

NF-Features – No-Feature-Features for Representing Non-textured Regions

Ralf Dragon, Muhammad Shoaib, Bodo Rosenhahn, and Joern Ostermann

Institut fuer Informationsverarbeitung
Leibniz Universitaet Hannover
30167 Hannover, Germany
{dragon,shoaib,rosenhahn,ostermann}@tnt.uni-hannover.de

Abstract. In order to achieve a complete image description, we introduce no-feature-features (NF-features) representing object regions where regular interest point detectors do not detect features. As these regions are usually non-textured, stable re-localization in different images with conventional methods is not possible. Therefore, a technique is presented which re-localizes once-detected NF-features using correspondences of regular features. Furthermore, a distinctive NF descriptor for non-textured regions is derived which has invariance towards affine transformations and changes in illumination. For the matching of NF descriptors, an approach is introduced that is based on local image statistics.

NF-features can be used complementary to all kinds of regular feature detection and description approaches that focus on textured regions, i.e. points, blobs or contours. Using SIFT, MSER, Hessian-Affine or SURF as regular detectors, we demonstrate that our approach is not only suitable for the description of non-textured areas but that precision and recall of the NF-features is significantly superior to those of regular features. In experiments with high variation of the perspective or image perturbation, at unchanged precision we achieve NF recall rates which are better by more than a factor of two compared to recall rates of regular features.

1 Introduction

During the past years, the combination of interest point detector and local descriptor has been successfully applied in a high number of computer vision problems. The main reason for that is the fact that establishing local image

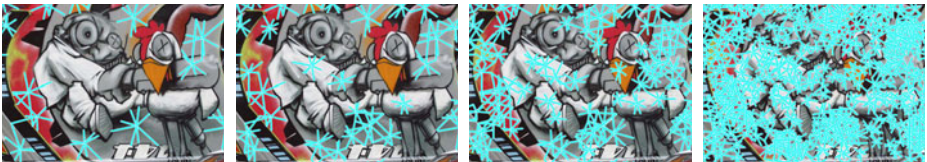


Fig. 1. Detected NF-features with increasing density using SIFT as regular features. The lines denote the extents that the feature descriptor is built from.

correspondences, which is one of the main computer vision problems, can be solved by inexpensive descriptor matching. As high repeatability under various external influences like changes in perspective or illumination is important for a stable matching, textured image regions with intensity variations in scale and space are chosen during the detection. Thereby, the local descriptor is created from high-entropy input data which results in distinctive descriptors.

However, broad categories of real-world objects have non-textured regions. Regular detectors very likely miss stable feature locations there. Additionally, most kinds of descriptors are built using local image gradients and thus lose their distinctiveness when built on non-textured areas. Hence, non-textured regions usually are not considered for detection and description.

In this paper we present a new feature type named *No-Feature features* (NF-features) that has the purpose to explicitly model regions without any features. This is inspired from the field of physics, in which the absence of electrons in conductors is modeled as the positively charged particle ‘electron hole’ with effective properties like mass and mobility. Likewise, NF-features are located on all those regions where any regular interest point detector left a ‘feature hole’. Thus NF-features results in a complete description for all regions of an image which enhances detection and classification.

The **contributions** of this paper are

- the development of a detection method for NF-features ensuring there is a minimal density in the extracted features,
- the derivation of a new descriptor in which contrast and intensity-shift invariant image content is stored,
- the derivation of a statistical descriptor matching method which evaluates the local image noise variance, and
- comprehensive evaluation of NF in combination with SIFT, MSER & GLOH, Hessian-Affine & GLOH and SURF as regular features.
- For further evaluation of NF we provide binaries [1].

The remaining sections are structured as follows: In Section 2, we give an overview of related work and explain the differences to our approach. In Section 3, the algorithms used for NF detection, description and matching are described. In Section 4, we show experiments and give a conclusion in Section 5.

2 Related Work

For stability and repeatability, all commonly-used interest point detectors detect image content that contains high entropy. By evaluating the second moment matrix, the Harris Corner Detector [2] detects interest points with intensity gradients that vary in two directions and thus are precisely locatable. In [3] points with extremal intensities are detected and by the evaluation of surrounding image contents with rays, an affine orientation is assigned. Maximally Stable Extremal Regions (MSER) [4] are detected by finding connected components which are extremal as they either have lower or higher intensity values than all surrounding pixels. These regions can be considered homogeneous or non-textured, but in

order to be detected, they necessarily have to exhibit a significant contour. The following detectors which are based on the scale-space, detect blob-like features not only in the spatial but also in the scale domain. The Harris-Affine detector [5], which is based on the Harris Corner Detector, evaluates the second moment matrix at a given scale and thus locates anisotropic blobs in the image. The Hessian-Affine (HAff) [6] detector works similar but the Hessian matrix is evaluated instead. The famous SIFT (Scale Invariant Feature Transform) detector [7] uses the Difference of Gaussians operator to locate features that correspond to isotropic blobs in the unscaled image. In [8], generalized junction-type features were proposed as interest points which are detected at different scales. The idea of detecting all kinds of maxima that exhibit spatial unpredictability is exploited in [9], where regions with maximal salience are detected. For a stable localization in the scale-space, all these methods detect only significant maxima. As non-textured regions with non-elliptic shape result in blurred maxima in the scale-space, they are usually not considered as keypoint location.

