible scales. Particularly suitable for calculating the scale-space representation of the image (or any other linear differential entity of the image) is the Gaussian kernel [9]

$$\phi_\sigma(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{1}{2}(x^2+y^2)/\sigma^2}. \tag{1}$$

The input function can now be calculated at any scale by convolution with the Gaussian

$$u(x, y, \sigma) = (\phi_\sigma * u)(x, y). \tag{2}$$

Derivatives of the input function can be calculated at any scale by

$$\mathcal{D}u(x, y, \sigma) = (\mathcal{D}\phi_\sigma * u)(x, y), \tag{3}$$

where $\mathcal{D}$ is any linear derivative operator with constant coefficients.

2.2   Catastrophe Theory

Critical points are points at any fixed scale at which the gradient vanishes. Catastrophe theory studies how such points change as certain control parameters change, in our case scale.

In the case of a generic 2D input function the catastrophes occurring in Gaussian scale space are creations and annihilations of critical points with opposite Hessian signature [10, 11], i.e. extrema and saddles. The movement of critical points through scale induces critical paths. Each path consists of one (or multiple) saddle branch(es) and extremum branch(es). The point at which a creation or annihilation occurs is referred to as a *top-point*[1]. A typical set of critical paths and top-points of an image is shown in Fig. 1. In a top-point the determinant of the Hessian of the input function becomes zero. A top-point is thus defined as a point for which

$$\begin{cases} u_x & = 0\,, \\ u_y & = 0\,, \\ u_{xx}u_{yy} - u_{xy}^2 = 0\,. \end{cases} \tag{4}$$

The extrema of the normalized Laplacean scale space as introduced by Lindeberg [2], and used by Lowe [3] in his matching scheme, lie on the critical paths of the Laplacean image. Multiple of such extrema may exist on the extremum branch of a critical path, whereas there is only one top-point per annihilating extremum/saddle pair, Fig. 2a.

---

[1] This misnomer is reminiscent of the 1D case [12], in which only annihilations occur generically, so that a top-point is only found at the top of a critical path.
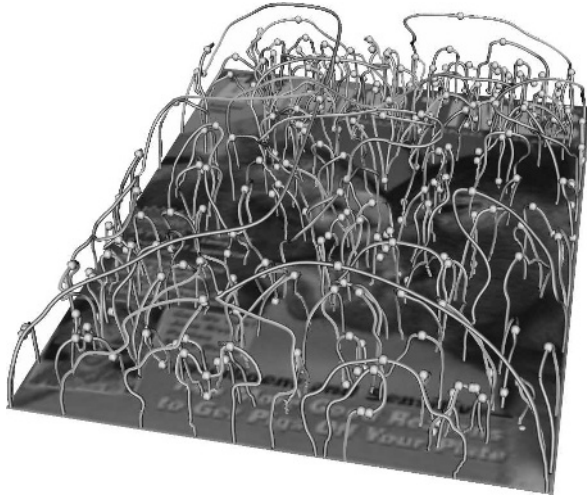
Fig. 1. Selection of critical paths and top-points of a magazine cover image

2.3   Invariance

Interest points are called invariant to transformation if they are preserved by the trans-
formation. From their definition (4), it is apparent that top-points are invariant under
gray value scaling and offset. Next to this the top-points are also invariant to scale-
Euclidean transformations (rotation, scaling, translation).

The top-points however are in theory not invariant to affine or projective transforma-
tions just like the interest point detectors mentioned earlier, but in practice they show to
be invariant under small affine or projective transformations.

2.4   Detection Versus Localization

Critical paths are detected by following critical points through scale. Top-points are
found as points on the critical paths with horizontal tangents.

