Increasing the Accuracy
of Feature Evaluation Benchmarks
Using Differential Evolution

Kai Cordes, Bodo Rosenhahn, Jörn Ostermann

Institut für Informationsverarbeitung (TNT)

Leibniz Universität Hannover, Germany

http://www.tnt.uni-hannover.de

Email: {cordes, rosenhahn, ostermann}@tnt.uni-hannover.de

Abstract—The accuracy evaluation of image feature detectors is done using the repeatability criterion. Therefore, a well-known data set consisting of image sequences and homography matrices is processed. This data serves as ground truth mapping information for the evaluation and is used in many computer vision papers.

An accuracy validation of the benchmarks has not been done so far and is provided in this work. The accuracy is limited and evaluations of feature detectors may result in erroneous conclusions.

Using a differential evolution approach for the optimization of a new, feature-independent cost function, the accuracy of the ground truth homographies is increased. The results are validated using comparisons between the repeatability rates before and after the proposed optimization. The new homographies provide better repeatability results for each detector. The repeatability rate is increased by up to 20%.

I. INTRODUCTION

The detection of image features is a requirement for many computer vision challenges such as object recognition, mobile robot navigation, or scene reconstruction. In the latter case, the accuracy of the detected points or regions is of key interest. Only small errors in the feature localization in the image can lead to large reconstruction errors, depending on the distance between the camera and the geometry of the scene.

Features are used to establish correspondences between different images containing the same scene captured by a camera. To obtain reasonable and comparable feature points, the image signal surrounding the feature position is analyzed and distinctive characteristics of this region are extracted. These characteristics are usually assembled to a vector which describes the feature. This vector, called descriptor [1], is used to establish correspondences by calculating a similarity measure (L_2 distance) between the current descriptor and the feature descriptors of a second image. Feature detection and correspondence analysis is more challenging if the baseline between the cameras capturing the images is large. The region description of corresponding features before and after the viewpoint change should be the same. Several methods have been proposed in the literature to model invariants, such as brightness change and rotation [2], [3], scale change [1], [4],
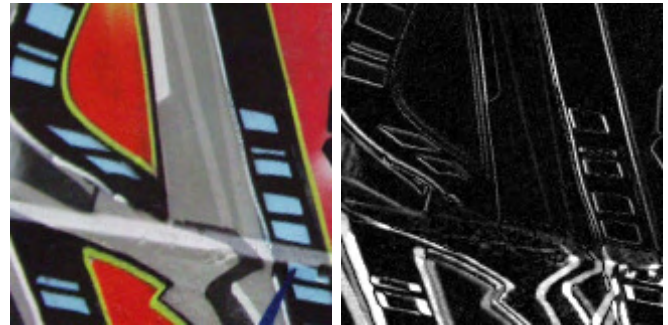

Fig. 1: Mapping error of a part of the image pair 1-3 of the *Graffiti* sequence. Left: Part of the third image I_3 of the sequence, right: Mapping error represented by color differences between I_3 and the mapped image I_1 using the *ground truth* homography from the reference data set. The errors result from an inaccurate homography (top part of the image) as well as from non-planar scene content (bottom part).

[5], [6], affine transformation [7], [8], [9], [10], [11], [12], [13], [14], [15], and perspective transformation [16].

For viewpoint changes, the mapping between the local regions of two images can be modeled by an affine transformation. This is motivated by the assumption that the scene surface is locally planar and perspective effects are small on local regions [12].

Extensive work has been done on evaluating feature detectors and descriptors [10], [11], [12], [17], [18], [19], [20], [21]. Commonly, the most important criterion for the accuracy of affine invariant detectors is the repeatability criterion [6], [10], [12], [20], [22], [23], [24] using the reference test set in [12]. The data set contains sequences of still images with changes in illumination, rotation, perspective, and scale. The mapping from one image to the next is restricted to a homography. For the benchmark test, the ground truth homography matrices are also provided in [12]. They are computed using manually selected as well as automatically detected point correspondences. For both sets of correspondences, a standard homography

estimation algorithm is used and the results are combined[1].

The most accurate localization of affine invariant features [25] is achieved by the MSER [8] detector, followed by Hessian-Affine [12]. However, their performances depend on the type of image transformation of the specific test sequence and the type of texture present in the images. It is also shown [25], [26] that the spatial feature distributions are different for each detector, which motivates to use complementary feature sets for computer vision challenges.

Although the authors comment that the homographies are *not perfect* [9], the data set is used as ground truth for high-accuracy evaluations [6], [9], [12]. In addition to inaccurate homography matrices, the image data of the set is not suitable in several cases. In the example image part shown in Fig. 1, the scene structure is not restricted to one plane, resulting in large mapping errors in the bottom part of the images. Haja et al. [20] evaluate another data set [21] for accuracy evaluation, but use the same estimation procedure as Mikolajczyk et al. [12] for computing the ground truth homographies. Moreels et al. [21] use a trifocal tensor-like double epipolar constraint for the evaluation, but loose a large amount of features which have to be identified in three images.

The ground truth data set [12] is used as a benchmark for many feature evaluations. With inaccurate homographies and inadequate image content, an evaluation may provide misleading results and conclusions. An accuracy validation of the data set has not been done so far and is provided in this work. Furthermore, it is shown that the accuracy of the homography matrices can be increased using a feature independent cost function and a Differential Evolution optimization technique. The detachment of the cost function from a feature detector is favorable because the resulting homography data is used for the evaluation of feature detectors. For the validation of the new homographies, the same feature detectors as in [12] are evaluated. The detectors are: MSER [8], Hessian-Affine [10], Harris-Affine [9], intensity extrema-based regions (IBR) [14], and edge-based regions (EBR) [13].

