Detection of Copy−Move Image Forgery Applying Robust Matching with K-D Tree Sorting

Chapter · September 2022

DOI: 10.1007/978-981-19-1520-8_22

Partha Chakraborty
Comilla University
61 PUBLICATIONS **653** CITATIONS

Mahmuda Khatun
Comilla University
12 PUBLICATIONS **95** CITATIONS

Md. Abu Sayed
Comilla University
2 PUBLICATIONS **6** CITATIONS

Md. Sabab Zulfiker
Bangabandhu Sheikh Mujibur Rahman University, Kishoreganj
16 PUBLICATIONS **216** CITATIONS

Detection of Copy–Move Image Forgery Applying Robust Matching with K-D Tree Sorting

Partha Chakraborty, Sabakun Nahar Tafhim, Mahmuda Khatun, Md. Abu Sayed, Sabab Zulfiker, Priyanka Paul, Md. Farhad Hossain, and Tanupriya Choudhury

Abstract Digital images contribute significantly to the field of visualization. Using stronger technology, digital image forgery is easier. The most common method of image forgery is to re-create a portion of a person's location or to conceal a portion of an image. In our paper, we worked on detecting region duplication forgery using COMOFORD databases by utilizing the discrete cosine transform (DCT), k-dimensional tree (k-d tree) for sorting efficiently, and a robust matching method. Here, the size of the block will be 16 × 16, and it will be divided into four blocks. This study can detect forged portions for PNG images with better performance by

P. Chakraborty (✉) · S. N. Tafhim · M. Khatun · Md. Abu Sayed
Department of Computer Science and Engineering, Comilla University, Cumilla 3506, Bangladesh
e-mail: partha.chak@cou.ac.bd

S. N. Tafhim
e-mail: sn.tafhim@gmail.com

M. Khatun
e-mail: mahmuda@cou.ac.bd

Md. Abu Sayed
e-mail: sayeed.cse.bd@gmail.com

S. Zulfiker
Department of Computer Science and Engineering, Daffodil International University, Dhaka, Bangladesh
e-mail: sabab.rumc@gmail.com

P. Paul · Md. Farhad Hossain
Department of Statistics, Comilla University, Cumilla 3506, Bangladesh
e-mail: priyanka.paulbd@gmail.com

Md. Farhad Hossain
e-mail: farhad390ju@gmail.com

T. Choudhury
Department of Informatics, School of Computer Science, University of Petroleum and Energy Studies, Dehradun, India
e-mail: tanupriya1986@gmail.com

D. Gupta et al. (eds.), *Pattern Recognition and Data Analysis with Applications*,
Lecture Notes in Electrical Engineering 888,
https://doi.org/10.1007/978-981-19-1520-8_22

highlighting the images with a quality factor of 0.5 and a threshold value of 10, as well as gives good results for JPEG images.

Keywords Image detection · Robust matching · K-D tree sorting · Image forgery · DCT

1 Introduction

We are living in such an era where people are able to share any kind of information with each other, located on every side of the world, with the least amount of cost by using the Internet. Widespread accessibility of technology and the negligible cost of equipment make it very convenient for daily life. The photography system provides a sophisticated medium for image modification with a top-notch appearance. For this reason, we are highly at risk of facing numerous threats related to our identity, financial security, as well as many national safety issues. We cannot completely avoid or be free of these dangers. People can share images on a variety of different Internet platforms. Digital images are assigned to various information and can be skillfully modified as a result of a limited but adequate protection system. The easy accessibility of editing software tools on all devices means that image editing has become an easy-going job at present. Edited images alter the original affection provided by the real image, which may contain threats to information security for people.

Many procedures exist to detect forgery in images, making it difficult to find more practical and perfect implementation procedures. Which algorithm has the best performance? It could have a high rate of false-positive detection. Furthermore, runtime capabilities are a critical component in determining how efficiently an algorithm works and ensuring the algorithm's usability. They are different in performance. Some provide good real-time performance, while others provide better results through modifications, and still others detect different geometrical modifications. The purpose of the research work is to inspect the existing forgery detection procedures for images for complexity reduction [1].

