Proceeding of the Electrical Engineering Computer Science and Informatics
Vol. 7, October 2020 **ISSN: 2407-439X**

Image Restoration Effect on DCT High Frequency Removal and Wiener Algorithm for Detecting Facial Key Points

Adhi Kusnadi
Department of Informatics
Universitas Multimedia Nusantara
Tangerang, Indonesia
adhi.kusnadi@umn.ac.id

Vincent Anderson Ngadiman
Department of Informatics
Universitas Multimedia Nusantara
Tangerang, Indonesia
vincent.ngadiman@student.umn.ac.id

Vincent Anderson Ngadiman
Department of Informatics
Universitas Multimedia Nusantara
Tangerang, Indonesia
vincent.ngadiman@student.umn.ac.id

Syarief Gerald Prasetya
Department of Accountancy
STIE Binaniaga
Bogor, Indonesia
er7et70@gmail.com

***Abstract*—This study aims to figure out the effect of using Histogram Equalization and Discrete Cosine Transform (DCT) in detecting facial keypoints, which can be applied for 3D facial reconstruction in face recognition. Four combinations of methods comprising of Histogram Equalization, removing low-frequency coefficients using Discrete Cosine Transform (DCT) and using five feature detectors, namely: SURF, Minimum Eigenvalue, Harris-Stephens, FAST, and BRISK were used for test. Data that were used for test were obtained from Head Pose Image and ORL Databases. The result from the test were evaluated using F-score. The highest F-score for Head Pose Image Dataset is 0.140 and achieved through the combination of DCT & Histogram Equalization with feature detector SURF. The highest F-score for ORL Database is 0.33 and achieved through the combination of DCT & Histogram Equalization with feature detector BRISK.**

Keywords—DCT, wiener filtering, feature detectors, key points, f-score

I. INTRODUCTION

Face recognition is a technology used to recognize people based on their facial characteristics, and with/without any prior knowledge. The numerous advantages of the technology become the reason of implementation by the government, private, and public sectors [1]. This technology works by analyzing and comparing the keypoints on a human face, extracted by a predefined method [2]. Compared to the traditional method of identification, the face recognition technology is more reliable [3]. Generally, human's biological pattern like signature, mode of walking and speech, and keystroke tend to change with time [4]. However, the physical part, such as face, fingerprints, and iris tend to remain unchanged for a lifetime [5], [6]. Over the decades, finger-prints have been used as a mean of identification [7]. One of the advantages of the face recognition method is that the observed person does not need to be approached to perform the identification process. Furthermore, the human face image is obtainable even from a cheap camera compared to the other biometric methods that require expensive tools to carry out biometric analyses such as the retina and iris [4]. However, the result is commonly affected by the noise due to the camera's defocus, inconsistency associated with the brightness, contrast levels, and other components that may disrupt the image. The noise component needs to be removed from the image due to its ability to degrade the quality, regardless of its intensity level, which will lead to diminished performance [8], [9].

Various studies have been conducted on the removal of noise by taking out the high-frequency band using the DCT method [10], [8], [11], [12], [13]. However, in these studies, eliminating the high frequency made the images blur because high frequency storages edges information [14]. This research was conducted to evaluate the performance of five feature detectors in determining the key points without using the image processing method [15]. The feature detectors used are Harris-Stephens, Speeded Up Robust Features (SURF), Features from Accelerated Segment Test (FAST), Binary Robust Invariant Scalable Keypoints (BRISK), and Minimum Eigenvalue. Detecting facial keypoints is meant to reconstructing 3D models of face [16].

The images used as subjects were acquired from ORL dataset, which is currently known as the AT&T Database of Faces and Head Pose Image Dataset. The obtained images were transformed from spatial to the frequency domain using the Discrete Cosine Transform (DCT) to remove the high-frequency band. After that, the images were transformed back to the spatial domain using the inverse DCT, then followed by the application of Wiener Filtering to deblurring the images and lastly the keypoints were detected by using five feature detectors [17]. The results were evaluated by comparing the acquired F-Score value from original grayscale images and the processed images. The F-Score values were obtained by evaluating the 15 facial keypoints to obtain accurate results [18]. The results show increase in F-score value which is advantageous in reconstruction of 3D modeling for 3D face recognition. Method using DCT scored the highest F-score value of 0.373 for ORL dataset and 0.200 for Head Pose dataset. Method using DCT & Wiener Filtering scored the highest F-score value of 0.339 for ORL dataset and 0.170 for Head Pose dataset.