The **contributions** of this paper are:
- the analysis of the ground truth data set used for accuracy evaluation of feature detectors,
- the improvement of ground truth data using a feature independent cost function, and
- the construction of new, highly-accurate data sets;
- the new evaluation data is provided at:
  www.tnt.uni-hannover.de/project/feature_evaluation/

In the following section, the basics of homography estimation used for the computation of ground truth homography matrices are explained. The optimization using Differential Evolution is derived in Section III. In Section IV, the image data in analyzed and deficiencies of the reference images are revealed. Section V shows experimental results using the proposed cost function and optimization for the reference data set. Furthermore, the construction of new, highly-accurate data sets is explained. In Section VI, the paper is concluded.
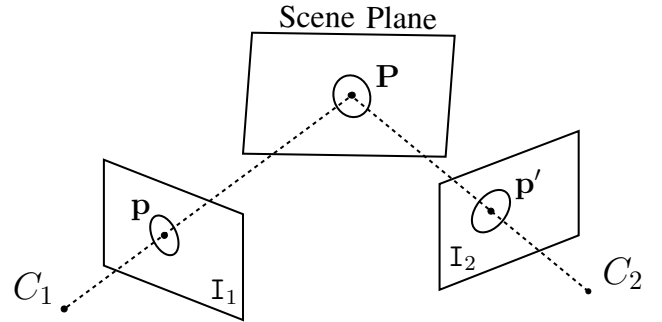
---

[1]www.robots.ox.ac.uk/~vgg/research/affine/det_eval_files/DataREADME

Fig. 2: Projection of an ellipse surrounding the scene point $\mathbf{P}$ to the images $\mathtt{I}_1$ and $\mathtt{I}_2$ of cameras C_1 and C_2, respectively. The image features $\mathbf{p}$ and $\mathbf{p}'$ are corresponding.

II. HOMOGRAPHY ESTIMATION

The homography between two images determines the mapping between a point $\mathbf{p}$ of the first image and the corresponding point $\mathbf{p}'$ in the second image. Corresponding points $\mathbf{p}, \mathbf{p}'$ result from the same scene point $\mathbf{P}$ through perspective projection into the image planes $\mathtt{I}_1, \mathtt{I}_2$. The image planes are associated with the Cameras C_1, C_2 as shown in Fig. 2. In case of a homography mapping, the scene is subject to the following restrictions:

A) Intrinsic camera parameters of C_1, C_2 are identical
B) The scene consists of rigid geometry
C) The observed scene geometry is planar **or** the camera movement from C_1 to C_2 is purely rotational

Then, each point $\mathbf{p}_j$ of the first image can be identified in the second image with coordinates $\mathbf{p}'_j$ with the homography matrix $\mathtt{H} \in \mathbb{R}^{3 \times 3}$ using homogeneous coordinates [27]:

$$\mathtt{H} \cdot \mathbf{p}_j = \mathbf{p}'_j \qquad (1)$$

It follows that the corresponding RGB values $\mathtt{I}_1(\mathtt{H} \cdot \mathbf{p}_j)$ and $\mathtt{I}_2(\mathbf{p}'_j)$ in the images $\mathtt{I}_1, \mathtt{I}_2$ should be identical:

$$\mathtt{I}_1(\mathtt{H} \cdot \mathbf{p}_j) = \mathtt{I}_2(\mathbf{p}'_j) \qquad (2)$$

A homography is defined only up to scale and has 8 degrees of freedom. Thus, it can be computed with four corresponding points in the images. As the images are influenced by noise, the problem to solve is to compute the best possible solution from an over-determined set of equations evaluating more than four correspondences. Usually, the Direct Linear Transformation (DLT) algorithm is used [27]. To create ground truth homographies for a given image pair, in [12] manually selected and automatically detected point correspondences are used for computation. However, the results are still subject to errors as shown in Fig. 1.

In this work, the homographies are estimated by minimizing the normalized sum of pixel color differences between the first image $\mathtt{I}_1$ mapped by the homography $\mathtt{H}$ and the second image $\mathtt{I}_2$ using each pixel $\mathbf{p}_j$, $j \in [1; J]$, of image $\mathtt{I}_1$ which is also visible in image $\mathtt{I}_2$. This approach leads to the following objective function with a homography error $E_{\mathtt{H}}$ to

Fig. 3: The edited *Graffiti* sequence with masked bottom parts where the homography violations occur. The radial distortion is roughly compensated using a line in the images.

be minimized:

$$E_{\mathtt{H}} = \frac{1}{J} \sum_{j=1}^{J} ||\mathtt{I}_1(\mathtt{H} \cdot \mathbf{p}_j) - \mathtt{I}_2(\mathbf{p}_j')|| \qquad (3)$$

We argue that the homography mapping function is optimal if the color distance between the mapped first image and the second image is minimal. Due to lighting and perspective changes between the images, the cost function is likely to have several local minima. Thus, a Differential Evolution (DE) optimizer is used for the minimization of $E_{\mathtt{H}}$ with respect to $\mathtt{H}$ in the cost function (3). Evolutionary optimization methods have proved impressive performance for camera parameter estimation challenges [28], [29], [30], finding the global optimum in a parameter space with many local optima. The homographies given by the reference data set can be used as an initially known approximate solution.

III. DIFFERENTIAL EVOLUTION OPTIMIZATION

For optimizing the cost function (3), the Differential Evolution (DE) algorithm [31], [32] is used. The optimization objective, the matrix $\mathtt{H}$, is defined only up to scale. Hence it is described by 8 parameters:

$$\mathtt{H} = \begin{pmatrix} h_1 & h_2 & h_3 \\ h_4 & h_5 & h_6 \\ h_7 & h_8 & 1 \end{pmatrix} \qquad (4)$$

It is common to set the ninth entry of $\mathtt{H}$ to 1. The remaining eight matrix elements are assembled to a vector $\mathbf{h} = (h_1, \dots, h_8) \in \mathbb{R}^8$ which holds the parameters to be optimized by the DE optimizer.

