Computer Systems Science & Engineering
DOI:10.32604/csse.2022.022400
Article

Tech Science Press

Keypoint Description Using Statistical Descriptor with Similarity-Invariant Regions

Ibrahim El rube'* **and Sameer Alsharif**

Department of Computer Engineering, College of Computers and Information Technology, Taif University, P.o. Box 11099, Taif, 21944, Saudi Arabia
*Corresponding Author: Ibrahim El rube'. Email: ibrahim.ah@tu.edu.sa
Received: 06 August 2021; Accepted: 09 September 2021

Abstract: This article presents a method for the description of key points using simple statistics for regions controlled by neighboring key points to remedy the gap in existing descriptors. Usually, the existent descriptors such as speeded up robust features (SURF), Kaze, binary robust invariant scalable keypoints (BRISK), features from accelerated segment test (FAST), and oriented FAST and rotated BRIEF (ORB) can competently detect, describe, and match images in the presence of some artifacts such as blur, compression, and illumination. However, the performance and reliability of these descriptors decrease for some imaging variations such as point of view, zoom (scale), and rotation. The introduced description method improves image matching in the event of such distortions. It utilizes a contourlet-based detector to detect the strongest key points within a specified window size. The selected key points and their neighbors control the size and orientation of the surrounding regions, which are mapped on rectangular shapes using polar transformation. The resulting rectangular matrices are subjected to two-directional statistical operations that involve calculating the mean and standard deviation. Consequently, the descriptor obtained is invariant (translation, rotation, and scale) because of the two methods; the extraction of the region and the polar transformation techniques used in this paper. The description method introduced in this article is tested against well-established and well-known descriptors, such as SURF, Kaze, BRISK, FAST, and ORB, techniques using the standard OXFORD dataset. The presented methodology demonstrated its ability to improve the match between distorted images compared to other descriptors in the literature.

Keywords: Keypoint detection; descriptors; neighbor region; similarity invariance

1 Introduction

Image matching is a challenging task in modern computer vision problems. Some applications, such as image structure from multiple frames and mapping creation for robotic applications, essentially use image matching for visualization purposes. For instance, multimodal biomedical images that are generated by different techniques such as X-ray, Magnetic resonance imaging (MRI), computerized tomography

(CT-scan), and ultrasonic are registered to acquire the privilege of each image and give advanced visualization. Therefore, image matching can be defined as finding the projection of a given point from one image to another based on some features. Such features can be described based on texture, color, or shape. They can be detected and extracted using different descriptors, such as scale-invariant feature transform (SIFT), (SURF), wavelet local feature descriptor (WLFD), and binary robust appearance and normal descriptors (BRAND). Nevertheless, the projection of points might be deceivable and incorrect due to illumination, object reflection, image deformation, noise, and other distractions.

Descriptors should be characterized with some features to give an effective result. They should not be affected by the noise, computationally efficient, and invariant to transformation, rotation, illumination, and scaling. Generally, descriptors can be categorized into two types, edge detection-based descriptors and keypoints-based descriptors. Although the former does not cost memory requirements, they lack detection accuracy because of occlusion and image perspectives. Besides, it is not robust for rotation, scaling, and other artifacts resulting in unstable edge detection. On the other hand, keypoints-based descriptors are based on selecting some points to compare them relative to their neighbors. Relativity metrics are based on morphological or statistical tools resulting in robustness to the noise and invariance to the said image deformations.

The common descriptors of such types are SIFT, SURF, and WLFD. These descriptors are competitively accurate and suitable for many applications. However, a high dimension descriptor is computationally exhaustive and requires memory and high-performance digital computers. Nevertheless, many research attempts have been made to mitigate the adverse effects. Such attempts include enhancing keypoints extraction, using optical flow to track keypoints, minimizing descriptor dimension, and using binary descriptors.

This paper presents a key point-based descriptor that statistically compares key points in two images based on their corresponding neighbor regions. The suggested algorithm selects the strongest key points and allocates the surrounding neighbors within a given window. It finds the corresponding neighbor regions from the nearby key points close to each key point of interest. For the descriptor, the suggested algorithm converts the disk-like area using the polar-to-rectangular transform to a rectangular matrix. Then, for each row and column in the matrix, it computes the mean and the standard deviation to produce means and standard deviations vectors.

The main contributions of this paper are as follows:

– The proposed descriptor is similarity (rotation, scale, and translation) invariant since it is computed from regions with rotated and scaled windows based on the local neighborhoods.
– The computation of the descriptor is simple because it is based on fundamental statistics (mean and standard deviation), which may lead to an acceptable degree of robustness to noise and image artifacts.

The rest of this paper is organized as follows: Section 2 summarizes the work in the literature that is related to the proposed algorithm in this paper. Section 3 describes the suggested methodology and gives details of its computation and implementation. Experiments and the results are shown in Section 4, with a discussion of the results. Finally, Section 5 concludes the paper with recommendations for future work.

2 Related Works

Image matching still attracts many interests. Among technical literature, general similarity measures methods are based on local or global features extraction. In this section, we review the latest keypoints image matching works based on local features.

Binary descriptors have been developed for many applications due to their implementation simplicity compared to other descriptors. The paper [1] used a binary keypoint-based descriptor to estimate the

relative pose of satellite from sequential images. The proposed low dimensional descriptors as the conventional keypoint detectors generate high dimensional descriptors complicating the computational process. Particularly, they reduced descriptors dimension using Harris detector to select some keypoints from enormous keypoints number. 27-D binary descriptor represents each keypoint based on local neighborhood points that were divided into 12 sectors with 6 binary bits each. They showed that the proposed method is suitable for keypoints matching using SIFT and ORB with less memory and efficient computation. The authors of [2] introduced a biologically inspired binary descriptor based on V1 cortical cell response. The motivation of such descriptor is to mitigate computational complexity in real-time application. They used two different datasets containing matching and nonmatching pairs where one of the datasets is used for training and the other for testing. The performance of the proposed descriptor is compared against other biologically inspired descriptors like BRISK and BRIEF, and the results verify its superiority.

On the other hand, keypoint-based descriptors can measure binary image similarity. The authors in [3] devised a binary image matching algorithm based on salient local keypoints. They developed the conventional color image descriptor for binary images. Mainly, they used the image background information to extract key points based on the image pixels' properties in a given local area. For each keypoint, contour points distribution in such area implements the feature vector. The authors tested the proposed algorithm for matching and retrieving hand-written and sketching images. The results proved that the proposed method competes against other conventional methods, especially for image retrieval.