In [10], the fusion of complementary information similar to our method has been proposed. They use a contour descriptor combined with a local descriptor and get improved results for the combination. However, non-textured regions are still not covered with this approach. As they contain no contours and no texture, no interest points may be described.

Local descriptors are usually built from statistical parameters around the detected location. The SIFT descriptor [7] is built from histograms of gradient orientation at the detected scale. [11] proposed GLOH (Gradient Location and Orientation Histogram) which extends SIFT by changing the location grid and using PCA for compression. [12] proposed SURF (Speeded Up Robust Features), an efficient variation of SIFT by using integral images for a more-efficient computation while making additional approximations towards SIFT. For invariance towards monotonic changes in intensity, SMD (Stable Monotonic Change Invariant Descriptor) [13] was introduced, which analyzes intensity order changes.

The idea of sampling lines for the descriptor that originate from the keypoint location is also used for Spiders [14] and for the intensity-based region detector [3]. In [14], the lines are used to determine the extents of a feature by evaluating the intensity run along a line and choosing that border location, where the intensity falls below a threshold for the first time. Likewise in [3], that location is selected where an intensity expression becomes maximal. Thus in contrast to NF, in both works the lines are used to determine the extents of a feature to make it affine invariant.

3 NF Features

Creating NF-features in an image I_1 follows the same paradigm as state-of-the-art local features: First the feature is detected and then a descriptor is built from the local image content. As our approach is always complementary to features like SIFT that we call *regular features*, the NF detection has to be performed after regular features have been detected. To match NF-features between images I_1

and I_2 , we introduce a technique called second level matching, for which regular correspondences, NF features of I_1 , and the image I_2 have to be given. Second level matching consists of second level detection and second level description which are both explained at the end of the following two sections.

In the derivation, x_c denotes the location of a feature of set F_c , features of images I_1 and I_2 are distinguished by $x_c^{(1)}$ and $x_c^{(2)}$, and $D(x, y)$ denotes the Euclidean distance between two vectors x and y .

3.1 NF Detection

NF-features should be complementary to regular features. Thus we detect an NF-feature at every location x_{nf} where all regular features at $x_{\text{reg}} \in F_{\text{reg}}$ are far, thus $D(x_{\text{nf}}, x_{\text{reg}}) > d_{\text{far}}$. We define d_{far} as a constant factor c of the median nearest neighbor Euclidean distance d_{mnn} of the regular features:

$$d_{\text{far}} = c \cdot d_{\text{mnn}}(F_{\text{reg}}) = c \cdot \text{median}_{F_{\text{reg}}} [D(x_{\text{reg}}, \text{nearest}_{F_{\text{reg}}}(x_{\text{reg}}))] . \quad (1)$$

Choosing $c = 3$ yields to a good trade-off between dense NF coverage and computation speed. Likewise, we clip d_{mnn} if it falls below 10 pel. In Fig. 1, detections with increasing NF density using d_{far} from $5d_{\text{mnn}}$ down to $2d_{\text{mnn}}$ are displayed.

The detection is performed iteratively using an algorithm similar to the Farthest Point Sampling [15]. Given F_{reg} and all already detected NF-features F_{nf} , we seek for the location which is farthest from all known features $F_{\text{kf}} = F_{\text{reg}} \cup F_{\text{nf}}$. In other words, we seek for the center of the largest hole in F_{kf} . To find that location efficiently, the Delaunay triangulation is built for F_{kf} (cf. Fig. 2(a)). The edges of its dual graph, the Voronoi diagram, cover all points to which the distance to the nearest two neighbors is identical. Thus, Voronoi vertices cover all points which are locally farthest to all known feature locations. From all Voronoi vertices we choose that point with maximal distance d_{max} to F_{kf} as NF location. Features are located iteratively until d_{max} falls below d_{far} . Using this algorithm, we ensure no hole remains with a radius larger than d_{far} .

Keypoint detection should be robust and repeatable. However, this detection method is only repeatable if after a detection in image I_1 , exactly the same regular features are detected in another image I_2 . To overcome this, when regular correspondences $C_{\text{reg}} = \{(F_{\text{reg}}^{(1)}, F_{\text{reg}}^{(2)})\}$ have been found between images I_1 and I_2 , second level detection is performed. For each NF location $x_{\text{nf}}^{(1)}$ from image I_1 , a local transformation T to image I_2 is estimated using the nearest n regular features with correspondences in I_2 (cf. Fig. 2(b)). The local transformation $x^{(2)} = T(x^{(1)})$ is then applied to localize the NF-feature in image I_2 as

$$x_{\text{nf}}^{(2)} = T \left(x_{\text{nf}}^{(1)} \right) . \quad (2)$$

Thus, the NF keypoint always fulfills the same local motion as the nearest regular correspondences. We call the involved regular features *anchor features*, as the NF keypoint is fixed to these features and performs the same local motion.

