J Math Imaging Vis (2008) 31: 121–132
DOI 10.1007/s10851-008-0079-0

Novel Similarity Measures for Differential Invariant Descriptors for Generic Object Retrieval

E. Balmashnova · L.M.J. Florack

Published online: 12 April 2008

Abstract Local feature matching is an essential component of many image and object retrieval algorithms. Euclidean and Mahalanobis distances are mostly used in order to quantify the similarity of two stipulated feature vectors. The Euclidean distance is inappropriate in the typical case where the components of the feature vector are incommensurable entities, and indeed yields unsatisfactory results in practice. The Mahalanobis distance performs better, but is less generic in the sense that it requires specific training data.

In this paper we consider two alternative ways to construct generic distance measures for image and object retrieval, which do not suffer from any of these shortcomings. The first approach aims at obtaining a (image independent) covariance matrix for a Mahalonobis-like distance function without explicit training, and is applicable to feature vectors consisting of partial image derivatives. In the second approach a stability based similarity measure (SBSM) is introduced for feature vectors that are composed of arbitrary algebraic combinations of image derivatives, and likewise requires no explicit training. The strength and novelty of SBSM lies in the fact that the associated covariance matrix exploits local image structure. A performance analysis shows that feature matching based on SBSM outperforms algorithms based on Euclidean and Mahalanobis distances.

Keywords Image retrieval · Image matching · Distance · Differential invariants · Interest point · Stability · Scale-space

E. Balmashnova (✉) · L.M.J. Florack
Technical University Eindhoven, Den Dolech, 2, Eindhoven
5600 MB, Netherlands
e-mail: e.balmashnova@tue.nl

1 Introduction

Local descriptors are widely used in object-recognition and image retrieval due to their robustness under occlusion, certain types of image transformations (translation, rotation, zooming, up to some extent view point changes), and due to their discriminative power. In order to avoid a large number of parameters, required by most of the popular descriptors, we concentrate on differential invariant descriptors [7–9, 12, 31]. These have additional conceptual advantages, such as mathematical simplicity, and the possibility to construct complete systems in a precise sense, i.e. sets of differential invariants that provably capture all differential structure up to some predefined order given some invariance group, cf. Olver [28] and Florack [6].

In this paper we concentrate on what is perhaps the most essential ingredient in any local feature based image and object retrieval algorithm, viz. the construction of an effective (pseudo-)distance measure to quantify the similarity between two feature vectors. The effectiveness of a stipulated feature set cannot be assessed without the consideration of an associated similarity measure.

Since we consider sparse sets of local features, we first need to define the interest points (in space and scale) that serve as anchor points for these features. A wide range of interest points has been proposed, such as Harris points [16], Harris-Laplace regions [26], Hessian-Laplace regions [26], DoG [24], Top-Points [29], etc. The second step is to build a descriptor that characterizes (the immediate neighborhood of) each interest point, which should be discriminative and invariant to certain image transformations. This multi-component local descriptor is what we call the feature vector. We will require (at least) invariance under scale-Euclidean transformations, so that the proper setting for interest points and feature vectors will be a scale space representation of the image, see e.g. Koenderink [17], Witkin [33], Lindeberg [21, 22], and Florack [6].

There are numerous ways to compute descriptors based on local image properties, such as pixel intensities, edges, textures, etc. The most straightforward one, a vector of neighboring image pixel values suffers from high computational complexity and low robustness to common intensity changes. One group of approaches uses histograms to represent some local characteristics of shape and appearance. A straightforward one is the histogram of neighboring pixel values. A more sophisticated descriptive histogram, based on so-called intensity-domain spin images, was proposed by Lazebnik et al. [20]. The authors consider a two-dimensional histogram encoding the distribution of image brightness values, where the two dimensions of the histogram are the distance from the center point and the intensity value. Similar local histogram based approaches have been proposed to capture shape characteristics, more precisely the edge distribution in the immediate neighborhood of a pixel, such as the so-called geometric histogram proposed by Ashbrook et al. [1], and the shape context by Belongie et al. [3].

Lowe [24] proposed the so-called scale invariant feature transform (SIFT), a local descriptor also represented by a histogram, in this case encoding the contextual gradient distribution. There are several more or less successful modifications: rotation-invariant feature transform, or RIFT, by Lazebnik et al. [20], PCA-SIFT by Yan and Sukthankar [34] (which takes advantage of Principal Component Analysis), Speeded Up Robust Features, or SURF, by Bay et al. [2], all improving performance over the original approach in specific cases. This group of local and differential descriptors (which will be described in more detail below), has been evaluated by Mikolajczyk and Schmid [27].

Another major group of techniques is based on spatial-frequency properties, such as Laws' filter mask [19], discrete cosine transform (DCT), wavelet transform, cf. Unser [32], and Gabor filters [11]. These descriptors are widely used in texture classification and their evaluation is done by Grigorescu et al. [14], and by Randen and Husoy [30].

In our research we focus on the feature vectors based on derivatives computed at an interest point, presented in Sect. 2.

In Sect. 4 we address distance. We show that in one particular case (using a specific set of differential invariants), similar performance as for the Mahalanobis distance can be reached without any training (Sect. 4.1). A more general approach to compute a distance measure is applicable to any feature vector constructed from Gaussian derivatives taken at the interest point [18], which shows improvement as compared to Mahalanobis and Euclidean distances used in evaluations done by Mikolajczyk and Schmid [27]. This is the subject of Sect. 4.2. The so-called *stability based similarity measure* (SBSM) proposed in this section is based on the analysis of local structure at the interest point, and therefore uses a more appropriate covariance matrix than in case of the globally defined Mahalanobis distance. The symmetry property intrinsic to a genuine distance function is lost (although this could easily be repaired by symmetrization, with a computational prize), but this does not affect the matching results. In fact, symmetry is not required conceptually, as we have an asymmetry in the role of matched pairs; the query object is considered a "ground truth" object, whereas the object to be retrieved is allowed to exhibit some variability relative to this. Despite asymmetry of our measure, we will adhere to the terminology of a "distance" for the sake of simplicity.

The experimental results are presented in Sect. 5, using the database and validation criterion discussed in Sect. 3.

We end with summary and conclusions in Sect. 6.

2 Differential Invariant Descriptors

Local image structure can be captured by the so-called *local jet* [17], roughly speaking the set of image derivatives computed up to some order. For brevity we indicate the various image derivatives by u_k, $k = 1, \ldots, n$, so that e.g. up to second order we have $n = 5$ and $u_1 = u_x, u_2 = u_y, u_3 = u_{xx}, u_4 = u_{xy}, u_5 = u_{yy}$. Differential feature vectors can then be expressed as functions on the local jet $\{u_1, \ldots, u_n\}$, with n determined by the jet's order and the dimension of space, as follows:

$$d_i = d_i(u_1, \ldots, u_n), \quad i = 1 \ldots m. \tag{1}$$

So each feature vector has m components, and, in principle, each component depends on all n derivatives up to the prescribed order. We consider several ways to build invariants as functions of this type, and give their interpretations.

