# Image Augmentation for Keypoint Detection and Matching Assessments

**Ibrahim El rube`**,

Computer Engineering Department, Taif University, KSA

**Abstract**
Keypoint detection and matching algorithms are frequently compared in the literature using datasets of real-world images that have a range of geometric and non-geometric variations; these include viewpoints, illuminations, visual content, and distortions. Homography (H) matrices often describe geometric variations when utilizing these image datasets. However, models for non-geometric differences between these images are rarely offered, resulting in inaccurate and misleading comparisons. This study presents a methodology for objectively comparing classical keypoint detection and matching algorithms by eliminating implicit non-geometric influences from assessments, therefore, offering a step towards limiting the comparison between an image pair to the geometric transformations between them. This proposed technique uses the H matrix provided by the image dataset to generate an augmented image that resembles one of the images in each image group. The performance of the proposed technique was evaluated using several traditional keypoint detections and matching techniques using image groups from well-known datasets to determine the impact of excluding non-geometric changes. The assessments are conducted using the performance measures of repeatability, precision, and recall rates.

***Keywords:***
*Augmented image; geometric transformation; homography matrix; keypoints.*

## 1. Introduction

Numerous academics and researchers have investigated image matching by detecting and comparing visual features like corners and keypoints [1]. In image-based applications, keypoints are the most often applied features (sometimes called interest points) which are used in various applications such as robotic navigation [2], image registration [3], object recognition [4], visual tracking [5], augmented reality [6], and many more [7]. Typically, the extracted or generated keypoints are desired to be insensitive to differences between similar images captured from the same real-world scene. These differences may include geometric transformations such as rotation, scale, skew, and perspective changes and photometric variations such as illumination changes. In addition, other non-geometric alterations must be resolved when dealing with the keypoints of these images.

Recently, several keypoint detectors and descriptors have been published, ranging from handcrafted keypoint algorithms to the most current neural networks and deep learning-based algorithms [1]. Although modern deep learning algorithms are more accurate and efficient in tackling issues involving similar images with geometry and photometric alterations, classical approaches remain competitive for certain image transformations [8]. Moreover, classical techniques may be involved or utilized in some deep learning algorithms for their operation and structure, with the former being less complex, generally faster, and needing no training phase to function. Because of this, these keypoint detectors and descriptors are still in use and high demand for several studies and applications.

Detection and matching algorithms are commonly compared using geometric and photometric variations datasets. Databases such as "VGG" [9] and "HPatches" [10] are often associated with homography (H) matrices to assist in the comparison of images. The H matrix needs to be given to aid in verifying the comparison, yet, models for implicit non-geometric adjustments, such as photometric changes, that have the potential to impact comparisons are rarely revealed. This problem suggests that keypoint-based algorithm comparisons may be inaccurate if the geometric transformation is challenging. An assessment approach based on image augmentation that discards implicit unprescribed non-geometric changes in evaluating keypoint-based algorithms is presented to address this issue. While works of literature generate new H matrices to produce augmented images for training purposes, this research augments images using the dataset's provided H matrix to avoid non-geometric modifications. The effect of removing non-geometric modifications is shown by comparing these image pairings' repeatability, accuracy, and recall rates to those of the originals using well-established keypoint detection and matching algorithms.

The following section describes picture augmentation and its role in image generation. Next, dataset artifacts that affect keypoint matching comparisons are described. The suggested solution is then presented after that. The fourth section compares the recommended approach to images with three keypoint detectors. Last, a conclusion is drawn with a look to future work.

Manuscript received November 5, 2022
Manuscript revised November 20, 2022
https://doi.org/10.22937/IJCSNS.2022.22.11.78

## 2. Image Augmentation

Image augmentation refers to the creation of new images from existing ones. Several ways have been used to build or grow image datasets, especially when training learning-based systems [11]. Inception, ResNet, and EfficientNet are a few well-known models that use image augmentation techniques during training [12]. In addition, as shown in [13] [16] [15] [16], image augmentation has also been used to match features in images with geometrical changes.

In [13], data augmentation was used as part of the training procedure for the conventional neural network (CNN). When the training images are replaced with counterparts that have undergone a geometric transformation, the trained network performs very well for image quality assessment (IQA), which is insensitive to geometric modifications. The authors of [6] noted that the environment is considered by camera-based systems when a keypoint detector or descriptor is used in practical applications like augmented reality. As a direct result, it is challenging to establish a technique suited for mobile apps that considers diverse viewing situations and resolves the noise issues associated with printed images [6]. Although the proposed augmentation strategy in [14] is somewhat light compared to the raw local feature description, it provides significant gains on large-scale benchmarks and in various contexts. This enhancement demonstrates the applicability and generalizability of geometric matching in many situations. Images of wild outdoor or indoor scenes, patch-level homography datasets, and application-level 3D reconstruction image sets are benchmarks [14]. During the training phase, the authors of [15] used various strategies to augment the data. Various cropping, scaling, and rotating methods were used to obtain the augmented images. In [16], the authors provide a system for locating and tracking various geometrically featured planar objects. Combining traditional keypoint detectors with Locally Likely Arrangement Hashing (LLAH) can achieve keypoint matching based on geometric features. If the keypoint matching inlier rate is above a predefined threshold, a pyramid level is chosen for the image-pyramid.

In the literature, the major goal of image augmentation for features and image matching was to train learning networks. This goal was achieved by delivering and integrating new images with varied scenarios do not present in the original collection. This study augments images differently than previous research. This study employs image augmentation to create images with only preset geometric differences (no non-geometric changes) for objective keypoint identification and matching assessments.

The generation of augmented images could be accomplished based on the H matrix provided in datasets, as indicated previously. For example, given the image pair $I_a$ and $I_b$, the second image is augmented from the first image.

$$\tilde{I}_b^a = I_a^T = I_a^{\mathrm{H}} \quad (1)$$

where $\tilde{I}_b^a$ represents the augmented image that is equivalent to the original image $I_b$ by transforming image $I_a$ using the transformation matrix T. Usually, the datasets are associated with the H matrix that transforms the first image (often called reference image) to the second image in the same image group.