2 Literature Review

For image forgery detection, many researchers have done related work at different times. This paper established a combined procedure for copy–movement forgery supported by scale invariant transformation features as well as the Fourier-Mellin technique [2]. This paper worked on a detection method named blind copy–move forgery by deploying the KS and SVD testing methods [3]. They proposed a DCT and cellular automata-based robust copy–movement fraud descriptor in [4]. Parveen et al. [5] suggested block-based copy–move picture fraud detection using DCT. Prakash et al. [6] worked on detecting copy–move forgeries using AKAZE and SIFT key point extraction. Hegazi et al. [7] discovered a density-based clustering approach with satisfied outlier dismissal based on a copy-moving forgery approach. Tan et al. [8] conducted a review of digital image copy forgery detection for localization using passive procedures. The paper worked on a copy-moving forgery approach based on modified key point extortion and pairing [9]. Mushtaq and Mir [10] worked on copy movement forgery detection for pictures. Mahmood et al. [11] suggested a robust stationary wavelet and DCT approach in order to detect and localize copy movement forgery. Ouyang et al. [12] developed a comprehensive copy movement forgery approach by combining Zernike moments and the pyramid model. Emam et al. [13] developed a two-stage key point detection system capable of detecting region duplication forgeries in digital images. Rasse [14] examine the detection of digital picture splicing forgeries using illumination color estimation. Kaur and Sharma [15] are working on improving the prevention of duplicate fraud technique. In this work, the author has experimented with the DCT and wavelet transformations [16]. This paper has conducted extensive research into various types of picture forgeries [17]. This paper investigated features to detect forgeries using a copy-moving forgery [18]. These study investigated the identification of copy–move frauds in digital pictures [19]. The emphasis of the research was on passive forensics for copy–move forgeries utilizing a DCT-based technique [20]. They investigated the copy–move forgery approach using cellular automata in [21]. This paper has worked on copy–move detection by merging cellular automata with regional binary patterns [22]. Create a system that uses template and HOG features for object detection in [23]. Create a method for robots to compute the degree of visual focus of human attention [24, 25], which tries to find object instances in unknown image sources. Make a system that uses face detection to obtain automatic student attendance [26].

3 Methodology

The proposed system used in this study is described in Fig. 1.

3.1 Taking Input Image

An input picture is a pixel-by-pixel modification of a source images. Detection of that input image in Fig. 2 has a higher value (less match) and a lower value (stronger match). In the input image, the threshold number of pitches appears to be copied together for it to be considered a forged region by the algorithm.

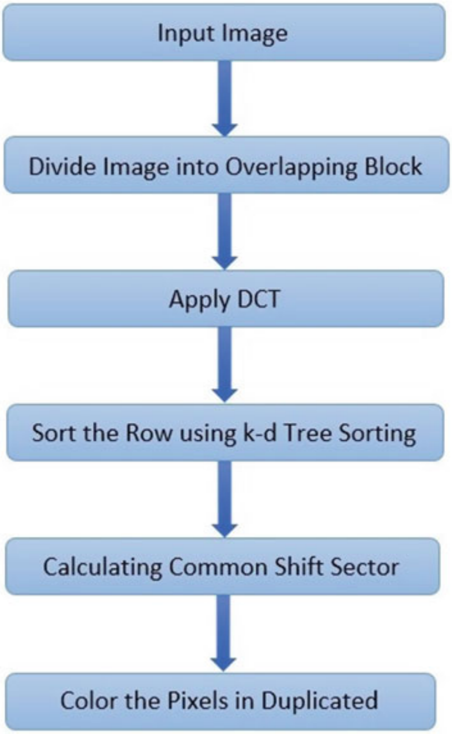

Input Image
Divide Image into Overlapping Block
Apply DCT
Sort the Row using k-d Tree Sorting
Calculating Common Shift Sector
Color the Pixels in Duplicated