The detection of top-points does not have to be exact, since, given an adequate initial
guess, it is possible to refine their position such that (4) holds to any desired precision. If
(x_0, y_0, t_0) denotes the approximate location of a top-point we can calculate the position
of the true top-point $(x_0 + \xi, y_0 + \eta, t_0 + \tau)$ in the neighborhood by:

$$\begin{bmatrix} \xi \\ \eta \\ \tau \end{bmatrix} = -\mathbf{M}^{-1} \begin{bmatrix} \mathbf{g} \\ \det \mathbf{H} \end{bmatrix}, \tag{5}$$

where

$$\mathbf{M} = \begin{bmatrix} \mathbf{H} & \mathbf{w} \\ \mathbf{z}^T & c \end{bmatrix}, \tag{6}$$

$$\mathbf{g} = \nabla u, \ \mathbf{H} = \nabla \mathbf{g}, \ \mathbf{w} = \partial_t \mathbf{g}, \ \mathbf{z} = \nabla \det \mathbf{H}, \ c = \partial_t \det \mathbf{H}, \tag{7}$$

in which **g** and **H** denote the image gradient and Hessian matrix, respectively, and in which all derivatives are taken in the point (x_0, y_0, t_0), cf. [11] for a derivation. This allows one to use a less accurate but fast detection algorithm.

2.5 Perturbative Approach in Scale Space

Given a set of measurements in scale space $v \in \mathbb{R}^n$ we can calculate the propagation of errors in a function $f : \mathbb{R}^n \to \mathbb{R}^m$ if the measurements are perturbed with noise n, $w = v + n \in \mathbb{R}^n$. The following equation describes how the perturbation affects f, using Einstein summation convention for repeated indices:

$$f_\alpha(w) - f_\alpha(v) \approx \delta f_\alpha \equiv \left. \frac{\partial f_\alpha}{\partial w_\beta} \right|_{w=v} n_\beta \tag{8}$$

The covariance matrix of f can be expressed as:

$$< \delta f_\alpha \delta f_\beta > = \frac{\partial f_\alpha}{\partial v_\gamma} \frac{\partial f_\beta}{\partial v_\delta} < n_\gamma n_\delta > \tag{9}$$

The noise matrix $< n_\gamma n_\delta >$ is given in [13] for the case when v denotes a partial derivative of the image obtained through convolution with a Gaussian derivative filter.

2.6 Stability

The stability of a top-point can be expressed in terms of the variances of spatial and scale displacements induced by additive noise. Since top-points are generic entities in scale space, they cannot vanish or appear when the image is only slightly perturbed. We assume that the noise variance is "sufficiently small" in the sense that the induced dislocation of a top-point can be investigated by means of a perturbative approach. By using eqn. (9) and substituting f with eqn. (5) we are able to calculate the effects of
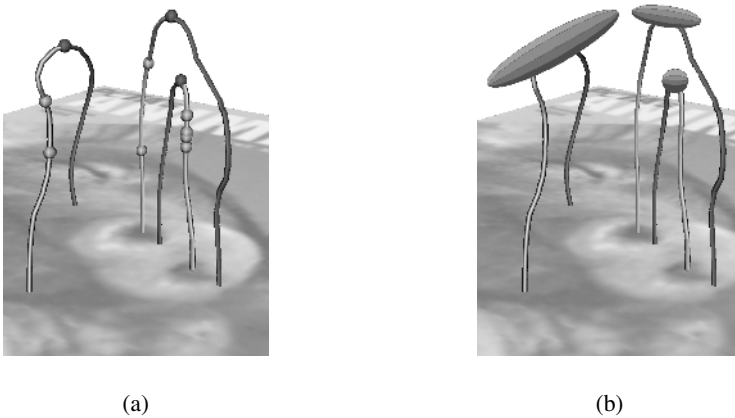

(a)   (b)

Fig. 2. a. A set of critical paths with corresponding top-points (topmost bullets), and extrema of the normalized Laplacian (remaining bullets). **b.** The ellipses schematically represent the variances of the scale-space displacement of each top-point under additive noise of known variance.

noise on the position of top-points in the form of a covariance matrix. It can be shown that the displacement depends on derivatives up to fourth order evaluated at the top-point, and on the noise variance. For detailed formulas (and experimental verifications) the reader is referred to [14].

The advantage of this approach is that variances of scale-space displacements can be predicted theoretically and in analytically closed-form on the basis of the local differential structure at a given top-point, cf. Fig. 2b for a schematic illustration. The ability to predict the motion of top-points under noise is valuable when matching noisy data (e.g. one may want to disregard highly instable top-points altogether).