Figure 1 is an example from the VGG dataset, the "Graffiti" image group (image pair $I_1$&$I_4$), which demonstrates a change of viewpoint across images of the same scene. One image is generated from the other image in the same pair using the appropriate transformation based on the H matrix of that pair. The symbols $I_a$ and $I_b$ represent the first and second images in each image pair for image groups within the examined datasets.

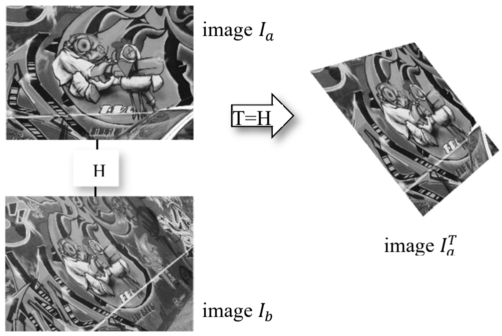



Fig. 1. Sample of image generation using the given homography matrix H for the image pair ($I_a = I_1$ & $I_b = I_4$) of the "Graffiti" group.

## 3. Viewpoint-Change Images Issues

Viewpoint-change datasets from real-world settings often include image group issues due primarily to violations of the homography assumption (i.e., images are not restricted to a plane), inaccurate homographies [12] [17], non-rectified distortions (e.g., the benchmark may not address the radial distortion [10]), and implicit non-geometric (e.g., photometric) changes. Even when the H matrix is supplied, which should aid in validating the comparisons, no model is often provided to explain any other discrepancies between these image pairs. Therefore, non-prescribed modifications, such as moving objects and photometric variations, may impact the comparisons' outcomes. For example, using the image pair from Fig. 1, the inverse of the H matrix is used to transform and project (align) the second image onto the first image, as shown in Fig. 2. This alignment of the original image pair ($I_1$ & $I_4$) in the "Graffiti" image group results in several differences and inaccuracies. These flaws cannot be rectified by adjusting the projection matrix, which is restricted to changes within the image plane. Images taken in real-world settings from various perspectives often include similar implicit variations. The imprecise H matrix and non-geometric alterations degrade the quality and accuracy of performance assessments for any image-matching process.

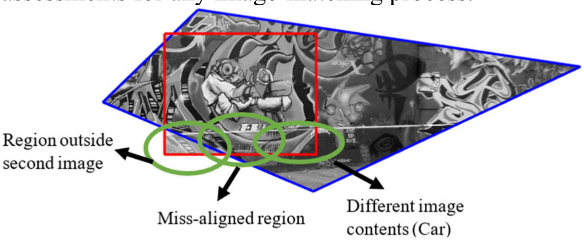



Fig. 2. Example of the differences that might affect the performance of an image-matching algorithm. Using the given homography matrix H, the figure depicts the projection of image ($I_4$) into image ($I_1$) for the Graffiti image group.

To further investigate the effect of these errors on images, the structural similarity index measure (SSIM) is utilized. The three steps of the SSIM method for evaluating similarity consist of brightness, contrast, and structural comparisons. Structural information extraction is independent of contrast and light effects [18]. The SSIM value ranges from 0 to 1, and it is defined by

$$SSIM(I_a, I_b) = \frac{(2\mu_a\mu_b + k_1)(2\sigma_{ab} + k_2)}{(\mu_a^2 + \mu_b^2 + k_1)(\sigma_a^2 + \sigma_b^2 + k_2)} \tag{2}$$

where $\mu_a$ is the average of image $I_a$, $\mu_b$ is the average of image $I_b$, $\sigma_a^2$ and $\sigma_b^2$ are the variances, and $\sigma_{ab}$ is the cross-covariance of the two images. $k_1$ and $k_2$ guarantee that the denominator is not zero.

The SSIM map, which presents a local image quality measure in the image's spatial domain, can show how similar the local areas of the two images are. Large (bright) values show similar local areas, while small (dark) values show different local areas. For example, figure 3 shows the SSIM maps for the "Graffiti" image pair ($I_1$&$I_4$) when projecting and aligning image $I_4$ to image $I_1$. The lower-intensity (darker) values in the SSIM indicate high errors between aligned images.

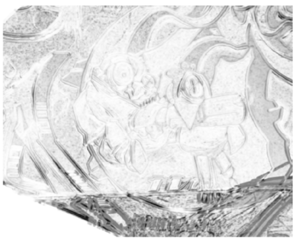


Fig. 3. Errors revealed by the SSIM map in the common region of image pair $I_1$ & $I_4$ of the "Graffiti" image group.

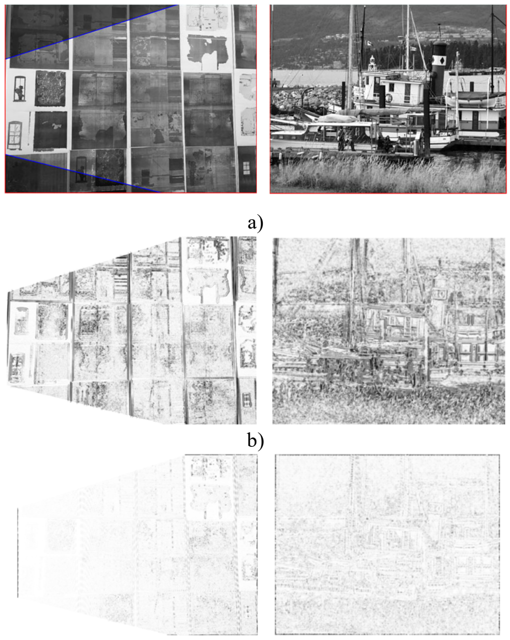
## 4. Comparisons for Keypoints-based Algorithms

The performance evaluation of the algorithms that use these pictures is impacted by the inaccuracies in the aligned image pairs of the perspective change datasets. This study offers objective comparisons by removing implicit non-geometric and non-described image modifications. $\tilde{I}_b^a$ is a mathematically transformed version of $I_a$, with no non-geometric alterations. Fig. 4 shows the SSIM map calculated for this newly generated image with its original image compared to Fig. 3 for the original image pair without augmentation.

a) b)

Fig. 4. a) The common region of the aligned "Graffiti" images $I_1$ and $\tilde{I}_4^1$, b) The SSIM map for a).