Fig. 1 Proposed system model

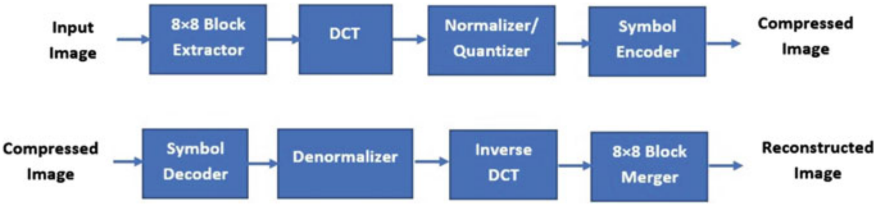


Fig. 2 Image compression block diagram using DCT

3.2 Divide Image into Overlapping Block

In order to detect copy–move fraud in pictures, the standardized shifting matrix counter C must be increased by one for each identical pairing of blocks:

$$C_{(d1,d2)} = C_{(d1,d2)} + 1$$

Displacement vectors are designed as well as in the ordered matrix A. Counter C is augmented in every combination of progressively fitting rows. Before this process begins, the displacement vector C is reset to null. Finally, in the pairing procedure, timer C represents the frequency through which separate scaled displacement vectors emerge. The program then searches for any standardized displacement vectors—$d(1)$, $d(2)$ … $d(n)$, wherein frequency surpasses a consumer threshold K:

$$C_{(d)}^{(r)} > K$$

For all $r = 1, n$.

The size of the smallest part that can be recognized by the method is proportional to the threshold K value. Larger numbers may cause the algorithm to overlook some blocks that aren't quite so tightly matched.

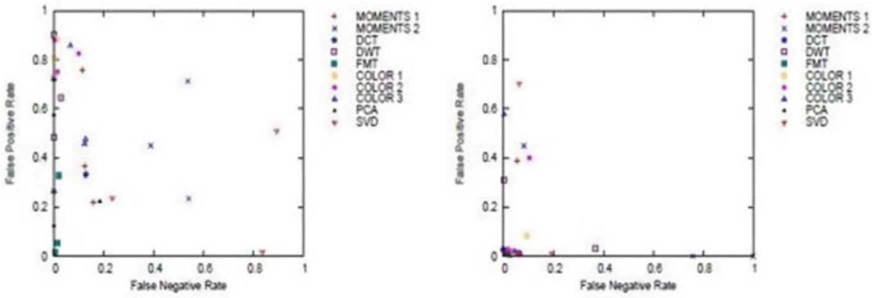
3.3 DCT Calculation

The DCT is a transform that is connected to the Fourier series. The Fourier series constants of a periodically and symmetrically long sequence are frequently related to the DCTs. There are eight standard DCT variations, with four of them being the most prevalent. The most common discrete cosine transform variation is the kind-II DCT, which is often referred as "essentially the DCT." The type-III DCT, as its inverse, is indeed referred to as the inverse DCT or the IDCT. Because of its solid power density, the DCT, and particularly the DCTII, is widely used in signal and image processing, particularly for overfitting compression. In typical applications, a significant amount of channel estimation is contained in a small group of low DCT processes [2].

3.4 K-D Tree Sorting Calculation

K-D trees are a suitable data structure for a wide range of purposes, particularly searches involving connecting multidimensional search keys. It has been discovered that the other methods' weakest point is that if too many blocks are incorrectly paired, this leads to incorrect hypotheses on the dominant shift vectors, rendering the result only usable large images as well as on minor images.

This performance completely changes when using a k-d tree. Here, the moment features detect operations in all images. Only, FMT and DCT crops have better detection in large image in Fig. 3. Among the color-based methods, COLOR 3 is the most effective. The structures of COLOR 1 and COLOR 2 have very high false-positive rates and cannot be constantly used for detection. Actually, the more balanced distance computation of the k-d tree makes the nature of most feature vectors better (Fig. 4).



