ABFT: ANISOTROPIC BINARY FEATURE TRANSFORM BASED ON STRUCTURE TENSOR SPACE

Seungryong Kim, Hunjae Yoo, Seungchul Ryu, Bumsub Ham and Kwanghoon Sohn

Digital Image Media Laboratory (DIML),
School of Electrical and Electronic Engineering, Yonsei University, Seoul, Republic of Korea
Email: khsohn@yonsei.ac.kr

ABSTRACT

Local feature matching is a fundamental step for many computer vision applications. Recently, binary feature transforms have been popularly proposed to improve the computational efficiency while preserving high matching performance. However, it is sensitive to noise and geometrical distortion such as affine transformation. In this paper, we propose ABFT framework, composed of a noise robust feature detection and affine invariant binary feature description based on a structure tensor space. Experimental results show that ABFT outperforms other state-of-the-art feature transforms in terms of the repeatability, recognition rate, and computational time.

Index Terms— Feature matching, binary feature, anisotropic, structure tensor.

1. INTRODUCTION

Local feature matching for finding a correspondence of the salient image regions is a fundamental step for many computer vision applications such as motion detection, 3-D modeling, panorama stitching, object tracking and object recognition [1].

Many robust algorithms have been proposed for the reliable matching in many literatures. The Scale Invariant Feature Transforms (SIFT) proposed by Lowe [1] has been the most popular approach due to high robustness to many of the variations and distortions. For the computational efficiency, Bay *et al.* proposed the Speeded-Up Robust Features (SURF) [2] which approximates to SIFT and outperforms other methods. Although these conventional algorithms show the competitive performance, they are difficult to be applied for the mobile applications or low-power devices because of the high computational complexity. Recently, Rosten *et al.* proposed the Features from Accelerated Segment Test (FAST) feature detector [3] by intensity segment test and Calonder *et al.* proposed the Binary Robust Independent Elementary Features (BRIEF) feature descriptor [4] by simple intensity difference tests. By combining FAST detection and BRIEF description, the binary feature transforms have been popularly proposed in order to overcome the computational limitation of conventional algorithms [5]. In these spectrums, Rublee *et al.* proposed the Oriented FAST and Rotated BRIEF (ORB) [6] which addresses the rotation variant problem of BRIEF. Leutenegger *et al.* also proposed the Binary Robust Invariant Scalable Keypoints (BRISK) [7] which is scale-space FAST detector in combination with bit-string descriptors. Alahi *et al.* proposed the Fast Retina Keypoint (FREAK) [8] inspired by the human visual systems.

However, the performance of binary feature transforms is highly affected by the conditions imposed by many real applications since it is sensitive to noise and geometrical distortion such as affine transformation. In this paper, we propose a novel feature matching framework combining a noise robust feature detection and affine invariant binary feature description called ABFT. Our approach detects the reliable key-points by rejecting false ones such as noise or corner-like structure. From these key-points, we build an anisotropic binary feature descriptor by estimating the orientation and local structure around key-points based on the structure tensor space. This paper is organized as follows. Section 2 reviews conventional binary feature transforms and defines the limitations of them. Section 3 introduces the ABFT in detail. Experimental results are presented in Section 4 and we conclude this paper in Section 5.

2. MOTIVATION AND OVERVIEW

2.1. Scale-Space FAST Feature Detection

FAST corner detection [3] and its variants are popularly used due to very low computational time while preserving the high detection performance. However, the FAST has several limitations. For example, the localization performance of the FAST decreases dramatically as the amount of noise increases. Since it detects the key-points based on the simple pixel-by-pixel intensity comparison, the degradation of pixels by noise, especially impulse noise, may change the difference of intensity, which increases a probability of false corner detection. In addition, in the scale-space FAST employed to ORB, BRISK or FREAK [6, 7, 8], corner-like structures such as the ramp edges appear in a coarse scale, which may be false corners. In other words, the scale-space FAST violates the scale-space causality criteria which means no new key-points should be generated in the coarse scale [9].

2.2. BRIEF Binary Feature Description

Binary feature descriptors such as BRIEF [4] or its variants are sensitive to geometrical distortion such as rotation or affine transformation. As mentioned in many literatures, the performance of BRIEF decreases dramatically for in-plane rotation variants. In order to overcome this problem, there have been several approaches such as steering the comparison pattern to orientation of features [6] or finding the best comparison pattern based on the machine-learning [4]. However, there are no binary feature descriptors which are invariant to affine transformation. Since affine transformation makes different scale change according to the directions, the conventional isotropic comparison pattern for binary descriptors cannot summarize the local structure for neighborhood of the feature properly.