Multi-sources images can be registered based on region homogeneity measures. Multimodal image construction is still an attractive area due to challenges that originate from various sources [4]. The paper [5] composes multimodal images. It proposed using histogram information to extract the angle and orientation of edges accurately from their distribution. It suggests extracting features from the said distribution based on the contour segments and Fréchet distance metric, leading to an accurate matching between multimodal images. Besides, it matches contour segments using a dual matching rule to filter mismatching points between contour segments. Another paper [6] developed a nonlinear SIFT-based descriptor that combines multispectral analysis to match synthetic aperture radar and optical images. Nonlinear diffusion constructs multiscale of both image types to preserve edge information. The paper exploited Harris detectors to find stable and repeatable keypoints. Eventually, it utilized Log Gabor histograms to construct their descriptor, proving its robustness against scaling, translation, and rotation of multimodal image matching. Another paper [7] used local keypoint to predict multispectral image matching initially. Then, it exploited the global keypoint information to enhance matching probability. The authors in [8] provide a comparative study for stereo image construction under different detectors and descriptors [9]. evaluated five keypoint-based descriptors performance at the laser scanning microscopy. Another paper [10] explored such descriptors performance at longitudinal retina registration.

Many studies are trying to improve descriptor's efficiency or reduce matching errors. The paper [11] proposed invariant wavelet-based descriptors. It extracts features in three scale pyramids key points to construct such descriptors. Particularly, it uses Harr wavelet transform to build such scale and obtain invariant translation and rotation using keypoints. The results show that the proposed descriptor's performance is similar to SIFT in discrimination but more efficient computationally. Besides, such descriptors outperform SURF and are much easier to be implemented. Another paper [12] proposed taking each pixel as a local pattern and binarizing gradient information to be rotationally invariant. Then, concatenating all sub-descriptors by combining statically local features into each bin. The demonstrated results show that the proposed descriptor is robust against different image transformations. Moreover, the article [13] developed a new differentiable version of SIFT. In contrary to SIFT algorithm, it used higher-order derivatives to create different layers of scale space. The proposed algorithm subtracts successive layers to construct the higher-order scale space. This technique extracts more features, but it increases

computational time. The authors in [14] proposed corner-based keypoint detector. The proposed detector is scale-invariant and efficiently reduces computational complexity. It depends on the gradient direction of a given closed contour in an integral image. First, it detects the edges in image and scale-spaces. Then, it subtracts the edge vectors from left and right of the targeted position to determine the corner. This method can rapidly decrease the complexity of keypoint detection and hence reduce processing time. Another paper [15] introduced a new feature extraction method so-called kaze (which means wind in Japanese) features. The proposed method aims to cover the shortages of gaussian scale-space features. In the blurred image, gaussian might ruin the boundaries and average the noise and some details resulting in inaccurate localization. Therefore, the authors developed a nonlinear scale-space descriptor using diffusing filters to adapt the noise to the real data and preserve the accuracy of object localization.

In [16], the proposed method aims to improve keypoints mismatching errors. It triangulates the key points in the source image and connects the points at the edges inside the triangular area. Then, it finds the corresponding formed shapes in the targeted image. Uncommon edges in both images are detected, and their corresponding points are removed. Such a method improves matching and increases descriptors robustness. Another article [17] uses Dempster Shafer theory to increase the confidence of keypoints similarity. It constructs evidence distribution from different matching descriptors information. Such distribution is sampled by the said theory resulting in reduced descriptor's matching error. In [18], the authors used the two-sided check technique to find optimal matching between keypoints in the reference and test images. They proposed to find all possible matching points in the test image for each point in the reference image. Similarly, they repeated the process backward to find all possible matching points in the reference image for each point in the test image. In each phase, they applied Hamming distance to measure the similarity between paired points. The persistent pairs in both phases are nominated to be correct matching, resulting in reduced matching error. In [18], the authors used the two-sided check technique to find optimal matching between keypoints in the reference and test images. They proposed to find all possible matching points in the test image for each point in the reference image. Similarly, they repeated the process backward to find all possible matching points in the reference image for each point in the test image. In each phase, they applied Hamming distance to measure the similarity between paired points. The persistent pairs in both phases are nominated to be correct matching, resulting in reduced matching error.

Machine learning is extensively exploited to enhance keypoint-based image matching. The paper [19] proposed a learning-based algorithm to find optimal keypoints correspondence for a local descriptor. Instead of a conventional neural network, it developed a bag matching dataset containing matching pairs of keypoints from similar objects and nonmatching pairs from different objects. The paper proposed differentiable score matching formula to minimize a Euclidean distance between descriptors. The method is learned from unlabeled video and three-dimensional models and applied to face matching. Similarly, [20] proposed self-supervised learning to match cloud keypoints.

In contrast, the paper [21] uses dynamic genetic programming to overcome the consequent variation of image rotation. Instead of machine learning that usually requires a large training dataset which is very expensive, it applies simple arithmetic and statistical operations to create descriptors. The results show that the proposed algorithm enhanced descriptor efficiency with fewer features.

Many comparison studies among literature have been done on different descriptors to build a good background of their performances. The authors in [22–24] conducted a comprehensive study of different types of famous keypoint-based detectors and descriptors to evaluate image matching under different imaging conditions such as rotation, perspective, blur, and zoom. Similarly, the paper [25] compared the performance of different types of affine region detectors such as Harris detector, region detector based on edges, region detector based on intensity, and salient region detector.

3 The Proposed Methodology

3.1 Background

Image matching can be categorized into area and feature based methods [26]. The area-based matching method uses a particular transformation model to measure similarity based on the global image intensity. Besides, some measure metrics and optimization algorithms are used to estimate the parameters of such a model. The measure metrics can be statistical such as the sum of squared error, cross-correlation, and mutual information. Nevertheless, the latter is based on statistical information between two images, and sometimes it is difficult for mutual information to find the global maxima in a given searching space [27]. Generally, the said metrics can be affected by deformations and the contents of matched images, and consequently, their reliability is reduced.