Assuming that the anchor features as well as the NF-feature are coplanar, T is a homography. For a stable estimate, RANSAC is used with the nearest

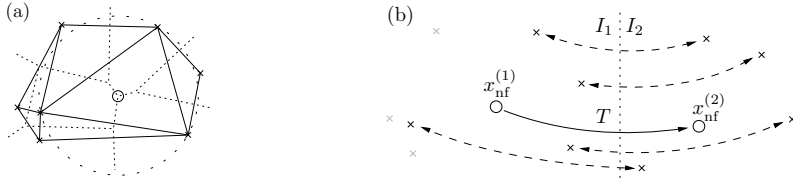


Fig. 2. (a) NF detection using the Delaunay triangulation (solid) and the Voronoi diagram (dotted). The NF-feature (small circle) is located at x_{nf} , which is that Voronoi vertex with maximum distance to all regular features (crosses). (b) Second level detection. Corresponding regular features from images I_1 and I_2 are used as anchor features in order to compute the local transformation $x^{(2)} = T(x^{(1)})$.

$n = 8$ correspondences, combined by a normalization step [16] and least-squares fitting. If RANSAC does not find a reasonable solution, there is no NF-feature located and thus no NF correspondence established. Likewise, any $x_{\text{nf}}^{(2)}$ too close to a regular feature is not considered, as I_2 is assumed to be textured at that location. By this second level detection, we obtain candidates for the matching.

3.2 NF Description

The NF descriptor should specify the contents of the region around its location. As there was no regular feature detected in that region, it is very likely that it is non-textured. However, the hull of the area which is built from the nearest regular points is textured. We exploit this transition from dull to featured within the NF descriptor: The descriptor is created by analyzing intensity runs from the NF location to the nearest regular features. To describe the whole area around the NF location, it is divided into 8 segments of same angle (Fig. 3(a)). In each segment, a line is sampled which runs towards the nearest regular feature inside the segment. If no regular feature has been detected or it is very far, we sample along the segment middle with a distance of $5d_{\text{mnn}}$, as we are unsure about the dimensions of the feature. By this sampling method, a hull around the NF location is formed with one hull point in each segment. We use the parametrization $t \in [0, 1]$ for the line starting at the NF location.

$$t_i = \frac{i}{N+1}, \quad i = 1 \dots N. \quad (3)$$

To extract the deviation from a smooth transition from dull to featured, the difference between $I(t)$ and the linear transition $I_{\text{lin}}(t) = I(0) + t \cdot (I(1) - I(0))$ is extracted for the descriptor (cf. Fig. 3(b)). For the j th segment, $j = 1 \dots 8$, we receive N samples δ_{ij} along the line towards the hull point:

$$\delta_{ij} = I(t_i) - I_{\text{lin}}(t_i) \quad (4)$$

$$= I(t_i) + (t_i - 1) \cdot I(0) - t_i \cdot I(1). \quad (5)$$

δ_{ij} is not contrast invariant, as $I^{(c)} = \alpha I$ yields $\delta^{(c)} = \alpha \delta$. As a stable measurement for normalization, we use the standard deviation of δ of all samples in the

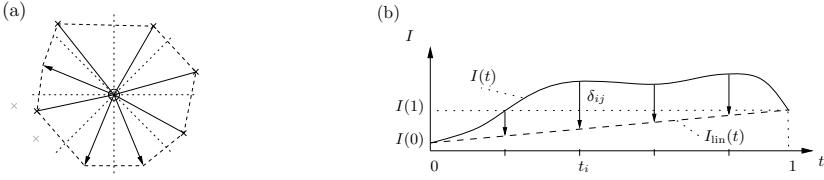


Fig. 3. (a) NF description by sampling lines (solid) from x_{nf} towards the hull points. The dotted lines denote the 8 segment borders. For every segment the nearest regular feature is selected as hull point. If no such feature exists, a point in the segment middle (arrows) with specific distance to x_{nf} is chosen as hull point. (b) Sampling δ_{ij} according to (5). The difference of the image intensity run $I(t)$ to the linear transition $I_{lin}(t) = I(0) + t \cdot (I(1) - I(0))$ is extracted.

NF hull. For further processing, a zero-mean descriptor d_{ij} is required:

$$d_{ij} = \frac{\delta_{ij} - \text{mean } \delta}{\sqrt{\text{var } \delta}}. \quad (6)$$

The NF descriptor d_{ij} , which has variance 1 and mean 0, specifies the contents of the convex region spanned by the hull points. To store it in a memory-efficient way, each element is quantized. We evaluated that a uniform quantization of 8 bit with clipping at ± 2 does not change the descriptors significantly.

As the image texture is assumed to be non-textured near to the NF sampling locations, small localization errors during the sampling can be neglected. Besides, we can sample the low-passed image signal with acceptable loss in precision to reduce the noise variance by a factor of s_{lp} . We use a Gaussian-shaped filter of size 7×7 and a variance of 1 that performs a suppression of $s_{lp} = 4.1$.