This paper is further divided into five sections. The second section provides explanations on the related works, and the third presents the setup for the experiment. Meanwhile, section four contains the results and discussion of the experiment, while the last section provides the experimental conclusion.

II. RELATED WORKS

A. *Discrete Cosine Transform (DCT)*

Discrete Cosine Transform (DCT) is a mathematical method used to transform an image from spatial to the frequency domain, by partitioning its pixel matrix into blocks of $N \times N$ size. In this research, the two-dimensional DCT is performed to processing the images using Equation 1 [19]:

$$C(u,v) = a(u) * a(v) \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} f(x,y) * cos\left[\frac{\pi(2x+1)u}{2N}\right] * \cos\left[\frac{\pi(2y+1)v}{2N}\right] \quad (1)$$

After the high-frequency component was removed by conducting feature extraction [10,11,18,20] the image was transformed back to the spatial domain using the inverse DCT as shown in Equation 2:

$$f(x,y) = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} a(u) * a(v) * c(u,v) * cos\left[\frac{\pi(2x+1)u}{2N}\right] * \cos\left[\frac{\pi(2y+1)v}{2N}\right] \quad (2)$$

B. *Wiener Filtering*

Wiener Filtering is a restoration method used to minimize the Mean Square Error (MSE) between the original and restored images. In the frequency domain, Equation 3 [20] is applied :

$$H_w(u,v) = \frac{H^*(u,v)}{|H(u,v)|^2 + K} \quad (3)$$

After creating the filter, it is then applied to the degraded image.

C. *Facial Keypoints*

Key points represent the local feature from human faces, which are substantial for 3D reconstruction. This study consists of a total of 15 facial key points, as shown in Table 1 [18]:

TABLE I. 15 FACIAL KEYPOINTS

Left eye center	Right eye center
Left eye inner corner	Right eye inner corner
Left eye outer corner	Right eye outer corner
Left eyebrow inner end	Right eyebrow inner end
Left eyebrow outer end	Right eyebrow outer end
Mouth left corner	Mouth right corner
Mouth center top lip	Mouth center bottom lip
Nose tip	

Figure 1 below shows the location of facial keypoints on the subject images as described in Table 1.

(a) (b)

Fig. 1. Location of facial keypoints: (a) subject 1 from ORL image database (b) person 06 from head pose image database.

D. *ORL Database*

The dataset consists of face images taken from April 1992 to April 1994 at the lab using 40 different subjects, with each consisting of ten different images. All images had a dark homogeneous background, while the subjects are in an upright, frontal position with varying lighting and facial expressions or details [21].

E. *Head Pose Database*

The Head Pose Image Database was created by capturing the face images of 15 different persons with varying pan and tilt angles ranging from -90° to +90°. Each person has two series of 93 images with a different pose, culminating in 2790 monocular face images. However, some people wear glasses or have a different skin color [22].

III. METHOD

The images used for the test come from ORL Database and Head Pose Image Dataset. The selected images of subjects from the dataset are subjects that do not wear glasses and do not have beard or moustaches. The chosen images then will be converted into grayscale images.

In this experiment, three methods will be applied into the images:

1. Not applying DCT or Wiener Filtering into the image.
2. Applying DCT to remove the high frequency component.
3. Applying both DCT and Wiener Filtering.

Through these three methods, there will be three different image as the outcomes. Then, feature detectors will be applied to the images to detect the facial keypoints. Lastly, the method will be evaluated by using F-score.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

To evaluate the results of the methods, F-Score, which is the harmonic mean value between precision and recall, was used to represent the test accuracy [22]. The recall is defined as the total number of correctly detected facial key points divided by the total feature points detected, as shown in Equation 4:

$$recall = \frac{Number\ of\ correct\ facial\ key\ points}{Total\ feature\ points\ detected} \quad (4)$$

While precision is the number of the correctly detected facial key points divided by the total number of point in human faces as shown in Equation 5:

$$precision = \frac{Number\ of\ correct\ facial\ key\ points}{15} \quad (5)$$

To balance the value between precision and recall, F-Score calculation is needed, which is represented in Equation 6:

$$F\ Score\ = 2 \times \frac{(Recall \times Precision)}{(Recall + Precision)} \quad (6)$$

Fig. 2 shows and compares the images that had been processed through the three different methods mentioned on the previous part.

(a) (b) (c)

Fig. 2. Image processed in 3 different methods: (a) the grayscale original image, (b) high-frequency band is removed, and (c) high-frequency band is removed and applied with wiener filtering.