Similarly, inaccuracies might be found in other datasets. For example, Fig. 5 displays errors in one of the viewpoint-change image groups "Abstract" in the "HPatches" dataset. Even though the dataset was intended to resolve the issues of the previous dataset, VGG, in which the homography constraint was not adequately enforced, it is evident that the error is due to the inaccuracy of the H matrix and the influence of the implicit variations.



a)

b)

c)

Fig. 5. Additional illustrations of the errors in the common region of the image pair in (left column) the “Abstract” group (from dataset “HPatches”) and (right column) “Boat” group (from dataset “VGG”) a) projection of image $\tilde{I}_4^1$ onto image $I_1$, b) the SSIM map for the

original image pair ( $I_1$ & $I_4$ ), c) the SSIM map, using data augmentation, for the image pair ($I_1$&$\tilde{I}_4^1$).

Figures 3-5 illustrate, as instances, the SSIM maps for the "Graffiti," "Abstract," and "Boat" image pairs ($I_1$&$I_4$) from two different datasets. These differences may be caused by missing parts, image contents change, and image segments outside the comparable region, as seen in Fig. 2. Moreover, scene illumination change, and resampling/interpolation process may cause the differences. Such variations may be present implicitly in several datasets used to evaluate and verify detection and matching algorithms, but primarily the assessment procedure disregards them.

### 4.1 Keypoint-based Methods

Keypoints are essential visual characteristics that have lately gained widespread use in the literature. Some of these methods were insensitive to fundamental geometric transformations like rotation and scaling, while others were designed to be invariant under higher-order transformations such as affine and projective transformations.

While deep learning has provided a new approach to image matching, it still faces challenges during training, such as using CNN models on sparse point data for registration, the estimate of transformations, and matching [1]. According to [8], some conventional methods compete with and, in some situations, outperform deep learning methods. However, deep-learned algorithms outperform conventional approaches in other situations by a small margin. For example, although deep learning is favoured in illumination invariance, conventional detectors remain competitive [19]. In [20], while proven, deep learning's potential to tackle matching challenges in photogrammetry and remote sensing is not entirely exposed. Although automation and deep learning reduce the need for domain expertise, utilizing them as "black boxes" without checking outcomes is risky. Although CNN-based descriptors outperform traditional handcrafted features, according to the authors of [21], an appropriate adjustment of features may considerably improve the results of all methods and reduce the difference between SIFT (Scale Invariant Feature Transform [22]) and the existing learning-based descriptors. However, for handcrafted descriptors, this also includes determining the suitable normalization for the target domain [21].

Despite extensive previous studies, as in [23] [24] [25], no single detector or descriptor is best for all image geometrical and photometric changes [6] [26]. For example, in [27], it was shown that SIFT is one of the most accurate algorithms; however, in [22], both the SURF (Speeded Up Robust Features) [28] and KAZE [29] approaches significantly surpassed SIFT. Moreover, KAZE is superior to SURF [30]. Therefore, as prototypes, these three keypoint detectors have been chosen for utilization in the experiments of this paper.

### 4.2 Performance Measures

In the literature, a range of measurements is used to evaluate and contrast traditional keypoint-based methods' computational and statistical performance, including repeatability, accuracy, and recall rates. The repeatability rate is computed for evaluating the performance of keypoint detectors, while the precision and recall rates are applied to descriptors and matching stages. For a keypoint detector applied to two similar images, $I_a$ and $I_b$, with $N_a$ detected keypoints in the first image and $N_b$ detected keypoints in the second image that lie in the common region of both images, the-repeatability rate is computed similar to [31].

$$Rep = \frac{N_{rp}}{2}\left(\frac{1}{N_a} + \frac{1}{N_b}\right)\ \% \qquad (3)$$

where $N_{rp}$ is the repeated keypoints count computed when projecting image $I_b$ onto image $I_a$.

The precision and recall rates are calculated by,

$$P = \frac{N_c}{N_{mt}}\ \% \qquad (4)$$

and

$$R = \frac{N_c}{N_{cr}}\ \% \qquad (5)$$

where $N_c$, $N_{mt}$, and $N_{cr}$ are the counts of correctly matched keypoint, the total matches, and the correspondences computed by projecting image $I_b$ onto image $I_a$ , respectively.

## 5. Experimental Results

### 5.1 Viewpoint-Change Image Groups

The image groups used in the experiments are taken from two well-known datasets in the computer vision field: the VGG (Oxford) [9] and HPatches [10] datasets. The images in these datasets were divided into groups based on their geometric and photometric changes. Since the interest in this paper is in geometric transformations, six image groups are chosen to represent the transformation; three groups, "Graffiti," "Boat," and "Bricks," belong to the first dataset, while the other three groups, "Abstract," "Bird," and "Laptop," are from the second dataset. The primary purpose behind forming all these image groups was to evaluate detection and matching algorithms for changes in geometrical transformations across images. The changes in these groups range from rotation+scale to projective transformations, as seen in Fig. 7, which displays all the image groups used in this study's experiments. All experiments throughout this paper were implemented using MATLAB 2022 software and its associated toolboxes.