In contrast, the feature-based matching method comprises three complementary stages, feature detection, feature description, and feature matching. Feature detection is the process of finding a special structure or an object in the image, and can be classified as a corner, blob, or edge. To reduce computational complexity, some keypoints are used to realize feature detectors such as Harris, SIFT, SURF, etc., and return distinguishable points from their surrounding [28]. Subsequently, the surroundings of such points should be mapped into a discriminative vector to ease the matching method. This process is so-called feature description, where some descriptors are used to achieve it. The descriptors can be classified into float, binary, or learning-based descriptors. The float descriptor is a gradient-based descriptor that depends on statistical tools to be created. Such a descriptor stores the orientation of gradient in a vector to be used for feature description [29], and the SIFT is an example of this kind of descriptor. In contrast, the binary descriptor compares the local intensities in images using some strategies such as concentric circle. Although this descriptor is computationally cheap, it might not be robust for some applications. The last descriptor type is the machine learning-based descriptor, where matching the keypoints is based on data learning from the patch of image. Performance of such a descriptor is promising and suitable for different applications. For instance, the study in [30] verifies the efficiency of such descriptors in multimodal image matching. The last stage of feature-based matching method is refining the matching points in what so-called feature matching process. Feature matching might not be absolutely perfect due to similarity of some points intensities that are not belonging to the same area. The refining method can be classified into graph matching, registration, and statistical-based refinement. In graph matching, the set of points create graphs with nodes and edges, and the similarity between the images is based on the created graphs. By contrast, in the set of points registrations, we assume that a predefined transformation model is known, and iteratively, we try to estimate the model parameters and find the best matching. However, a preferable method is sampling or statistical-based method, where some tools such as random sample consensus (RANSAC) or maximum likelihood (ML) are used to remove the mismatched points.

We developed our proposed descriptor based on the keypoints feature-based method. Particularly, the proposed descriptor is generated using statistical tools such as the standard deviation and the mean to infer the correspondences in the reference and the query images based on the neighbor keypoints within a specific window. The distance and the direction between any key point of interest and its neighbor key point determine the size and orientation of the neighborhood region that computes the statistical-based descriptor. The following subsection discusses the procedure of descriptor generation and matching.

3.2 Keypoints Description and Marching Method

The general procedure of the keypoints feature-based image matching method is represented as a block diagram, as in Fig. 1. This method applies a keypoint detector to detect remarkable key points from the targeted images. Some key points are not strong enough, and they cannot provide relevant and significant

information. Therefore, it is very appropriate to exclude them before proceeding with the rest of the procedure. After identifying the strongest key points, we determine the neighbors of such key points within a specific window. Then, the descriptors concerning the neighbors in the selected window are generated for each key point. Based on these descriptors, the distances between key points are computed for matching. Finally, we refine the mismatched key points using RANSAC to remove inconsistent key points.

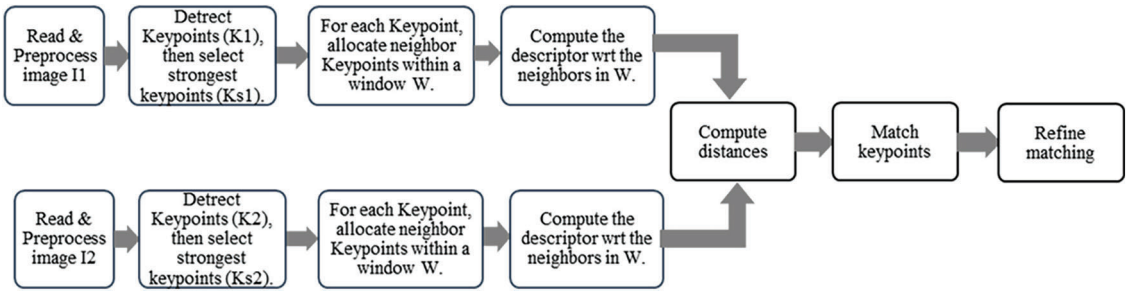

Figure 1: Keypoints detection, description, and matching algorithm's block diagram

3.2.1 Similarity Invariant Neighbor-Based Regions (SINR)

Given two images, a reference image I1 and a query image I2, the proposed algorithm starts by preprocessing the images to eliminate any artifacts in the image. It detects the key points (K1) for image I1 and (K2) for image I2 independently using the contourlet-based keypoint detectors as in [31]. This key point detector proved to have good repeatability compared to the state-of-the-art methods in this field [31]. From the detected key points, the proposed algorithm selects the strongest key points (Ks1) for the reference image I1 and the strongest key points (Ks2) for the query image I2. The number of the strongest key points (NKs2) in the second image is determined by the following equation:

$$NKs2 = \frac{NK2}{NK1}(NKs1) \tag{1}$$

NK1 represents the number of initially detected key points for image 1, NK2 is the number of detected key points in image 2, and NKs1 is the number of strongest key points in image 1. In this paper, due to the large number of initially detected key points in each image in the dataset, NKs1 is set to equal 1200 key points for all reference images.

The following procedure elucidates the descriptor computation for each key point:

– For each interest key point in Ks1 and Ks2, the neighbor key points are determined within an annulus-formed window W centered at the interest keypoint with inner radius equals d1 and outer radius equals d2. In this paper, d1 = 4 and d2 = 32, which, theoretically, allows the descriptor to be invariant to relative scale changes between images.
– For each interest-neighbor keypoint pair, a neighbor region is determined, as illustrated in Fig. 2. This neighborhood region, which is centered at the mid-point between the key point of interest and its nearby key point, is enclosed by these two key points, which scale the region's size according to the distance L between the two key points.
– A rectangular matrix from the pixel values on the polar grid from the formed neighbor region is computed by utilizing the polar-to-rectangular transformation, as shown in Fig. 3. This matrix is first normalized then re-sampled to have the same number of rows (# of circles = 12) and columns (# of angles = 72) for all extracted regions.

– The descriptor values are formed by the means and standard deviations vectors calculated from the rectangular matrix. The means vector consists of the mean of each row (i.e., circle) and the means of the columns (i.e., per angle), as demonstrated in Fig. 3. Both vectors are similarity invariant because they are computed from translation, scale, and rotation invariant regions. The same is repeated for the standard deviation vector for both the rows and the columns of the matrix.