2.1 Cartesian Invariants

Once the local jet has been calculated, the differential information up to Nth order (say) is available. However, the jet's components expressed relative to a coordinate system are not invariant to mere changes of coordinates. One of the ways to avoid this problem is to choose one particular, geometrically meaningful coordinate system, and to compute the local jet components (which are then invariants by construction) as partial image derivatives relative this system. Blom [4] and Florack et al. [8] proposed to use a *gauge coordinate system* as a right handed local frame, in which one axis (ordinate component: w) points in the same direction as the local image gradient, and the other axis (abscissa component: v) is tangential to the isophote.

The N-jet at a given base point has a finite number of independent degrees of freedom.

In 2D, e.g., the 3-jet generically consist of 9 Cartesian invariants (intensity at the point is considered a 0-th order differential invariant, so in the (v,w)-gauge system we have $u, u_w, u_{vv}, u_{vw}, u_{ww}, u_{vvv}, u_{vvw}, u_{vww}, u_{www}$), the 4-jet captures 14 invariants (viz. the foregoing plus $u_{vvvv}, u_{vvvw}, u_{vvww}, u_{vwww}, u_{wwww}$), and so forth. Note that (in 2D) this number of Cartesian invariants is always 1 less than the number of independent partial derivatives in an arbitrary coordinate system, since $u_v = 0$ identically, yet u_x and u_y are generically independent. However, the selection of interest points may, and will by definition, reduce this number further by virtue of defining constraints among these invariants, as we will see.

We construct the differential feature vectors in such a way that they are invariant to certain transformations, notably translation and rotation (the "trivial" prerequisite for Cartesian coordinate invariance already considered above), as well as zooming and linear intensity changes. In the experimental part we consider sets of differential invariants, evaluated at a top-point [29] of the image Laplacian Δu (at some implicit scale and position). For top-points of the Laplacian image the following set of equations holds:

$$\begin{cases} \partial_x \Delta u = u_{xxx} + u_{xyy} = 0, \\ \partial_y \Delta u = u_{xxy} + u_{yyy} = 0, \\ \det \mathbf{H}(\Delta u) = (u_{xxxx} + u_{xxyy})(u_{xxyy} + u_{yyyy}) \\ \qquad\qquad - (u_{xxxy} + u_{xyyy})^2 = 0. \end{cases} \tag{2}$$

Note that these identities are Cartesian invariant, and therefore also hold in the (v,w)-gauge, obtained by formal replacement $(x,y) \to (v,w)$. This demonstrates the dependencies among local jet components alluded to previously. More precisely, the local 3-jet is reduced by 2, and the local 4-jet by 3 degrees of freedom if anchored at these top-points. A further reduction by 1 degree of freedom is obtained by normalizing the local jets in such a way that invariance under grey-scale scalings by a constant factor is realized. Thus in particular we are left with $6 = 9 - 2 - 1$ independent invariants for the local 3-jet if we consider only Laplacian top-points and insist on scale-Euclidean, linear grey-scale invariance.

We collect the non-trivial, scaled and normalized scale-Euclidean, linear grey-scale invariant differential invariants up to third order into a feature vector. One possible representation is given by (3), using Einstein's summation convention:[1]

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

Here ε is the *Levi-Civita* or *permutation tensor* defined in d spatial dimensions as follows:

$$\varepsilon_{i_1 i_2 \dots i_d} = \begin{cases} 0, & \text{if any two labels are the same,} \\ 1, & \text{if } i_1, i_2, \dots, i_d \text{ is an even} \\ & \text{permutation of } 1, \dots, d, \\ -1, & \text{if } i_1, i_2, \dots, i_d \text{ is an odd} \\ & \text{permutation of } 1, \dots, d. \end{cases} \tag{4}$$

So $\varepsilon_{11} = \varepsilon_{22} = 0, \varepsilon_{12} = -\varepsilon_{21} = 1$. This set is complete, in the sense that there exists no other third order invariant (at a Laplacian top-point!) that is independent of the entries of (3). (One may alternatively consider the set of partial derivatives in the (v,w)-gauge taking into account (2) and linear grey-scale invariant normalization.)

2.2 General Grey-Scale Invariants

The invariants proposed by Florack et al. [9] are based on the local isophote structure of an image. Isophotes are curves (in 2D) of constant grey-value u in an image, and their shape is invariant under the group of invertible intensity transformations, $u \mapsto \gamma(u)$ with $\gamma' \neq 0$.

By construction of the (v,w)-frame, the isophote can be locally represented by the implicit function $w(v)$ such that $u(v,w(v))$ is constant, whence there exists an open neighborhood Ω of the interest point, the origin $(v,w) = (0,0)$ say, such that

$$\frac{\mathrm{d}^n}{\mathrm{d}v^n} u(v,w(v)) = 0 \quad \text{for all } (v,w) \in \Omega, \tag{5}$$

for all orders $n \geq 1$. Differential isophote structure can now be captured by the local Taylor coefficients $w'(0) \equiv 0$, $w''(0), w'''(0), w''''(0), \dots$ of the isophote function, which can be solved from (5). By construction these are invariant under general (invertible) grey-scale transformations. The gauge condition, $w'(0) \equiv 0$, implies $u_v(0,0) = 0$, i.e. to first order grey-scale does not vary in isophote-tangent direction. Solving (5) order by order produces the following

[1] That is, a sum over a spatial index of the type $\sum_{i=1}^{d} X_{ii}$ is condensed into X_{ii}.

system of general grey-scale invariants up to fourth order:

$$
\begin{aligned}
w'(0) &\equiv 0 \quad \text{(definition of } (v,w)\text{-gauge)},\\
w''(0) &= -\frac{u_{vv}}{u_w},\\
w'''(0) &= 3\frac{u_{vv}u_{vw}}{u_w^2} - \frac{u_{vvv}}{u_w},\\
w''''(0) &= -3\frac{u_{vv}(4u_{vw}^2 + u_{vv}u_{ww})}{u_w^3}\\
&\quad + \frac{6u_{vv}u_{vvw} + 4u_{vvv}u_{vw}}{u_w^2} - \frac{u_{vvvv}}{u_w}.
\end{aligned}
\tag{6}
$$

(Extension beyond fourth order is straightforward, but won't be needed. It is tacitly understood that the image derivatives on the right hand side are evaluated at the base point of interest, $(v, w) = (0, 0)$. Gauge derivatives can be transformed back to an arbitrary Cartesian coordinate system by a gradient dependent rotation, cf. [9] for details.) The invariant $w''(0)$ is the well-known isophote curvature.

A second set of general grey-scale invariants can be obtained by considering flow lines (gradient integral lines, orthogonal to the isophotes), which are also invariant to invertible grey-scale transformations. Both sets are mutually dependent, for the flow lines are completely fixed by the isophotes, vice versa. However, for *fixed* differential order they do provide independent invariants.

