An Evaluation of Local Feature Detectors and Descriptors for Infrared Images

Johan Johansson[1], Martin Solli[2], and Atsuto Maki[1]

[1] Royal Institute of Technology (KTH), Sweden
[2] FLIR Systems AB, Sweden
{johanj4,atsuto}@kth.se, martin.solli@flir.se

Abstract. This paper provides a comparative performance evaluation of local features for infrared (IR) images across different combinations of common detectors and descriptors. Although numerous studies report comparisons of local features designed for ordinary visual images, their performance on IR images is far less charted. We perform a systematic investigation, thoroughly exploiting the established benchmark while also introducing a new IR image data set. The contribution is two-fold: we i) evaluate the performance of both local float type and more recent binary type detectors and descriptors in their combinations under a variety (6 kinds) of image transformations, and ii) make a new IR image data set publicly available. Through our investigation we gain novel and useful insights for applying state-of-the art local features to IR images with different properties.

Keywords: Infrared images, local features, detectors, descriptors

1 Introduction

Thermography, also known as infrared (IR) imaging or thermal imaging, is a fast growing field both in research and industry with a wide area of applications. At power stations it is used to monitor the high voltage systems. Construction workers use it to check for defective insulation in houses and firefighters use it as a tool when searching for missing people in buildings on fire. It is also used in various other contexts for surveillance.

In the field of image analysis, especially within computer vision, the majority of the research have focused on regular visual images. Many tasks there comprise the usage of an interest point or feature detector in combination with a feature descriptor. These are, for example, used in subsequent processing to achieve panorama stitching, content based indexing, tracking, reconstruction, recognition etc. Hence, local features play essential roles there, and their development and evaluations have, for many years, been an active research area, resulting in rich knowledge on useful detectors and descriptors.

A research question we pose in this paper is how we can exploit those local features in other types of images, in particular, IR spectral band images, which has been less investigated. The fact that IR images and visual images have different characteristics, where IR images typically contain less high frequency information, necessitates an independent study on the performances of

common local detectors in combination with descriptors in IR images. In this context, the contributions of this paper are:
1) the systematic evaluation of local detectors and descriptors in their combinations under six different image transformations using established metrics, and
2) a new IR image database (http://www.csc.kth.se/~atsuto/dataset.html).

1.1 Related Work

For visual images several detectors and descriptors have been proposed and evaluated in the past. Mikolajczyk and Schmid [1] carried out a performance evaluation of local descriptors in 2005. The local descriptors were then tested on both circular and affine shaped regions with the result of GLOH [1] and SIFT [2] to have the highest performance. They created a database consisting of images of different scene types under different geometric and photometric transformations, which later became a benchmark for visual images.

A thorough evaluation of affine region detectors was also performed in [3]. The focus was to evaluate the performance of affine region detectors under different image condition changes: scale, view-point, blur, rotation, illumination and JPEG compression. Best performance in many cases was obtained by MSER [4] followed by Hessian-Affine [5, 6] and Harris-Affine [5, 6].

Focusing on fast feature matching, another evaluation [7] was performed more recently for both detectors and descriptors: the comparison of descriptors shows that novel real valued descriptors LIOP [8], MRRID [9] and MROGH [9] outperform state-of-the-art descriptors of SIFT and SURF [10] at the expense of decreased efficiency. Our work is partly inspired by yet another recent evaluation [11] involving exhaustive comparisons on binary features although those are all for visual images.

IR images have been studied in problem domains such as face recognition [12, 13], object detection/tracking [14, 15], and image enhancement of visual images using near-infrared images [16], to name a few. With respect to local features, a feature point descriptor was addressed for both far-infrared and visual images in [17] while a scale invariant interest point detector of blobs was tested against common detectors on IR images in [18]. This work however did not use the standard benchmark evaluation framework which kept itself from being embedded in comparisons to other results.

The most relevant work to our objective is that of Ricaurte et al. [19] which evaluated classic feature point descriptors in both IR and visible light images under image transformations: rotation, blur, noise and scale. It was reported that SIFT performed the best among several considered descriptors in most of their tests while there is not a clear winner. Nevertheless, unlike their studies on visual images, the evaluation was still limited in that it did not test different combinations of detectors and descriptors while also opting out view-point changes. Nor was it based on the standard evaluation framework [1, 3].

