

Feature vector similarity based on local structure

Citation for published version (APA):

Balmachnova, E., Florack, L., & Romeny, B. M. T. H. (2007). Feature vector similarity based on local structure. In F. Sgallari, A. Murli, & N. Paragios (Eds.), *Proceedings of the First International Conference on Scale Space and Variational Methods in Computer Vision (SSVM 2007) 30 May - 2 June 2007, Ischia, Italy* (pp. 386-393). (Lecture Notes in Computer Science; Vol. 4485). Springer. https://doi.org/10.1007/978-3-540-72823-8_33

DOI:

[10.1007/978-3-540-72823-8_33](https://doi.org/10.1007/978-3-540-72823-8_33)

Document status and date:

Published: 01/01/2007

Document Version:

Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.tue.nl/taverne

Take down policy

If you believe that this document breaches copyright please contact us at:

openaccess@tue.nl

providing details and we will investigate your claim.

Feature Vector Similarity Based on Local Structure*

Evgeniya Balmachnova, Luc Florack, and Bart ter Haar Romeny

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

Abstract. Local feature matching is an essential component of many image retrieval algorithms. Euclidean and Mahalanobis distances are mostly used in order to compare two feature vectors. The first distance does not give satisfactory results in many cases and is inappropriate in the typical case where the components of the feature vector are incommensurable, whereas the second one requires training data. In this paper a stability based similarity measure (SBSM) is introduced for feature vectors that are composed of arbitrary algebraic combinations of image derivatives. Feature matching based on SBSM is shown to outperform algorithms based on Euclidean and Mahalanobis distances, and does not require any training.

1 Introduction

Local descriptors evaluated at certain interest points in scale space are widely used in object-recognition and image retrieval due to their robustness under occlusion, certain image transformations (rotation, zooming, up to some extent view point changes) and distinctive power. First, the interest points should be localized in space and scale. There is wide range of interest points [1],[2],[3], such as Harris points, Harris-Laplace regions, Hessian-Laplace regions, DoG, Top-Points. The second step is to build a description of the interest point, which should be discriminative and invariant to certain image transformations.

There are many different techniques to describe local image properties. Mikolajczyk and Schmid [4] compared several local descriptors, such as steerable filters, differential invariants, moment invariants, complex filters, cross-correlation of different types of interest points, and SIFT [1]. In this paper we focus only on differential type of descriptors, particularly on improvements of discriminative power of descriptors by introducing a sensible distance measure for such descriptors.

Our approach is applicable to any feature vector constructed from Gaussian derivatives taken at the interest point and it shows improvement compared to Mahalanobis and Euclidean distances. The proposed *stability based similarity measure* (SBSM) is based on analysis of local structure at the interest point and therefore uses a more appropriate covariance matrix than in case of the global Mahalanobis distance. The symmetry property of a true distance function is lost, but this does not affect matching results.

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

2 Feature Vectors

In this article we consider only differential features in scale-space, e.g. differential invariants proposed by Florack et al. [5] or steerable filters [6]. We collect local features with the same base point in scale-space into a “feature vector”. Under perturbation (noise, rotation, jpeg-compression and so on) every feature vector behaves differently, depending on the local image structure at the interest point. Our intention is to take this fact into account when comparing features.

Local image structure can be captured by the so-called *local jet*[7], i.e. a set of image derivatives computed up to some order. As our features are differential descriptors feature vectors can be expressed as functions on the local jet $\{u_1, \dots, u_n\}$ (for brevity, we have indicated the various image derivatives by $u_k, k = 1, \dots, n$), as follows

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

So each feature vector has m components.

We construct the differential feature vectors in such a way that they are invariant to certain transformations, notably, rotation, zooming and linear intensity changes. In the experimental part we consider one particular set of differential invariants up to third order, evaluated at a top-point of the image Laplacian Δu , viz.

$$\mathbf{d} = \begin{pmatrix} \sigma \sqrt{(u_x^2 + u_y^2)}/u \\ \sigma(u_{xx} + u_{yy})/\sqrt{u_x^2 + u_y^2} \\ \sigma^2(u_{xx}^2 + 2u_{xy}^2 + u_{yy}^2)/(u_x^2 + u_y^2) \\ \sigma(u_x^2 u_{xx} + 2u_x u_y u_{xy} + u_y^2 u_{yy})/(u_x^2 + u_y^2)^{3/2} \\ \sigma^2((-3u_x^2 + u_y^2)u_y u_{xxx} + (-u_x^2 + 3u_y^2)u_x u_{yyy})/(u_x^2 + u_y^2)^2 \\ \sigma^2(-u_x^3 u_{xxx} + 3u_x u_{xxx} u_y^2 + 3u_x^2 u_y u_{yyy} - u_y^3 u_{yyy})/(u_x^2 + u_y^2)^2 \end{pmatrix} \tag{2}$$