Computing the hull is not repeatable if other nearest regular features are used. Thus, we have to distinguish between first and second level descriptions again. When regular correspondences have been established and the NF-feature is located in image I_2 , we also transform the hull points $x_{hull}^{(1)}$ from I_1 to I_2 using $x_{hull}^{(2)} = T(x_{hull}^{(1)})$. Sampling is then performed analog to first level matching.

3.3 Descriptor Matching Using Local Noise Estimation

As the descriptor is invariant with respect to small offsets, the two possibilities for two descriptors to differ are that either the image content differs or the presence of image noise. The two cases are to be classified to consider two descriptors to differ or to match. Because the content of the NF hull is non-textured and thus very unlikely to contain any high-frequency patterns, it can be assumed that the local image variance is due to the noise. We use the high-pass filtered image signal around the NF location to estimate the variance of Gaussian-distributed image noise. For a more robust estimate, we collect estimations of the variance at all sampling locations and take the median value of all estimations.

To analyze if two descriptors match, the NF descriptor $d^{(1)}$ is compared element-wise with $d^{(2)}$ using the difference

$$e_{ij} = d_{ij}^{(1)} - d_{ij}^{(2)}. \quad (7)$$

If we assume that two segment descriptors do not match, $d^{(1)}$ and $d^{(2)}$ are two independent random variables which were each normalized in (6) to have $\text{var } d = 1$. Assuming d_{ij} being Gaussian-distributed, e_{ij} is also Gaussian-distributed with $\text{var } e = 2$. On the other side if we assume that the descriptors match perfectly, then $\text{var } e = 0$. Among all perturbations which lead to non-perfect descriptor matches, image noise is the only one that is not due to image contents changes. Thus we further estimate the influence of image noise on e_{ij} .

First, we look at the influence of additive zero-mean Gaussian-distributed image noise with variance σ^2 on the samples $I(t)$ from (5). $I(0)$, $I(1)$ and $I(t_i)$ become independent random variables. During the sampling, their variance was reduced from the low-pass filter by s_{lp} . Thus, we have

$$\text{var } I(t) = \frac{\sigma^2}{s_{\text{lp}}}. \quad (8)$$

As the three random variables are scaled in (5) by factors of 1, $t_i - 1$ and t_i respectively, we get

$$\text{var } \delta_{ij} = \frac{\sigma^2}{s_{\text{lp}}} (1^2 + (t_i - 1)^2 + t_i^2) = 2 \frac{\sigma^2}{s_{\text{lp}}} (1 - t_i + t_i^2). \quad (9)$$

We assume that during the normalization of the descriptor in (6), the influence of the image noise on $\text{var } \delta$, which was created from the whole descriptor, is negligible compared to the influence on δ_{ij} . The influence of σ^2 on d_{ij} is thus

$$\text{var } d_{ij} = \frac{2\sigma^2}{s_{\text{lp}} \text{var } \delta} (1 - t_i + t_i^2) = 2\bar{\sigma}^2 (1 - t_i + t_i^2), \quad (10)$$

where we introduce $\bar{\sigma}^2$ as normalized image noise variance. When we compare two independent descriptor elements according to (7), the variance of e_{ij} is

$$\text{var } e_{ij} = 2(\bar{\sigma}_1^2 + \bar{\sigma}_2^2)(1 - t_i + t_i^2). \quad (11)$$

We can assume that the image noise variance is constant in the sampled area. Thus we can use (3) to derive the expected variance $E[\text{var } e_{ij}]$ of the random variable e_{ij} over all realizations (i, j) . An estimation of $E[\text{var } e_{ij}]$ can be found by computing the element-wise mean square distance (MSD) between two zero-mean descriptors:

$$E[\text{var } e_{ij}] = D_{\text{MSD}}(d^{(1)}, d^{(2)}) = 2(\bar{\sigma}_1^2 + \bar{\sigma}_2^2) \frac{1}{N} \sum_{i=1}^N (1 - t_i + t_i^2) \quad (12)$$

$$= 2(\bar{\sigma}_1^2 + \bar{\sigma}_2^2) r_N, \quad (13)$$

where r_N is a scale factor ($\frac{3}{4} \leq r_N < \frac{5}{6}$) that increases with growing number of samples:

$$r_N = \frac{5N + 4}{6N + 6}. \quad (14)$$

With an a-priori probability of p for two matching line segments, we set the classification border to the weighted middle between the expected values.

$$b = 2(1 - p) + 2p(\bar{\sigma}_1^2 + \bar{\sigma}_2^2)r_N \quad (15)$$

As there is no a-priori information about the area between the features, p is set to 0.5.

$$b = 1 + (\bar{\sigma}_1^2 + \bar{\sigma}_2^2)r_N \quad (16)$$

The theoretical limit where NF-features are not classifiable is at a mean image noise variance of

$$\sigma_{\max}^2 = \frac{s_{lp} \text{var } \delta}{2 r_N}. \quad (17)$$

It seems we can handle all noisy images with a high noise suppression s_{lp} or with a high number of samples N , but the variance of δ will also decrease by this: As the image content of the sampled line is non-textured, the difference to the linear run sampled in (5) mainly contains low-frequency patterns. So the theoretical limit depends on the image contents and thus cannot be derived here.