The optimization is based on a population of Np solution candidates $\mathbf{h}_{n,i}$, $n \in [1; Np]$, at iteration i where each candidate is a position in the 8-dimensional search space. The search space intervals are determined using an initially known approximate solution, which is given by the homography of the reference data set. This homography determines the center of the search space. The solution candidates are randomly generated within the search space intervals. The population improves by generating new 8-tupels iteratively for each candidate. New positions for the next iteration step $i + 1$ are calculated by

$$\mathbf{h}_{n,i+1}' = \mathbf{h}_{k,i} + F \cdot (\mathbf{h}_{l,i} - \mathbf{h}_{m,i}) \qquad (5)$$
$$\mathbf{h}_{n,i+1} = c\left(\mathbf{h}_{n,i},\ \mathbf{h}_{n,i+1}'\right) \qquad (6)$$

where k, l, m are random integers from the interval $[1; Np]$, $F \in [0; 1]$ is a weighting scalar, $\mathbf{h}_{n,i+1}'$ is a displaced $\mathbf{h}_{k,i}$ by a weighted difference vector, and $c\left(\mathbf{h}_{n,i},\ \mathbf{h}_{n,i+1}'\right)$ is a crossover operator copying coordinates from both $\mathbf{h}_{n,i}$ and $\mathbf{h}_{n,i+1}'$ in order to create $\mathbf{h}_{n,i+1}$. The crossover operator c is provided with a value $Cr \in [0; 1]$ specifying the probability to copy values either from $\mathbf{h}_{n,i}$ or $\mathbf{h}_{n,i+1}'$ to $\mathbf{h}_{n,i+1}$ for each component of the vector. Only if the new candidate $\mathbf{h}_{n,i+1}$ proves to have lower cost using the cost function (3) it replaces $\mathbf{h}_{n,i}$, otherwise it is discarded.

In this application, the initial search space intervals are provided by the homographies of the reference data set or can easily be done manually for the construction of new data sets (Section V-B). Common DE settings are used: $Np = 120$, $Cr = 0.8$, $F = 0.8$. The DE converges after approximately 400 iterations with significantly increased accuracy as shown in the results section V-A. The calculation of the cost function is computationally expensive and mainly dependent on the image size. The estimation of the homography for the first image pair of the *Graffiti* sequence (800×640 Pixels) needs about two hours (385 iterations) using common PC hardware. For the application of the calculation of ground truth homographies, this computational expense is not critical.

IV. REFERENCE DATA ANALYSIS

The reference data set[2] consists of several image sequences (6 images each) and matlab code for the computation of the *Repeatability* of the features detected in image pairs. The *Repeatability Criterion* is explained in Section IV-A. In Section IV-B, the image data is analyzed.

A. Repeatability Criterion

The *Repeatability* is the most often used criterion to evaluate the localization accuracy of feature detectors. Matlab code is provided together with a data set consisting of an image sequence and *ground truth* homography matrices that represent

---

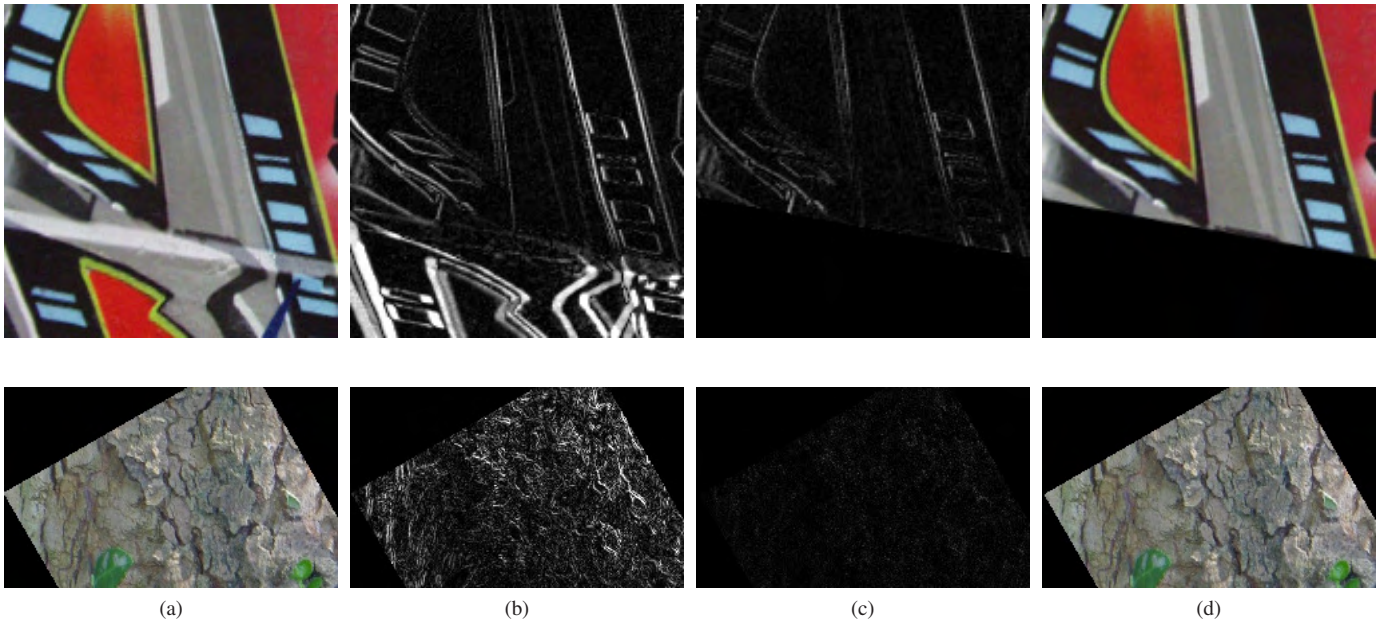[2]www.robots.ox.ac.uk/~vgg/research/affine/index.html