2.7 Local Invariant Features

For matching it is important that a set of distinctive local invariant features is available in the interest points. It is possible to use any set of invariant features in the top-points. Mikolajcyck and Schmid [6] give an overview of a number of such local descriptors.

For our experiments we have used a complete set of differential invariants up to third order. The complete sets proposed by Florack et al. [8] are invariant to rigid transformations. By suitable scaling and normalization we obtain invariance to spatial zooming and intensity scaling as well, but the resulting system has the property that most low order invariants vanish identically at the top-points of the original (zeroth order) image, and thus do not qualify as distinctive features. Thus when considering top-points of the original image other distinctive features will have to be used. In [15] the embedding of a graph connecting top-points is used as a descriptor. This proved to be a suitable way of describing the global relationship between top-points of the original image. In this paper we use the Laplacian of the input function as input for our top-point detector. For this case the non-trivial, scaled and normalized differential invariants up to third order are collected into the column vector given by (10), again using summation convention:

$$
\begin{pmatrix}
\sigma \sqrt{u_i u_i}/u \\
\sigma u_{ii}/\sqrt{u_j u_j} \\
\sigma^2 u_{ij} u_{ij}/u_k u_k \\
\sigma u_i u_{ij} u_j/(u_k u_k)^{3/2} \\
\sigma^2 u_{ijk} u_i u_j u_k/(u_l u_l)^2 \\
\sigma^2 \varepsilon_{ij} u_{jkl} u_i u_k u_l/(u_m u_m)^2
\end{pmatrix} . \tag{10}
$$

Here ε_{ij} is the completely antisymmetric epsilon tensor, normalized such that $\varepsilon_{12} = 1$. Note that the derivatives are extracted from the original, zeroth order image, but evaluated at the location of the top-points of the image Laplacian. This is, in particular, why the gradient magnitude in the denominator poses no difficulties, as it is generically nonzero at a top-point.

The resulting scheme (interest point plus differential feature vector) guarantees manifest invariance under the scale-Euclidean spatial transformation group, and under linear gray value rescalings.

2.8 Similarity Measure in the Feature Space

To compare features of different interest points a distance or similarity measure is needed. The most often used measures in literature are the Euclidean and Mahalanobis

distance. If x_0 and x are two points from the same distribution which has covariance matrix Σ, then the Mahalanobis distance is given by

$$d(x_0, x) = (\mathbf{x} - \mathbf{x_0})^T \Sigma^{-1} (\mathbf{x} - \mathbf{x_0}) \tag{11}$$

and is equal to the Euclidean distance if the covariance matrix Σ is the identity matrix. The advantage of the Mahalanobis distance is that it can be used to measure distances in non-Euclidean spaces. The drawbacks however are that the covariance matrix has to be derived by using a large training set of images, and that the covariance matrix is the same for every measurement. By using the perturbative approach from sec. 2.5 and using the set of differential invariants from (10) as functions f_α and the set of third order derivatives as v_β we can now calculate a covariance matrix for every single feature vector. This enables us to use (11) to calculate the similarity between two feature vectors using the covariance matrix Σ_{x_0} derived specifically for feature vector x_0, where $d(x_0, x)$ close to zero means very similar, and $d(x_0, x) \gg 0$ very dissimilar. Note that this makes the similarity measure asymmetric: $d(x_0, x) \neq d(x, x_0)$. Therefore we cannot speak of a distance measure. This however does not pose problems since we are only matching unidirectionally, viz. object to scene.

3   Experiments

3.1   Database

For the experiments we use a data set containing transformed versions of 12 different magazine covers. The covers contain a variety of objects and text. The data set contains rotated, zoomed and noisy versions of these magazine covers as well as images with perspective transformations. For all transformations the ground truth is known, which enables us to verify the performance of different algorithms on the database. Mikolajczyk's data set used in [4, 6] is not suitable for our purposes, as we require ground truth for genuine group transformations not confounded with other sources of image changes, such as changes in field of view. To our knowledge Mikolajczyk's data set does not provide this.