(a) Large images

(b) Small images

Fig. 3 Feature test K-D tree sorting

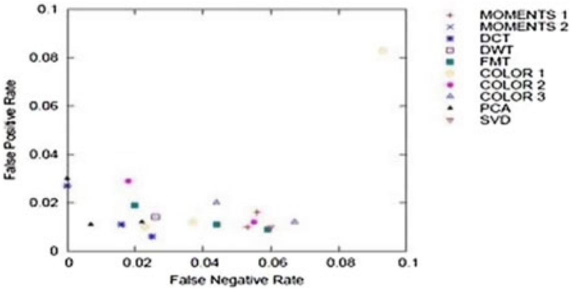


Fig. 4 Feature test k-d tree representation with small images

In Fig. 3, every feature can be used to detect copy–move forgeries with a k-d tree. Only, MOMENTS 2 and COLOR 3 presented difficulties with a single small image. Upon closer examination, one can see that the error rates of the small images exhibit a very low false-positive rate for all methods [3].

3.5 *Common Shift Sector Analysis*

All AC frequency components for 16 × 16 blocks are 2.5 times greater on average than for 8 × 8 blocks, according to experiments, and the DC factor is two times as large. As a consequence, in the following of the 16 × 16 blocks, the frequency modulation matrix (for the Q-factor Q) required to compute the frequency domain has a different shape in Fig. 5.

I is an 8 × 8 unit matrix, and q_{ij} is a standard JPEG normalization matrix with a center frequency of Q where all components are equivalent to one. This is understandable because it is an impromptu test, yet the matrix functioned well throughout the practical tests and subtle tweaks to the matrix had a small impact on the outcomes. We stopped looking into the quantization matrix selection [4].

$$Q_{16} = \begin{pmatrix} Q'_8 & 2.5q_{18}I \\ 2.5q_{81}I & 2.5q_{88}I \end{pmatrix}, where\ Q'_8 = \begin{pmatrix} 2q_{00} & 2.5q_{12}\dots\dots & 2.5q_{18} \\ 2.5q_{21} & 2.5q_{22}\dots\dots & 2.5q_{28} \\ \dots\dots & \dots\dots\ \dots\dots & \dots\dots \\ 2.5q_{81} & 2.5q_{82}\dots\dots & 2.5q_{88} \end{pmatrix}$$

Fig. 5 Common shift sector analysis

3.6 Forgery Detection Using Pixels

Full steps in Fig. 6 are described below:

I. Image splicing is a technique that uses a mixture of multiple or perhaps more mutual photos to generate a fake image, or a method that uses a mixture of two or perhaps more mutual photos to make a convincing photograph.

II. The images are modified less in image retouching. It only highlights a few of the image's many facets.

III. The replica movement forgery is one of the best solid forgeries. This is the most major form of picture tampering, in which information is added or removed by coating a section of an image. A copy–move operation is one in which a segment of a picture is cut and pasted into another segment of a comparable image. In the following, there is an overview of copy–move forgery: (1) Copy–move without reflection; (2) Copy–move with different scaling; (3) Copy–move with different scaling; (4) Rotate the copy–move. According to the literature review, copy–move and forgery sequence information are classified into two types. (1) A method based on blocks, (2) A method focused on key points.

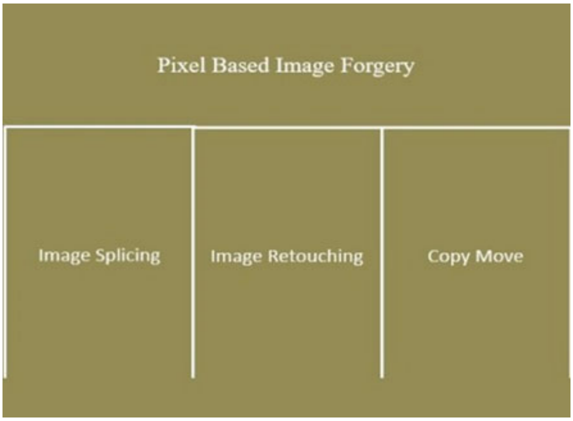


Fig. 6 Forgery detection using pixels

3.7 Color the Pixels in Duplicated Region

The following requirements for the detection algorithm can be directed at:

I. An approximate comparison of image object segments must be possible with the detecting technique.
II. There must work for a fair length of time and produce a minimal false positive rate, such as locating erroneously linked regions.
III. It is also worth mentioning that instead of a mixture of tiny patches or single pixels, the fabricated portion will most likely be a linking element.