Fig. 4: Top row: part of image pair 1-3 of the *Graffiti* sequence, bottom row: part of image pair 1-4 of the *Bark* sequence. Image column (a) shows the image mapping $\mathtt{H}_{13}^{graf} \odot \mathtt{I}_1^{graf}$ and $\mathtt{H}_{14}^{bark} \odot \mathtt{I}_1^{bark}$ using the homographies from the reference data set. Column (d) shows the mapping with the improved homographies. The images (b) and (c) show the color differences to the images $\mathtt{I}_3^{graf}$ and $\mathtt{I}_4^{bark}$, respectively. Compared to our results (c), the homographies from the data set (b) show large errors. Several error structures in the difference images (b) indicate a localization error > 2 pel. Comparing the images (a) and (d), a shift is noticeable.

the mapping between image pairs. In [9], the following error measures to calculate the repeatability of a detector evaluating an image pair are presented. The matrix $\mathtt{H}$ is the a priori known homography relating two images of the data set.

A) For the evaluation of the point position accuracy, the error in *Relative Location* d is computed:

$$d = ||\mathtt{H} \cdot \mathbf{p} - \mathbf{p}'|| < d_{thres} \qquad (7)$$

If the *Relative Location* error for a feature pair $\mathbf{p}, \mathbf{p}'$ is below a threshold $d_{thres} = 1.5$ pel, the feature pair is deemed corresponding.

B) For the evaluation of detected regions in the images (in general ellipses), the *Overlap Error* ϵ_O is computed:

$$\epsilon_O = 1 - \frac{R_{\mu_{\mathbf{p}}} \cap R_{(\mathtt{H}^\top \mu_{\mathbf{p}'} \mathtt{H})}}{(R_{\mu_{\mathbf{p}}} \cup R_{(\mathtt{H}^\top \mu_{\mathbf{p}'} \mathtt{H})})} \qquad (8)$$

where R_μ represents the elliptic region defined by $\mathbf{x}^\top \mu \mathbf{x} = 1$ [12]. The *Overlap Error* is a percentage value that is minimal if a detected ellipse region $R_{\mu_{\mathbf{p}}}$ is exactly matching the ellipse $R_{(\mathtt{H}^\top \mu_{\mathbf{p}'} \mathtt{H})}$ determined by the ground truth homography $\mathtt{H}$. A feature pair is deemed corresponding, if the *Overlap Error* between the ellipses of the best possible match is below a threshold ϵ_{thres}. In [9], the threshold is chosen to $\epsilon_{thres} = 20\%$, in [10], [12] it is $\epsilon_{thres} = 40\%$, and in [20] $\epsilon_{thres} = 50\%$ is chosen.

The *Repeatability* value is the ratio of corresponding feature pairs and the minimal number of detected feature points in the first and the second image. While the *Relative Location* $d_{thres} = 1.5$ pel gives an impression of the needed accuracy for feature detectors and ground truth homographies, the *Overlap Error* is its generalization to detected ellipses. The error structures in the difference images in Fig. 4 (b) indicate that for several regions in the images the resulting *Relative Location* error is above the threshold of 1.5 pel just because of inaccurate homographies. In these regions, perfect as well as wrong feature pairs will be misclassified. For the evaluation of affine invariant feature detectors, the *Overlap Error* is used having the same problems.

B. Reference Image Analysis

Some of the sequences of the reference data set contain scene structures that violate the homography assumption. The *Graffiti* sequence, for example, contains a second plane in the bottom part of the image as shown in Fig. 1. Each detected feature in this region is not suitable for the evaluation, but has an effect on the repeatability value. Furthermore, the images suffer from radial distortion. Several image regions that do not fulfill the conditions of a homography mapping can be seen in the other sequences as well. In the *Wall* sequence, a second plane is visible, too (non-planar geometry). Significant movements of scene structures occur in the sequences *Boat* (boats, persons, leaves), *Bark* (leaves), and *Trees* (leaves). Unfortunately, these image regions are highly textured. Thus,
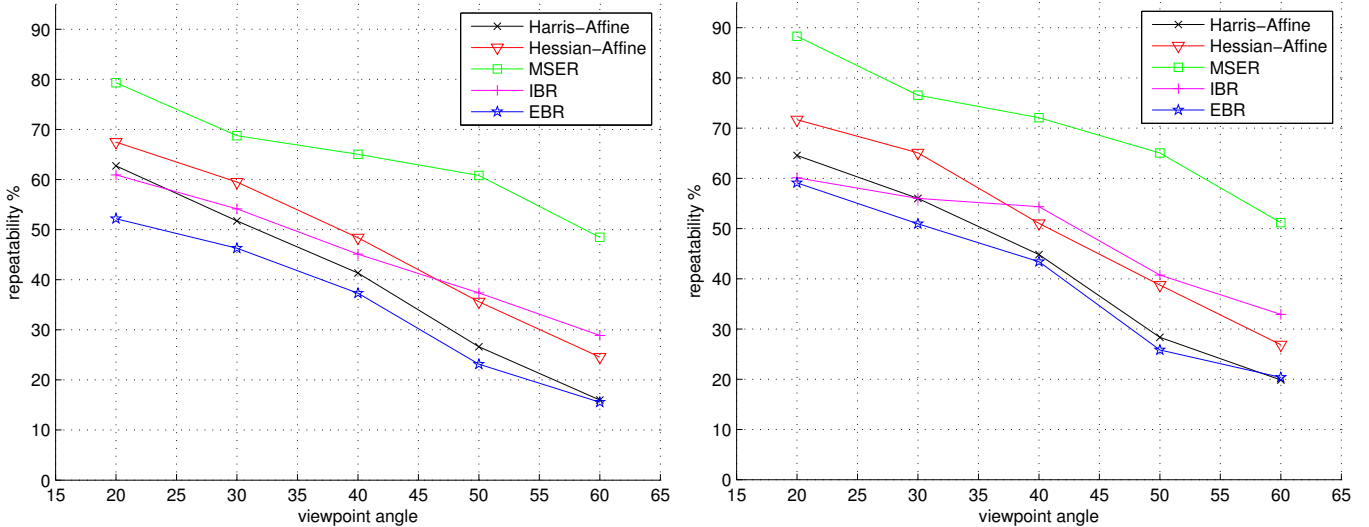
Fig. 5: *Graffiti* sequence: results with old homographies (left) and our new homographies (right) using a threshold of $\epsilon_{thres} = 40\%$. The MSER and IBR repeatability scores are increased by up to 9%.