To the best of the authors' knowledge, a thorough performance evaluation for combinations of detectors and descriptors was yet to be made on IR images.

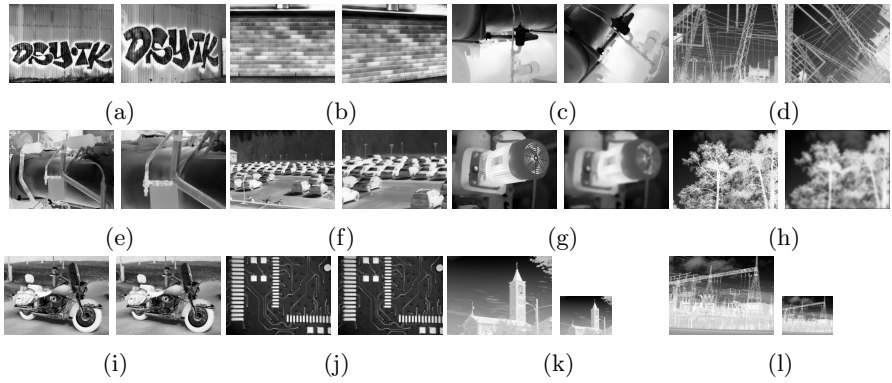

Fig. 1: Examples of images, under various deformations, that are included in the data set. Each image pair consists of a reference image, left, and a test image, right. (a,b) Viewpoint, (c,d) Rotation, (e,f) Scale, (g,h) Blur, (i,j) Noise, (k,l) Downsampling.

2 Evaluation Framework

The benchmark introduced in [1,3] made well established evaluation frameworks for measuring the performance of detectors and descriptors. We thus choose to use those to ensure the reliability and comparability of the results in this work.

2.1 Matching

To obtain matching features we use nearest neighbours (NN). To become a NN match the two candidates have to be the closest descriptors in descriptor space for both descriptors. The distance between features are calculated with the Euclidean distance for floating point descriptors whereas the Hamming distance is applied to binary descriptors.

Further, a descriptor is only allowed to be matched once, also known as a putative match [11]. Out of the acquired matches, correct matches are identified by comparing the result to the ground truth *correspondences*. The correspondences are the correct matching interest points between a test image and a reference image. For insight in how the ground truth is created see [1].

2.2 Region Normalization

When evaluating descriptors a measurement region larger than the detected region is generally used. The motivation is that blob detectors such as Hessian-Affine and MSER extract regions with large signal variations in the borders. To increase the distinctiveness of extracted regions the measurement region is increased to include larger signal variations. This scale factor is applied to the extracted regions from all detectors. A drawback of the scaling would be the risk of reaching outside the image border.

In this work we implement an extension of the region normalization used in [3] (source available) to expand the image by assigning values to the unknown area by bilinear interpolation on account of the border values.

As detected regions are of circular or elliptical shape all regions are normalized to circular shape of constant radius to become scale and affine invariant.

2.3 Performance Measures

Recall Recall is a measure of the ratio of correct matches and correspondences (defined in [1]). The measure therefore describes how many of the ground truth matches were actually found.

$$Recall = \frac{\#Correct\ matches}{\#Correspondences} \tag{1}$$

1−Precision The 1−Precision measure portraits the ratio between the number of false matches and total number of matches (defined in [1]).

$$1 - Precision = \frac{\#False\ matches}{\#Putative\ matches} \tag{2}$$

Matching Score MS is defined as the ratio of correct matches and again the number of *detected features* visible in the two images.

$$Matching\ Score = \frac{\#Correct\ matches}{\#\ Detected\ features} \tag{3}$$

2.4 Database

We have generated a new IR image data set for this study. The images contained in the database can be divided into the categories *structured* and *textured* scenes. A textured scene has repetitive patterns of different shapes while a structured scene has homogeneous regions with distinctive edges. In Figure 1, examples of structured scenes are presented in the odd columns of image pairs and those of textured scenes in the even columns of image pairs. Out of the standard images, captured by the cameras, the database is created by synthetic modification to include the desired image condition changes. An exception is for view-point changes where all images (of mostly planar scenes) were captured with a FLIR T640 camera without modification. The database consists of 118 images in total.

Deformation Specification The image condition changes we include in the evaluation are six-fold: view-point, scale, rotation, blur, noise and downsampling.