3.8 Exact and Robust Match

The user requires the smallest segment size that should be considered for the match at first. Assume this segment is a *BB* pixel square. A matrix has rows which are lexicographically sorted (as *BB* numeric data points) in order to classify the identical rows. In $MN\log_2 (MN)$ phase, this can be done. $MN\log_2 (MN)$ phases can be used to accomplish this. The matching rows in Fig. 7 can be found by searching for the sorted matrix *A* presenting *MN* rows for two comparable concurrent rows.

Robust match detection is the same as targeted search detection in that we sort and compare the blocks' robust representation, which includes quantized DCT coefficients, rather than their pixel representation. The algorithm also considers each matching block pair's shared positions and only produces an exact block pair when there are several other corresponding combinations in about the same reciprocal location (shift vector). The technique maintains the positions of matching blocks in a different section. As an example—the dimensions of a block's uppermost left pixel are used to determine its own location and increase a displacement vector counter *C* if two following rows of the ordered matrix *A* are detected. Let's denote the locations of the two identical blocks $(i1, i2)$ and $(j1, j2)$, respectively.

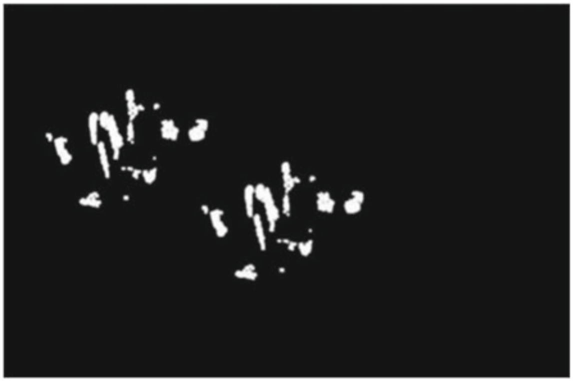


Fig. 7 Results of the experimental algorithm

$$D = (d1, d2) = (i1 - j1, i2 - j2)$$

is the displacement vector between the two matched blocks.

The displacement vectors d is regularized since the displacement vectors—d as well as d correlated to the same displacement.

4 Experimental Details

In this experiment, we used an AMD A8 processor. It is a very simple processor but works smoothly. For problem solving, we use a 64-bit operating system, 2.00 GHz graphics, and the MATLAB software; the three parts of the coding section are as follows:

I. In the first section, a color image known as a suspected image is printed. The first section is linked to the second and third sections.
II. In the second section, overlapping blocks are divided using a robust matching DCT matrix design. The size of the block will be a 16 × 16 matrix, and it will be divided into four blocks and a compute DCT matrix.
III. In the third part, we detect a forged region. Before computing the shift vector, the data are sorted using the k-d tree. We convert the detected part into an RGB color image.

In the test results, the sample result is displayed shown in Figs. 8 and 9. Here, three test results are found by applying the above methods. The image has been tempered with a quality factor and a threshold for locating common shift vectors. This JPEG image was divided into overlapping blocks, DCT'd, and then sorted using a k-d tree. The shift vector is then computed. In this PNG-formatted image, two common shift vectors are obtained.

The detection of the input image has a higher value (less match) and a lower value (stronger match) in each of the three input images. The threshold number of pitches appears to be copied together in the given picture of this computation to consider it a forged region divided into overlapping blocks. Here, we use MATLAB software.

The size of the block in the second portion of the input picture will be a 16 × 16 matrix, split into four blocks. This work is able to detect forged portions after modification. In this case, we are going to use a quality factor of 0.5 and a threshold of 10. We computed the DCT matrix here.

Here, we detect forged regions. The k-d tree sorts and then computes the shift vector. We make an RGB image out of it shown in Fig. 9. In this PNG-formatted image, two common shift vectors are obtained.