they are likely to be detected by a feature detector. While the violation of planar scene structure in *Graffiti* and *Wall* can be eliminated by masking the image parts (see Fig. 3), a correction of the image data *Boat* and *Trees* is not possible. Using a robust cost function [29], [30] improves the results for the usable parts, but this does not lead to significantly increased repeatability rates.

Ignoring the regions with homography violations as well as the optimization technique explained in Section II allow for increasing the accuracy of the computed homography. This is shown for the examples *Graffiti* and *Bark* in the next section.

V. EXPERIMENTAL RESULTS

The presented approach aims to improve the homography matrices from the ground truth data set [12]. The increased accuracy is shown using the error measure E_H introduced in Section II as well as the *Repeatability Criterion* described in Section IV-A. The latter is the most often used criterion for the accuracy evaluation of feature detectors. The *Repeatability* evaluation technique is the same as used in literature (e.g. [12]). The comparisons are shown in Section V-A.

The proposed method can also be used to generate new highly-accurate data sets. This is demonstrated for the example of a *viewpoint change* scenario in Section V-B.

A. Reference Image Data Set

The image data set provides image pairs with associated homography matrices that map each point in one image to the corresponding point in the other image. However, visible errors occur using the provided homographies when the first image of a pair is mapped and compared to the second image. Two examples are shown in Fig. 4. Consider the top row. The left image (a) shows an image part of the third image of the

Graffiti sequence $\mathtt{I}_3^{graf}$. The second image (b) indicates the color differences $\|\mathtt{H}_{13}^{graf} \odot \mathtt{I}_1^{graf} - \mathtt{I}_3^{graf}\|$ between the image $\mathtt{I}_1^{graf}$ mapped to the second by $\mathtt{H}_{13}^{graf}$ and the second image $\mathtt{I}_3^{graf}$ for each pixel. The bottom part of the *Graffiti* sequence represents a violation of the planar assumption of the scene and therefore displays large errors. To exclude this region from the homography estimation, this part is masked (image (d) of Fig. 4, top row). As shown in Table I, the masking decreases the error E_H significantly. The last row in Table I proves that the proposed optimization method provides a further decrease of the error E_H. This is visualized in Fig. 4 (c).

TABLE I: *Graffiti* sequence: cost function values E_H for the homographies of the data set and our new homographies. A visualization is shown in Fig. 4, top row.

Image Pair $\mathtt{I}^{graf}$	1-2	1-3	1-4	1-5	1-6
$E_H^{old}(original)$	11.16	16.36	20.44	24.47	29.16
$E_H^{old}(masked)$	5.89	9.92	15.82	20.98	25.00
$E_H^{new}(masked)$	4.62	8.71	14.95	19.94	24.30

Minimizing the proposed cost function (3) results in more accurate homographies as validated in Fig. 5. The left diagram shows the repeatability evaluation as presented in [12] while the right shows the same evaluation using the edited images and the new homographies. The repeatability values increase for each detector using the data resulting from the new approach. For this sequence, the MSER detector and the intensity based region (IBR) detector gain up to 9% in repeatability. The reason for the differing gain of the detectors is the different spatial distribution of their features.
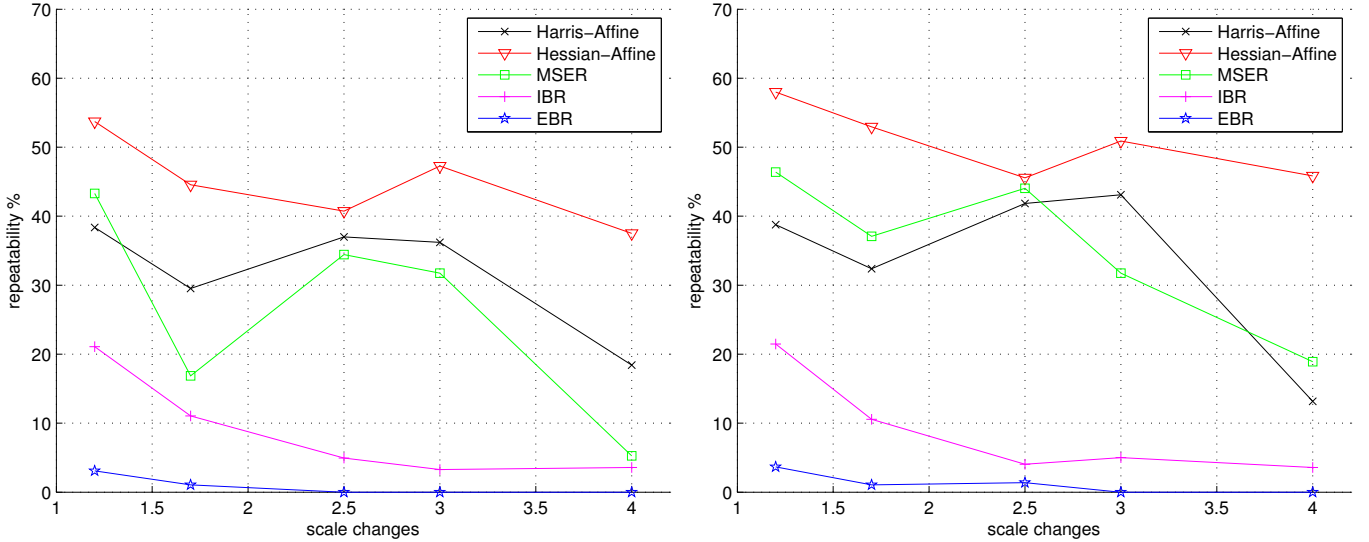
Fig. 6: *Bark* sequence: results with old homographies (left) and our new homographies (right) using a threshold of $\epsilon_{thres} = 20\%$. While in the left diagram Harris-Affine seems superior to MSER, the right diagram depicts a different result. The repeatability score for MSER is increased by up to 20%.