A. *Experiment with ORL Database*

1) BRISK

A total of 5, 7, and 5 facial key points were detected from 17, 21, and 13 keypoints using BRISK feature detector as shown in Fig. 3 (a), (b), and (c).

(a) (b) (c)

Fig. 3. BRISK detection results on face images from ORL database: (a) the grayscale original image, (b) high-frequency band is removed, and (c) high-frequency band is removed and applied with wiener filtering.

2) Harris-Stephens

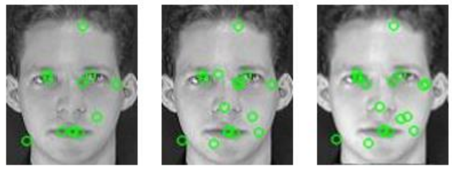
By using Harris-Stephens feature detector, 33, 27, and 21 keypoints were detected in Fig. 4 (a), (b), and (c). Although some keypoints were detected, none of the detected keypoints could be registered as facial keypoints due to the location of the detected keypoints.

(a) (b) (c)

Fig. 4. Harris-Stephens detection results on face images from ORL database: (a) the grayscale original image, (b) high-frequency band is removed, and (c) high-frequency band is removed and applied with wiener filtering.

3) SURF

By using SURF feature detector, 1, 2, and 2 facial keypoint were detected from 12, 16, and 18 keypoints, as shown in Fig. 5 (a), (b), and (c).



(a) (b) (c)

Fig. 5. SURF detection results on face images from ORL database: (a) the grayscale original image, (b) high-frequency band is removed, and (c) high-frequency band is removed and applied with wiener filtering.

4) FAST

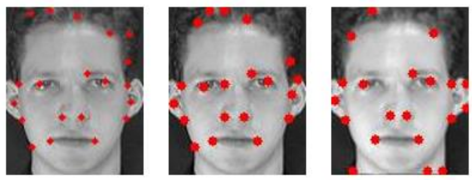
FAST feature detector successfully detected 38, 45, and 32 keypoints with 5, 9, and 9 facial keypoints as shown in Fig. 6 (a), (b), and (c).

(a) (b) (c)

Fig. 6. FAST detection results on face images from ORL database: (a) the grayscale original image, (b) high-frequency band is removed, and (c) high-frequency band is removed and applied with wiener filtering.

5) Minimum Eigenvalue

The maximum keypoints detected for Minimum Eigenvalue feature detector is set to 20 with the detection of 3 facial keypoints in the same position. The location of detected keypoints are shown in Fig.7 (a), (b), and (c).



(a) (b) (c)

Fig. 7. Minimum eigenvalue detection results on face images from ORL database: (a) the grayscale original image, (b) high-frequency band is removed, and (c) high-frequency band is removed and applied with wiener filtering.

Table 2 below shows the average F-score from five images used from ORL Dataset and Fig. 8 shows the comparison of the average score in graph.

TABLE II. AVERAGE F-SCORE FROM ORL IMAGE DATABASE

Feature Detectors	**ORIGINAL**	**DCT**	**DCT & Wiener**
SURF	0.151	0.157	0.155
FAST	0.244	0.282	0.339
Harris-Stephens	0.005	0.005	0.006
BRISK	0.282	0.373	0.244
Minimum Eigenvalue	0.206	0.251	0.206

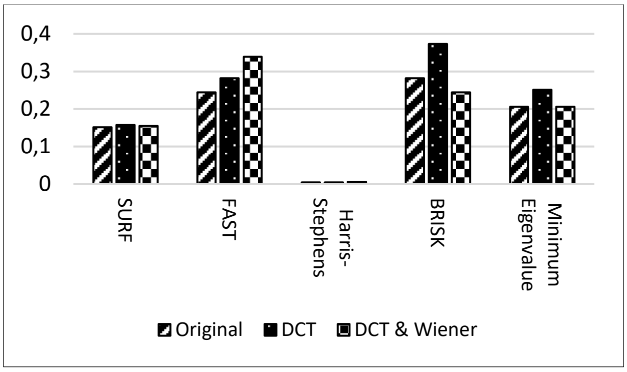


Fig. 8. Comparison of average f-score from methods applied on orl dataset.

As shown in Fig. 8, the highest average F-Score achieved in ORL Dataset is 0.373 and obtained by using DCT method and BRISK feature detector.