- Images are taken from different view-points starting at 90° angle to the object. The maximum view-point angle is about 50-60° relative to this.
- Zoom is imitated by scaling the height and width of the image using bilinear interpolation. The zoom of the image is in the range $\times$[1.25-2.5] zoom.
- Rotated images are created from the standard images by 10° increments.
- Images are blurred using a Gaussian kernel of size 51$\times$51 pixels and standard deviation up to 10 pixels.
- White Gaussian noise is induced with increasing variance from 0.0001 to 0.005 if the image is normalized to range between 0 and 1.
- Images are downsampled to three reduced sizes; by a factor of 2, 4 and 8.

2.5 Implementation Details

Local features are extracted using OpenCV [20] version 2.4.10 and VLFeat [21] version 0.9.20 libraries. OpenCV implementations are used for SIFT, SURF, MSER, FAST, ORB, BRISK, BRIEF [22] and FREAK [23] while Harris-Affine, Hessian-Affine and LIOP are VLFeat implementations. Unless explicitly stated the parameters are the ones suggested by the authors.

IR images are loaded into MATLAB R2014b using FLIR's Atlas SDK. When loaded in MATLAB the IR images contain 16 bit data which are quantized into 8 bit data and preprocessed by histogram equalization.

To calculate the recall, MS and 1−precision, this work utilizes code from [3].

Parameter Selection VLFeat includes implementations of Harris-Laplace and Hessian-Laplace with the possibility of affine shape estimation. To invoke the detectors functions there are parameters to control a peak and edge threshold.

The peak threshold affects the minimum acceptable cornerness measure for a feature to be considered as a corner in Harris-Affine and equivalently a blob by the determinant of the Hessian matrix in Hessian-Affine. According to the authors of [5] the used value for the threshold on cornerness was 1000. As no similar value is found to the Hessian-Affine threshold it is selected to 150. With the selected threshold the number of extracted features is in the order of magnitude as other detectors in the evaluation. The edge threshold is an edge rejection threshold and eliminates points with too small curvature. It is selected to the predetermined value of 10.

Regarding the region normalization, we choose a diameter of 49 pixels whereas 41 pixels is chosen arbitrary in [1]. The choice of a larger diameter is based on the standard settings in the OpenCV library for the BRIEF descriptor.

3 Evaluation Results

This section presents the results of combinations of the detectors and descriptors which are listed in Table 1. The evaluation is divided into floating point and binary point combinations, with the exceptions Harris-Affine combined with ORB and BRISK, and SURF combined with BRIEF and FREAK. The combination of Harris-Affine and ORB showed good performance in [11], while BRISK is combined with Harris-Affine as the descriptor showed good performance throughout this work. SURF is combined with binary descriptors as it is known to outperform other floating point detectors in computational speed. Combinations which are also tested in the evaluation on visual images in [7].

Evaluated combinations are entitled by a concatenation of the detector and descriptor with `hes` and `har` being short for Hessian-Affine and Harris-Affine. In case of no concatenation, e.g. `orb`, both ORB detector and descriptor are applied.

The performances are presented in precision-recall curves for the structured scenes, Figure 2, and for the textured scenes, Figure 3. We also present the

average results of both scene types in recall, precision and MS for each transformation. Here the threshold is set to accept all obtained matches as a threshold would be dependent on descriptor size and descriptor type.

3.1 Precision-Recall Curve

Recall and 1−Precision are commonly combined to visualize the performance of descriptors. It is created by varying an acceptance threshold for the distance between two NN matched features in the descriptor space. If the threshold is small, one is strict in acquiring correct matches which leads to high precision but low recall. A high threshold means that we accept all possible matches which leads to low precision, due to many false positives, and a high recall since all correct matches are accepted. Ideally a recall equal to one is obtained for any precision. In real world applications this is not the case as noise etc. might decrease the similarity between descriptors. Another factor arises while regions can be considered as correspondences with an overlap error up to 50%, hence descriptors will describe information in areas not covered by the other region. A descriptor with a slowly increasing curve indicates that it is affected by the image deformation.

3.2 Results

View-Point The effect of view-point changes on different combinations is illustrated in Figure 2a and Figure 2b for the structured scene in Figure 1a, while the results of the textured scene in Figure 1b are presented in Figure 3a and Figure 3b. The average of the performances against perspective changes in the structured and textured scenes are presented in Table 2a.