Two common shift vectors are obtained in this image. In the detected forged region of the image, a color is assigned to the forged region. The image is highlighted with a quality factor of 0.5 and a threshold value of 10, which gives better performance in detection. We used the COMOFORD database in our research, which included 60 tempered photos. In these photos, this technique correctly and efficiently detects

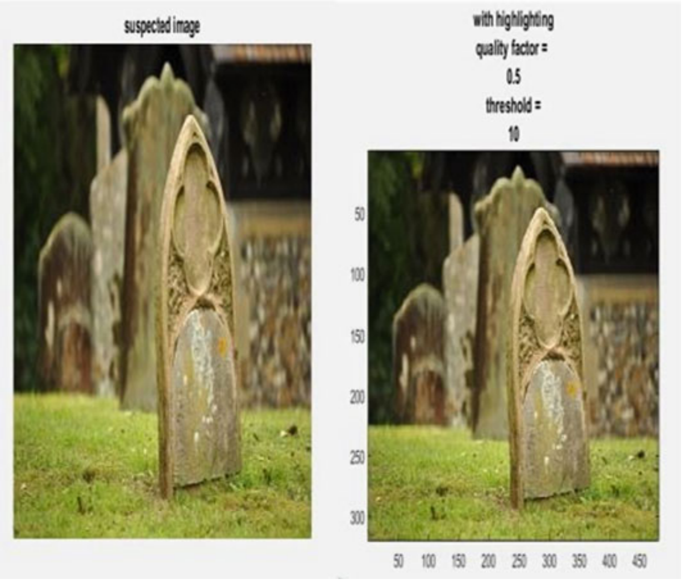


Fig. 8 Tempered image and highlighting with quality factor and threshold for finding common shift vector

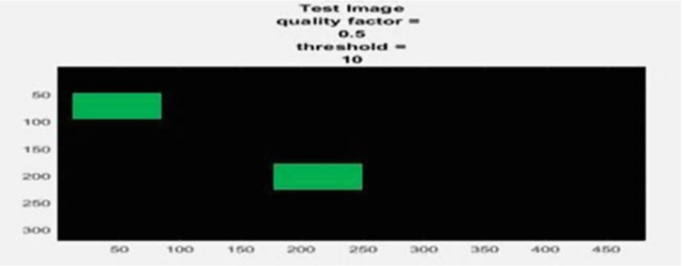


Fig. 9 Forged region detection

most copy–move forgeries. We can see that the previous approach failed to detect fraud in the jpeg format of images. However, there is another method to detect forgery in jpeg format images. DCT may also be used to identify image forgeries, with good results for jpeg images, as seen in the figure. Detecting forged digital photographs is being accomplished in a variety of ways. In this case, we discovered a region of copy picture forging, which is a technique for duplicating and pasting a section of an image.

5 Discussion

Copy–move forging is one of the most frequent counterfeit techniques. Several researchers have defined a variety of methods for detecting altered photos.

However, before being pasted, the duplicated portions are sometimes rotated or flipped. In our paper work, we adopted an effective method for digital images in order to perceive the identical area in the image. To begin, the image is subdivided into adjoining rectangular blocks, which are then used to generate overlapping blocks. Second, the DCT transformation is used to restrict the search area and makes the search unit more resistant to post-processing operations like compression and rotation. Finally, the feature vectors are sorted using a k-d tree after they have been transformed. The output is shown in Figs. 10 and 11.

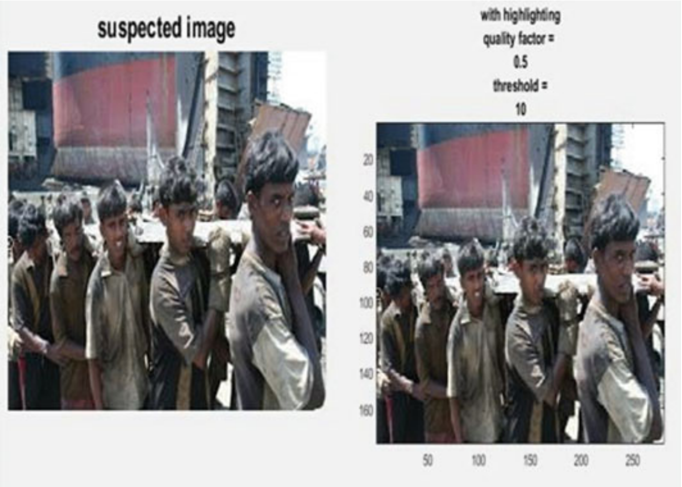


Fig. 10 Tempered image and highlighting with quality factor and threshold for finding common shift vector