Let the flow line be parameterized by

$$
\mathbf{v}(\lambda) = \begin{pmatrix} v(\lambda)\\ w(\lambda) \end{pmatrix} \quad (\lambda \in \mathbb{R}),
\tag{7}
$$

such that $\mathbf{v}(0)$ coincides with the origin, i.e. our base point of interest. The unit tangent vector of a flow line is, by definition, aligned with the gradient (a dot indicates differentiation w.r.t. λ; we suppress the parameter λ henceforth in the notation):

$$
\dot{\mathbf{v}} = \begin{pmatrix} \dot v\\ \dot w \end{pmatrix} = \frac{1}{\sqrt{u_v^2 + u_w^2}} \begin{pmatrix} u_v\\ u_w \end{pmatrix}.
\tag{8}
$$

Note that $\|\dot{\mathbf{v}}\| = 1$ is a trivial invariant, unlike higher order derivatives, which can be computed by using (8):

$$
\|\ddot{\mathbf{v}}\| = {u_w}^{-1}\,|u_{vw}|\,,
\tag{9}
$$

$$
\begin{aligned}
\|\dddot{\mathbf{v}}\| &= u_w^{-2}\big(u_{vw}^4 + (u_w u_{vww}\\
&\quad + u_{vw}(u_{vv} - 2u_{ww}))^2\big)^{\frac{1}{2}},
\end{aligned}
\tag{10}
$$

$$
\begin{aligned}
\|\ddddot{\mathbf{v}}\| &= u_w^{-3}\big(9u_{vw}^2(u_w u_{vww} + u_{vw}(u_{vv} - 2u_{ww}))^2\\
&\quad + (u_{vw}(6u_{ww}^2 + u_{vv}^2) - 7u_{vw}^3 + u_w^2 u_{vwww}\\
&\quad - 5u_{ww}u_{vw}u_{vv} + u_w(u_{vww}(u_{vv} - 3u_{ww})\\
&\quad + 3u_{vw}(u_{vvw} - u_{www})))^2\big)^{\frac{1}{2}},
\end{aligned}
\tag{11}
$$

and so forth. The second order invariant $\|\ddot{\mathbf{v}}\|$ is the flow-line curvature. Note that the third order derivative of the isophote, recall (6), is expressed in terms of both isophote and flow-line curvatures, as well as one third-order grey-value invariant. This illustrates the dependencies that generally exist between the isophote and flow-line induced systems of differential invariants.

As a third order feature vector we may choose (again, the base point of interest, corresponding to $\lambda = 0$, is implicit in the notation henceforth)

$$
(\sigma w'', \sigma^2 w''', \sigma^2\|\ddot{\mathbf{v}}\|^2, \sigma^3\|\dddot{\mathbf{v}}\|^2),
$$

and as a fourth order feature vector we may choose

$$
(\sigma w'', \sigma^2 w''', \sigma^3 w'''', \sigma^2\|\ddot{\mathbf{v}}\|^2, \sigma^3\|\dddot{\mathbf{v}}\|^2, \sigma^4\|\ddddot{\mathbf{v}}\|^2).
$$

In the terminology of (1), these feature vectors have lengths $m = 4$ and $m = 6$, respectively.

2.3 Steerable Filters

We showed how to construct rotation invariant feature vectors using gauge coordinates. Another approach was proposed by Freeman and Adelson [10] in terms of steerable filters.

The nth order directional derivative in the direction indicated by the angle θ with respect to the x-axis is given by

$$
u^{(n)}(\theta) = \partial_\theta^n u = (\cos\theta\,\partial_x + \sin\theta\,\partial_y)^n u.
\tag{12}
$$

To obtain a set of nth order, we compute $(n + 1)$ directional derivatives oriented in the directions given by the angles $\theta_{n,i}$, in which $i = 0, \dots, n$ labels the $n + 1$ directions. The directions are

$$
\theta_{n,i} = i\pi/(n + 1) + \theta_g,
\tag{13}
$$

where θ_g is any fixed orientation at the point (in our case, the gradient direction). Invariance to linear intensity changes is obtained by dividing the higher order derivatives by the gradient magnitude, i.e. $\|\nabla u\| = u'(\theta_g)$. The nth order feature vector is

$$
\left(\frac{\sigma u''(\theta_{2,0})}{u'(\theta_g)}, \frac{\sigma u''(\theta_{2,1})}{u'(\theta_g)}, \frac{\sigma u''(\theta_{2,2})}{u'(\theta_g)}, \dots, \frac{\sigma^{n-1} u^{(n)}(\theta_{n,n})}{u'(\theta_g)}\right).
\tag{14}
$$

In case of Laplacian top-points third order features are linearly dependent, therefore two of them should be dropped, recall a previous argument. Again, scale factors have been incorporated to ensure spatial scale invariance.

3 Validation

In foregoing sections we have proposed different ways to construct complete systems of differential invariants (up to some predefined differential order) given an invariance group. By virtue of completeness the specificities of a given system (within the context of a stipulated invariance group) are not important (complete systems can be represented in many equivalent ways and all capture the same information), but combined with a distance concept the choice of a particular one does become relevant. Apart from this, imposing different invariance groups will obviously affect the potential power of a set of invariants for a particular retrieval task. In this section we therefore propose a criterion for evaluating the various systems of differential invariants under various distance measures. The actual validation is postponed, and is carried out in the respective subsections of Sect. 4, after we have introduced the various distance measures.

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 (Fig. 1). 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 [26, 27] is, although more realistic, not suitable for our validation 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 Evaluation Criterion

We use a criterion proposed by Yan and Sukthankar [34]. It is based on the number of correct matches and the number of false matches obtained for an image pair. With a "match" we generally indicate an established coupling between a query and a scene object, which may or may not be correct, i.e. the term as such is used here without the implicit connotation of being correct. For the sake of definiteness we therefore also refer to matched pairs in general as *possible matches*, to be distinguished from the disjoint subsets of *correct matches*, respectively *false matches* (the latter two according to some available ground truth):

$$\begin{aligned}\#\text{possible matches} &= \#\text{correct matches}\\ &\quad + \#\text{false matches}.\end{aligned}\tag{15}$$

The operational criterion for calling a match correct will be discussed below. A false match is a possible match that is not correct.

We couple interest points, i.e. we establish a (possible) match, if the distance between their feature vectors is below a certain threshold d. Note that since we know the transformations we also know the ground truth for the matches. Each feature vector from the reference image is compared to each vector from the transformed image, and the number of correct matches as well as the number of false matches is counted. The threshold d is varied to obtain curves as detailed in the next section. The results are presented with *recall* versus 1 − precision. Recall is the number of correctly matched points relative to the number of ground truth *correspondences* between two images of the same scene. A correspondence refers to what we know from ground truth, and indicates a pairing of an object and a scene point that has either been found as a correct match, or should ideally have been found but has been overlooked as such.