From the result it is clear that the performance varies depending on the scene and the combination. In the structured scene all combinations show dependency to perspective changes by a slow continuous increase in recall. Best performances among floating point descriptors are obtained by `mser-liop` and `hes-liop`.

Among binary combinations the best performance is obtained by `orb-brisk`, for both scenes, with results comparable to the best performers in the floating point family of combinations. Consecutive in performance are `orb` and `orb-freak` indicating how combinations based on the ORB detector outperform other binary combinations based on BRISK and FAST.

Table 1: Included binary and floating point detectors and descriptors. Binary types are marked with (*).

Detectors		Descriptors	
Hessian-Affine	SURF	LIOP	BRISK*
Harris-Affine	BRISK*	SIFT	FREAK*
MSER	FAST*	SURF	ORB*
SIFT	ORB*	BRIEF*	

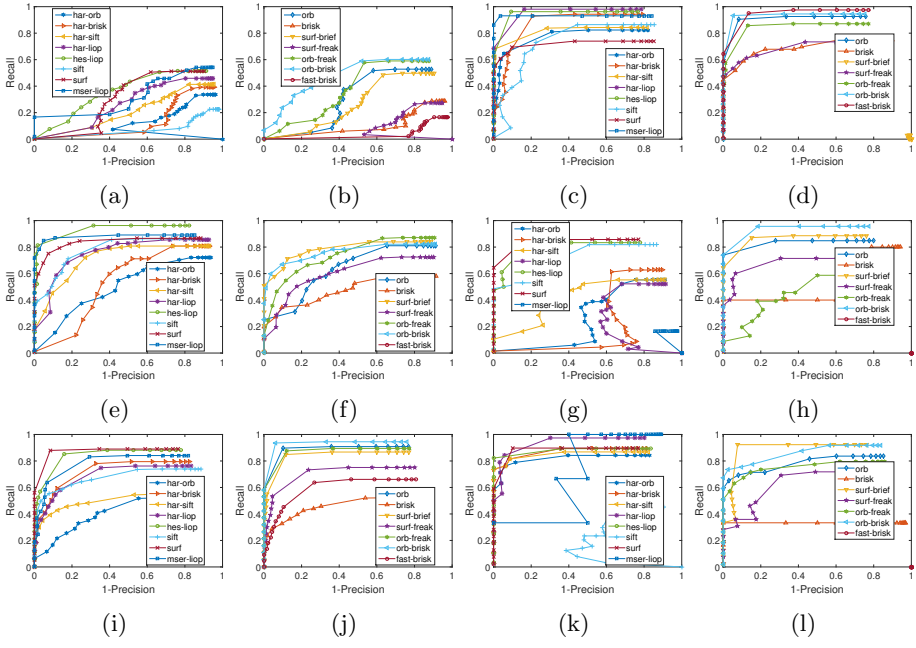

Fig. 2: Performance against view-point (a) & (b), rotation (c) & (d), scale (e) & (f), blur (g) & (h), noise (i) & (j), downsampling (k) & (l) in *structured scenes.*

Rotation The results of the combinations due to rotation are illustrated in Figure 2c and Figure 2d, for the structured scene in Figure 1c, and in Figure 3c and Figure 3d, for the textured scene in Figure 1d. The average results are presented in Table 2b.

We observe that the overall performance is much higher for rotation than for view-point changes. The majority of combinations has high performance in both the structured and textured scene. Figure 2d shows an illustrative example of how different detectors and descriptors perform in different setups. For example `surf-brief`, with BRIEF known to be sensitive to rotation, has a poor performance while `surf-freak` and `surf`, still indicating a dependence to rotation, have a greatly improved performance. The poor performance of `surf-brief` is shown by its fixed curve at a low precision and recall in the lower right corner.

Overall best performance is achieved by `hes-liop` followed by `har-liop` among floating point combinations, while for binary methods best performance is obtained by `orb-brisk`, `orb` and `orb-freak`.

Scale The effects from scaling are shown in Figure 2e and Figure 2f for the structured scene in Figure 1e, and in Figure 3e and Figure 3f for the textured scene in Figure 1f. The average results are presented in Table 2c.