One can show that this set is complete, in the sense that there exists no other third order invariants (at a Laplacian top-point) that is independent of the entries of \mathbf{d} , Eq. (2). Another set, used for testing, is the set of steerable filters up to third order as described by Freeman [6].

3 Distance Measure

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 does not make much sense in case of Eq. (2), and indeed performs poorly in practice as we will illustrate in section 5. The Mahalanobis distance gives better matching results but has 3 disadvantages, viz

- it requires a covariance matrix estimated from training data;
- performance results will depend on the training set used;
- it is a global measure, not optimally adapted to the local structure at any feature point of interest.

We propose a more generic measure similar to the Mahalanobis distance that obviates training and takes into account the local structure at the feature points of interest. 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. In the next section we present the details of this approach.

4 Feature Vector Perturbation

We use a perturbation approach for estimation of the covariance matrix for each feature vector as alluded to in the previous section.

We generically model changes in the image due to rendering artifacts induced by transformations, jpeg-compression effects and noise as additive random perturbations. This distribution of the random value is assumed pixel to be the same for all pixels. Due to linearity of scale-space the perturbed local jet in the point is

$$\{v_1, \dots, v_n\} = \{u_1, \dots, u_n\} + \{N_1, \dots, N_n\}, \tag{3}$$

where the last term models the perturbation.

Let us rewrite Eq. (1) for the unperturbed image as

$$d_i = d_i(\mathbf{u}) \tag{4}$$

and for the perturbed image as

$$\tilde{d}_i = \tilde{d}_i(\mathbf{v}) \tag{5}$$

Then the difference between these two descriptors can be approximated by a Taylor expansion of Eq. (5) around \mathbf{u} up to first order:

$$\Delta d_i = \tilde{d}_i - d_i = \sum_{k=1}^n \frac{\partial d_i}{\partial v_k} \Big|_{v_k=u_k} N_k \tag{6}$$

Therefore, the approximate covariance matrix Σ is given by ($\langle N_i \rangle$ is assumed to be zero)

$$\Sigma_{ij} = \langle \Delta d_i \Delta d_j \rangle = \sum_{k=1}^n \sum_{l=1}^n \frac{\partial d_i}{\partial v_k} \Big|_{v_k=u_k} \frac{\partial d_j}{\partial v_l} \Big|_{v_l=u_l} \langle N_k N_l \rangle \tag{7}$$

The only step left is to estimate this covariance matrix for the feature perturbation, which is the subject of the next section.

4.1 Gaussian Correlated Noise

The momentum $M_{m_x, m_y, n_x, n_y}^2 = \langle N_{m_x, m_y} N_{n_x, n_y} \rangle$ 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 is much smaller than scale t is given by [8]

$$M_{m_x, m_y, n_x, n_y}^2 \simeq N^2 \left\langle \left(\frac{\tau}{2t} \right) \left(\frac{-1}{4t} \right)^{\frac{1}{2}(m_x+m_y+n_x+n_y)} Q_{m_x+n_x} Q_{m_y+n_y} \right\rangle \tag{8}$$

with Q_k given by Table 1.

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

Let us take the correlation kernel to be roughly of one pixel width corresponding to $\tau = \epsilon^2/4$, where ϵ denotes pixel size. For Gaussian derivatives of first and second order we obtain the following correlation matrix:

$$C = (\langle N_i N_j \rangle)_{ij} = \begin{pmatrix} 4t & 0 & 0 & 0 & 0 \\ 0 & 4t & 0 & 0 & 0 \\ 0 & 0 & 3 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 3 \end{pmatrix} \frac{\epsilon^2 \langle N^2 \rangle}{(4t)^3}, \tag{9}$$

where $(N_1, \dots, N_5) = (N_x, N_y, N_{xx}, N_{xy}, N_{yy})$, and the matrix entries in Eq. (9) are labeled accordingly. This correlation matrix together with Eq. (7) gives an approximation of the covariance matrix of each local feature vector for given perturbation variance and pixel size.