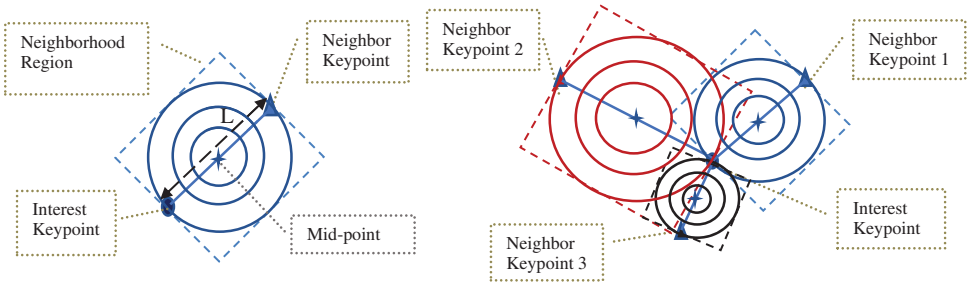

Figure 2: (left) A neighborhood region with concentric circles centered at the mid-point between a key point of interest and one of its nearby key points enclosed by the two key points. (Right) Example of three neighboring regions shaped by the key point of interest and its three neighboring key points

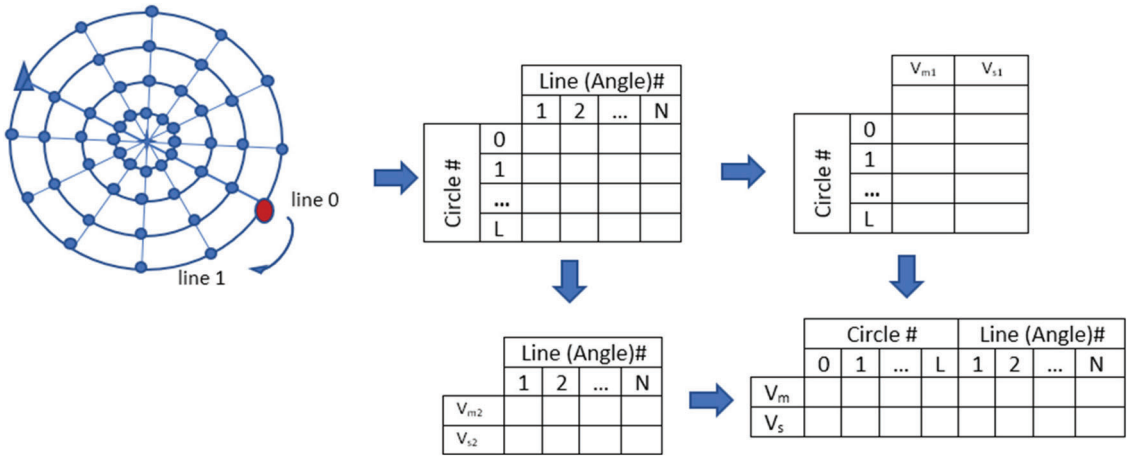

Figure 3: Descriptor computation based on the means and standard deviations of the rectangular matrix transformed from the polar grid of one of the neighbor regions

Combining the mean and the standard deviation vectors in such a way gives the descriptor the desired properties of invariance to similarity (translation, rotation, and scale) transformation, the needed level of discrimination, and the robustness to some extent of variations. When comparing a key point in image I1 with another key point in image I2, a distance matrix is computed from the descriptors defined by the neighborhood regions of each key point; then, the best three matches are averaged and stored as the distance between these two points. The "city-block" measure is used in this paper as the primary distance measure, where the matches are found by considering the matching problem as a linear assignment that minimizes the global cost of the distance matrix [32], which is already implemented in a MATLAB toolbox. The inliers are obtained by applying the RANSAC algorithm to the matches found so far. To further enhance the output of this method, another refining stage could be used by estimating the geometric transformation from the keypoint matches obtained from the RANSAC algorithm.

3.2.2 Extended Neighborhood Region

The method described so far suffers from the small regions formed from neighbors close to the interest keypoint. Another problem may occur because some of the neighborhood regions formed by the interest-neighbor keypoints pair could have uniformly distributed intensity values or small variations intensities. One solution to overcome this problem is to increase the value of d1 (inner radius of the annulus-shaped window). However, it will eliminate more neighbor key points, limiting the algorithm's scale invariance capability.

Another preferred solution to these problems is to expand the neighboring regions formed by interest-neighbor key points pairs, as shown in Fig. 4. The region described in the previous section is region (a), and the newly extended area is region (b). In region (b), the radius of the largest circle will be (L) rather than (L/2), as was the case for the region (a). All other steps for computing and matching the descriptors and are kept the same as described previously.

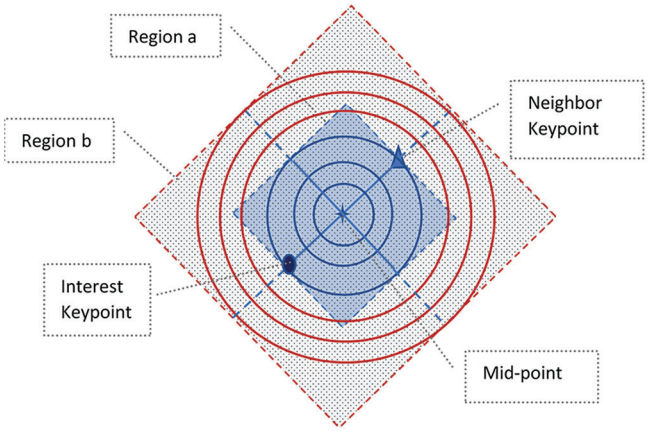

Figure 4: Extended neighboring region (b) with concentric circles (blue and red) delimited with a double distance between the key point of interest and a neighboring key point. Region (a) is defined as per Fig. 1

4 Results and Validation

4.1 Dataset

The dataset of Oxford [33], which consists of 8 groups of images with different variations, is used in the experiments to evaluate the performance of the proposed method and compare it to other well-known methods found in the literature. The eight groups have different image deformations such as viewpoint change, scale, and rotation transformations, in addition to blur, illumination, and JPEG compression artifacts, as shown in Fig. 5.

The efficiency of matching the key points is calculated by two measures: precision and recall.