So,

$$\text{recall} = \frac{\#\text{correct matches}}{\#\text{correspondences}}.\tag{16}$$

Note also that

$$\text{recall} = \frac{\#\text{possible matches} - \#\text{false matches}}{\#\text{correspondences}}.\tag{17}$$

The number of false matches relative to the number of possible matches is expressed, by definition, by 1 − precision:

$$1 - \text{precision} = \frac{\#\text{false matches}}{\#\text{possible matches}}.\tag{18}$$

In other words, using the previous definitions, (15) and (18):

$$\text{precision} = \frac{\#\text{correct matches}}{\#\text{possible matches}}.\tag{19}$$

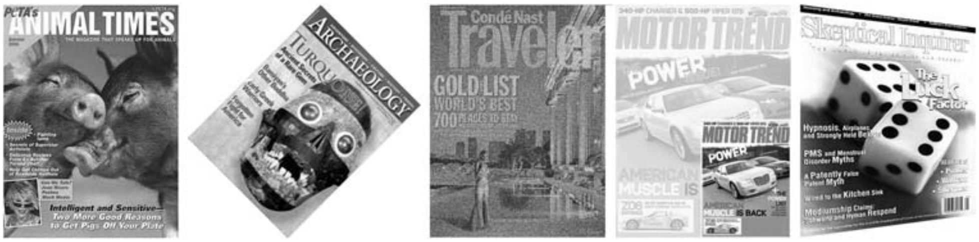

Fig. 1 A selection of data set images. From *left to right*: unchanged, rotated, added noise, scaled, changed perspective

(Note that recall and precision are independent quantities.) The number of correct matches and correspondences is obtained with an *overlap error*, ε, so as to allow for some tolerance in the true position of the target point and the one obtained by transformation of the query point, for these (subpixel) positions will in practice hardly ever be exactly equal. The overlap error measures how well the points correspond under a transformation H. It is defined by the ratio of the intersection and union of two disks S_1 and S_2 with centers in the interest points, x_1 and x_2, and radii given by the scales of the points, σ_1 and σ_2,

$$\varepsilon = 1 - \frac{S_2 \cap HS_1}{S_2 \cup HS_1}, \tag{20}$$

where $HS_1 = \{Hx | x \in S_1\}$.

In case of transformations close to scale-Euclidean ones, HS_1 can be approximated by a disk, and areas of intersection and union can be computed analytically. We call a match correct if the error ε in the image area covered by two corresponding regions is less than 50% of the region union. The number of correspondences in order to compute recall in (16) is determined with the same criterion.

A perfect descriptor gives a recall equal to 1 for any precision. In practice, due to noise and transformations, the distance between two descriptors is almost never exactly zero, so that the recall starts from some low value and increases with increasing threshold. Horizontal curves indicate that the recall is attained with a high precision and is limited by the specificity of the scene. A slowly increasing curve shows that the descriptor is more sensitive to image degradation. If curves corresponding to different descriptors are far apart and have different slopes, then the discriminative power and robustness of the descriptors is different for a given image transformation or scene type.

4 Distance

All definitions for the evaluation criteria in the previous section require a quantitative descriptor for the similarity between two features, and this "distance" measure will crucially affect evaluation results. Therefore we focus on an operational distance concept in this section. We begin by a discussion of some common distance measures and list a number of deficiencies. Subsequently we make an attempt to overcome these by introducing novel measures, and finally we subject these to a performance evaluation.

The space of features is a vector space, but it is not obvious how to introduce a norm because of the incommensurability of the components. Similarity between descriptors is usually computed with either the Euclidean or the Mahalanobis distance measure. The Euclidean distance,

$$\rho_{\text{Euclidean}}(\mathbf{d}^{(1)}, \mathbf{d}^{(2)})^2 = \big(\mathbf{d}^{(1)} - \mathbf{d}^{(2)}\big)^T (\mathbf{d}^{(1)} - \mathbf{d}^{(2)}), \tag{21}$$

makes little sense in view of the heterogenic nature of feature vector components. In particular it does not take into account the fact that the components of (1) may be correlated, nor that they are entities of possibly different dimensionalities, as is e.g. the case with (3). Indeed, the naive Euclidean distance performs very poorly in practice, as we will illustrate in Sect. 5.

A sensible similarity measure should take the correlation of features into account. Recall that the Mahalanobis distance has been introduced precisely in order to achieve this:

$$\begin{aligned}\rho_{\text{Mahalanobis}}(\mathbf{d}^{(1)}, \mathbf{d}^{(2)})^2 &= \big(\mathbf{d}^{(1)} - \mathbf{d}^{(2)}\big)^T \\ &\quad \times \mathbf{C}^{-1}(\mathbf{d}^{(1)} - \mathbf{d}^{(2)}).\end{aligned} \tag{22}$$

The covariance matrix $\mathbf{C}$ is obtained from training data. In principle, it is always possible to reduce computational time by transforming the Mahalanobis distance into a Euclidean distance via a suitably chosen system of Cartesian coordinates:

$$\begin{aligned}&\mathbf{C}^{-1} = \mathbf{R}^T\mathbf{D}^{-1}\mathbf{R},\\ &\mathbf{d}_{\text{norm}} = \mathbf{D}^{-1/2}\mathbf{R}\mathbf{d},\\ &\rho_{\text{Mahalanobis}}(\mathbf{d}^{(1)}, \mathbf{d}^{(2)}) = \rho_{\text{Euclidean}}(\mathbf{d}^{(1)}_{\text{norm}}, \mathbf{d}^{(2)}_{\text{norm}}),\end{aligned} \tag{23}$$

where $\mathbf{D}$ is a diagonal matrix and $\mathbf{R}$ is an orthogonal matrix.

The Mahalanobis distance gives better matching results, but has three disadvantages, viz.

- need for supervised initialization: the Mahalanobis distance requires a covariance matrix $\mathbf{C}$ to be estimated from training data;
- lack of genericity: the matrix $\mathbf{C}$, and therefore performance results will depend on the training set used;
- non-locality: the matrix $\mathbf{C}$, and consequently the Mahalanobis distance, is an image independent entity and therefore not optimally adapted to the local structure at any feature point of interest.

We propose two ways to overcome these deficiencies, one exploiting the local Taylor expansions at the base points of the features (Sect. 4.1), and one which is akin to the conventional Mahalanobis distance but obviates training, and likewise takes into account the local structure at the feature points of interest (Sect. 4.2). In this case the covariance matrix is obtained directly from the differential structure at each interest point. The matrix can be obtained in analytical form and reflects the actual behavior of the descriptor due to small perturbations. One could say that training by explicit examples is replaced by a *Gedanken* experiment in which the effect on the feature components of all hypothetical, local, additive noise perturbations with fixed variance is taken into account in the realization ("analytical training") of a local covariance matrix. In the next sections we present the details of these two approaches.

4.1 Taylor Expansion

We follow and slightly adapt an approach previously proposed by Griffin [13] and Loog [23] to compare neighborhoods of two interest points. Let us represent the structure of the image in a neighborhood of a point by the Taylor expansion in a gauge coordinate system using a Gaussian window:

$$f(0,0)^{-1}f(x,y)=1+\sum_{n=1}^{\infty}\sum_{i=0}^{n}\frac{1}{i!(n-i)!}f_{n,i}x^{i}y^{n-i}, \tag{24}$$

in which

$$f_{n,i}=f(0,0)^{-1}\frac{\partial^{n}f(0,0)}{\partial x^{i}\partial y^{n-i}}, \tag{25}$$

x, y are gauge coordinates in the neighbourhood of the origin, i.e. the point of interest. In order to make it scale invariant, we scale the local coordinate system at a feature point by its scale coordinate in scale space:

$$(x',y')=\sigma^{-1}(x,y).$$

The distance between two feature point neighbourhoods can now be defined as the $\mathbb{L}_2$-norm of the difference of the two corresponding Taylor expansions at the respective points, windowed by a Gaussian aperture. If the scales corresponding to the feature points' scale coordinates are given by σ_1 and σ_2, respectively, and the images are denoted by f and g, with scaled local Taylor polynomials $f_1^{(N)}$ and $g_2^{(N)}$, then (omitting primes on dummy variables)