4.2 Distance

We define the similarity between feature descriptors \mathbf{u} and \mathbf{u}_0 in a similar way as the well-known Mahalanobis distance, except that for every point u_0 we insert its own covariance matrix

$$\rho(\mathbf{u}; \mathbf{u}_0) = (\mathbf{u} - \mathbf{u}_0)^T \Sigma_{\mathbf{u}_0}^{-1} (\mathbf{u} - \mathbf{u}_0) \tag{10}$$

Consequently, the function $\rho(\mathbf{u}; \mathbf{u}_0)$ is not symmetric, therefore it is not a distance in the strict sense. The reference image u_0 is considered to be the ‘‘ground truth’’. Covariance matrix and, as a consequence, distance are proportional to the constant $\epsilon^2 \langle N^2 \rangle$, i.e. the product of the noise variance and the 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 point to the set of all points of the reference image, therefore the constant can be omitted.

5 Experiments

5.1 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 [2,4] 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 or lighting conditions. To our knowledge Mikolajczyk’s data set does not provide this.

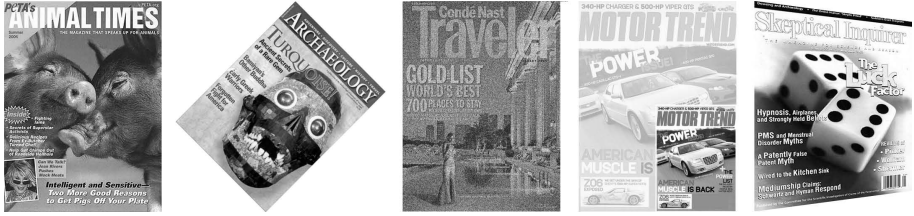


Fig. 1. A selection of data set images. From left to right: unchanged, rotated, added noise, scaled, changed perspective.

5.2 Validation

We use a criterion proposed in [9]. It is based on the number of correct matches and the number of false matches obtained for an image pair. We couple interest points if the distance between their feature vectors is below a certain threshold d (such a pair is called *a possible match*). Note that since we know the transformations we also know ground truth for the matches. Each feature vector from the reference image is compared to each vector from the transformed one, and the number of *correct matches* (according to the ground truth, out of all possible matches) as well as the number of *false matches*, is counted.

$$\#possible\ matches = \#correct\ matches + \#false\ matches \tag{11}$$

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:

$$recall = \frac{\#correct\ matches}{\#correspondences} \tag{12}$$

The number of false matches relative to the number of possible matches is presented by 1–precision

$$1 - precision = \frac{\#false\ matches}{\#possible\ matches} \tag{13}$$

The number of correct matches and correspondences is determined with the overlap error. 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 ,

$$\varepsilon = 1 - \frac{S_2 \cap HS_1}{S_2 \cup HS_1}, \tag{14}$$

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

In case of transformations closed to Scale-Euclidean ones, HS_1 can be also approximated by a disk, and areas of intersection and union can be computed analytically. A match is 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 Eq. (12) is determined with the same criterion.

5.3 Results

In our experimental setting the distance between every point from the reference image and every point from the transformed one is calculated. Two points are considered to be matched if the distance ρ between their feature vectors is below a threshold d . The result obtained by varying d is presented by a curve. The curve presents 1-precision versus recall as a function of d . For experiments with the Mahalanobis distance training data are required. We use part of the data base for training and the rest for evaluation.

Experiments were conducted with different choices of image transformations, feature vectors and interest points. For every pair of images a recall vs 1-precision curve is built, and then the mean curve over 12 pairs is computed. In all the experiments usage of

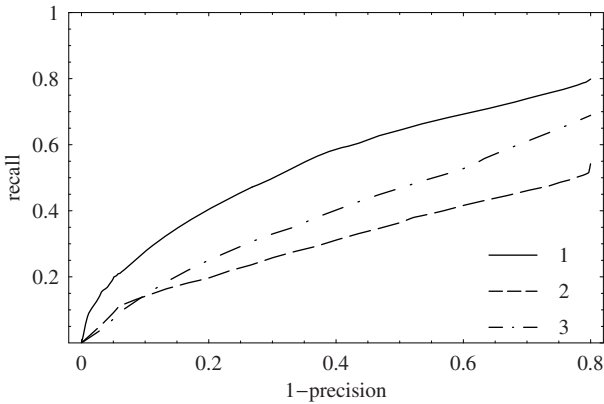


Fig. 2. Evaluation of different distances in case of DoG points, differential invariants for 5% noise. 1- SBSM, 2 - Euclidean distance, 3- Mahalanobis distance.