Precision: measures the quality of the descriptor and the matching technique used to extract the correct matched features from the overall matched features. It can be computed as follows:

$$Precision = \frac{\#\ correct\ \ matches}{\#\ total\ \ matches} \tag{2}$$

Recall: measures the capability of the descriptor and the matching algorithm to allocated correct matches from all possible correspondences between the two images. It can be computed as:

$$Recall = \frac{\#\ correct\ \ matches}{\#\ correspondences} \tag{3}$$

Higher precision and recall values indicate better performance of the descriptor and the matcher.

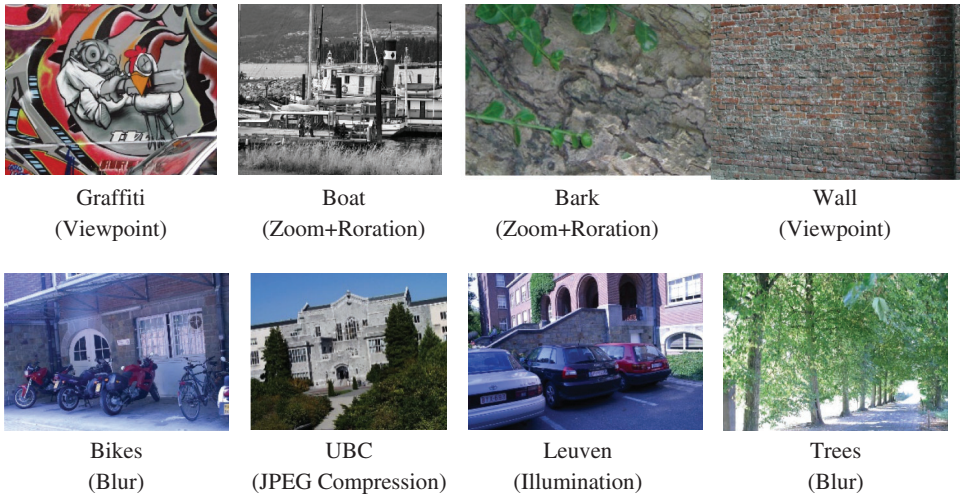

Figure 5: Reference image of each image group within the Oxford dataset

4.2 Comparing Region (a) with Region (b)

The average precision measure on all image groups in Oxford's data set is computed To assess the descriptors' performance in regions (a) and (b). The descriptors with RANSAC in region (a) are denoted by DaR, while DbR denotes the same descriptors in region (b). The descriptors with the RANSAC followed by inlier extraction using the geometric transformation estimation method for the region (a) are denoted by DaRG, while DbRG denotes the same descriptors in region (b), as seen in Fig. 6.

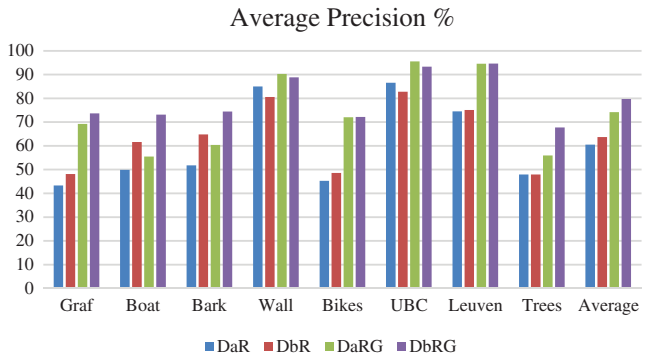

Figure 6: Average Precision comparison between descriptors DaR, DbR, DaRG, and DbRG

Fig. 6 shows that expanding the descriptor computation area beyond the key points (i.e., from region (a) to region (b)) improves the overall results for most of the image groups in the dataset. However, the Wall and UBC image groups have average precision values of Da better than the Db descriptor. Nevertheless, the overall average of all precisions shows this improvement among all the dataset groups.

4.3 Comparison with Other Methods

The method's performance presented in this paper is compared to the following detector-descriptor methods: SURF, ORB, FAST, kaze, and BRISK. Experiments are conducted using MATLAB's functions (ver 2021a) with their default settings. For a fair comparison, all methods are limited to their strongest 1200 key points on the reference image of each group. The number of the key points in the query image is calculated as indicated in Eq. (1). Furthermore, the RANSAC method is applied to eliminate outliers and retrieve inlier's key points after each method. The precision and recall defined by Eq. (2) and Eq. (3) are utilized for the comparison, as can be seen in the plots of Figs. 7a–7h.

Figs. 7a–7h indicates that the proposed descriptor's distance (DbRG) is well-suited and outperformed the other tested methods when it comes to measuring the quality of viewpoint change images (Graffiti and Wall), although it is not meant to be invariant to weak perspective or affine transformations. It also performs well and competes well with the SURF approach for images with rotation and zoom changes (Boat and Bark groups). In contrast, the other algorithms exhibit unstable and changing behavior in these groups. The ORB and kaze algorithms, for example, show good precision scores for the Bark group of images, except for the last image, but their performance drops for the Boat group. On the contrary, BRISK method has a better performance for Boat group than Bark group of images. All techniques attain high precision on JPEG compressed images and illumination variation images due to their implemented methods and the preprocessing approach employed on these images before conducting keypoint detection and matching. In Bikes images, which has more structures than the Tree images, all method, except FAST and some extend the proposed algorithm with (DbR), show a high level of robustness to the blur effect. For the natural images of the Trees group, kaze has the best performance for this type of variation, followed by SURF and then our suggested method with (DbRG). The suggested approach with (DbRG) is affected by higher blur effect due to the change of the intensity values, especially for the last image in the group two group. However, it achieves a satisfactory precision for the Bike image group and most of the Trees group, except for the last image.

The proposed descriptor computed from the extended region b (DbR & DbRG) has better recall measures, as seen in Figs. 7a–7h, than all other methods on all image groups. As indicated before in Eq. (2), the recall values depend on the correctly detected key points and the relative correspondences between the images subject to the comparison. The correspondences are found by the exhaustive search of key points with distances within 3 pixels apart from the key point of interest with the help of the homography matrix found with the data set. The high recall values are due to the efficiency of the overall suggested algorithm to detect and obtain a higher number of correct matched key points between these images. Fig. 8 shows the average number of correctly matched key points and the average F1-measure for the whole dataset for the proposed method compared with the adopted methods in the experiment. F1-measure combines the recall and precision values and is computed as follows:

$$F1 - measure = \frac{2 * Recall * Precision}{(Recall + Precision)} \tag{4}$$

It can be observed from this figure that the proposed method outperformed all methods in terms of the average number of correct matches and F1-measure.