The second example, the *Bark* sequence is evaluated without masking any image parts. The results for E_H are summarized in Table II. The repeatability comparisons before and after the optimization are shown in Fig. 6 and for the image pair 1-3 for different thresholds ϵ_{thres} in Fig. 7. As shown in Table II, the error E_H is reduced significantly, especially for the image pairs 1-4, 1-5, 1-6. The remaining errors in the image pairs 1-2 and 1-3 are due to moving scene content (leaves) occurring for these two image pairs. Using the new homographies, an evaluation may lead to different results as demonstrated in Fig. 6 and Fig. 7, respectively. In Fig. 6, a threshold $\epsilon_{thres} = 20\%$ (see Section IV-A) is used. In Fig. 7 the threshold is varied for a specific image pair 1-3 of the sequence. Again, the MSER detector shows higher gains (up to 20%) in repeatability values than most of the others. In the new evaluation, the MSER results in better performance than the Harris-Affine detector.

TABLE II: *Bark* sequence: cost function values E_H for the homographies of the data set and our new homographies. A visualization is shown in Fig. 4, bottom row.

Image Pair $\mathtt{I}^{bark}$	1-2	1-3	1-4	1-5	1-6
E_H^{old}	11.20	13.36	9.19	6.98	10.29
E_H^{new}	10.14	10.04	2.57	3.21	3.34

The homography matrices are re-estimated and improved for the other sequences of the data set [12] as well. In this paper, we limit the evaluation to *Graffiti* (*viewpoint change* scenario) and *Bark* (*scale change* scenario).

We can infer that our estimation technique clearly outperforms the method used in [12]. The new scheme for estimating

homographies could be of great value for the community, resulting in improved validation. It can also be used to produce new, highly-accurate data sets as shown in the next section.

B. Construction of Natural Image Data Set

To construct a new sequence data set for the most challenging scenario *viewpoint change*, a planar scene is captured by a Canon EOS350d camera from different viewpoints with constant intrinsic parameters. The image size is 1536×1024. As a first post processing step, the radial distortion in the images is compensated. Initial estimates for the homographies are done manually. Then, the proposed optimization scheme is used to estimate the new homographies. No features are used for the homography determination. Images and error results E_H of this data set, called *Grace*, are shown in Fig. 8 and Table III, respectively. The results demonstrate a small mapping error compared to the *viewpoint change* image sequence *Graffiti* (Table I) of the reference data set.

TABLE III: *Grace*: Cost function values E_H for the homographies of the new data set and the homographies resulting from the proposed approach.

Image Pair $\mathtt{I}^{grace}$	1-2	1-3	1-4	1-5	1-6
E_H	3.44	4.62	6.02	8.21	9.99

To validate the *viewpoint change* scenario [12], the results of the image pair 1-3 of the new data set *Grace* are compared to the original evaluation of *Graffiti* in Fig. 9. The highest score for this new sequence is obtained by MSER, followed by Hessian-Affine and Harris-Affine. The dominance of the
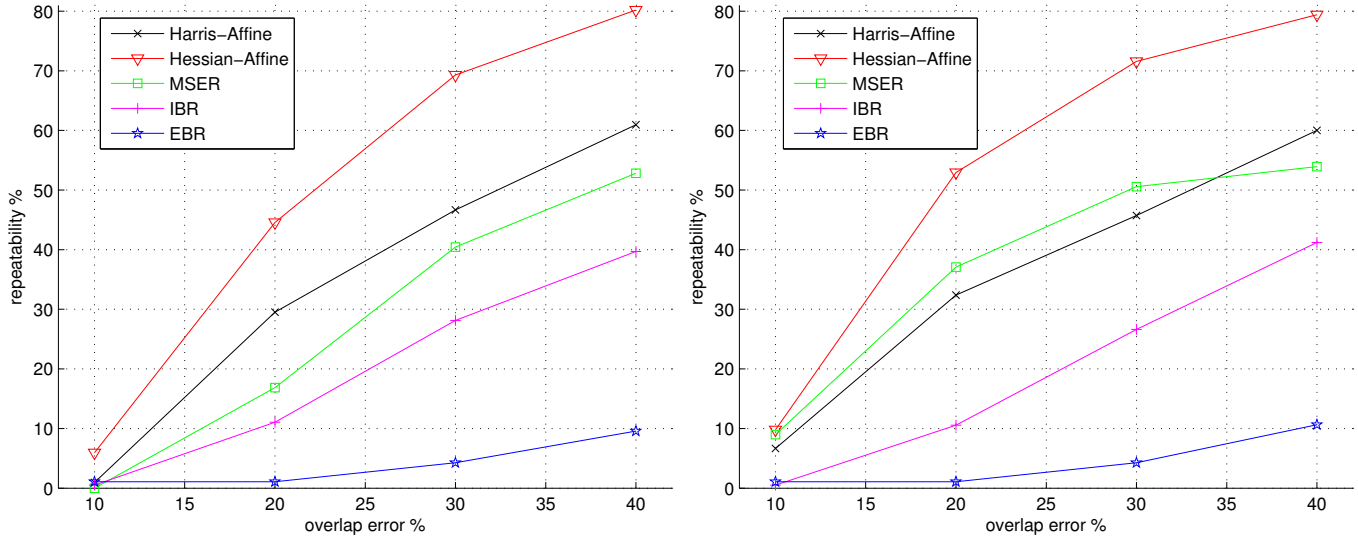
Fig. 7: *Bark* image pair 1-3: results with old homographies (left) and our new homographies (right) with various overlap error thresholds $\epsilon_{thres} = 10\%, \ldots, 40\%$. While in the left diagram Harris-Affine seems superior to MSER in high-accurate feature positions, the ranking is different in the right diagram.