B. *Maintaining with Head Pose Image Database*

This part will show the example result acquired by applying the methods and feature detectors to images from Head Pose Image Database.

1) *BRISK*

The use of the BRISK feature detector showed a total of 4, 5, and 4 facial keypoints detected from 34, 57, and 39 detected keypoints as shown in Fig. 9 (a), (b), and (c).

(a) (b) (c)

Fig. 9. BRISK detection results on face images from head pose database: (a) the grayscale original image, (b) high-frequency band is removed, and (c) high-frequency band is removed and applied with wiener filtering.

2) *Harris-Stephens*

By using Harris-Stephens feature detector, 12 keypoints were detected in both Fig. 10 (a) and (b), with a slight increase in Fig. 10 (c) by 15 keypoints, which is similar to the ORL image database. But, there was not any facial keypoints detected in the image.

(a) (b) (c)

Fig. 10. Harris-Stephens detection results on face images from head pose database: (a) the grayscale original image, (b) high-frequency band is removed, and (c) high-frequency band is removed and applied with wiener filtering.

3) *SURF*

By using SURF feature detector, a total of 10, 9, and 6 facial keypoints from 46, 70, and 67 keypoints were detected as shown in Fig. 11 (a), (b), and (c).

(a) (b) (c)

Fig. 11. Comparison SURF detection results on face images from head pose database: (a) the grayscale original image, (b) high-frequency band is removed, and (c) high-frequency band is removed and applied with wiener filtering.

4) *FAST*

In FAST feature detector method, 26, 67, and 53 keypoints were detected and 2, 5, and 5 facial keypoints were detected as well as shown in Fig. 12 (a), (b), and (c).

(a) (b) (c)

Fig. 12. FAST Detection Results on Face Images from Head Pose Database: (a) the grayscale original image, (b) high-frequency band is removed, and (c) high-frequency band is removed and applied with Wiener Filtering.

5) *Minimum Eigenvalue*

The maximum amount of keypoints is set to 20, as carried out in the experiment with ORL database. The feature detector detected a total of 2 facial keypoints at different positions on each images as shown in Fig. 13 (a), (b), and (c).

(a) (b) (c)

Fig. 13. Minimum eigenvalue detection results on face images from head pose database: (a) the grayscale original image, (b) high-frequency band is removed, and (c) high-frequency band is removed and applied with wiener filtering.

Table 3 shows the average F-score from five images used from Head Pose Image Database and Fig. 14 shows the comparison in table. For Head Pose Image Database, the highest average F-score is 0.200. The highest average score is acquired through DCT method and using SURF feature detector.

TABLE III. AVERAGE F-SCORE FROM HEAD POSE IMAGE DATABASE

Feature Detectors	**ORIGINAL**	**DCT**	**DCT & Wiener**
SURF	0.180	0.200	0.170
FAST	0.044	0.084	0.084
Harris-Stephens	0	0	0
BRISK	0.097	0.108	0.092
Minimum Eigenvalue	0.057	0.069	0.080

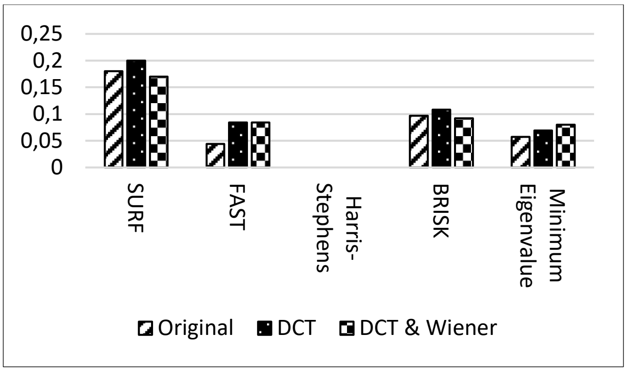


Fig. 14. Comparison of average f-score from methods applied on head pose image dataset.

According to the results of 5 different images used from the Head Pose Image Database, the best performing method is DCT and combined with SURF feature detector. On the other hand, for ORL dataset, the best performing method is DCT and combined with BRISK feature detector.

SURF and BRISK feature detector work by considering the pixel value and applying Gaussian kernel as well as the Wiener Filtering, which smoothens the image [23], [24]. Gaussian kernel affected the pixel value when the image was smoothened once which later made the pixel became undetected and the value decreased even further when DCT & Wiener were applied. This was also proven in both datasets with the decreasing number of detected keypoints when Wiener was applied with BRISK as the feature detector. This explains why both feature detector works better in DCT-processed images rather than DCT & Wiener Filtering-processed images.