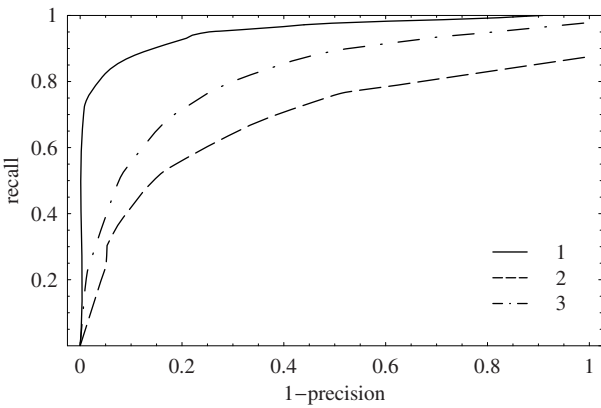


Fig. 3. Evaluation of different distances in case of Top-Points points, differential invariants and 45 degree rotation. 1- SBSM, 2 - Euclidean distance, 3- Mahalanobis distance.

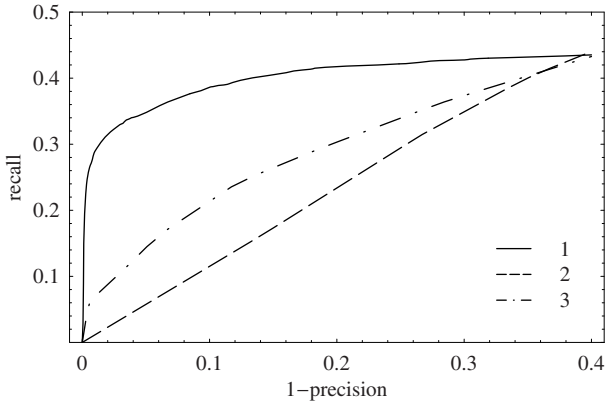


Fig. 4. Evaluation of different distances in case of Top-Points points, steerable filters and rotation+zooming. 1- SBSM, 2 - Euclidean distance, 3- Mahalanobis distance.

SBSM improved the performance. In the paper we present three examples. Figure 2 depicts SBSM, Euclidean and Mahalanobis curves in case of 5% noise, where differential invariants are used in Difference-of-Gaussian points. In Fig 3 Top-Points are used as interest points and differential invariants as features for 45 degree rotation case. Figure 4 depicts results of using steerable filters at Top-Points for image rotation and zooming.

6 Summary and Conclusions

In this paper we have introduced a new stability based similarity measure (SBSM) for feature vectors constructed by means of differential invariants. The algorithm is based on a perturbation approach and uses properties of noise propagation in Gaussian scale-space.

The advantage of this approach is that a local covariance matrix describing the stability of the feature vector can be predicted theoretically on the basis of the local differential structure, and that no training data are required. In fact the analytical noise model underlying the SBSM replaces the role of training. The drawback of using SBSM is a necessity to store a covariance matrix for every point of the reference image.

Another advantage of SBSM is 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 criterium. This allows one to reduce the amount of data stored as well as computational time needed for matching.

The experiments show improvement in performance for different choices of interest points, different combinations of derivatives and several transformations.

References

1. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vision* **60** (2004) 91–110
2. Mikolajczyk, K., Schmid, C.: Scale and affine invariant interest point detectors. *International Journal of Computer Vision* **60** (2004) 63–86

3. Platel, B., Balmachnova, E.G., Florack, L.M.J., ter Haar Romeny, B.M.: Top-points as interest points for image matching. In: Proceedings of the Ninth European Conference on Computer Vision, Graz, Austria (2006) 418–429
4. Mikolajczyk, K., Schmid, C.: A performance evaluation of local descriptors. Submitted to PAMI (2004)
5. 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** (1993) 327–348
6. Freeman, W., Adelson, E.: The design and use of steerable filters. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **13** (1991) 891–906
7. Koenderink, J.J.: The structure of images. *Biological Cybernetics* **50** (1984) 363–370
8. 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** (1993) 1–13
9. Ke, Y., Sukthankar, R.: Pca-sift: A more distinctive representation for local image descriptors. In: Proceedings of the Conference on Computer Vision and Pattern Recognition. (2004) 511–517