$$\begin{aligned}I_N=\iint\left(\sum_{n=1}^{N}\sum_{i=0}^{n}\frac{1}{i!(n-i)!}(\sigma_1^{n}f_{n,i}-\sigma_2^{n}g_{n,i})x^{i}y^{n-i}\right)^{2}\\ \times\Phi(x,y)\,dxdy\end{aligned} \tag{26}$$

with

$$\Phi(x,y)=\exp\bigl(-(x^2+y^2)\bigr). \tag{27}$$

The introduced distance is invariant to rotation, zooming and grey-value scaling, for if we denote the right hand side of (26) by

$$I_N=\|f_1^{(N)}-g_2^{(N)}\|_{\Phi}^{2}, \tag{28}$$

then, by change of variables, respectively isotropy of Φ,

$$\begin{aligned}I_N^{R}&\equiv\|f_1^{(N)}\circ\mathbf{R}^{-1}-g_2^{(N)}\circ\mathbf{R}^{-1}\|_{\Phi}^{2}\\&=\|f_1^{(N)}-g_2^{(N)}\|_{\Phi\circ\mathbf{R}}^{2}\\&=\|f_1^{(N)}-g_2^{(N)}\|_{\Phi}^{2}=I_N,\end{aligned} \tag{29}$$

proving rotational invariance. Scale invariance follows in a similar fashion (note that scaling spatial variables affects the scale factors $\sigma_{1,2}$ proportionally), whereas grey-value invariance is manifest by dividing out the features' zeroth order Taylor coefficients in (25).

It is possible to compute the distance, (26), analytically, and to recast it in a form akin to the conventional Mahalanobis distance:

$$\begin{aligned}I_N=\sum_{i_1=0}^{N}\sum_{i_2=0}^{N}\sum_{j_1=0}^{N-i_1}\sum_{j_2=0}^{N-i_2}(\sigma_1^{i_1+j_1}f_{i_1+j_1,i_1}-\sigma_2^{i_1+j_1}g_{i_1+j_1,i_1})\\ \times C_{i_1j_1i_2j_2}(\sigma_1^{i_2+j_2}f_{i_2+j_2,i_2}-\sigma_2^{i_2+j_2}g_{i_2+j_2,i_2}).\end{aligned} \tag{30}$$

Straightforward computation yields

$$C_{i_1j_1i_2j_2}=\begin{cases}\frac{2\pi(j_1+j_2-1)!!(i_1+i_2-1)!!}{i_1!j_1!i_2!j_2!}\\ \quad\text{if } i_1+i_2\in2\mathbb{N}\wedge j_1+j_2\in2\mathbb{N},\\ 0\quad\text{otherwise.}\end{cases} \tag{31}$$

Therefore comparing two neighborhoods by their windowed Taylor series turns out to be equivalent to comparing two feature vectors consisting of gauge derivatives scaled by intensity at the center using a modified Mahalonobis distance. But unlike the original Mahalonobis distance it does not require training data, and is therefore of a more generic nature.

The performance of differential descriptors consisting of gauge coordinates is compared for three different distance measures, Euclidean, the conventional, trained Mahalanobis distance, and the modified one with the analytical covariance matrix given by (31). The experiments were conducted for different interest points, namely DoG points and top-points, and for different image transformations. For all the cases Euclidean distance performs very poorly, whereas both Mahalanobis distances perform approximately the same, although neither does sufficiently well. The typical results for one of the experiments, in which DoG points have been taken as interest points and the image has been rotated over 45 degrees (worst case scenario), are shown in Fig. 2.

Because neither distance yields satisfactory performance we turn to a novel, alternative distance measure in the next section.

4.2 Stability Based Similarity Measure

The main deficiency that the newly proposed analytically obtained Mahalanobis distance from the previous section has in common with the conventional, experimentally obtained Mahalanobis distance is that the covariance matrix, recall (31), is a global measure, i.e. it does not depend on local image structure at the feature points of interest. It is also less versatile, since it is applicable to only one particular (albeit complete) representation of image structure, viz. the local gauge derivatives up to some order, recall (30).

In this section we construct a second type of generically applicable, analytical Mahalanobis-like distance which

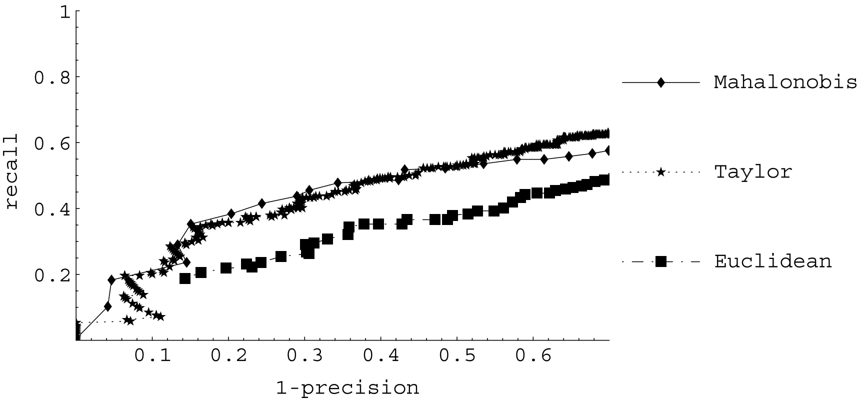

Fig. 2 Evaluation of different distances in case of DoG points and differential invariants for an experiment of matching image pairs under a 45 degree rotation

overcomes these drawbacks, but likewise does not require any training. In this way we arrive at a distance measure that meets all three requirements that were itemized in the beginning of this section. To this end we introduce a stability based similarity measure (SBSM) for feature vectors. In this case the feature vectors may be composed of arbitrary algebraic combinations of image derivatives. Despite the fact that no training is involved in the SBSM, feature matching based on SBSM is shown to outperform algorithms based on Euclidean and the previously studied conventional and analytical Mahalanobis distances.

4.2.1 Feature Vector Perturbation

We use a perturbation approach for the estimation of a covariance matrix for each feature vector. We generically model changes in the image due to rendering artifacts induced by transformations, jpeg-compression effects, and other sources of noise, as a zero-mean additive random image perturbation. The distribution of the random value is assumed to be the same for all pixels,[2] and may be pixel-correlated. The only thing we will ultimately need is the variance of this distribution.

Recall the notational convention for local jet components (partial image derivatives) introduced in (1). Due to linearity of scale-space the perturbed local jet in the point is

$$\{v_1, \dots, v_n\} = \{u_1, \dots, u_n\} + \{n_1, \dots, n_n\}, \tag{32}$$

in which the first term on the right hand side models the unperturbed, and the last term the perturbations of the various local jet components.