3.2   Repeatability

Schmid et al. [5] have introduced the so-called repeatability criterion to evaluate the stability and accuracy of interest points and interest point detectors. The repeatability rate for an interest point detector on a given pair of images is computed as the ratio between the number of point-to-point correspondences and the minimum number of interest points detected in the images ($\times 100\%$).

If the interest point in the perturbed image has moved less than a distance of ϵ pixels away from the position where it would be expected when following the transformation, we mark the point as a repeatable point (typically we set $\epsilon \approx 2$ pixels).

Experiments show the repeatability of top-points under image rotation (Fig. 4a) and additive Gaussian noise (Fig. 4b). Image rotation causes some top-points to be lost or created due to the resampling of the image. In the Gaussian noise experiment we

Fig. 3. A selection of data set images. From left to right: unchanged, rotated, added noise, scaled, changed perspective.
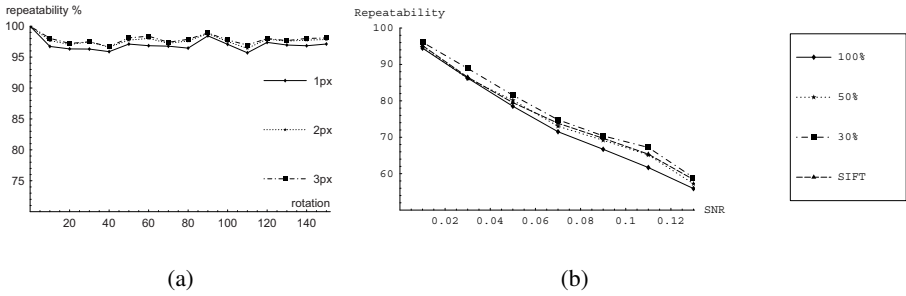

(a)   (b)

Fig. 4. a. The repeatability rate of the top-points for different angles of rotation for different ϵ. **b.** The repeatability rate of the top-points and SIFT interest points for additive Gaussian noise expressed in signal to noise ratio.

demonstrate that by using the stability variances described in Sec. 2.6 the repeatability of the top-points can be increased. The top-points are ordered on their stability variances. From this list 100%, 50% and 30% of the most stable top-points are selected for the repeatability experiment respectively. From Fig. 4b it is apparent that discarding instable points increases the repeatability significantly. We compare the repeatability of our interest point detector to the SIFT interest point detector by Lowe [3]. In Fig. 4b can be seen that when we apply a threshold on our stability measure (the SIFT keypoints have already been thresholded on stability) we slightly outperform the SIFT interest point detector for the noise case. Both algorithms perform worst for a rotation of 45 degrees. On the average taken over the entire database of 45 degree rotated images the repeatability of the SIFT interest points is 78%. Our top-point interest point detector showed a repeatability rate of 85% when thresholded on stability.

The high repeatability rate of the top-points enables us to match images under any angle of rotation and under high levels of noise.

3.3   Receiver Operator Characteristics

For the performance evaluation of the similarity measure we use a similar criterion as the one used in [6]. This criterion is based on Receiver Operating Characteristics (ROC) of detection rate versus false positive rate. Two points are said to be similar if

the distance between their feature vectors is below a threshold t. The value of t is varied to obtain the ROC curves.

Given two images representing the same object the True Positive Rate TPR is the number of correctly matched points with respect to the number of possible matches:

$$TPR = \frac{\#\text{correct matches}}{\#\text{possible matches}} \qquad (12)$$

The condition for calling a match correct is the same as in sec. 3.2. The False Positive Rate FPR as defined in [6] is calculated as:

$$FPR = \frac{\#\text{incorrect matches}}{(\#\text{object points})(\#\text{scene points})} \qquad (13)$$

where the object is the original image and the scene a transformed version of the original image.