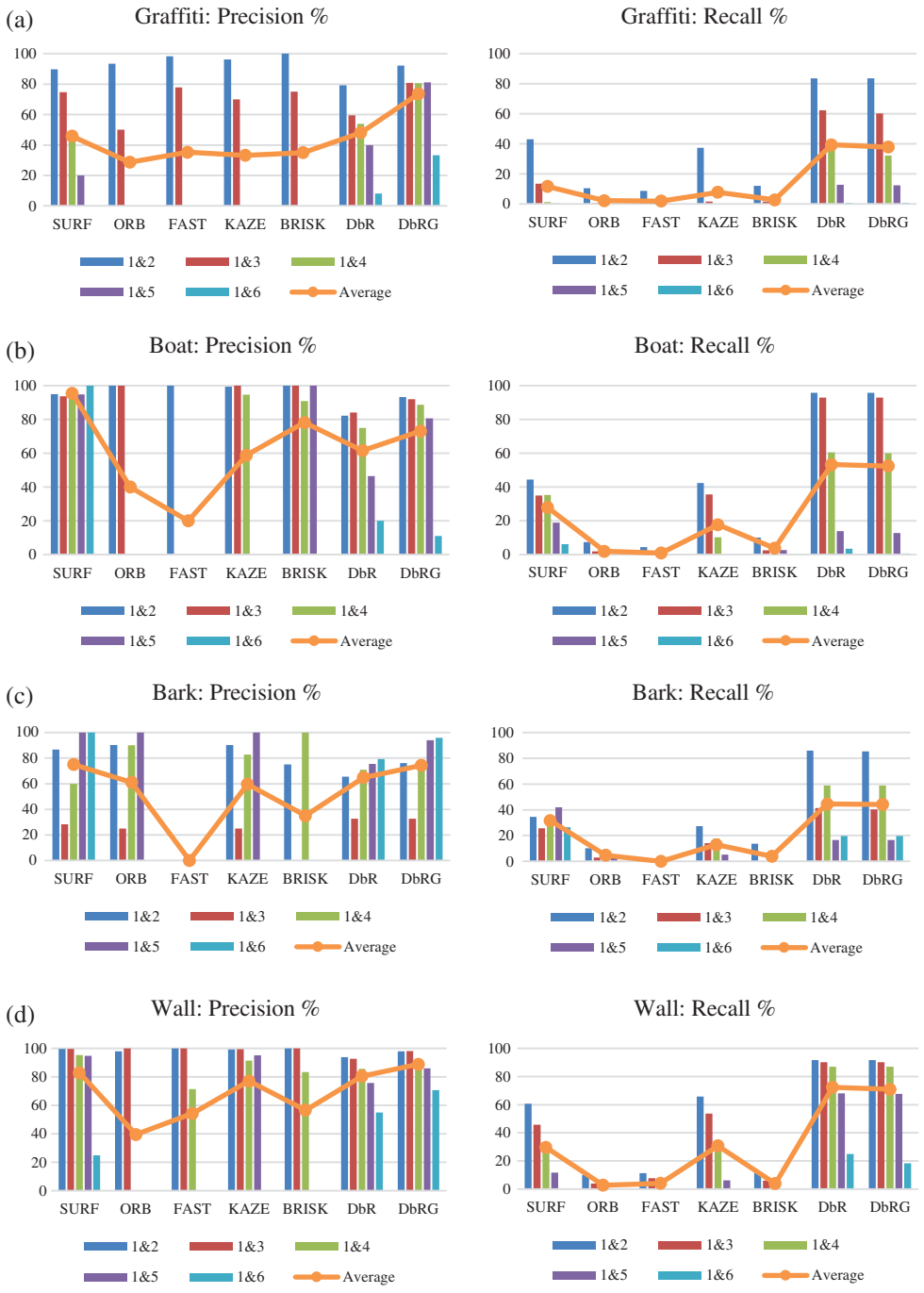

Figure 7: continued

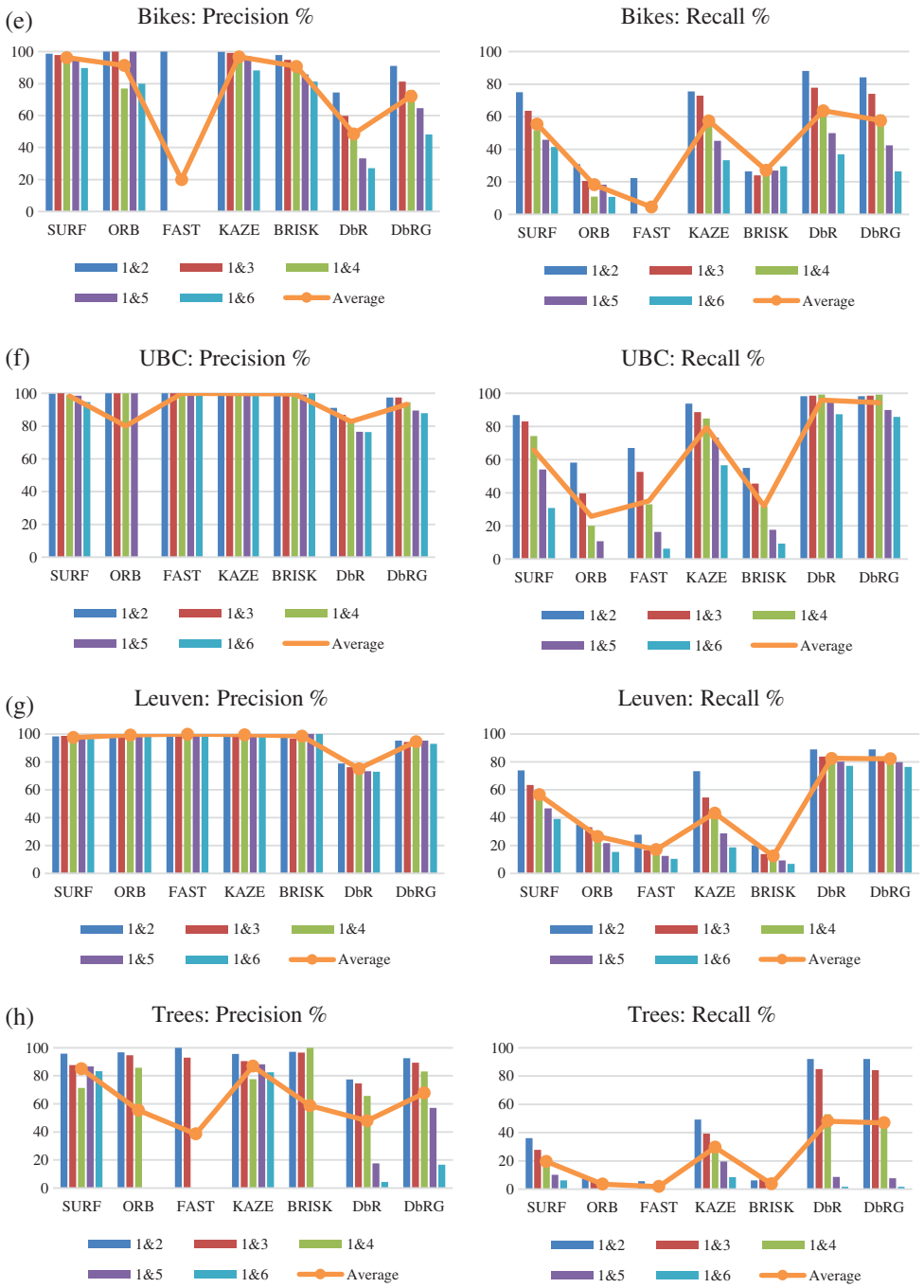

Figure 7: Precision and recall measures for different images in Oxford dataset. (a) Precision (left) and recall (right) for structured images with viewpoint change (Graffiti). (b) Precision (left) and recall (right) for rotation and zoom (scale) change images (Boat). (c) Precision (left) and recall (right) for rotation and zoom (scale) change images (Bark). (d) Precision (left) and recall (right) for viewpoint change images with texture (Wall). (e) Precision (left) and recall (right) for blurred structured images (Bikes). (f) Precision (left) and recall (right) for images with JPEG compression noise (Ubc). (g) Precision (left) and recall (right) for images with illumination change (Leuven). (h) Precision (left) and recall (right) for blurred natural images (Trees)

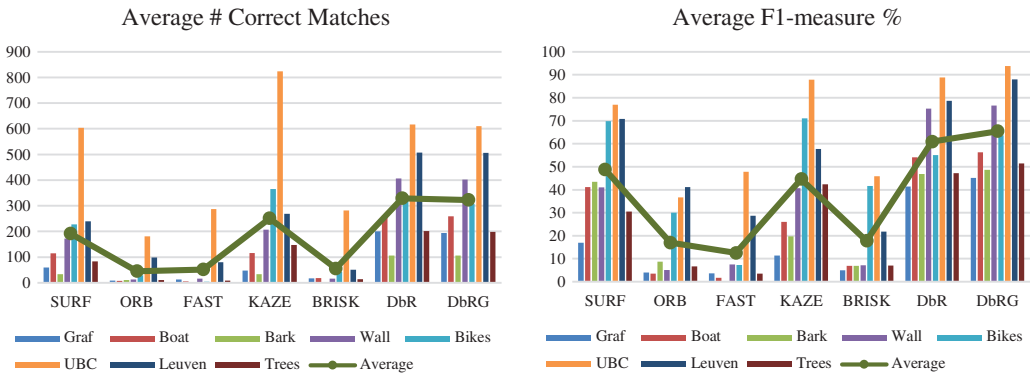

Figure 8: (left) Average number of correct matches and (right) Average F1-measure for all eight image groups

5 Conclusions and Future Recommendations

Although several state-of-the-art keypoint detectors and descriptors have been introduced in the last decades, there is still room for enhancement in this field. We introduced a simple descriptor based on statistical operations within the local neighborhood. The results show that the suggested key point description approach provides a convenient method to produce scale and rotation invariant descriptors from the neighbor regions of key points of interest. In general, the key point matching algorithm based on this descriptor outperforms the tested key point descriptor methods for images that were affected by geometric transformations (i.e., first four groups of Oxford's dataset), whereas it shows comparable results with SURF, kaze, and BRISK methods in most of the rest of image groups. The benefit of the suggested method includes its uses of simple statistics, mean and standard deviation, for calculation, and its invariance to rotation