The combinations show a stable behavior with similar performance in both scene types. Among floating point combinations, best performance is achieved

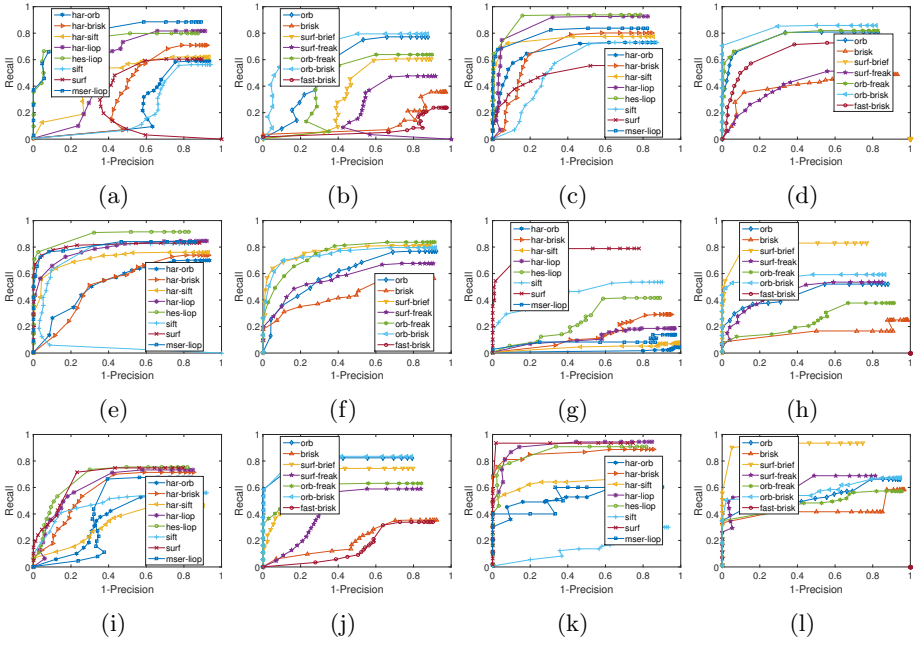

Fig. 3: Performance against view-point (a) & (b), rotation (c) & (d), scale (e) & (f), blur (g) & (h), noise (i) & (j), downsampling (k) & (l) in *textured scenes.*

by `mser-liop` succeeded by `hes-liop`.The top performers within binary combinations are `surf-brief`, `orb-freak` and `orb-brisk`.

Blur The results of combinations applied to images smoothed by a Gaussian kernel can be seen in Figure 2g and Figure 2h for the scene in Figure 1g, and in Figure 3g and Figure 3h, for the scene in Figure 1h. Combinations average results are presented in Table 2d.

Best performance among floating point combinations is attained by `surf`, outperforming other combinations in stability, which is visualized by a horizontal precision-recall curve. It is followed by `sift` and `hes-liop`. Overall best performance can be found in the category of binary combinations, with `surf-brief` as the top performer, outperforming floating point combinations. The consecutive performers are `orb-brisk` and `orb` which achieve best performance among corner based combinations, with comparable or better results than blob based combinations.

Noise The performance of combinations applied to images with induced white Gaussian noise is presented in Figure 2i and Figure 2j for the structured scene in Figure 1i. The corresponding results for the textured scene in Figure 1j are

Table 2: Average performance when using NN as matching strategy by the measures: precision, recall and Matching Score (MS).

(a) Viewpoint

Combination	Precicion	Recall	MS
har-orb	19.6	40.8	19.6
har-brisk	22.9	47.3	22.9
har-sift	21.5	45.1	21.5
har-liop	26.3	54.9	26.3
hes-liop	41.8	62.8	41.8
sift	14.6	39.6	14.6
surf	31.7	44.0	31.7
mser-liop	30.1	71.1	30.1
orb	39.2	56.2	39.2
brisk	20.7	42.2	20.7
surf-brief	37.6	51.8	37.6
surf-freak	27.8	38.2	27.8
orb-freak	38.1	55.3	38.1
orb-brisk	43.0	62.1	43.0
fast-brisk	15.4	28.6	15.4

(b) Rotation

Combination	Precicion	Recall	MS
har-orb	59.1	76.8	59.1
har-brisk	63.7	82.5	63.7
har-sift	61.3	79.6	61.3
har-liop	74.1	96.4	74.1
hes-liop	84.9	96.6	84.9
sift	44.2	76.4	44.2
surf	62.7	75.3	62.7
mser-liop	67.0	88.0	67.0
orb	76.6	87.4	76.6
brisk	37.7	63.5	37.7
surf-brief	7.2	8.4	7.2
surf-freak	54.7	65.7	54.7
orb-freak	71.4	81.5	71.4
orb-brisk	79.2	90.4	79.2
fast-brisk	61.6	75.0	61.6