3.4   Performance of the Similarity Measure

To evaluate the performance of the similarity measure defined in sec. 2.8 we have calculated the ROC curves as described sec. 3.3 for a set of experiments. For comparison we have included the ROC curves for the Mahalanobis and Euclidean distance measures. The covariance matrix for the Mahalanobis distance was obtained by training on the data set itself. In Fig. 5 the mean ROC curves for three experiments are shown. In experiment a. the images in the database are matched to a 50% scaled down version of the same images. In experiment b. the images in the database are matched to noisy versions of the same images. In experiment c. the images in the database are matched to the 45 degree rotated versions of the same images. In all the experiments it is obvious that the new similarity measure greatly improves the performance of the matching algorithm.

3.5   Performance of the Descriptors

To evaluate the performance of the differential invariant features defined in sec. 2.7 we have calculated the ROC curves as described sec. 3.3 for a set of experiments. For comparison we have included the ROC curves of the SIFT algorithm for which a precompiled program is publicly available. The SIFT features consist of a 128 feature long vector containing information about the gradient angles in the neighborhood of the interest points. The experiments in Fig. 6 show superior performance of our differential invariant features over the SIFT features. The difference becomes even more evident if only stable top-points are used.

In a different set of experiments we have tested the performance of both algorithms under perspective change. For small perspective changes our algorithm performs slightly better than the SIFT algorithm. However this performance rapidly decreases for larger perspective changes. The SIFT features outperform our features in this case. This is probably due to the higher order information used in our feature vector which is more affected by perspective or affine changes than the first order information used in the SIFT feature vector.
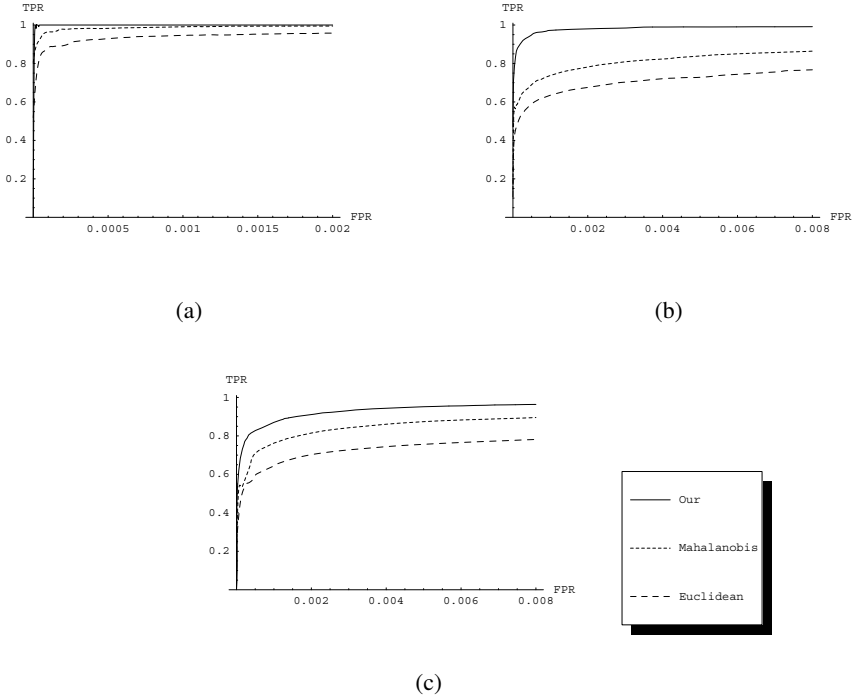
(a)

(b)

(c)

Fig. 5. a. mean ROC curve for 50% scaling. **b.** mean ROC curve for 5% additive Gaussian noise. **c.** mean ROC curve for 45 degree rotation.
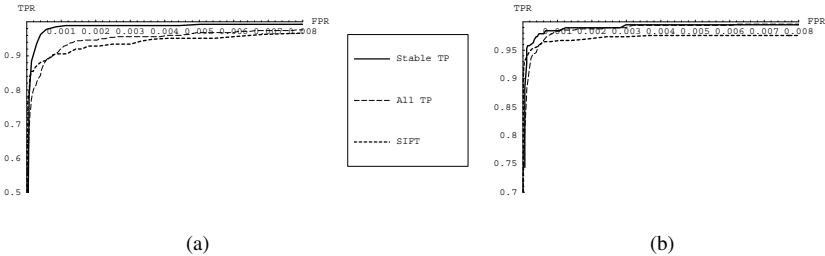

(a)

(b)

Fig. 6. a. ROC curve for 45 degree rotation. **b.** ROC curve for 5% additive Gaussian noise.