Let us rewrite (1) for the unperturbed and perturbed images in condensed form as

$$d_i = d_i(\mathbf{u}), \tag{33}$$

$$\tilde{d}_i = d_i(\mathbf{v}). \tag{34}$$

Noting that according to (32) $\mathbf{v} = \mathbf{u} + \mathbf{n}$, in self-explanatory notation, the difference between the two descriptors, (33–34), can be approximated by a Taylor expansion of (34) around $\mathbf{u}$ up to first order in $\mathbf{n}$:

$$\Delta d_i = \tilde{d}_i - d_i \approx \sum_{k=1}^{n} \left.\frac{\partial d_i}{\partial v_k}\right|_{v_k=u_k} n_k. \tag{35}$$

Therefore, the approximate covariance matrix $\mathbf{\Sigma}$ is given by (note that $\langle n_i \rangle$ vanishes)

$$\begin{aligned}\Sigma_{ij} &= \langle \Delta d_i \Delta d_j \rangle \\ &= \sum_{k=1}^{n}\sum_{l=1}^{n} \left.\frac{\partial d_i}{\partial v_k}\right|_{v_k=u_k} \left.\frac{\partial d_j}{\partial v_l}\right|_{v_l=u_l} \langle n_k n_l \rangle.\end{aligned} \tag{36}$$

The covariance matrix $\mathbf{C} = \langle n_k n_l \rangle_{1 \le k, l \le n}$ of the noise derivatives is given in the following Section. A statistical approach to obtain this matrix is considered by Markussen et al. [25].

4.2.2 Gaussian Correlated Noise

For convenience, instead of linear indexing of the set of derivatives n_i in (32–36) we consider a more explicit double index (n_x, n_y), where n_x and n_y correspond to powers of derivatives with respect to x and y.

The momentum $M^2_{m_x,m_y,n_x,n_y} = \langle n_{m_x,m_y} n_{n_x,n_y} \rangle$ of Gaussian derivatives of orders (m_x, m_y) and (n_x, n_y) of correlated noise in case the spatial noise correlation distance τ (or rather $\sqrt{\tau}$) is much smaller than scale t is given by

$$\begin{aligned}M^2_{m_x,m_y,n_x,n_y} &\simeq \langle n^2 \rangle \left(\frac{\tau}{2t}\right) \left(\frac{-1}{4t}\right)^{\frac{1}{2}(m_x+m_y+n_x+n_y)} \\ &\quad \times Q_{m_x+n_x} Q_{m_y+n_y},\end{aligned} \tag{37}$$

with Q_k given by Table 1. The proportionality constant $\langle n^2 \rangle$ is the variance of the noise function for the zeroth order image. We refer to Blom et al. for a derivation and further

[2] This need not be quite realistic, but it is the order of magnitude of perturbation that concerns us.

Table 1 Some values of Q_n ($Q_n = 0$ if n is odd)

n	0	2	4	6
Q_n	1	1	3	15

details [5]. Equation (37) summarizes the effect of zero-mean additive, pixel-correlated noise on image derivatives in Gaussian scale space, and thus implicitly on any algebraic combination of these. Thus it can be used to express the sensitivity (degree of robustness) of the components of a differential feature vector, which is what we are about to exploit in order to arrive at the SBSM.

Let us take the correlation kernel to be roughly of one pixel width corresponding to $\tau = \delta^2/4$, where δ denotes pixel size. Going back to the linear indexing of the n-jet, for Gaussian derivatives of first and second order we then obtain the following correlation matrix:

$$\mathbf{C} = \langle n_i n_j \rangle_{1 \le i,j \le 5} = \begin{pmatrix} 4t & 0 & 0 & 0 & 0 \\ 0 & 4t & 0 & 0 & 0 \\ 0 & 0 & 3 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 3 \end{pmatrix} \frac{\delta^2 \langle n^2 \rangle}{(4t)^3}, \tag{38}$$

where $(n_1, \dots, n_5) = (n_x, n_y, n_{xx}, n_{xy}, n_{yy})$, and the matrix entries in (38) are labeled accordingly. This correlation matrix together with (36) gives an approximation of the covariance matrix of each local feature vector for given perturbation variance and pixel size.

4.2.3 Similarity Measure

We define the similarity between feature descriptors $\mathbf{d}$ and $\mathbf{d}_0$ in a similar way as for the Mahalanobis distance, except that for every point $\mathbf{d}_0$ we insert its associated covariance matrix, recall (36):

$$\rho_{\text{SBSM}}(\mathbf{d}; \mathbf{d}_0) = (\mathbf{d} - \mathbf{d}_0)^T \, \Sigma_{\mathbf{d}_0}^{-1} (\mathbf{d} - \mathbf{d}_0). \tag{39}$$

Consequently, the function $\rho_{\text{SBSM}}(\mathbf{d}; \mathbf{d}_0)$ is not symmetric, therefore it is not a distance in the strict sense. The reference image $\mathbf{d}_0$ is considered to be the "ground truth". The covariance matrix and, as a consequence, the distance are proportional to the constant $\delta^2 \langle n^2 \rangle$, i.e. the product of noise variance and pixel size. This constant is the same for all points of the reference image and hence does not change the ordering of distances from some fiducial object point to the set of all points of the reference image, whence the constant can be omitted.

In the next section we subject the SBSM to an experimental test, and compare performances relative to all distance measures introduced.

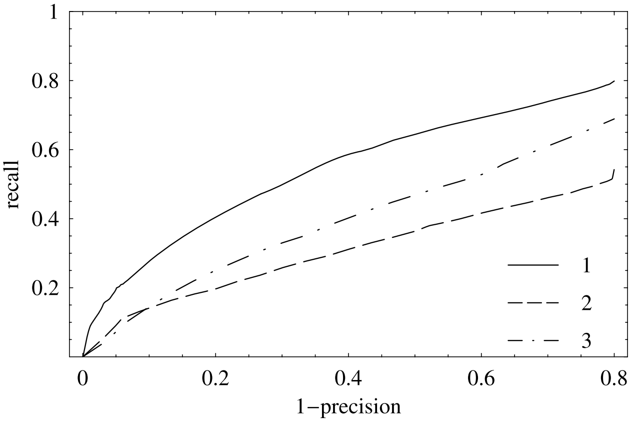

Fig. 3 Evaluation of different distances in case of DoG points, differential invariants for 5% noise. *1*: SBSM; *2*: Euclidean distance; *3*: Mahalanobis distance

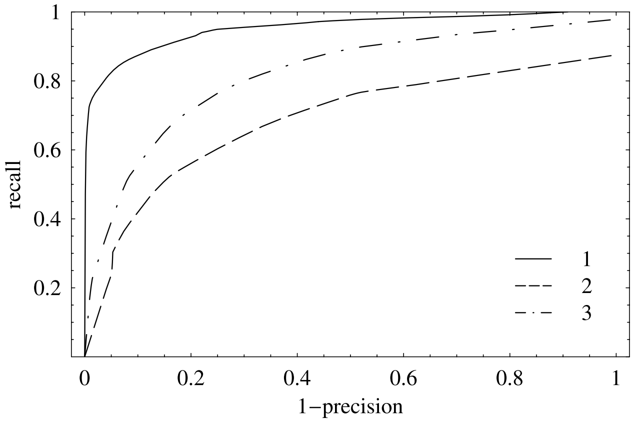

Fig. 4 Evaluation of different distances in case of top-points, differential invariants and 45 degree rotation. *1*: SBSM; *2*: Euclidean distance; *3*: Mahalanobis distance

5 Experiments

In our experimental setting the distance between every point from the reference image and every point from the transformed one is calculated for the database presented in Sect. 3.1. Two points are considered to be matched if the distance ρ_{SBSM}, (39), between their feature vectors is below a certain threshold d. The result obtained by varying d is presented by a curve. The curve presents recall versus $1 -$ precision as a function of d. The covariance matrix for the conventional Mahalanobis distance was obtained by (intentionally) training on the data set itself, so that it may be regarded "optimal" in the sense that no better results are likely to be obtained in practice when using a different training set in the proper way.