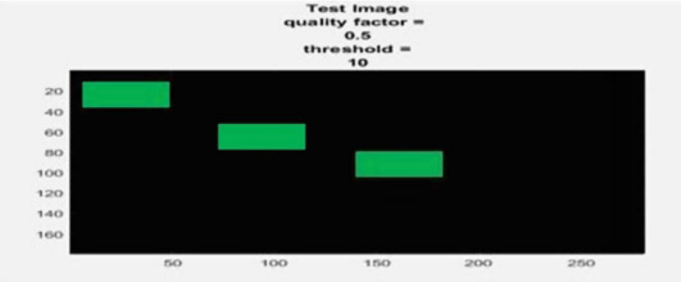


Fig. 11 Forged region detection

We are now working on detecting copy–move image fraud utilizing robust matching and k-d tree sorting. It only works with compressed photos; original images are not supported. To improve query performance and accuracy, we want to modify the data structures even more. Even if the pasted region has been rotated or translated, this method still works.

6 Conclusion

On the COMOFORD database, we utilized a DCT transformation technique and a robust matching method to identify region replication forgery and obtain efficient results. The obtained result is shown in Figs. 12 and 13.

Our future target is to work on our own large dataset and modify the DCT and k-d tree sorting data structures for jpeg images to acquire noticeable results. We will also try to modify the model in such a way that it can work with the original images.

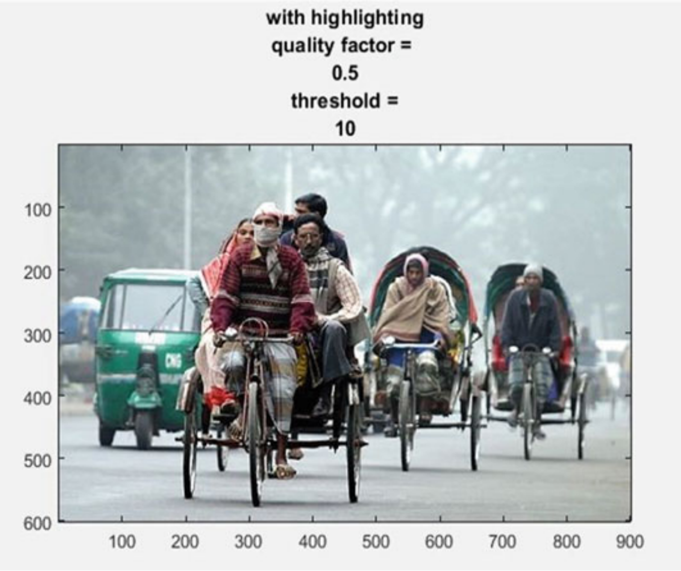


Fig. 12 Tempered image and highlighting with quality factor and threshold for finding common shift vector