We empirically determined $N = 4$ samples per line for regular camera images. Thus, we achieve the following descriptor size: when using the here-proposed parameters (8 segments with 4 samples per segment, 8 bit descriptor quantization) the descriptor only occupies 32 byte. With this approach, we have to additionally store the intra-segment angles (8 bit) and the distances of the 8 hull points (8 bit) as well as the normalized image variance with high precision (32 bit). Thus, the NF descriptor size is 52 byte.

4 Experiments

We first demonstrate the properties of NF in a cluttered environment (Fig. 4). Using NF-SIFT, we match 3 T-shirts worn by 6 different persons under different illumination conditions. It can be seen that regular features match only at the T-shirt logos whereas NF-features match on most of the T-shirt area. Further, occluded or changed image contents like the faces is not matched.

To show the performance of our algorithm, we use natural image pairs of sequences for the evaluation of affine invariant features from [11]. To demonstrate illumination invariance, we use the Memorial sequence from [17] originally used to create high dynamic range images. For a larger experiment, we use the Amsterdam Library of Object Images (ALOI) [18] which includes 1000 images under varying illumination conditions. To demonstrate NF-features are useful in combination with different kinds of feature types, we use SIFT, MSER, Hessian-Affine (HAff) and SURF as regular features, where for MSER and Hessian-Affine GLOH is used as descriptor.

As NF uses second level matching and as the NF descriptor comparison is not based on a nearest neighbor similarity, evaluating recall vs. precision graphs by varying thresholds is not suitable here. Instead we focus on the common way of establishing regular correspondences using second nearest neighbor (2NN)

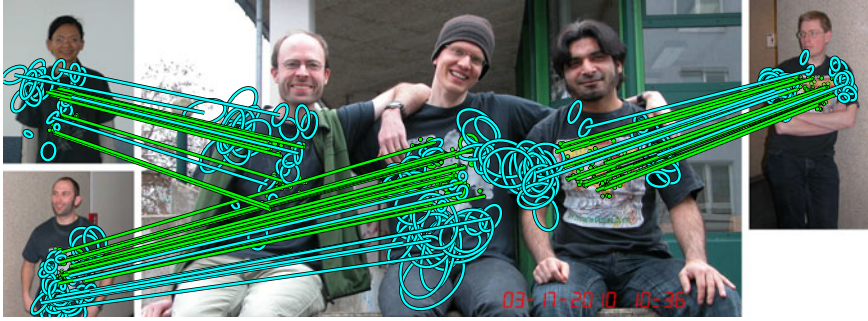


Fig. 4. NF (cyan) and SIFT (green) correspondences between 3 T-shirts worn by 6 different persons in a cluttered environment. The NF features are located on non-textured T-shirt regions. The correspondence lines are thinned out by a factor of 5.

matching as proposed in [7]. We then build NF correspondences, estimate them separately from the regular correspondences and compare all eight cases.

To measure performance, we use precision and recall of the extracted keypoints of the first image I_1 of every image pair (I_1, I_2) according to (18), where true positives (tp), false positives (fp) and false negatives (fn) are counted.

$$\text{recall} = \frac{\text{tp}}{\text{tp} + \text{fn}}, \quad \text{precision} = \frac{\text{tp}}{\text{tp} + \text{fp}}. \quad (18)$$

To verify the correspondences in the case of sequences with a moving camera, homography matrices supplied with the test material are used as ground truth. As we focus on dense object description for object recognition, we want to count an imprecise localization as inlier in contrast to ‘real’ outliers with false correspondences. For true positive correspondences we thus accept a maximum deviation of 15 pel from the ground truth which is approximately the average d_{mnn} of the ALOI image database. Correspondences with a higher distance are classified as false positives. All features detected only in the first image of the illumination image pair are counted as false negatives. True negatives are not analyzed as no significant occlusions exist in the sequences.

The runtime for processing NF-features using our non-optimized code depends on the image contents. If there are many regular features near each other (e.g. Fig. 6(b)) and large areas are unsampled, the iterative sampling algorithm from Section 3.1 covers large unsampled regions. In such cases, the runtime is up to 10 times the processing time of regular SIFT features. In images, where holes between regular features are filled (e.g. Fig. 1) NF matching needs roughly twice the processing time.