and scale variations without the need to search for hypothetical angles or scales. Still, some enhancements are to be done for this descriptor, such as increasing its invariance to higher transformations such as affine and perspective image deformations and enhancing its robustness to blur and noise effects. This invariance can be achieved by involving another neighboring key point in computing the area surrounding the key point of interest. Another suggestion is to use more statistics such as the measurement of skewness to increase the discrimination capability of the descriptor.

Acknowledgement: The authors would like to thank Taif University for supporting this work.

Funding Statement: The authors received no specific funding for this study.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] Y. Huang, Z. Zhang, H. Cui and L. Zhang, "A low-dimensional binary-based descriptor for unknown satellite relative pose estimation," *Acta Astronautica*, vol. 181, pp. 427–438, 2021.

[2] M. Saleiro, K. Terzić, J. Rodrigues and J. H. du Buf, "BINK: Biological binary keypoint descriptor," *Biosystems*, vol. 162, pp. 147–156, 2017.

[3] H. Chatbri, K. Kameyama, P. Kwan, S. Little and N. E. O'Connor, "A novel shape descriptor based on salient keypoints detection for binary image matching and retrieval," *Multimedia Tools and Applications*, vol. 77, no. 21, pp. 28925–28948, 2018.

[4] X. Jiang, J. Ma, G. Xiao, Z. Shao and X. Guo, "A review of multimodal image matching: Methods and applications," *Information Fusion*, vol. 73, pp. 22–71, 2021.

[5] G. Xu, Q. Wu, Y. Cheng, F. Yan, Z. Li *et al.*, "A robust deformed image matching method for multi-source image matching," *Infrared Physics & Technology*, vol. 115, p. 103691, 2021.

[6] Q. Yu, D. Ni, Y. Jiang, Y. Yan, J. An *et al.*, "Universal SAR and optical image registration via a novel SIFT framework based on nonlinear diffusion and a polar spatial-frequency descriptor," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 171, pp. 1–17, 2021.