From the research [25], it can be inferred that noise also affects keypoints detection, because noise is formed through random variation of intensity of pixels in an image. This also means that when there is noise, there may be changes in the value of pixel which can affect the detection result of feature detector.

When images were detected by Harris-Stephens feature detector, the result was inconsistent in both datasets. In the test with ORL dataset, any changes in detected keypoints were dependent on the processed image while in the test with Head Pose Image Database, there was an increase from when detecting the original grayscale image with images that had been processed with DCT & Wiener Filtering. According to the research by [26], it is known that the Harris-Stephens do not have a definite way to describe the threshold value, which is necessary to define a descriptor [27].

According to the research in [28], it is known that the FAST feature detector is very sensitive to noise, thus causing the feature detector to detect more keypoints when there are a lot of noise on the image. In both ORL and Head Pose dataset, FAST feature detector detects more keypoints only when DCT was used to process the images compared to when both methods were utilized. As for the F-Score values, FAST feature detector scored higher when images from the ORL database were processed with both DCT and Wiener method and the score even decreased in Head Pose using the same methods. Through visual observation on ORL database, the images had more noise rather than in Head Pose Image Database. According to the research conducted by [26], due to the ability of FAST to detect actual corner-points, the greater the noise, the more likely it is to detect the facial key points. However, because F-Score calculation is based on total of all keypoints detected, FAST feature detector fails to deliver the best results.

V. Conclusion

From the experiment, it can be concluded that the proposed method with DCT & Wiener Filtering do not strongly affect the F-score of facial keypoint detection as much as the method with DCT only to remove the high-frequency coefficients. The feature detectors that were used for detecting the facial keypoints also affecting the F-score result of each database. For Head Pose Image Dataset, the highest F-score is achieved with DCT and SURF feature detector with value of 0.2 which has the improvement rate of 11.11% from the original. The method with DCT & Wiener Filtering with the same feature detector scores value of 0.170. For ORL dataset, the highest F-score is achieved with DCT and BRISK with value of 0.373 which has the improvement rate of 32.27% from the original. The method with DCT & Wiener Filtering with the same feature detector scores value of 0.282. According to the result, it can be concluded that DCT-only method provides better result rather than method that involves DCT & Wiener Filtering due to the smoothing effect in Wiener Filtering. Furthermore, each feature detector has a different way of recognizing keypoints which may lead to different results depending on the condition of images from different databases.

Acknowledgment

The authors gratefully acknowledge the support from Universitas Multimedia Nusantara and Ristek-Dikti for grant PDUPT at the year of 2020. And Our students at Universitas Multimedia Nusantara for their help to finish the research.

References

[1] M. G. Galterio, S. A. Shavit, and T. Hayajneh, "A review of facial biometrics security for smart devices," *Computers*, vol. 7, no. 3, Sep. 2018.

[2] S. Zhou and S. Xiao, "3D face recognition: a survey," *Human-centric Comput. Inf. Sci.*, vol. 8, no. 1, pp. 35, Dec. 2018.

[3] O. A. Esan, T. Zuva, S. M. Ngwira, and K. Zuva, "Performance Improvement of Authentication of Fingerprints Using Enhancement and Matching Algorithms," *Int. J. Emerg. Technol. Adv. Eng.*, vol. 3, no. 2, 2013.

[4] L. Masupha, T. Zuva, and S. Ngwira, "A review of gait recognition techniques and their challenges," *Third international conference on digital information processing, e-business and cloud computing.*, 2015, pp. 63–69.

[5] H. A. Aboalsamh, "Vein and fingerprint biometrics authentication-future trends," *Int. J. Comput. Commun.*, vol. 3, no. 4, pp. 67–75, 2009.

[6] O. A. Esan, S. Ngwira, T. Zuva, and L. Masupha, "Health care infrastructure security using bimodal biometrics system," *Health Care (Don. Mills)*, vol. 3, no. 02, 2014.

[7] A. W. Senior and R. M. Bolle, "Improved fingerprint matching by distortion removal," *IEICE Trans. Inf. Syst*, vol. 84, no. 7, pp. 825–832, 2001.

[8] F. Wang *et al.*, "The Devil of Face Recognition is in the Noise," *ECCV 2018 LNCS* , pp. 765–780, 2018.