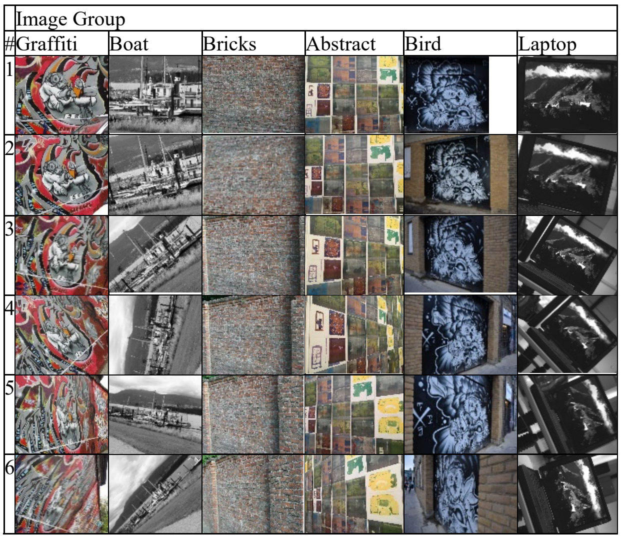
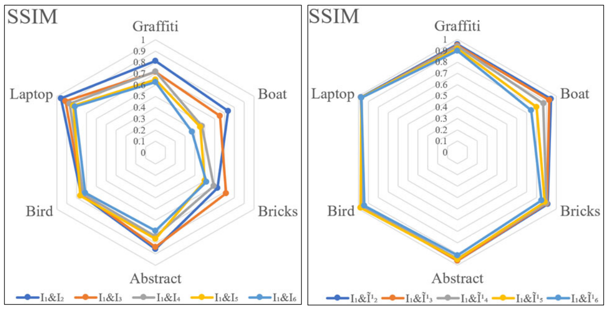



Fig. 6. The image groups utilized in the experiments with their names. Each of the displayed groups contains six images with varying geometric transformations.

## 5.2 SSIM Measures

As stated in the earlier section, the tested datasets include the homography H matrix that specifies the geometric change between the group's first image (reference) and the remaining images. However, the images presented in the studied datasets, including the image groups in Fig. 6, suffer from erroneous alignment and non-geometric variances that impact the keypoint matching algorithms' measurements with no explicit, detailed information on the non-geometric changes in these groups. This is evident from Fig. 7, which depicts the SSIM measurements (2) of the six tested image groups. The top plot is obtained when the second original image, $I_b$, is transformed and superimposed over the original first image, $I_a$, within its pair of images. The bottom plots show the SSIM measurements when image augmentation is used for generating the image $\tilde{I}_a^b$using the supplied H matrix for each image pair in the tested datasets. The results of Fig. 7 demonstrate that the image augmentation technique used in this study significantly enhances the similarity between images in each image pair assessed for geometric differences. However, the considerable scale differences between the images in the "Boat" image group provide variable SSIM findings. These significant scale changes induce additional errors due to the accompanying resampling and interpolation processes.




a) b)

Fig. 7. The SSIM measurements for a) original image pairs ($I_a$, $I_b$) when $I_b$ is projected onto image $I_a$; b) image pairs ($I_a$,$\tilde{I}_b^a$)using image augmentation augmented image $\tilde{I}_b^a$ projected into image $I_a$. $I_a$ and $I_b$ represent the first and second images for each image pair examined.

## 5.3 Performance Assessment

Three traditional handcrafted keypoint detectors and descriptors were adopted as examples in this paper for performance comparisons using the proposed method; KAZE, SURF, and SIFT. Figures 8-10 show the repeatability, accuracy, and recall rates for the original image pairs ($Rep_0$, $P_0$, and $R_0$) and use the augmentation method for the second image in each r ($Rep_1$, $P_1$, and $R_1$). These rates are determined for a distance of two pixels in all cases.

The results of Figs. 8-10 illustrate that, in general, all measures, especially the precision, of the three applicable keypoint approaches, KAZE, SURF, and SIFT, improved when implicit unprescribed (non-geometric) modifications were excluded from comparisons. Since the three studied methods are rotation- and scale-invariant, this benefit is particularly noticeable for similarity and moderate affine transformation. The precision rates of the augmented modified image pairs for the "Laptop" and "Boat" image groups are much higher than the original image pairs' rates. The increase in precision rates varies between the "Bricks" and "Abstract" image sets, which exhibit affine and weak perspective transformations. KAZE performs better with dense keypoints in these image groups' augmented images. The limited number of keypoints and a greater degree of modifications in the last image pair of the "Bricks" group present difficulties for the performance of the other techniques, SURF and SIFT. Despite improvements, the last two pairs of images in perspective transformation, "Graffiti" and "Bird" image groups, continue to be challenging for all three methods due to the greater degree of modification than the first images in the groups.