3. ANISOTROPIC BINARY FEATURE TRANSFORM

In this section, we propose an Anisotropic Binary Feature Transform (ABFT), as it transforms the local salient image regions into robust binary descriptors. ABFT is composed of feature detection and feature description steps. In detection step, ABFT detects the noise robust key-points on corner candidate region. From these key-points,

978-1-4799-2341-0/13/$31.00 ©2013 IEEE

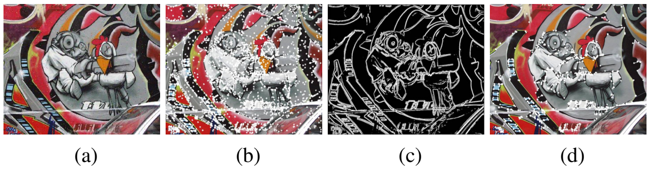


Fig. 1. Results of key-points detection. (a) the noise 'Graffiti' image, (b) BRISK [7], (c) Corner Candidate Region, (d) ABFT

ABFT builds the affine invariant binary descriptor by considering the local structure around key-points in description step. Especially, we use the structure tensor space for estimating the characteristic scale and orientation of key-points and for constructing anisotropic concentric patterns to build affine invariant binary descriptor.

3.1. Feature Detection

In the ABFT framework, we propose the scale-space corner candidate region FAST for noise robustness. First, we construct a Haar-wavelet gradient map similar to SURF [2]. The Haar-wavelet gradient is computed by box filtering using integral image as follows:

$$\begin{aligned} I_x &= (\sum_{(s,t)\in P_R} I(s,t) - \sum_{(s,t)\in P_L} I(s,t))/(\sum_{(s,t)\in P} I(s,t)) \\ I_y &= (\sum_{(s,t)\in P_U} I(s,t) - \sum_{(s,t)\in P_D} I(s,t))/(\sum_{(s,t)\in P} I(s,t)) \end{aligned} \quad (1)$$

where I_x and I_y are the derivatives of image I along the x and y direction. P_R, P_L, P_U and P_D are the sub-patch toward to right, left, up and down sides on 9×9 patch P for neighborhood of key-points. Since the summation of sub-patch intensity and normalization terms reduce noise effects, the Haar-wavelet gradient magnitude clearly represents structure silhouette in the image. Then, we define the corner candidate region as pixels whose Haar-wavelet gradient magnitude is large enough to be corner by proper threshold based on the fact that gradient magnitude of corner is the local maximum. Finally, we detect key-points on the corner candidate region by FAST corner detection. It is very advantageous that it detects the reliable structure corners by rejecting the false ones such as the noise or corner-like structure. In addition, the detection computational time decrease since the corner search area is restricted not on the homogeneous or edge regions but on the corner candidate regions.

We construct an image scale-space I^i and gradient scale-space ∇I^i for scale selection. Each scale-space consists of the N octave for $i = \{0, 1, ..., N-1\}$ by bilinear interpolation down-sampling by a factor of 1.5 similar to BRISK [7]. From the gradient scale-space, the structure tensor space is computed as follows.

$$S_i = K_\rho * \nabla I^i (\nabla I^i)^T = K_\rho * \begin{bmatrix} (I_x^i)^2 & I_x^i * I_y^i \\ I_x^i * I_y^i & (I_y^i)^2 \end{bmatrix} \quad (2)$$

where K_ρ is a Gaussian kernel for weighting neighborhood of the key-points. Note that the structure tensor space is selectively computed not on the overall image octaves but on the each key-point. In addition, if the structure tensor is calculated once, it can be used for estimating scale, orientation and anisotropic pattern of features in overall process, which reduces the additional computational times for these tasks dramatically.

We use the Harris corner measure [10] from the structure tensor space in order to determine a characteristic scale to key-points. First, we convert the coordinates of all coarse image key-points into the corresponding coordinates in the original image. Then, we find a local extrema with the Harris corner measure among the nearby key-points existing on a 3×3 neighborhood in the original image.