4.1 Descriptor Invariance in Image Sequences

First we examine the influence of changes in global illumination (‘Leuven’), in changes of internal (‘Bark’, ‘Boat’) and external camera parameters (‘Graf’,

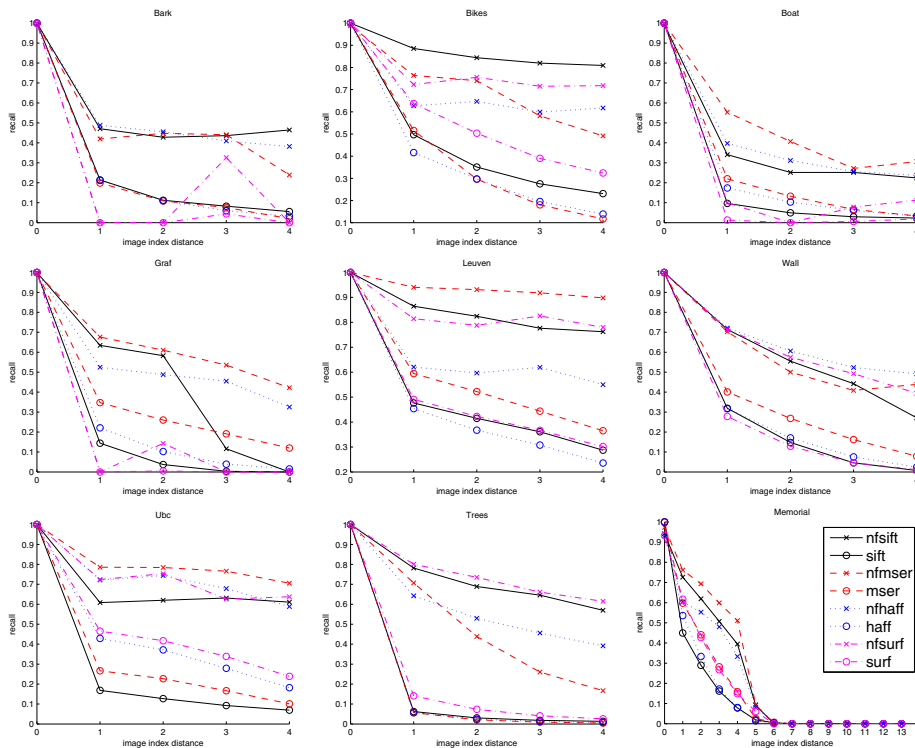


Fig. 5. Recall over the image index distance in the sequences ‘Bark’ (zoom and rotation), ‘Bikes’ (blur), ‘Boat’ (zoom and rotation), ‘Graf’ (viewpoint), ‘Leuven’ (illumination), ‘Wall’ (viewpoint), ‘Ubc’ (JPEG compression), ‘Trees’ (blur) and ‘Memorial’ (dynamic range). Crosses denote NF features, circles regular features.

‘Wall’), in adding blur (‘Bikes’, ‘Trees’) and JPEG artifacts (‘Ubc’) and in variations of the dynamic range (‘Memorial’). For each series, correspondences between all image pairs are established using 2NN for regular features and the here-presented methods for NF. To analyze the descriptor invariance, we reduce the effect of wrong second level detection due to false regular correspondences (We further analyze this in Section 4.3). We enforce high precision by a loose outlier filtering of the regular correspondences using RANSAC to estimate a homography from the unfiltered correspondences. By this the precision values are similar (almost all above 0.9). So we can compare the approaches using the recall value, which we average between all image pairs of the same image index distance (Fig. 5).

It can be seen that for all sequences NF always get higher recall values than regular features, often with a factor of more than two. This is positively surprising as regular features serve as anchor points for the second level detection. Thus we can deduce that NF descriptors have better invariance properties towards illumination changes, blurring, JPEG compression and affine transformations than regular features.

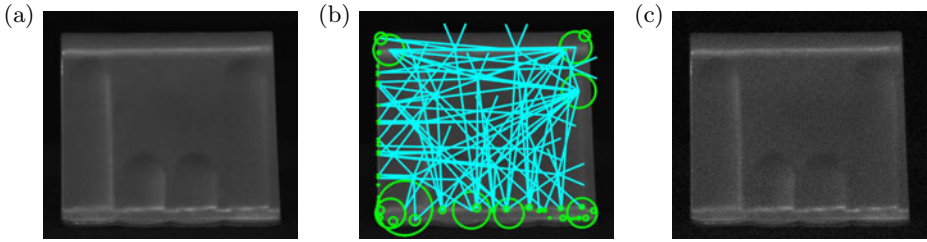


Fig. 6. (a) Original image 256 of the ALOI database. (b) The sampled lines of the detected NF-features (cyan) and SIFT anchor points (green circles denoting the extent). (c) The examined image with added image noise variance of 100.

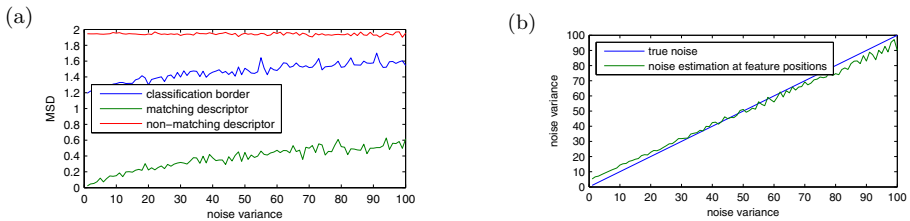


Fig. 7. (a) Influence of image noise on the descriptor difference variance from image 256 of the ALOI database. (b) Noise variance estimation at the feature locations vs. true noise variance.