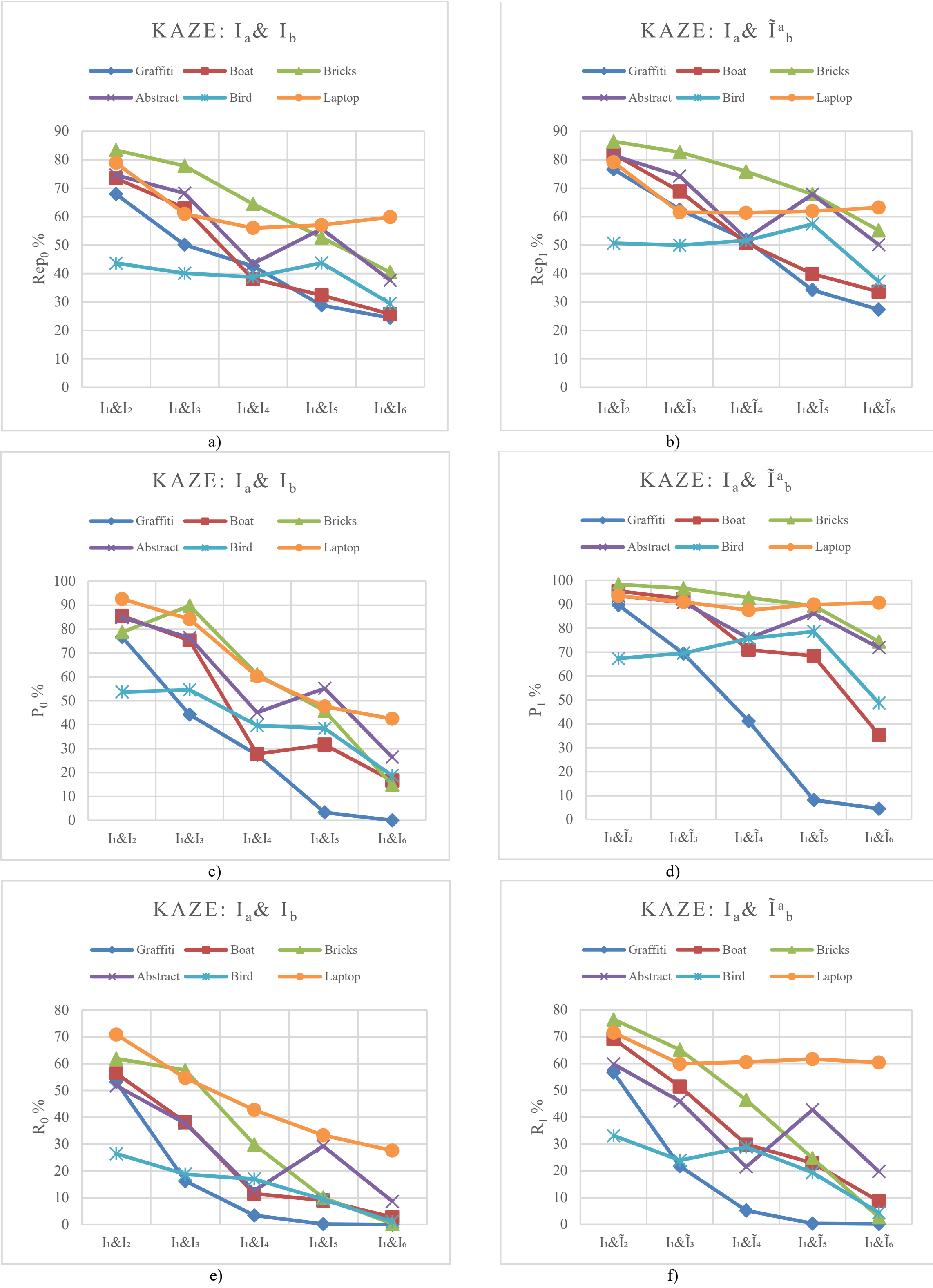



Fig. 8. Repeatability (Rep%), precision (P%), and recall (R%) rates using the KAZE method for the original image pairs (a, c, and e in the left column) and when the second image is augmented in each image pair (b, d, and f on the right column).