Fig. 8: Example images of the *Grace* sequence with rectified images and fully planar geometry.

MSER detector in this field is verified by this result. The localization accuracy ranking of the other detectors is clearer than in the original evaluation. These results are confirmed using the other image pairs of the *Grace* sequence.

VI. Conclusion

In this work, the accuracy of a widely used benchmark data set for feature detector evaluation is analyzed. The accuracy of the homography matrices is improved using a feature-independent cost function and the Differential Evolution approach for its optimization. New highly-accurate data sets are constructed using the presented approach.

The increased accuracy is validated using the *Repeatability*, a criterion commonly used for the detector evaluation. Using the reference data set and the proposed approach for the homography estimation, the repeatability rates are increased by up to 9% using an *Overlap Error* threshold of 40% and up to 20% with a threshold of 20%. Furthermore, it is demonstrated that using the reference data set for high-accuracy evaluation may lead to false conclusions.

The results of the proposed approach provide improved, more accurate validation of feature detectors.

The evaluation data is available at the project page:
www.tnt.uni-hannover.de/project/feature_evaluation/

References

[1] M. Brown and D. G. Lowe, "Invariant features from interest point groups," in *British Machine Vision Conference (BMVC)*, 2002, pp. 656–665.

[2] J. Canny, "A computational approach to edge detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, vol. 8, no. 6, pp. 679–698, 1986.

[3] C. Harris and M. Stephens, "A combined corner and edge detector," in *Alvey Vision Conference*, 1988, pp. 147–151.

[4] T. Lindeberg, "Feature detection with automatic scale selection," *International Journal of Computer Vision (IJCV)*, vol. 30, pp. 79–116, 1998.

[5] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision (IJCV)*, vol. 60, no. 2, pp. 91–110, 2004.

[6] W. Förstner, T. Dickscheid, and F. Schindler, "Detecting interpretable and accurate scale-invariant keypoints," in *IEEE International Conference on Computer Vision (ICCV)*, 2009, pp. 2256–2263.

[7] T. Lindeberg and J. Garding, "Shape-adapted smoothing in estimation of 3-d shape cues from affine deformations of local 2-d brightness structure," *Image and Vision Computing*, vol. 15, no. 6, pp. 415–434, 1997.
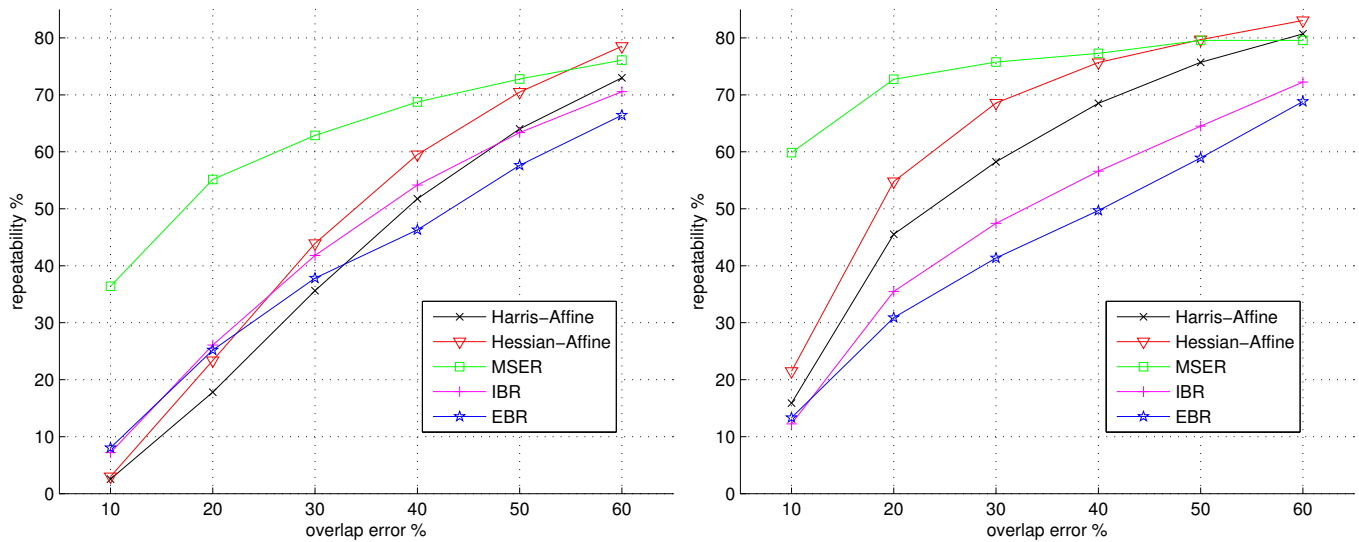
Fig. 9: Comparison of *Graffiti* image pair 1-3 (left, reference data) with *Grace* image pair 1-3 (right, new data) using various overlap error thresholds $\epsilon_{thres} = 10\%, \ldots, 60\%$. In the right diagram, the ranking of the detectors is much clearer.