4.2 Influence of Image Noise onto the Descriptor Distinctiveness

In this experiment, the distinctiveness of the NF descriptor is analyzed. Topological information from regular correspondences is not considered. In contrast to all kinds of regular state-of-the-art descriptors, NF descriptors are directly built from image intensity values. Thus the influence of image noise on the distinctiveness of the descriptor seems to be crucial. We now analyze the behavior of NF towards noise. Therefore we use the center view of object 256 of the ALOI database which shows a non-textured surface that has regular features at the borders only (Fig. 6). This means there is no transition from non-textured to textured during the line sampling for the descriptor. This is highly-crucial as the NF descriptor has low variance (cf. (17)) and by this image noise has a high impact on the matching result. We detect NF locations on the object and compute the descriptor differences while adding Gaussian image noise. Using the descriptor MSD, we compare descriptors which should match and those which should not match (Fig. 7(a)).

It can be seen that the descriptor difference variance runs as expected: For non-matching features it is independent from the image noise and reliably at approximately 2, where it grows from 0 with increasing noise if the features match. However, we have small systematic deviations from the derived model concerning the estimation of the image noise (cf. Fig. 7(b)). Besides, the MSD of non-matching features is slightly but significantly smaller than 2 which means

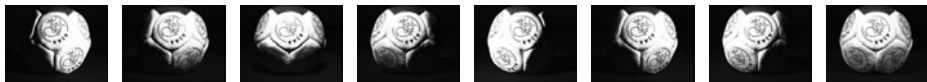


Fig. 8. ALOI object 103 with varying illumination $l_1 \dots l_8$ viewed from camera $c1$



Fig. 9. Matching regular (green ellipses) and NF-features (cyan stars) of object 113 from the ALOI image database using NF-SIFT, NF-MSER, NF-HAff and NF-SURF

that they are statistically dependent. However these deviations are small and likely to be overestimated in this experiment because of the large untextured area. So we can deduce that NF descriptors are distinctive under the influence of image noise, even for small changes in image contents.

4.3 Image Database

We use the series ‘Illumination Direction Collection’ from ALOI with camera $c1$ in which one object is observed by a static camera during 8 different illumination conditions $l_1 \dots l_8$ (Fig. 8). Illumination $l_1 \dots l_5$ were taken at angles of $-60^\circ \dots 60^\circ$ in steps of 30° . l_6 and l_7 were taken combining the side illuminations $l_1 + l_2$ and $l_4 + l_5$ respectively. l_8 is all illuminations combined.

We establish correspondences of each object illumination setting with each other illumination setting of the same object. To measure the impact of false NF correspondences due to false anchor point correspondences, we do not filter any correspondence like in Section 4.1. However, we allow the correspondence cluster filtering which is performed in SIFT, as it is an essential part of the algorithm. We compute recall and precision values according to (18) for all illumination pairs of the same object and average them. The results are plotted as precision and recall matrices over the eight illumination setting in Fig. 10 and in the form of precision and recall graphs over the angle of illumination change in Fig. 11.¹ In Fig. 9 we show a comparison for the detection of all examined NF feature combinations.

Generally two tendencies can be observed: Concerning the precision, images with similar illumination, e.g. (l_1, l_6) , have higher precision values for regular features. With increasing variations of the illumination, NF outperforms regular features in precision, e.g. (l_1, l_5) . However, NF-HAff show inferior results

¹ Please note that NF recall of identical images does not necessarily have to equal to 1, as the estimation of the local motion model T may fail if there are too few suitable anchor points available.

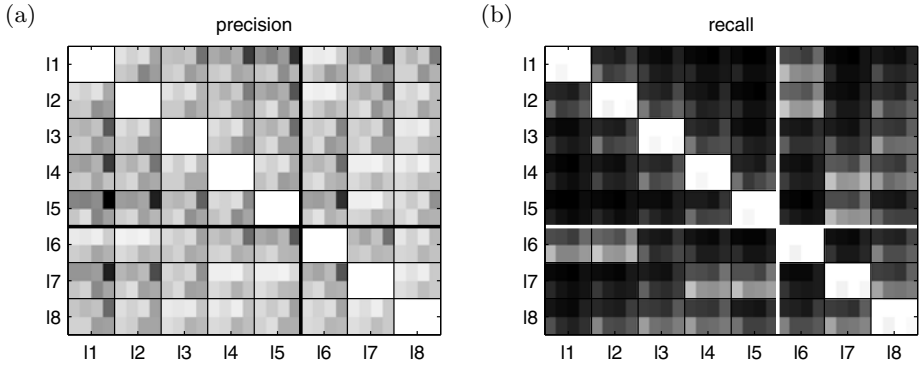


Fig. 10. Precision (a) and recall (b) matrix for the illumination conditions $l_1 \dots l_8$, where $l_1 \dots l_5$ are single-illuminated images and l_6, l_7 and l_8 are illuminated with a combinations thereof. In each square, regular (top) and NF features (bottom) with are opposed using (starting left) SIFT, MSER, HAff and SURF. White denotes precision and recall of 1, black a precision of 0.72 and a recall of 0.06.