(c) Scale

Combination	Precicion	Recall	MS
har-orb	44.6	75.7	44.6
har-brisk	48.8	82.9	48.8
har-sift	48.0	81.5	48.0
har-liop	53.6	90.9	53.6
hes-liop	65.0	94.0	65.0
sift	57.5	83.4	57.5
surf	60.7	81.2	60.7
mser-liop	74.2	97.0	74.2
orb	60.5	82.1	60.5
brisk	35.1	59.6	35.1
surf-brief	64.8	86.6	64.8
surf-freak	54.8	73.2	54.8
orb-freak	62.8	85.5	62.8
orb-brisk	62.7	84.9	62.7
fast-brisk	11.2	13.3	11.2

(d) Blur

Combination	Precicion	Recall	MS
har-orb	19.5	30.9	19.5
har-brisk	29.8	47.8	29.8
har-sift	20.2	32.0	20.2
har-liop	24.5	38.4	24.5
hes-liop	53.7	57.3	53.7
sift	35.7	60.9	35.7
surf	63.4	66.7	63.4
mser-liop	24.4	33.0	24.4
orb	60.4	71.0	60.4
brisk	21.5	53.4	21.5
surf-brief	82.4	86.5	82.4
surf-freak	62.8	66.0	62.8
orb-freak	38.7	45.6	38.7
orb-brisk	67.9	79.6	67.9
fast-brisk	33.1	34.2	33.1

(e) Noise

Combination	Precicion	Recall	MS
har-orb	51.9	65.2	51.9
har-brisk	67.9	85.8	67.9
har-sift	51.1	64.1	51.1
har-liop	68.7	86.8	68.7
hes-liop	76.5	88.8	76.5
sift	45.7	72.9	45.7
surf	73.5	80.7	73.5
mser-liop	69.2	86.5	69.2
orb	81.7	91.0	81.7
brisk	47.5	62.3	47.5
surf-brief	79.1	86.9	79.1
surf-freak	69.6	76.5	69.6
orb-freak	76.1	84.7	76.1
orb-brisk	83.8	93.4	83.8
fast-brisk	63.0	65.1	63.0

(f) Downsampling

Combination	Precicion	Recall	MS
har-orb	47.6	68.2	47.6
har-brisk	59.4	86.2	59.4
har-sift	51.0	73.5	51.0
har-liop	60.8	88.6	60.8
hes-liop	65.3	88.3	65.3
sift	26.9	36.7	26.9
surf	78.2	85.8	78.2
mser-liop	39.5	72.0	39.5
orb	28.3	53.2	28.3
brisk	15.9	42.7	15.9
surf-brief	82.4	90.5	82.4
surf-freak	61.5	67.8	61.5
orb-freak	27.2	50.5	27.2
orb-brisk	29.8	55.8	29.8
fast-brisk	0.0	0.0	0.0

shown in Figure 3i and Figure 3j. The average results of the two scenes against noise are presented in Table 2e.

The overall performance for various combinations is relatively high. Best performance among floating point combinations are attained by `hes-liop` and `surf`, stagnating at about the same level of recall in the precision-recall curves for both scenes and in Table 2e. The overall best performance in the case of induced noise is achieved by `orb-brisk` followed by `orb` and `surf-brief` showing better performance than the floating point category.

Downsampling Last, we evaluate the effect on combinations caused by downsampling and present the results in Figure 2k and Figure 2l for the structured scene in Figure 1k. For the textured scene in Figure 1l the results are presented in Figure 3k and Figure 3l. The obtained average results for NN matching are presented in Table 2f.

Studying the precision-recall curves of floating point methods and Table 2f, the best performers on downsampled images are `surf`, `hes-liop` and `har-liop` and `har-brisk`. Among binary methods the best performance is obtained by `surf-brief`, with better results than `surf`, with `surf-freak`, `orb-brisk` and `orb` to come after. MSER does in Figure 2k reach a 100 % recall which can be explained by that very few regions are detected.

4 Comparisons to Results in Earlier Work

4.1 IR images

The most related work in the long wave infrared (LWIR) spectral band [19] shows both similarities and differences to the results in this work. Best performance against blur is in both evaluations obtained by SURF. For rotation and scale best performance is achieved by SIFT in the compared evaluation while LIOP, not included in mentioned evaluation, shows highest robustness to the deformation in this work.