[8] J. Matas, O. Chum, M. Urban, and T. Pajdla, "Robust wide baseline stereo from maximally stable extremal regions," in *British Machine Vision Conference (BMVC)*, vol. 1, 2002, pp. 384–393.

[9] K. Mikolajczyk and C. Schmid, "An affine invariant interest point detector," in *European Conference on Computer Vision (ECCV)*, ser. Lecture Notes in Computer Science, A. Heyden, G. Sparr, M. Nielsen, and P. Johansen, Eds., vol. 2350. Springer Verlag, 2002, pp. 128–142.

[10] ——, "Scale & affine invariant interest point detectors," *International Journal of Computer Vision (IJCV)*, vol. 60, no. 1, pp. 63–86, 2004.

[11] ——, "A performance evaluation of local descriptors," *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, vol. 27, no. 10, pp. 1615–1630, 2005.

[12] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. V. Gool, "A comparison of affine region detectors," *International Journal of Computer Vision (IJCV)*, vol. 65, no. 1-2, pp. 43–72, 2005.

[13] T. Tuytelaars and L. V. Gool, "Content-based image retrieval based on local affinely invariant regions," in *Visual Information and Information Systems*, ser. Lecture Notes in Computer Science, D. Huijsmans and A. Smeulders, Eds., vol. 1614. Springer Berlin / Heidelberg, 1999, pp. 493–500.

[14] ——, "Wide baseline stereo matching based on local, affinely invariant regions," in *British Machine Vision Conference (BMVC)*, 2000, pp. 412–425.

[15] G. Yu and J.-M. Morel, "A fully affine invariant image comparison method," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2009, pp. 1597–1600.

[16] K. Köser and R. Koch, "Perspectively invariant normal features," *IEEE International Conference on Computer Vision (ICCV)*, pp. 1–8, Oct 2007.

[17] Y. Ke and R. Sukthankar, "Pca-sift: A more distinctive representation for local image descriptors," in *IEEE International Conference on Computer Vision and Pattern Recognition (ICCV)*, 2004, pp. 506–513.

[18] C. Schmid, R. Mohr, and C. Bauckhage, "Comparing and evaluating interest points," *IEEE International Conference on Computer Vision and Pattern Recognition (ICCV)*, pp. 230–235, 1998.

[19] ——, "Evaluation of interest point detectors," *International Journal of Computer Vision (IJCV)*, vol. 37, no. 2, pp. 151–172, 2000.

[20] A. Haja, B. Jähne, and S. Abraham, "Localization accuracy of region detectors," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, jun 2008, pp. 1 –8.

[21] P. Moreels and P. Perona, "Evaluation of features detectors and descriptors based on 3d objects," in *IEEE International Conference on Computer Vision (ICCV)*, vol. 1, oct 2005, pp. 800 – 807.

[22] H. Bay, T. Tuytelaars, and L. V. Gool, "Surf: Speeded up robust features," in *European Conference on Computer Vision (ECCV)*, A. P. A. Leonardis, H. Bischof, Ed., vol. 3951. Berlin / Heidelberg, Germany: Springer, 2006, pp. 404–417.

[23] M. Grabner, H. Grabner, and H. Bischof, "Fast approximated sift," in *Asian Conference of Computer Vision (ACCV)*, ser. Lecture Notes in Computer Science, P. Narayanan, S. Nayar, and H.-Y. Shum, Eds., vol. 3851. Springer Berlin / Heidelberg, 2006, pp. 918–927.

[24] K. Cordes, O. Müller, B. Rosenhahn, and J. Ostermann, "Bivariate feature localization for sift assuming a gaussian feature shape," in *Advances in Visual Computing, 6th International Symposium (ISVC)*, ser. Lecture Notes in Computer Science, G. Bebis, Ed., vol. 6453, no. 1, nov 2010, pp. 264–275.

[25] T. Tuytelaars and K. Mikolajczyk, *Local invariant feature detectors: a survey*. Foundations and Trends in Computer Graphics and Vision, 2008, vol. 3.

[26] T. Dickscheid, F. Schindler, and W. Förstner, "Coding images with local features," *International Journal of Computer Vision (IJCV)*, pp. 1–21, 2010.

[27] R. I. Hartley and A. Zisserman, *Multiple View Geometry*, 2nd ed. Cambridge University Press, 2003.

[28] O. Urfalıoğlu, "Robust estimation of camera rotation, translation and focal length at high outlier rates," in *Proceedings of Canadian Conference on Computer and Robot Vision (CRV)*, may 2004, pp. 464 – 471.

[29] A. Weissenfeld, O. Urfalıoğlu, K. Liu, and J. Ostermann, "Robust rigid head motion estimation based on differential evolution," in *IEEE International Conference on Multimedia & Expo (ICME)*, jul 2006, pp. 225 – 228.

[30] K. Cordes, P. Mikulastik, A. Vais, and J. Ostermann, "Extrinsic calibration of a stereo camera system using a 3d cad model considering the uncertainties of estimated feature points," in *The 6th European Conference on Visual Media Production (CVMP)*, nov 2009, pp. 135–143.

[31] K. V. Price, R. Storn, and J. A. Lampinen, *Differential Evolution - A Practical Approach to Global Optimization*, ser. Natural Computing Series. Berlin, Germany: Springer, 2005.

[32] R. Storn and K. Price, "Differential evolution - a simple and efficient adaptive scheme for global optimization over continuous spaces," *Technical Report TR-95-012, ICSI*, 1995.