All key-points with the Harris corner measure less than the threshold are discarded in order to eliminate the false corners on the homogeneous or edge regions. In addition, if a key-point in the coarse image is too far from nearby key-points in the original image, it is considered the false corner-like features. Thus, we also discard the coarse scale key-points to satisfy the causality criteria in scale-space or reject corner-like features. Fig. 1 shows the results of key-points detection in the noise 'Graffiti' image. While the conventional scale-space FAST detects the false key-points such as noise or corner-like features, ABFT detects the explicit key-points on structure regions.

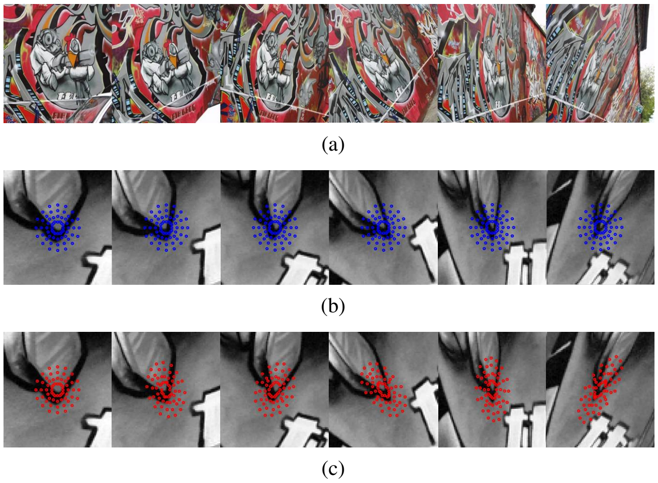


Fig. 2. Concentric circle patterns as the affine transformation varies. (a) the 'Graffiti' image sequences, (b) isotropic concentric pattern, (c) anisotropic concentric pattern

3.2. Feature Description

The ABFT builds a binary descriptor using the results of randomly sampled intensity comparison similar to BRIEF [4]. These random sampling pairs are located on the concentric pattern around the key-points. In contrast to an isotropic concentric pattern employed to BRISK [7] or DAISY descriptor [11], an anisotropic concentric pattern is used to build affine invariant and distinctive descriptor in ABFT description.

Fig. 2 shows the comparison of the concentric pattern types in 'Graffiti' sequences varying affine transformation degree. As shown in Fig. 2 (b), the isotropic concentric pattern is imprecise to describe the deformed neighborhood of the key-point. By contrast, the anisotropic concentric pattern in Fig. 2 (c) distinctly summarizes the local structure around key-points since the directions of affine transformation is considered properly.

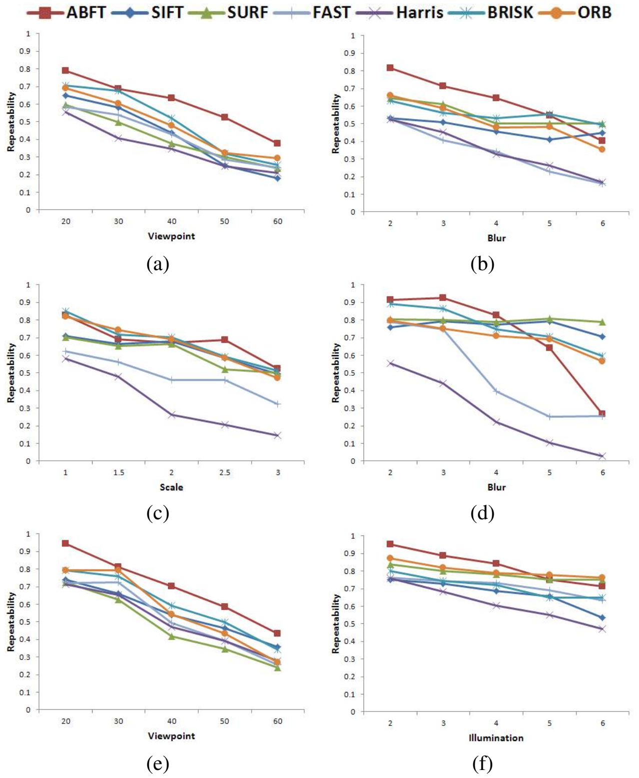
The isotropic concentric pattern Φ is defined as follows:

$$\Phi_{i,j} = [r_i \cos\theta_j, r_i \sin\theta_j]^T, \quad 0 \le i < n_r, 0 \le j < n_\theta \quad (3)$$

where r_i is the concentric circle radius for scale-normalized, θ_i is the angles of each point on concentric pattern.