Experiments were conducted with different choices of image transformations (rotation, perspective changes, noise, scaling), feature vectors, and interest points. For every pair of images a recall versus $1 -$ precision curve is constructed,

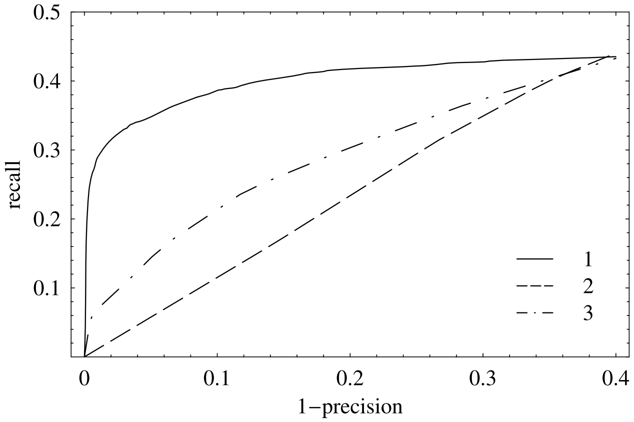

Fig. 5 Evaluation of different distances in case of top-points, steerable filters and rotation plus zooming. *1*: SBSM; *2*: Euclidean distance; *3*: Mahalanobis distance

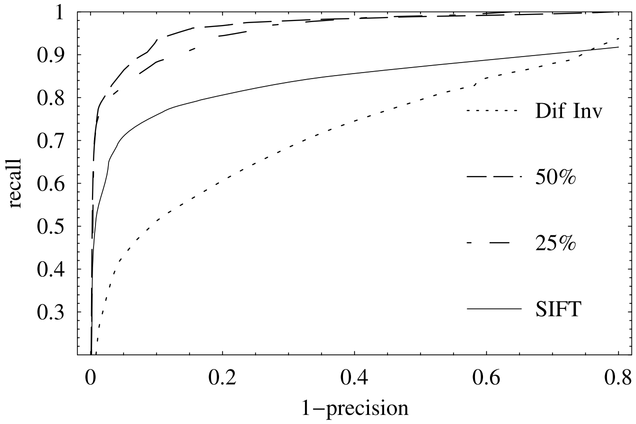

Fig. 6 Evaluation of differential invariants (with SBSM) and *SIFT* for top-points and 50% zooming. In case of differential invariants (*Dif Inv*) also a reduction to 50% and 25% of the most stable features has been considered, as indicated in the legend

and then the mean curve over 12 pairs of images is computed. In all the experiments usage of SBSM improved the performance. Here we present three typical examples. Figure 3 depicts SBSM, Euclidean and Mahalanobis curves in case of 5% noise, where differential invariants are used in Difference-of-Gaussian points [24]. In Fig. 4 top-points [29] are used as interest points and differential invariants as features for the worst-case 45 degree rotation experiment. Figure 5 depicts results of using steerable filters at top-points for image rotation and zooming. As might be expected, in all these cases the use of the inappropriate Euclidean distance yields the worst performance. SBSM, on the other hand, clearly improves performance of the feature vectors as compared to all other measures.

In order to indicate the practical use besides improved performance of differential invariant type of feature vectors, consider Fig. 6. An advantage of SBSM is the possibility of using it in order to threshold interest points with very unstable and therefore unreliable feature vectors. In this experiment we use the determinant of the covariance matrix as a criterion.

6 Summary and Conclusions

In this paper we have focused on descriptive feature vectors based on derivatives computed at a set of interest points. These descriptors have very small dimensionality, e.g. in comparison to the popular SIFT representation the dimensionality of a feature vector is an order of magnitude lower. Differential features admit a clear geometric interpretation, as they represent local image structure. They are easy to implement and fast to compute [15].

We have considered two novel ways to construct a generic distance measure for the quantitative comparison of feature vectors. The first approach yields a generic, analytical covariance matrix for a Mahalanobis-like distance function that obviates training and yet turns out to give the same performance as for the standard approach using experimental training on a restricted class of images. The resulting distance measure has, however, the disadvantage of being less versatile (it only applies to feature vectors consisting of image derivatives up to some order), and it does not exploit the particular local structure at each feature. This has led us to our second approach, in which we have introduced a new stability based similarity measure (SBSM) for feature vectors consisting of arbitrary algebraic combinations of image derivatives. The algorithm is based on a perturbation approach and uses properties of noise propagation in Gaussian scale-space. Besides being more versatile it exploits the local structure of the image at each feature point.

In comparison to the other distance measures, experiments confirm that the use of SBSM leads to a clear improvement in performance for different choices of interest points, different combinations of derivatives and several transformations.

The advantage of the proposed approach is that a local SBSM covariance matrix describing the stability of the feature vector can be predicted theoretically on the basis of the local differential structure, so that no training data are required. In fact one could say that the analytical noise model underlying the SBSM replaces the role of training. This makes SBSM generically applicable to a broad range of image and object retrieval tasks. Another advantage of SBSM is the possibility of using it in order to threshold interest points with very unstable and therefore unreliable feature vectors. One can think of eigenvalues of the covariance matrix as a criterion. This at the same time allows one to reduce the amount of data stored as well as computational time needed for matching. A drawback of

SBSM is the necessity to store a covariance matrix for every point of the reference image. But even with the necessity to store such a matrix, the dimensionality of the descriptive data per point remains significantly lower in comparison to SIFT.

As a final remark we note that, although a machine learning method trained on any specific set of training images may outperform any generic algorithm in object retrieval tasks, the latter can naturally cope with a general variety of images of which no examples are a priori available. It goes without saying that being in possession of additional knowledge about the object retrieval task might be exploited to improve performance, but recall that our intention has been explicitly not to account for any specific prior knowledge. It is, however, not clear to us how one should incorporate such prior knowledge into our locally defined Stability based Similarity Measure.

Acknowledgements The Netherlands Organisation for Scientific Research (NWO) is gratefully acknowledged for financial support.