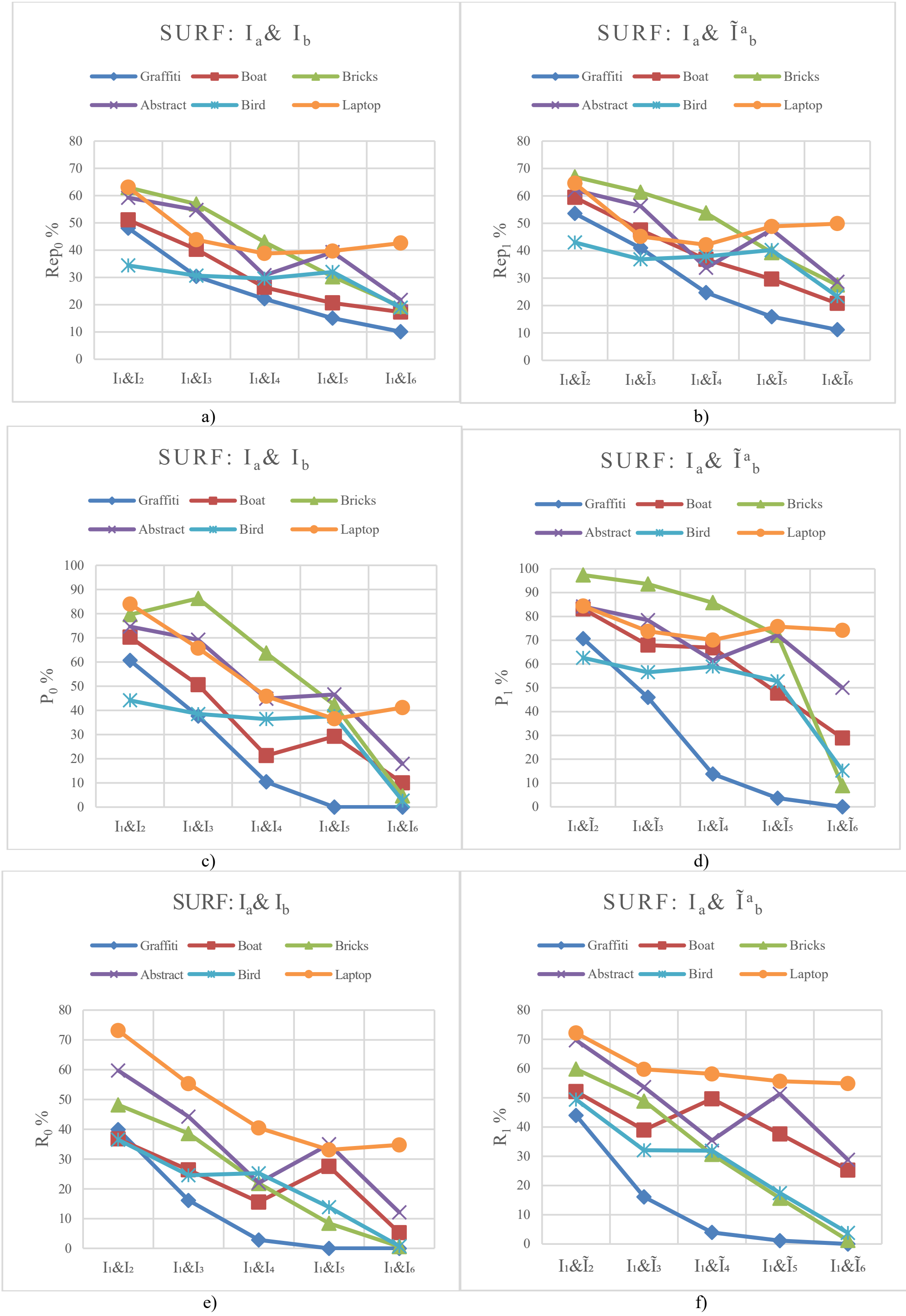



Fig. 9. Repeatability (Rep%), precision (P%), and recall (R%) rates using the SURF method for the original image pairs (a, c, and e in the left column) and when the second image is augmented in each image pair (b, d, and f on the right column).

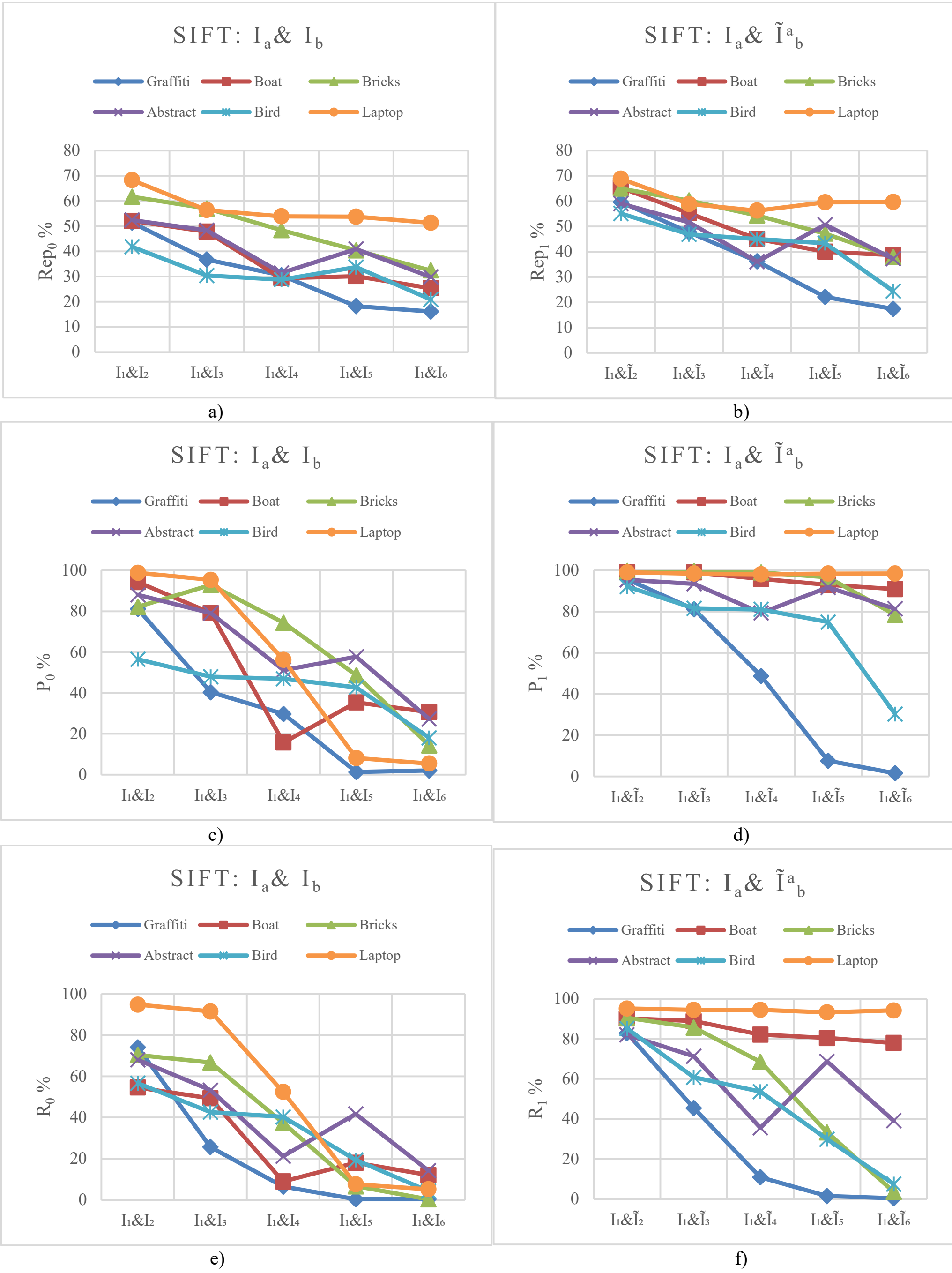



Fig. 10. Repeatability (Rep%), precision (P%), and recall (R%) rates using the SIFT method for the original image pairs (a, c, and e in the left column) and when the second image is augmented in each image pair (b, d, and f on the right column).