Among binary combinations [19] presents a low performance for ORB and BRISK with their default detectors. In this work the low performance of the combination of BRISK is observed while the combination of ORB is a top performer among binary methods. An important difference between these two evaluations is that we have performed a comparison of numerous detector and descriptor combinations, which have led to the conclusion of a good match of the ORB detector and BRISK descriptor.

4.2 Visual Images

In the evaluation of binary methods for visual images in [24], it is obvious how the combination of detector and descriptor might affect the performance. When evaluating descriptors with their default detectors, BRISK and FREAK perform much worse than when combined with the ORB detector. Best overall performance was obtained by ORB detector in combination with FREAK or BRISK descriptors and ORB combined with FREAK is the suggested combination to use. In this work we have observed that BRISK with its default detector performed worse, in most categories the worst, compared to when in combination with the ORB detector while the combination of ORB and FREAK has lower performance in the LWIR spectral band. With the high performance of the combination of ORB and BRISK we can conclude that the choice of combination has large effect on the performance both in visual images and IR images.

Another similarity is the high performance by Hessian-Affine with LIOP in [8] and in this work as well as by the combination of SURF which shows high performance in both spectral bands.

5 Conclusions and Future Directions

We have performed a systematic investigation on the performance of state-of-the-art local feature detectors and descriptors on infrared images, justified by the needs in various vision applications, such as image stitching, recognition, etc. While doing so, we have also generated a new IR image data set and made it publicly available. Through the extensive evaluations we have gained useful insight as to what local features to use according to expected transformation properties of the input images as well as the requirement for efficiency. It should

be highlighted that the combination of detector and descriptor should be considered as it can outperform the standard combination. As the consequence of our comparisons at large, Hessian-Affine with LIOP, and SURF detector with SURF descriptor have shown good performance to many of the geometric and photometric transformations. Among binary detectors and descriptors competitive results are received with the combination of ORB and BRISK.

Compared to the most relevant work by Ricaurte et al. [19] this work evaluated performances against viewpoint changes, the LIOP descriptor, float type detectors as Hessian-Affine and Harris-Affine including different combinations of detectors and descriptors, filling the gap of evaluations for IR images.

In future research we will extend the study from those hand crafted features to learning based representations such as RFD [25] as well as those [26, 27] obtained by deep convolutional networks which were shown to be very effective for a range of visual recognition tasks [28–30]. Fischer et al. [31] demonstrated that those descriptors perform consistently better than SIFT also in the low-level task of descriptor matching. Although the networks are typically trained on the ImageNet data set consisting of visual images, it will be interesting to see if such a network is applicable to extracting descriptors in IR images (via transfer learning), or one would need yet another large data set of IR images to train a deep convolutional network. Nevertheless, the study in this direction is beyond the scope of this paper and left for the subject of our next comparison.

References

1. Mikolajczyk, K., Schmid, C.: A performance evaluation of local descriptors. TPAMI **27**(10) (2005) 1615–1630
2. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. IJCV **60**(2) (2004) 91–110
3. Mikolajczyk, K., Tuytelaars, T., Schmid, C., Zisserman, A., Matas, J., Schaffalitzky, F., Kadir, T., Gool, L.: A Comparison of Affine Region Detectors. IJCV **65**(1-2) (2005) 43–72 (`http://www.robots.ox.ac.uk/~vgg/research/affine`).
4. Matas, J. and Chum, O. and Urban, M. and Pajdla, T.: Robust wide baseline stereo from maximally stable extremal regions. In: BMVC. (2002) 36.1–36.10
5. Mikolajczyk, K., Schmid, C.: An Affine Invariant Interest Point Detector. In: ECCV. (2002) 128–142
6. Mikolajczyk, K., Schmid, C.: Scale & Affine Invariant Interest Point Detectors. IJCV **60**(1) (2004) 63–86
7. Miksik, O., Mikolajczyk, K.: Evaluation of local detectors and descriptors for fast feature matching. In: ICPR. (2012) 2681–2684
8. Wang, Z., Fan, B., Wu, F.: Local Intensity Order Pattern for Feature Description. In: ICCV. (2011) 603–610
9. Fan, B., Wu, F., Hu, Z.: Rotationally Invariant Descriptors Using Intensity Order Pooling. TPAMI **34**(10) (2012) 2031–2045
10. Bay, H., Ess, A., Tuytelaars, T., Van Gool, L.: Speeded-Up Robust Features (SURF). Computer Vision and Image Understanding **110**(3) (2008) 346–359
11. Heinly, J., Dunn, E., Frahm, J.M.: Comparative evaluation of binary features. In: ECCV. (2012) 759–773