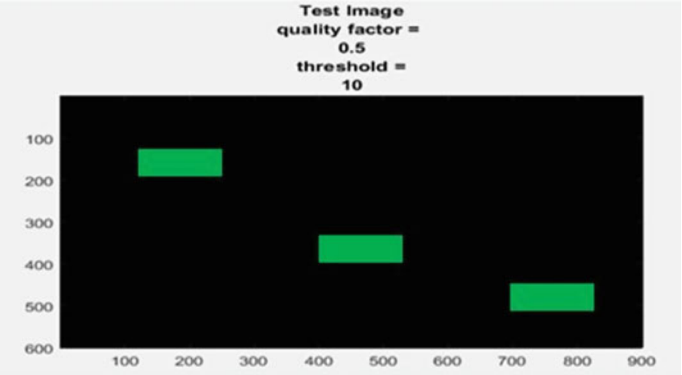


Fig. 13 Forged region detection

References

1. Head, J., Lai, Y.-K.: Image forgery detection (2015)
2. Meena, K.B., Tyagi, V.: A hybrid copy-move image forgery detection technique based on Fourier-Mellin and scale invariant feature transforms (2020)
3. Ahmed, B., Gulliver, T.A., Zahir, S.A.: Blind copy move forgery detection using SVD and KS test (2020)
4. Gani, G., Qadir, F.: A robust copy move forgery detection technique based on discrete cosine transform and cellular automata (2020)
5. Parveen, A., Khan, Z.H., Ahmad, S.N.: Block-based copy–move image forgery detection using DCT (2019)
6. Prakash, C.S., Panzade, P.P., Om, H.: Detection of copy move forgery using AKAZE and SIFT key point extraction (2019)
7. Hegazi, A., Taha, A., Selim, M.M.: An improved copy move forgery detection based on density based on clustering and guaranteed outlier removal (2019)
8. Tan, W., Wu, Y., Wu, P., Chen, B.: A survey on digital image copy move forgery localization using passive techniques (2019)
9. Yang, H.Y., Qi, S.R., Niu, Y., Niu, P.P., Wang, X.Y.: Copy move forgery detection based on adaptive keypoints extraction and matching (2019)
10. Mushtaq, S., Mir, A.H.: Image copy move forgery detection (2018)
11. Mahmood, T., Mehmood, Z., Shah, M., Saba, T.: A robust technique for copy move forgery detection and localization in digital images via stationary wavelet and discrete cosine transform (2018)
12. Ouyang, J., Liu, Y., Liao, M.: Robust copy-move forgery detection method using pyramid model and Zernike moments (2018)
13. Emam, M., Han, Q., Zhang, H.: Two-stage key point detection scheme for region duplication forgery detection in digital image (2017)
14. Rasse, S.G.: Review of detection of digital image splicing forgeries with illumination color estimation. Int. J. Emerg. Res. Manag. Technol. (2017)
15. Kaur, A., Sharma, R.: Optimization of copy-move forgery detection technique. Int. J. Adv. Res. Comput. Sci. Softw. Eng. (2017)
16. Telagarapu, P., Naveen, V.J., Prasanthi, A.L., Santhi, G.V.: Image compression using DCT and wavelet transformations (2017)
17. Chakraborty, P., et al.: A human-robot interaction system calculating visual focus of human's attention level. IEEE Access **9** (2021)

18. Christlein, V., Riess, C., Angelopoulou, E.: A study on features for the detection of copy move forgeries. In: Sicherheit (2017)
19. Chakraborty, P., Nawar, F., Chowdhury, H.A.: Sentiment analysis of Bengali Facebook data using classical and deep learning approaches. In: Innovation in Electrical Power Engineering, Communication, and Computing Technology, pp. 209–218. Springer, Singapore (2022)
20. Hasan, M.R., et al.: Reliable identity management system using Raspberry Pi. In: 2020 2nd International Conference on Sustainable Technologies for Industry 4.0 (STI). IEEE (2020)
21. Chakraborty, P., Sultana, S.: IoT-based smart home security and automation system. In: Micro-Electronics and Telecommunication Engineering, pp. 497–505. Springer, Singapore (2022)
22. Feroz, M., Sultana, M., Hasan, M., Sarker, A., Chakraborty, P., Choudhury, T.: Object detection and classification from a real-time video using SSD and YOLO models. In: Computational Intelligence in Pattern Recognition, pp. 37–47. Springer, Singapore (2022)
23. Sultana, M., Ahmed, T., Chakraborty, P., Khatun, M., Hasan, M.R., Uddin, M.S.: Object detection using template and HOG feature matching. Int. J. Adv. Comput. Sci. Appl. **11**(7), 233–238 (2020)
24. Chakraborty, P., Yousuf, M.A., Rahman, M.Z., Faruqui, N.: How can a robot calculate the level of visual focus of human's attention, pp. 329–342 (2020)
25. Muzammel, C.S., Chakraborty, P., Akram, M.N., Ahammad, K.M.: Zero-shot learning to detect object instances from unknown image sources. Int. J. Innov. Technol. Explor. Eng. **9**(4), 988–991 (2020)
26. Chakraborty, P., Muzammel, C.S., Khatun, M., Islam, S.F., Rahman, S.: Automatic student attendance system using face recognition. Int. J. Eng. Adv. Technol. **9**(3), 93–99 (2020)