First, the ABFT determines the dominant orientation of key-points using the structure tensor space. The eigenvector of structure tensor determines the dominant orientation of the local structure [15, 16]. Thus, in order to allow the descriptor to be invariant to in-plain rotation, the ABFT rotates the concentric pattern to the direction of the eigenvector to smallest eigenvalue of each structure tensor.

The ABFT also transforms an isotropic concentric pattern to an anisotropic concentric pattern. The structure tensor provides a method for estimating the affine shape of a local structure. It is proved that the neighborhood of features can be normalized by multiplying the square root matrix of structure tensor to the neighborhood [12]. In this paper, we apply this property inversely to trans-



(a) (b)

(c) (d)

(e) (f)

Fig. 3. Repeatability evaluation for Mikolajczyk's database [13]. (a) 'Graffiti' images, (b) 'Trees' images, (c) 'Boat' images , (d) 'Bikes' images, (e) 'Wall' images, (f) 'Leuven' images

form the isotropic concentric pattern into the anisotropic concentric pattern for the affine invariant descriptor.

Therefore, the anisotropic concentric pattern Λ is defined by multiplying the rotation matrix and the inverse of the square root matrix of structure tensor to isotropic concentric pattern as follows:

$$\Lambda_{i,j} = S^{-1/2} \cdot R_\theta \cdot \Phi_{i,j}, \quad 0 \le i < n_r, 0 \le j < n_\theta \tag{4}$$

where $S^{-1/2}$ is the inverse of the square root matrix of structure tensor and R_θ is the rotation matrix of the orientation of key-points.

In order to build a binary descriptor, a random intensity comparison test can be defined as follows [7]:

$$\tau(\Lambda; p_i, q_i) = \begin{cases} 1 & I(\Lambda, p_i) < I(\Lambda, q_i) \\ 0 & otherwise \end{cases} \tag{5}$$

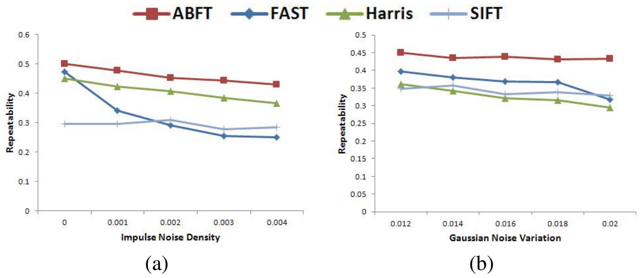
where $I(\Lambda, p_i)$ and $I(\Lambda, q_i)$ are the intensity of randomly sampled pair p_i and q_i on the anisotropic circle concentric pattern Λ. Therefore, the ABFT builds the anisotropic binary feature descriptor that corresponds to the decimal counterpart as follows:

$$\sum_{i=1}^{N} 2^{i-1} \tau(\Lambda; p_i, q_i) \tag{6}$$

In this paper, we determine the length of descriptors N to 512 which shows the suitable performance, but 128 or 256 are enough to proper matching performance. Matching costs of ABFT descriptor are very low because the Hamming distance, performed by a bitwise XOR, can be used for a measure of their similarity.

4. EXPERIMENTAL RESULTS

In this section, we extensively evaluate the ABFT with a variety of other methods such as SIFT [1], SURF [2], ORB [6] and BRISK [7] in terms of the repeatability, recognition rate and computational complexity. We evaluate these criteria for Mikolajczyk's database, which is popular evaluation frameworks in the field [13, 14]. The database consists of six image sequences and each sequence has the different image deformation condition: viewpoint and affine change (Graffti and Wall), scale and rotation change (Boat), blur change (Bikes and Trees) and illumination change (Leuven). The database also provides a ground truth homography that can be used for estimating the correspondence of key-points. The implementation of BRISK are obtained from the authors and the others such as SIFT, SURF and ORB come from OpenCV 2.3 implementation.



(a) (b)

Fig. 4. Repeatability evaluation for the noise 'Graffiti' image. (a) impulse noise, (b) Gaussian noise

4.1. Repeatability

The performance of the feature detector in terms of localization accuracy is measured by the repeatability as defined in [13]. The repeatability is the ratio of the correspondences to the minimum number of detected key-points in each image. The correspondences are identified by measuring the distance between a key-point in the one image and the projected key-point from the second image by the proper homography [13].