The effects of the chosen augmentation approach on the averages of the three metrics, repeatability, accuracy, and recall rates, are shown in Fig. 11. Each bar in Fig. 11's plots represents the change in the average of the stated measure following image augmentation, stacked for the three keypoint-based approaches. Figure 11 presents the values according to the following

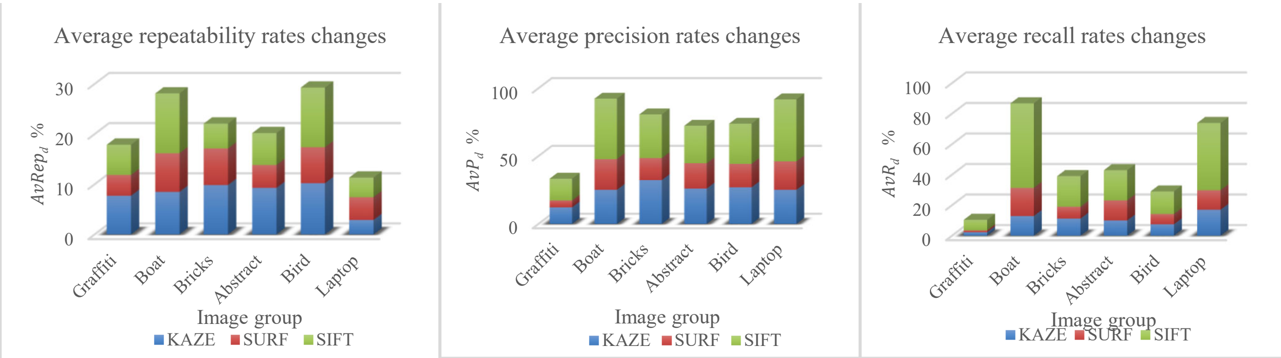
$$AvRep_d = AvRep_1 - AvRep_0 \% \qquad (6.a)$$

$$AvP_d = AvP_1 - AvP_0 \% \qquad (6.b)$$

$$AvR_d = AvR_1 - AvR_0 \% \qquad (6.c)$$

where $AvRep_0$, $AvP_0$, and $AvR_0$ are the averages of the original images' repeatability, precision, and recall rates per image group ($Rep_0$, $P_0$, and $R_0$, are illustrated in Figs. 8, 9, and 10, respectively). When applying image augmentation, the averages of the repeatability, precision and recall rates per image group are denoted as $AvRep_1$, $AvP_1$, and $AvR_1$, respectively ($Rep_1$, $P_1$, and $R_1$, respectively) are shown in Figs. 8, 9, and 10).

All three keypoint-based algorithms evaluated reacted differently to the augmented images with replaceable superiority, consistent with the earlier section's comments and the literature results. However, according to Fig. 11, the repeatability and accuracy rates of the KAZE and SIFT algorithms were more significantly impacted by the image augmentation than those of the SURF approach. The SIFT technique, on the other hand, performs better for the rotation and scale change image groups "Boat" and "Laptop" since it is supposed to be invariant to these sorts of changes. Furthermore, due to the high number of generated keypoints, the KAZE algorithm performs better in terms of repeatability than other methods.




a) b) c)

Fig. 11. Comparisons of the change rates for the three performance measures: a) Repeatability, b) Precision, and c) Recall rates.

## Conclusions

Accurate and fair performance evaluation of keypoint detectors and matching algorithms is essential in developing modern-day image processing and machine vision applications. These performance evaluations typically use datasets of real-world images that have a range of geometric and non-geometric variations. Unfortunately, ignoring non-geometric influences has reduced the accuracy and fairness of the performance evaluation outcomes. Therefore, a methodology is presented in this paper to address this issue. The proposed technique creates an augmented image for each pair for image datasets where the H-matrix is included, eliminating the non-geometric influences. The proposed method for augmenting images does not require replacement comparisons with the source datasets. Instead, it offers an additional way to analyze exact geometric transformations of real images using H matrices derived from actual perspective change situations. The performance of the proposed method was evaluated using three well-known traditional handcrafted keypoint detectors and descriptors, namely KAZE, SURF, and SIFT. In addition, the repeatability, precision, and recall rates were measured. Results demonstrate that the three keypoint detectors and descriptors behave as predicted when analyzing geometric modifications.

Future work should build on these experimental comparisons of the algorithms and conduct a more thorough review of them utilizing various evaluation techniques. In addition, this work might be expanded to explore the relationship between other particular combinations of augmenting data and assessing methods.

## References

[1] 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, Aug. 2020.

[2] F. Kallasi, D. L. Rizzini, and S. Caselli, “Fast Keypoint Features from Laser Scanner for Robot Localization and Mapping,” IEEE Robotics and Automation Letters, vol. 1, no. 1, pp. 176–183, Jan. 2016.

[3] C. Liu, J. Xu, and F. Wang, “A Review of Keypoints’ Detection and Feature Description in Image Registration,” Scientific Programming, vol. 2021, pp. 1–25, Dec. 2021.

[4] S. Kim and Y. Hwang, “A Survey on deep learning based methods and datasets for monocular 3D object detection,” Electronics, vol. 10, no. 4:517, pp. 657–666, Sep. 2021.

[5] K. Y. Kok and P. Rajendran, “A Descriptor-Based Advanced Feature Detector for Improved Visual Tracking,” Symmetry, vol. 13, no. 8, p. 1337, Jul. 2021.

[6] M. Mokhtar, F. Mohamed, M. Sunar, A. Abd Aziz, M. Arshad, and M. Sidik, “Image Features Detection and Tracking for Image Based Target Augmented Reality Application,” IEEE Conference on Graphics and Media (GAME), pp. 38–43, 2019.

[7] 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, Sep. 2021.

[8] K. Bartol, D. Bojanic, T. Pribanic, T. Petkovic, Yago Diez Donoso, and J. S. Mas, “On the Comparison of Classic and Deep Keypoint Detector and Descriptor Methods,” 11th International Symposium on Image and Signal Processing and Analysis (ISPA), pp. 64–69, 2019.