4   Retrieval Example

A simple example of an object retrieval task is demonstrated here. We have a set of magazine covers (of size 500×300 pixels) and a scene (of size 1000×700 pixels) containing a number of the magazines, distributed, rotated, scaled, and occluded.

The task is to retrieve a magazine from the scene image. For the query images we find approximately 1000 stable top-points per query image (which may be pre-computed

off-line). For the scene image we find approximately 5000 stable top-points. In Fig. 8 the interest points are shown (above a certain scale), 782 points are matched correctly for the left image and 211 for the down-scaled right image. The objects can now easily be extracted from the scene by using a clustering algorithm as described in [3].
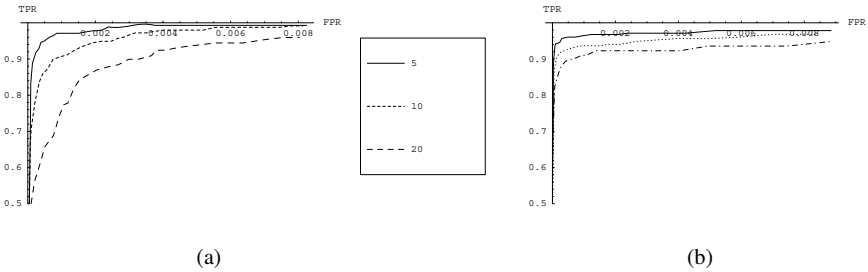

(a)   (b)

Fig. 7. ROC curves perspective change for 5, 10 and 20 degrees for: **a.** Our interest points and differential invariants **b.** SIFT interest points and features.
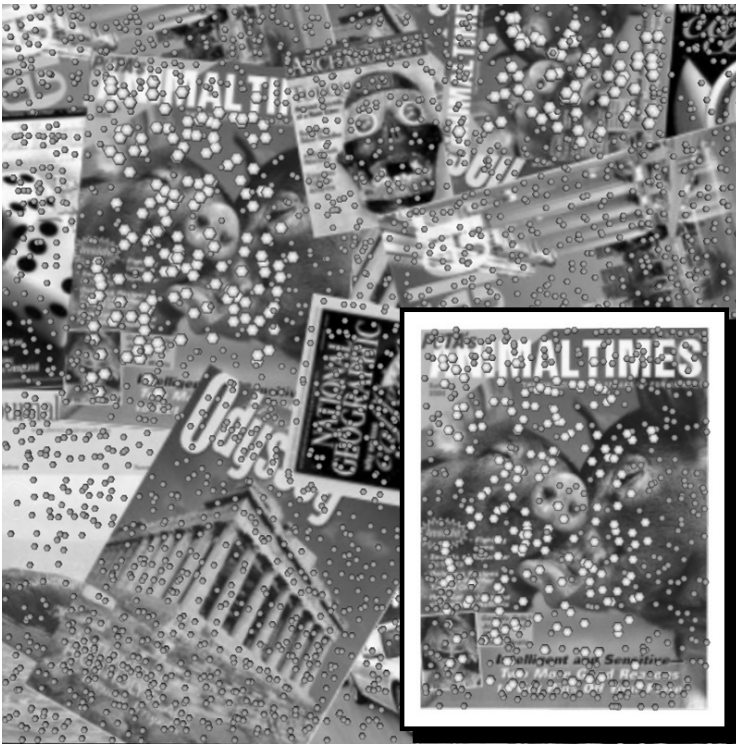

Fig. 8. Matching interest points (white) of a query object and a scene containing two rotated, scaled and occluded versions of the object. Interest points that do not match are shown in gray

5   Summary and Conclusions

We have introduced top-points as highly invariant interest points that are suitable for image matching. Top-points are versatile as they can be calculated for every generic function of the image.