As shown in Fig. 3, among the conventional methods, BRISK represents the highest repeatability in the overall image sequence except for some images. However, ABFT represents better repeatability than BRISK and other detectors. In 'Graffiti', 'Wall' and 'Leuven' sequences, ABFT represents the best detection performance compared with the others. In addition, while the performance of corner detector such as ORB, BRISK or FAST decrease in the 'Trees', 'Boat' and 'Bikes' images containing perspective deformation such as blur or scale change, ABFT outperforms other corner detectors although it also detects corner features. However, the performance of ABFT also decrease in high blur deformation degrees as shown in Fig. 3 (b), (d), which are limitation of corner detector compared with the blob detector such as SIFT, SURF.

In order to evaluate robustness to noise for ABFT, we measure the repeatability for the 'Graffiti' images corrupted by additive noise, the Gaussian and impulse noise. The corrupted Gaussian noise variance varies from 0.01 to 0.02 with zero mean and impulse noise density varies from 0 to 0.005 which are enough to verify the noise robustness. As mentioned before, FAST and its variant are very vulnerable to the noise, especially impulse noise. Fig. 4 shows that ABFT is more robust to both impulse and Gaussian noise than the conventional FAST or SIFT known as the noise robust algorithm.

4.2. Recognition Rate

The recognition rate is computed as the ratio of the correct matching number of descriptors to the number of descriptors defined in [4].

The performance of feature descriptor depends on the types of image deformation [14]. Fig. 5 shows that the binary descriptors such as ORB or BRISK outperforms the real-valued descriptors SIFT or SURF in the perspective deformation such as illumination, blur. On the other hands, in the geometrical distortion such as

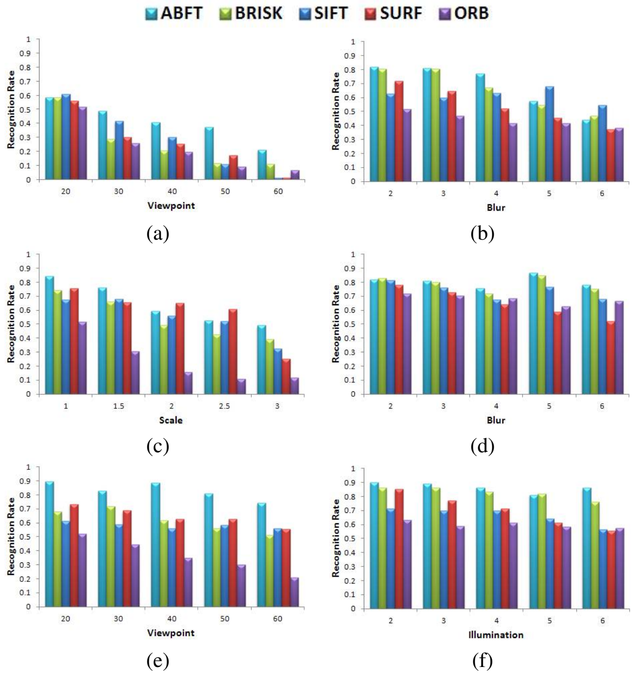


Fig. 5. Recognition rate evaluation for Mikolajczyk's database [13]. (a) 'Graffiti' images, (b) 'Trees' images, (c) 'Boat' images , (d) 'Bikes' images, (e) 'Wall' images, (f) 'Leuven' images

viewpoint or affine, SIFT or SURF outperforms ORB or BRISK. However, ABFT shows competitive performance with other descriptors representing the highest recognition rate in any image sequences and even outperforms. The recognition rate of most descriptors for the 'Graffiti' images which contains affine distortion decrease dramatically as the degree of distortion increases. However, ABFT shows consistently high recognition rate for the overall sequence, which clearly shows the robustness of ABFT to affine or geometrical distortion. In addition, under viewpoint changes, the SIFT or SURF generally represent higher performance than the conventional binary descriptors. However, the highest recognition rate of ABFT in 'Wall' image represents that it compensates the drawback of the other binary descriptors such as BRISK or ORB. In other image sequences, ABFT also outperforms other descriptors.