[7] Y. Li, J. Zou, J. Jing, H. Jin and H. Yu, "Establish keypoint matches on multispectral images utilizing descriptor and global information over entire image," *Infrared Physics and Technology*, vol. 76, pp. 1–10, 2016.

[8] A. J. Malekabadi, M. Khojastehpour and B. Emadi, "A comparative evaluation of combined feature detectors and descriptors in different color spaces for stereo image matching of tree," *Scientia Horticulturae*, vol. 228, pp. 187–195, 2018.

[9] D. Ünay and S. G. Stanciu, "An evaluation on the robustness of five popular keypoint descriptors to image modifications specific to laser scanning microscopy," *IEEE Access*, vol. 6, pp. 40154–40164, 2018.

[10] S. K. Saha, D. Xiao, S. Frost and Y. Kanagasingam, "Performance evaluation of state-of-the-art local feature detectors and descriptors in the context of longitudinal registration of retinal images," *Journal of Medical Systems*, vol. 42, no. 4, pp. 1–11, 2018.

[11] C. Barajas-García, S. Solorza-Calderón and E. Gutiérrez-López, "Scale, translation and rotation invariant wavelet local feature descriptor," *Applied Mathematics and Computation*, vol. 363, p. 124594, 2019.

[12] T. Song, F. Meng, Q. Wu, B. Luo, T. Zhang *et al.*, "L2SSP: Robust keypoint description using local second-order statistics with soft-pooling," *Neurocomputing*, vol. 230, pp. 230–242, 2017.

[13] U. Park, J. Park and A. K. Jain, "Robust keypoint detection using higher-order scale-space derivatives: Application to image retrieval," *IEEE Signal Processing Letters*, vol. 21, no. 8, pp. 962–965, 2014.

[14] B. Li, H. Li and U. Söderström, "Scale-invariant corner keypoints," in *Proc. of 2014 IEEE Int. Conf. on Image Processing (ICIP)*, Paris, France, pp. 5741–5745, 2014.

[15] P. F. Alcantarilla, A. Bartoli and A. J. Davison, "KAZE Features," in *Proc. of European Conf. on Computer Vision*, Florence, Italy, pp. 214–227, 2012.

[16] X. Zhao, Z. He and S. Zhang, "Improved keypoint descriptors based on delaunay triangulation for image matching," *Optik*, vol. 125, no. 13, pp. 3121–3123, 2014.

[17] V. M. Mondéjar-Guerra, R. Muñoz-Salinas, M. J. Marín-Jiménez, A. Carmona-Poyato and R. Medina-Carnicer, "Keypoint descriptor fusion with dempster–Shafer theory," *International Journal of Approximate Reasoning*, vol. 60, pp. 57–70, 2015.

[18] M. -L. Cheng and M. Matsuoka, "An enhanced image matching strategy using binary-stream feature descriptors," *IEEE Geoscience and Remote Sensing Letters*, vol. 17, no. 7, pp. 1253–1257, 2019.

[19] N. Markuš, I. Pandžić and J. Ahlberg, "Learning local descriptors by optimizing the keypoint-correspondence criterion: Applications to face matching, learning from unlabeled videos and 3D-shape retrieval," *IEEE Transactions on Image Processing*, vol. 28, no. 1, pp. 279–290, 2018.

[20] Y. Yuan, D. Borrmann, J. Hou, Y. Ma, A. Nüchter *et al.*, "Self-supervised point set local descriptors for point cloud registration," *Sensors*, vol. 21, no. 2, pp. 486, 2021.

[21] H. Al-Sahaf, M. Zhang, A. Al-Sahaf and M. Johnston, "Keypoints detection and feature extraction: A dynamic genetic programming approach for evolving rotation-invariant texture image descriptors," *IEEE Transactions on Evolutionary Computation*, vol. 21, no. 6, pp. 825–844, 2017.

[22] Ş. Işık, "A comparative evaluation of well-known feature detectors and descriptors," *International Journal of Applied Mathematics Electronics and Computers*, vol. 3, no. 1, pp. 1–6, 2014.

[23] S. A. K. Tareen and Z. Saleem, "A comparative analysis of sift, surf, kaze, akaze, orb, and brisk," in *Proc. of 2018 Int. Conf. on Computing, Mathematics, and Engineering Technologies (iCoMET)*, Sukkur, Pakistan, pp. 1–10, 2018.

[24] Z. Peng, "*Efficient matching of robust features for embedded SLAM*," *M.S. thesis*, University of Stuttgart, Germany, 2012.

[25] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas *et al.*, "A comparison of affine region detectors," *International Journal of Computer Vision*, vol. 65, no. 1, pp. 43–72, 2005.

[26] J. Ma, X. Jiang, A. Fan, J. Jiang and J. Yan, "Image matching from handcrafted to deep features: A survey," *International Journal of Computer Vision*, vol. 129, no. 1, pp. 23–79, 2021.

[27] A. Sotiras, C. Davatzikos and N. Paragios, "Deformable medical image registration: A survey," *IEEE Transactions on Medical Imaging*, vol. 32, no. 7, pp. 1153–1190, 2013.

[28] J. Fan, Y. Wu, M. Li, W. Liang and Y. Cao, "SAR and optical image registration using nonlinear diffusion and phase congruency structural descriptor," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 9, pp. 5368–5379, 2018.

[29] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. of 2005 IEEE Computer Society Conf. on Computer Vision and Pattern Recognition (CVPR'05)*, San Diego, CA, USA, vol. 1, pp. 886–893, 2005.

[30] S. Kim, S. Lin, S. R. Jeon, D. Min and K. Sohn, "Recurrent transformer networks for semantic correspondence," *Advances in Neural Information Processing Systems*, vol. 31, pp. 6126–6136, 2018.

[31] S. R. Saydam, I. A. E. Rube' and A. A. Shoukry, "Contourlet Based Interest Points Detector," in *Proc. of 2008 20th IEEE Int. Conf. on Tools with Artificial Intelligence*, Dayton, Ohio, USA, pp. 509–513, 2008.

[32] I. S. Duff and J. Koster, "On algorithms for permuting large entries to the diagonal of a sparse matrix," *SIAM Journal on Matrix Analysis and Applications*, vol. 22, no. 4, pp. 973–996, 2001.

[33] University of Oxford dataset, 2021. [Online]. Available at: https://kahlan.eps.surrey.ac.uk/featurespace/web/data.htm.