12. Li, S., Zhang, L., Liao, S., Zhu, X., Chu, R., Ao, M., He, R.: A Near-Infrared Image Based Face Recognition System. In: 7th International Conference on Automatic Face and Gesture Recognition. (2006) 455–460
13. Maeng, H., Choi, H.C., Park, U., Lee, S.W., Jain, A.: NFRAD: Near-Infrared Face Recognition at a Distance. In: International Joint Conference on Biometrics, IJCB. (2011) 1–7
14. Strehl, A., Aggarwal, J.: Detecting Moving Objects in Airborne Forward Looking Infra-Red Sequences. In: Proceedings IEEE Workshop on Computer Vision Beyond the Visible Spectrum: Methods and Applications. (1999) 3–12
15. Broggi, A., Fedriga, R., Tagliati, A.: Pedestrian Detection on a Moving Vehicle: an Investigation about Near Infra-Red Images. In: Intelligent Vehicles Symposium, IEEE. (2006) 431–436
16. Zhang, X., Sim, T., Miao, X.: Enhancing Photographs with Near Infra-Red Images. In: CVPR. (2008) 1–8
17. Aguilera, C., Barrera, F., Lumbreras, F., Sappa, A.D., Toledo, R.: Multispectral Image Feature Points. Sensors **12**(9) (2012) 12661–12672
18. Ferraz, L., Binefa, X.: A Scale Invariant Interest Point Detector for Discriminative Blob Detection. In: Pattern Recognition and Image Analysis. Volume 5524. (2009) 233–240
19. Ricaurte, P., Chilán, C., Aguilera-Carrasco, C.A., Vintimilla, B.X., Sappa, A.D.: Feature Point Descriptors: Infrared and Visible Spectra. Sensors **14**(2) (2014) 3690–3701
20. Bradski, G.: Opencv, open source computer vision. Dr. Dobb's Journal of Software Tools (2000)
21. Vedaldi, A., Fulkerson, B.: VLFeat: An Open and Portable Library of Computer Vision Algorithms. "`http://www.vlfeat.org/`" (2008)
22. Calonder, M., Lepetit, V., Strecha, C., Fua, P.: BRIEF: Binary Robust Independent Elementary Features. In: ECCV. Volume 6314. (2010) 778–792
23. Alahi, A., Ortiz, R., Vandergheynst, P.: FREAK: Fast Retina Keypoint. In: CVPR. (2012) 510–517
24. Figat, J., Kornuta, T., Kasprzak, W.: Performance evaluation of binary descriptors of local features. In: Computer Vision and Graphics. Volume 8671. (2014) 187–194
25. Fan, B., Kong, Q., Trzcinski, T., Wang, Z., Pan, C., Fua, P.: Receptive fields selection for binary feature description. IEEE Transactions on Image Processing **23**(6) (June 2014) 2583–2595
26. Balntas, V., Johns, E., Tang, L., Mikolajczyk, K.: PN-Net: Conjoined triple deep network for learning local image descriptors. CoRR **abs/1601.05030** (2016)
27. Simo-Serra, E., Trulls, E., Ferraz, L., Kokkinos, I., Fua, P., Moreno-Noguer, F.: Discriminative learning of deep convolutional feature point descriptors. In: ICCV. (2015)
28. Zeiler, M.D., Fergus, R.: Visualizing and understanding convolutional networks. In: ECCV. (2014)
29. Donahue, J., Jia, Y., Vinyals, O., Hoffman, J., Zhang, N., Tzeng, E., Darrell, T.: Decaf: A deep convolutional activation feature for generic visual recognition. In: ICLR. (2014)
30. Girshick, R.B., Donahue, J., Darrell, T., Malik, J.: Rich feature hierarchies for accurate object detection and semantic segmentation. In: CVPR. (2014)
31. Fischer, P., Dosovitskiy, A., Brox, T.: Descriptor matching with convolutional neural networks: a comparison to SIFT. **arXiv:1405.5769** (2014)