[9] "Visual Geometry Group - the University of Oxford," Visual Geometry Group - University of Oxford. https://www.robots.ox.ac.uk/~vgg/data/affine/

[10] V. Balntas, K. Lenc, A. Vedaldi, T. Tuytelaars, J. Matas, and K. Mikolajczyk, “HPatches: A benchmark and evaluation of handcrafted and learned local descriptors,” IEEE Transactions on Pattern Analysis and Machine Intelligence, pp. 1–1, 2019.
[11] C. Shorten and T. M. Khoshgoftaar, “A survey on Image Data Augmentation for Deep Learning - Journal of Big Data,” Journal of Big Data, vol. 6, no. 1, pp. 1–48, Jul. 2019.
[12] K. Cordes, B. Rosenhahn, and J. Ostermann, "High-Resolution Feature Evaluation Benchmark," In Wilson, R., Hancock, E., Bors, A., Smith, W. (eds) Computer Analysis of Images and Patterns. CAIP 2013. Lecture Notes in Computer Science, vol. 8047, pp. 327–334, 2013.
[13] K. Ma, Z. Duanmu, and Z. Wang, “Geometric Transformation Invariant Image Quality Assessment Using Convolutional Neural Networks,” IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Calgary, Canada, pp. 6732–6736, Apr. 2018.
[14] Z. LOU et al., "ContextDesc: Local Descriptor Augmentation with Cross-Modality Context,” IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, California, USA, pp. 2522–2531, 2019.
[15] Z. Guo and S. Ying, “Whole-Body Keypoint and Skeleton Augmented RGB Networks for Video Action Recognition,” Applied Sciences, vol. 12, no. 12, p. 6215, Jun. 2022.
[16] H. Uchiyama and E. Marchand, "Toward augmenting everything: Detecting and tracking geometrical features on planar objects," Proceedings of the 2011 10th IEEE International Symposium on Mixed and Augmented Reality (ISMAR '11). IEEE Computer Society, USA, pp. 17–25, Oct. 2011.
[17] K. Mikolajczyk and C. Schmid, “An Affine Invariant Interest Point Detector | SpringerLink,” In: Heyden, A., Sparr, G., Nielsen, M., Johansen, P. (eds) Computer Vision, ECCV 2002, Lecture Notes in Computer Science, Springer, Berlin, Heidelberg, vol. 2350, pp. 128–142, Jan. 2002.
[18] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, “Image Quality Assessment: From Error Visibility to Structural Similarity,” IEEE Transactions on Image Processing, vol. 13, no. 4, pp. 600–612, Apr. 2004.
[19] K. Lenc and A. Vedaldi, “Large scale evaluation of local image feature detectors on homography datasets,” British Machine Vision Conference, BMVC, Northumbria University, Newcastle, UK, pp. 1–13, 2018, [Online]. Available: https://dblp.org/rec/conf/bmvc/LencV18
[20] L. Chen, F. Rottensteiner, and C. Heipke, "Feature detection and description for image matching: from handcrafted design to deep learning," Geo-spatial Information Science, vol. 24, no. 1, pp. 58–74, Nov. 2020.
[21] V. Balntas, K. Lenc, A. Vedaldi, T. Tuytelaars, J. Matas, and K. Mikolajczyk, “HPatches: A benchmark and evaluation of handcrafted and learned local descriptors,” IEEE Transactions on Pattern Analysis and Machine Intelligence, pp. 1–1, 2019.
[22] D. G. Lowe, “Distinctive Image Features from Scale-Invariant Keypoints,” International Journal of Computer Vision, vol. 60, no. 2, pp. 91–110, Nov. 2004.
[23] A. Moghimi, T. Celik, A. Mohammadzadeh, and H. Kusetogullari, “Comparison of Keypoint Detectors and Descriptors for Relative Radiometric Normalization of Bitemporal Remote Sensing Images,” IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 14, pp. 4063–4073, 2021.
[24] Y. Ou, Z. Cai, J. Lu, J. Dong, and Y. Ling, “Evaluation of Image Feature Detection and Matching Algorithms,” 5th International Conference on Computer and Communication Systems (ICCCS), Shanghai, China, pp. 220–224, 2020.
[25] S. Tareen and Z. Saleem, “A comparative analysis of SIFT, SURF, KAZE, AKAZE, ORB, and BRISK,” International Conference on Computing, Mathematics and Engineering Technologies – iCoMET 2018, Sukkur, Pakistan, pp. 1–10, Mar. 2018.
[26] S. Gauglitz, T. Höllerer, and M. Turk, “Evaluation of Interest Point Detectors and Feature Descriptors for Visual Tracking,” International Journal of Computer Vision, vol. 94, no. 3, pp. 335–360, Mar. 2011.
[27] S. Dhatrak and D. Gharpure, “Comparative study of SURF, KAZE, and MSER for visual odometry in presence of noise,” AIP Conference Proceedings, vol. 2355, no. 1, 2021.
[28] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, “Speeded-Up Robust Features (SURF),” Computer Vision and Image Understanding, vol. 110, no. 3, pp. 346–359, Jun. 2008.
[29] P. F. Alcantarilla, A. Bartoli, and A. J. Davison, "KAZE Features," In Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y., Schmid, C. (eds) Computer Vision – ECCV 2012. ECCV 2012. Lecture Notes in Computer Science, vol. 7577, pp. 214–227, 2012.
[30] E. Karami, S. Prasad, and M. Shehata, “Image Matching Using SIFT, SURF, BRIEF and ORB: Performance Comparison for Distorted Images,” arXiv.org, Oct. 07, 2017. https://arxiv.org/abs/1710.02726 (accessed 2022).
[31] M. Awrangjeb and G. Lu, "Robust image corner detection based on the chord to point distance accumulation technique," IEEE Transactions on Multimedia, vol. 10. no. 6, pp. 1059–1072, 2008.

**IBRAHIM EL RUBE’** (IEEE SM’12) received the B.Eng. degree from the computers and electronics engineering department in 1992, the M.Sc. degree in electronics and communications engineering in 1999, respectively, from Arab Academy for Science and Technology, Egypt, and the Ph.D. degree in systems design engineering from the University of Waterloo, in 2006, Canada. He has held lecturing positions at the Department of Electronics and Communications Engineering, Arab Academy for Science and Technology, Egypt, during the years from 1993 to 2011. In Aug. 2006 he was an Assistant Professor and became an Associate Professor in 2010. In 2011, he joined the computer engineering department, Taif University, where he is currently an Associate Professor and has been the department's chair from 2012 to 2016. His current research interests include signal, image and video processing, computer vision, pattern recognition, and machine intelligence.