[9] R. Petrović, B. Stojanović, M. Pavlović, S. Puzović, and S. Stanković, "Thermal Imagery Noise Influence on Face Recognition Performance," *2019 18th International Symposium INFOTEH-JAHORINA (INFOTEH).*, 2019, pp. 1–5.

[10] R. R. Ahirwal and S. Nachankar, "A Novel Method Of Average Filtering For Removing Noise And Face Recognition," *Eur. J. Comput. Sci. Inf. Technol.*, vol. 4, no. 2, pp. 9–24, 2016.

[11] A. Kusnadi, R. Winantyo, and I. Z. Pane, "Removing DCT High Frequency on Feature Detector Repeatability Quality," *2019 5th International Conference on New Media Studies (CONMEDIA).*, 2019, pp. 238–243.

[12] A. Buemi, A. Bruna, M. Mancuso, A. Capra, and G. Spampinato, "Chroma Noise Reduction in DCT Domain Using Soft-Thresholding," *EURASIP J. Image Video Process.*, vol. 2010, no. 1, pp. 1-13, 2011.

[13] S. Chopra, E. Blias, L. Chavina, V. Alexeev, and V. Sudhakar, "June Technical Luncheon High Frequency Restoration of Surface Seismic Data," vol. 10, no. 3, pp. 353–359, 2002.

[14] M. Nixon and A. S. Aguado, "Feature Extraction & Image Processing for Computer Vision," *Featur. Extr. Image Process. Comput. Vis.*, 2012.

[15] A. Kusnadi, Wella, R. Winantyo, and I. Z. Pane, "Evaluation of Feature Detectors on Repeatability Quality of Facial Keypoints in Triangulation Method," *2018 International Conference on Smart Computing and Electronic Enterprise, ICSCEE 2018*, pp. 1-4, 2018.

[16] L. Alexander, A. Kusnadi, W. Wella, R. Winantyo, and I. Z. Pane, "Authentication System Using 3D Face With Algorithm DLT and Neural Network," *2018 Joint 10th International Conference on Soft Computing and Intelligent Systems (SCIS) and 19th International Symposium on Advanced Intelligent Systems (ISIS)*, 2018, pp. 186–189.

[17] D. S. Al-amri and D. S. Ali, "Restoration and Deblured Motion Blurred Images," *IJCSI Int. J. Comput. Sci.*, vol. 11, no.1, 2014.

[18] Y. Wang and Y. Song, "Facial Keypoints Detection," 2014.

[19] N. Agarwal and A. M. Khan, "Application of DCT in image processing," *Int. J. Eng. Res. Technol*, 2018.

[20] A. Zisserman, "Lecture 3: Image Restoration." University of Oxford, 2014.

[21] AT&T Laboratories Cambridge, "The Database of Faces," 2002. [Online]. Available: https://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html. [Accessed: 20-Mar-2019].

[22] N. Gourier, D. Hall, and J. L. Crowley, "Estimating Face orientation from Robust Detection of Salient Facial Structures," in *Proceedings of Pointing 2004, ICPR, International Workshop on Visual Observation of Deictic Gestures*, 2004.

[23] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-Up Robust Features (SURF)." in *Computer Vision-ECCV 2006*, 2006, pp. 404-417.

[24] S. Leutenegger, M. Chli, and R. Y. Siegwart, "BRISK: Binary Robust Invariant Scalable Keypoints," in *2011 IEEE International Conference on Computer Vision*, 2011, p. 2548-2555.

[25] A. Maity and R. Chatterjee, "Impulsive noise in images: a brief review," *Accent. Trans. Image Process. Comput. Vis.*, vol. 4, no. 10, pp. 6–15, Feb. 2018.

[26] Z. Jiang, Y. Wu, J. Wang, J. Lu, and L. Zhang, "The Improved Harris Operator based on Steerable Filter," in *ISME 2016 - Information Science and Management Engineering IV*, vol. 1, pp. 305–311, 2016.

[27] D. Tyagi, "Introduction to Harris Corner Detector - Analytics Vidhya - Medium," 2019. [Online]. Available: https://medium.com/analytics-vidhya/introduction-to-harris-corner-detector-32a88850b3f6. [Accessed: 18-Mar-2020].

[28] H. A. Kadhim and W. A. Araheemah, "A Comparative Between Corner-Detectors (Harris, Shi-Tomasi & FAST) in Images Noisy Using Non-Local Means Filter," *J. Al-Qadisiyah Comput. Sci. Math.*, vol. 11, no. 3, pp. 86-93, Sep. 2019.