We have pointed out that top-points are invariant under scale-Euclidean transformations as well as under gray value scaling and offset. The sensitivity of top-points to additive noise can be predicted analytically, which is useful when matching noisy images. Top-point localization does not have to be very accurate, since it is possible to refine its position using local differential image structure. This enables fast detection, without losing the exact location of the top-point.

The repeatability of the top-points has proven to be better than the widely used SIFT interest points in a set of experiments. In the future we strive to compare our top-points to other popular interest points like the Harris-Laplace points and descriptors like PCA-SIFT and GLOH.

As features for our interest points we use a feature vector consisting of only six normalized and scaled differential invariants. We have also introduced a similarity measure based on the noise behavior of our feature vectors. Thresholding on this similarity measure increases the performance significantly.

A similarity measure was derived based on the noise behavior of the differential invariant features. This measure significantly increases performance over the popular Mahalanobis and Euclidean distance measures.

For scale-Euclidean transformations as well as additive Gaussian noise our algorithm (6 features in vector) has proven to outperform the SIFT (128 features in vector) approach. However for large perspective changes the SIFT algorithm performs better probably due to the lower order derivatives used for the feature vector.

References

1. Harris, C., Stephens, M.: A combined corner and edge detector. In: Proc. 4th Alvey Vision Conf. (1988) 189–192
2. Lindeberg, T.: Scale-space theory: A basic tool for analysing structures at different scales. J. of Applied Statistics **21(2)** (1994) 224–270
3. Lowe, D.: Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vision **60**(2) (2004) 91–110
4. Mikolajczyk, K., Schmid, C.: Scale and affine invariant interest point detectors. International Journal of Computer Vision **60**(1) (2004) 63–86
5. Schmid, C., Mohr, R., Bauckhage, C.: Evaluation of interest point detectors. Int. J. Comput. Vision **37**(2) (2000) 151–172
6. Mikolajczyk, K., Schmid, C.: A performance evaluation of local descriptors. IEEE Transactions on Pattern Analysis & Machine Intelligence **27**(10) (2005) 1615–1630
7. Florack, L.M.J., Haar Romeny, B.M.t., Koenderink, J.J., Viergever, M.A.: Scale and the differential structure of images. Image and Vision Computing **10**(6) (1992) 376–388
8. Florack, L.M.J., Haar Romeny, B.M.t., Koenderink, J.J., Viergever, M.A.: Cartesian differential invariants in scale-space. Journal of Mathematical Imaging and Vision **3**(4) (1993) 327–348
9. Koenderink, J.J.: The structure of images. Biological Cybernetics **50** (1984) 363–370

10. Damon, J.: Local Morse theory for solutions to the heat equation and Gaussian blurring. Journal of Differential Equations **115**(2) (1995) 368–401
11. Florack, L., Kuijper, A.: The topological structure of scale-space images. Journal of Mathematical Imaging and Vision **12**(1) (2000) 65–79
12. Johansen, P., Skelboe, S., Grue, K., Andersen, J.D.: Representing signals by their top points in scale-space. In: Proceedings of the 8th International Conference on Pattern Recognition (Paris, France, October 1986), IEEE Computer Society Press (1986) 215–217
13. Blom, J., Haar Romeny, B.M.t., Bel, A., Koenderink, J.J.: Spatial derivatives and the propagation of noise in Gaussian scale-space. Journal of Visual Communication and Image Representation **4**(1) (1993) 1–13
14. Balmachnova, E., Florack, L., Platel, B., Kanters, F., Haar Romeny, B.M.t.: Stability of top-points in scale space. (In: Proceedings of the 5th International Conference on Scale Space Methods in Computer Vision (Germany, April 2005)) 62–72
15. Platel, B., Fatih Demirci, M., Shokoufandeh, A., Florack, L., Kanters, F., Dickinson, S.: Discrete representation of top points via scale space tessellation. (In: Proceedings of the 5th International Conference on Scale Space Methods in Computer Vision (Germany, April 2005))