References

1. Ashbrook, A.P., Thacker, N.A., Rockett, P.I., Brown, C.I.: Robust recognition of scaled shapes using pairwise geometric histograms. In: BMVC '95: Proceedings of the 6th British Conference on Machine Vision, Surrey UK, 1995, vol. 2, pp. 503–512. BMVA Press, Bristol (1995)
2. Bay, H., Tuytelaars, T., Van Gool, L.: Surf: Speeded-up robust features. In: 9th European Conference on Computer Vision, Graz, Austria
3. Belongie, S., Malik, J., Puzicha, J.: Shape matching and object recognition using shape contexts. IEEE Trans. Pattern Anal. Mach. Intell. **24**(4), 509–522 (2002)
4. Blom, J.: Topological and geometrical aspects of image structure. Ph.D. thesis, University of Utrecht, Department of Medical and Physiological Physics, Utrecht, The Netherlands (1992)
5. Blom, J., ter Haar Romeny, B.M., Bel, A., Koenderink, J.J.: Spatial derivatives and the propagation of noise in Gaussian scale-space. J. Vis. Commun. Image Represent. **4**(1), 1–13 (1993)
6. Florack, L.M.J.: Image Structure. Computational Imaging and Vision Series, vol. 10. Kluwer Academic, Dordrecht (1997)
7. Florack, L.M.J., ter Haar Romeny, B.M., Koenderink, J.J., Viergever, M.A.: Scale and the differential structure of images. Image Vis. Comput. **10**(6), 376–388 (1992)
8. Florack, L.M.J., ter Haar Romeny, B.M., Koenderink, J.J., Viergever, M.A.: Cartesian differential invariants in scale-space. J. Math. Imaging Vis. **3**(4), 327–348 (1993)
9. Florack, L.M.J., ter Haar Romeny, B.M., Koenderink, J.J., Viergever, M.A.: General intensity transformations and differential invariants. J. Math. Imaging Vis. **4**(2), 171–187 (1994)
10. Freeman, W., Adelson, E.: The design and use of steerable filters. IEEE Trans. Pattern Anal. Mach. Intell. **13**(9), 891–906 (1991)
11. Gabor, D.: Theory of communication. J. IEEE **93**, 429–457 (1946)
12. Gouet, V., Montesinos, P., Pele, D.: A fast matching method for color uncalibrated images using differential invariants (1998)
13. Griffin, L.D.: The second order local-image-structure solid. IEEE Trans. Pattern Anal. Mach. Intell. **29**(8), 1355–1366 (2007)
14. Grigorescu, S.E., Petkov, N., Kruizinga, P.: A comparative study of filter based texture operators using mahalanobis distance. ICPR **03**, 3897 (2000)
15. ter Haar Romeny, B.M.: Front-End Vision and Multi-Scale Image Analysis: Multi-Scale Computer Vision Theory and Applications, written in Mathematica. Computational Imaging and Vision Series, vol. 27. Kluwer Academic, Dordrecht (2003)
16. Harris, C., Stephens, M.: A combined corner and edge detector. In: Proc. 4th Alvey Vision Conf., pp. 189–192 (1988)
17. Koenderink, J.J.: The structure of images. Biol. Cybern. **50**, 363–370 (1984)
18. Koenderink, J.J., van Doorn, A.J.: Receptive field families. Biol. Cybern. **63**, 291–298 (1990)
19. Laws, K.I.: Rapid texture identification. In: Proc. SPIE Conf. Image Processing for Missile Guidance, pp. 376–380 (1980)
20. Lazebnik, S., Schmid, C., Ponce, J.: A sparse texture representation using local affine regions. IEEE Trans. Pattern Anal. Mach. Intell. **27**(8), 1265–1278 (2005)
21. Lindeberg, T.: Scale-space for discrete signals. IEEE Trans. Pattern Anal. Mach. Intell. **12**(3), 234–245 (1990)
22. Lindeberg, T.: Scale-Space Theory in Computer Vision. The Kluwer International Series in Engineering and Computer Science. Kluwer Academic, Dordrecht (1994)
23. Loog, M.: The jet metric. In: Sgallari, F., Murli, A., Paragios, N. (eds.) SSVM. Lecture Notes in Computer Science, vol. 4485, pp. 25–31. Springer, Berlin (2007)
24. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vis. **60**(2), 91–110 (2004)
25. Markussen, B., Pedersen, K.S., Loog, M.: A scale invariant covariance structure on jet space. In: Olsen, O.F., Florack, L.M.J., Kuijper, A. (eds.) DSSCV. Lecture Notes in Computer Science, vol. 3753, pp. 12–23. Springer, Berlin (2005)
26. Mikolajczyk, K., Schmid, C.: Scale and affine invariant interest point detectors. Int. J. Comput. Vis. **60**(1), 63–86 (2004)
27. Mikolajczyk, K., Schmid, C.: A performance evaluation of local descriptors. IEEE Trans. Pattern Anal. Mach. Intell. **27**(10), 1615–1630 (2005)
28. Olver, P.J.: Classical Invariant Theory. London Mathematical Society Student Texts, vol. 44. Cambridge University Press, Cambridge (1999)
29. Platel, B., Florack, L.M.J., Kanters, F.M.W., Balmachnova, E.G.: Using multiscale top points in image matching. In: Proceedings of the 11th International Conference on Image Processing, Singapore, pp. 389–392 (2004)
30. Randen, T., Husøy, J.H.: Filtering for texture classification: a comparative study. IEEE Trans. Pattern Anal. Mach. Intell. **21**(4), 291–310 (1999)
31. Schmid, C., Mohr, R.: Local grayvalue invariants for image retrieval. IEEE Trans. Pattern Anal. Mach. Intell. **19**(5), 530–535 (1997)
32. Unser, M.: Local linear transforms for texture measurements. Signal Process. **11**(1), 61–79 (1986)
33. Witkin, A.P.: Scale-space filtering. In: Proceedings of the International Joint Conference on Artificial Intelligence, Karlsruhe, Germany, pp. 1019–1022 (1983)
34. Yan, K., Sukthankar, R.: Pca-sift: a more distinctive representation for local image descriptors. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 2, pp. 506–513 (2004)

E. Balmashnova Evgeniya Balmashnova received her M.Sc. degree in mathematics at Novosibirsk State University, Russia, in 2000. In 2001–2003 she followed postgraduate programme Mathematics for Industry at the Stan Ackermans Institute in Eindhoven. She received her Ph.D. degree in 2007 from Eindhoven University of Technology. She is currently a postdoctoral fellow at the Department of Mathematics and Computer Science at the same university. Her research is focused on medical image analysis, high angular resolution diffusion imaging and diffusion tensor imaging in particular, multiscale approaches in image analysis.

L.M.J. Florack Luc Florack received his M.Sc. degree in theoretical physics in 1989, and his Ph.D. degree cum laude in 1993 with a thesis on image structure, both from Utrecht University, The Netherlands. During the period 1994–1995 he was an ERCIM/HCM research fellow at INRIA Sophia-Antipolis, France, and INESC Aveiro, Portugal. In 1996 he was an assistant research professor at DIKU, Copenhagen, Denmark, on a grant from the Danish Research Council. In 1997 he returned to Utrecht University, were he became an assistant research professor at the Department of Mathematics and Computer Science. In 2001 he moved to Eindhoven University of Technology, Department of Biomedical Engineering, were he became an associate professor in 2002. In 2007 he was appointed full professor at the Department of Mathematics and Computer Science, retaining a parttime professor position at the former department. His research covers mathematical models of structural aspects of signals, images, and movies, particularly multiscale and differential geometric representations, and their applications to imaging and vision, with a focus on cardiac cine magnetic resonance imaging, high angular resolution diffusion imaging, and diffusion tensor imaging, and on biologically motivated models of "early vision".