4.3. Computational Complexity

We compare the computational time of ABFT with that of SIFT, SURF, ORB and BRISK. The experiments run on Intel(R) Core(TM) 2 Quad CPU Q6600 at 2.40 GHz. The computational time is measured for the 'Graffiti' sequences which have 680×800 sizes by calculating the average of 20 runs. As shown in Table 1, the detection and description time of binary feature transforms such as BRISK, ORB and ABFT are an order of magnitude faster than SURF conventionally known as the most computationally efficient method. Especially, ABFT shows the best timing performance among the state-of-the-art binary descriptors. Although the computational performance of ABFT is similar to BRISK, ABFT shows better repeatability and recognition rate than BRISK.

5. CONCLUSION

In this paper, we proposed the noise robust feature detector and affine invariant feature descriptor called ABFT. It detects the reliable key-points based on corner candidate region FAST which is robust to noises or false corner-like structures. From these key-points, it builds the affine invariant binary feature descriptor by transform the isotropic concentric pattern to the anisotropic concentric pattern based on the structure tensor space. Experimental results show that ABFT is robust to noise and geometrical distortion compared with other state-of-the-art methods in terms of the repeatability and recognition rate. ABFT also shows the best computational efficiency. In future, we will extend the ABFT so that it would be applied to spatiotemporal matching.

Table 1. Computational time evalution results

Algorithms	***ABFT***	***BRISK***	***ORB***	***SURF***	***SIFT***
Detection(ms)	172	156	328	531	3984
Description(ms)	158	157	610	938	4875
Total(ms)	330	313	938	1469	8859
Key-points	1545	1435	1998	1444	1516
Time/Point(ms)	0.214	0.218	0.469	1.017	5.844

6. REFERENCES

[1] D.G.Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91-110, 2004

[2] H.Bay, A.Ess, T.Tuytelaars and L.V.Gool "Speeded-Up Robust Features (SURF)," *Computer Vision and Image Understanding*, vol. 110, no. 3, pp. 346-359, 2008

[3] E.Rosten, R.Porter and T.Drummond, "Faster and better : a machine learning approach to corner detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 1, pp. 105-119, 2010

[4] M.Calonder et al., "BRIEF : Computing a local binary descriptor very fast," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 7, pp. 1281-1298, 2011

[5] J.Heinly, E.Dunn and J.M.Frahm, "Comparative Evaluation of Binary Features," in *Proc. IEEE Conf. European Conference on Computer Vision*, 2012

[6] E.Rublee, V.Rabaud, K.Konolige and G.Bradski, "ORB : an efficient alternative to SIFT or SURF," in *Proc. IEEE Conf. International Conference on Computer Vision*, 2011

[7] S.Leutenegger, M.Chli and R.Y.Siegwart, "BRISK : Binary Robust Invariant Scalable Keypoints," in *Proc. IEEE Conf. International Conference on Computer Vision*, 2011

[8] A.Alahi, R.Ortiz and P.Vandergheynst, "FREAK : Fast Retina Keypoint," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2012

[9] T.lindeberg, "Feature Detection with Automatic Scale Selection," *Int. J. Comput. Vis.*, vol. 30, no. 2, pp. 79-116, 1998

[10] C. Harris and M. Stephens, "A combined Corner and Edge detector," *Proceedings of the 4th Alvey Vision Conference*, pp. 147-151, 1988

[11] E.Tola, V.Lepetit and P.Fua, "DAISY : An Efficient Dense Descriptor applied to Wide-Baseline Stereo," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, No. 5, pp. 815-830, 2010

[12] K.Mikolajczyk and C.Schmid, "Scale and Affine Invariant Interest Point Detectors," *Int. J. Comput. Vis.*, vol. 60, no. 1, pp. 63-86, 2004

[13] K.Midolajczyk et al., "A Comparison of Affine Region Detectors," *Int. J. Comput. Vis.*, vol. 65, no. 1, pp. 43-72, 2005

[14] K.Mikolajczyk and C.Shmid, "A Performance Evaluation of Local Descriptors," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, No. 10, pp. 1615-1630, 2005

[15] R.v.d.Boomgaard and J.v.d.Weijer, "Robust Estimation of Orientation for Texture Analysis," in *Proc. IEEE Conf. European Conference on Computer Vision*, 2002

[16] T.Brox, J.Weickert, B.Burgeth and P.Mrazek, "Nonlinear Structure Tensors," *Image and Vision Computing*, vol. 24, no. 1, pp. 41-55, 2006