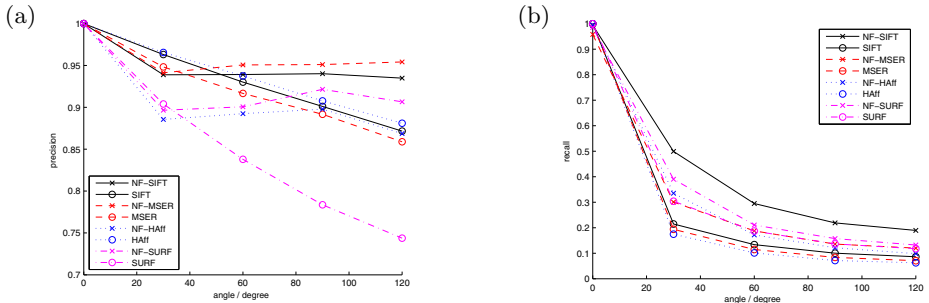


Fig. 11. Mean precision (a) and recall (b) over the angle of illumination change using NF-SIFT, SIFT, NF-MSER, MSER, NF-HAff, HAff, NF-SURF and SURF

compared to HAff. Concerning the recall rates, NF always outperforms regular features, where NF-SIFT is by far better than the other three NF combinations.

5 Conclusion

We derived a framework for NF-features which is complementary to every regular interest point detection approach with local descriptors. During detection, centers of regions unsampled by a regular feature detection are determined as NF locations. The second level matching algorithm re-locates suitable NF features in further images according to a local transformation which is extracted from already-established regular correspondences. The descriptor is built by sampling lines from the non-textured NF location to the nearest regular feature locations.

Using standard test material and enforcing high precision, we demonstrated that the repeatability of NF-features is significantly improved towards regular features, often by a factor of more than two. In a challenging experiment with high variations of the illumination without outlier filtering, we also achieved significantly better results concerning recall and precision. Thus, NF-features are not only useful for a complete description of the image contents but also improve recall and precision rates. For further evaluation, we provide binaries [1] that may be combined with any type of regular features.

References

1. NF Project Website, <http://www.tnt.uni-hannover.de/project/nf>
2. Harris, C., Stephen, M.: A combined corner and edge detector. In: Fourth Alvey Vision Conference, pp. 147–151 (1988)
3. Tuytelaars, T., van Gool, L.: Matching widely separated views based on affine invariant regions. *IJCV* 59(1), 61–85 (2004)
4. Matas, J., Chum, O., Urban, M., Pajdla, T.: Robust wide-baseline stereo from maximally stable external regions. *Image and Vision Computing* (2004)
5. Mikolajczyk, K., Schmid, C.: An affine invariant interest point detector. In: *ICCV* (2002)
6. Mikolajczyk, K., Schmid, C.: Scale and affine invariant interest point detector. *IJCV* 60, 63–86 (2004)
7. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. *IJCV* 60, 91–110 (2004)
8. Förstner, W., Dickscheid, T., Schindler, F.: Detecting interpretable and accurate scale-invariant keypoints. In: *ICCV*, pp. 2256–2263 (2009)
9. Kadir, T., Zisserman, A., Brady, M.: An affine invariant salient region detector. In: Pajdla, T., Matas, J(G.) (eds.) *ECCV 2004*. LNCS, vol. 3021, pp. 228–241. Springer, Heidelberg (2004)
10. Ferrari, V., Fevrier, L., Jurie, F., Schmid, C.: Groups of adjacent contour segments for object detection. *TPAMI* 30(1), 36–51 (2008)
11. Mikolajczyk, K., Schmid, C.: A performance evaluation of local descriptors. *TPAMI* 27(10), 1615–1630 (2005)
12. Bay, H., Tuytelaars, T., van Gool, L.: Surf: Speeded up robust features. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) *ECCV 2006*. LNCS, vol. 3951, pp. 404–417. Springer, Heidelberg (2006)
13. Gupta, R., Mittal, A.: Smd: A locally stable monotonic change invariant feature descriptor. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) *ECCV 2008*, Part IV. LNCS, vol. 5305, pp. 265–277. Springer, Heidelberg (2008)
14. Stanski, A., Hellwich, O.: Spiders as robust point descriptors. In: Kropatsch, W.G., Sablatnig, R., Hanbury, A. (eds.) *DAGM 2005*. LNCS, vol. 3663, pp. 262–268. Springer, Heidelberg (2005)
15. Eldar, Y., Lindenbaum, M., Porat, M., Zeevi, Y.Y.: The farthest point strategy for progressive image sampling. *TIP* 6(9), 1305–1315 (1997)
16. Hartley, R.I.: In defense of the eight-point algorithm. *TPAMI* 19, 580–593 (1997)
17. Debevec, P.E., Malik, J.: Recovering high dynamic range radiance maps from photographs. In: *SIGGRAPH*, pp. 369–378 (1997)
18. Geusebroek, J.M., Burghouts, G.J., Smeulders, A.W.M.: The Amsterdam Library of Object Images. *IJCV* 61(1), 103–112 (2005)