Similarity Checking Of Keypoints For Identification Of Copy-Move Forgery (SCK-ICMF)

Chandrakala¹, Dr M Sasikala²

¹Assistant professor GECW chandrakalmail@rediffmail.com
²Principal GECW sasi_mun@rediffmail.com

Abstract

In the advanced world of technologies, pictorial data is also one of the important data source to represent information to the world. A pictorial data can be an information about an incident took place; it can be an evidence to the judgmental courts and so on. As the technology is getting advanced for capturing those digital image data, the misuse of those data is also increasing by manipulating those digital image data. There are many techniques to modify or manipulate the digital image data such as copy-move, scaling and rotating, translation of the image and resizing of the image, etc. that can be used for falsifying the right information as wrong. Thus, to find out whether there is any manipulation in the image or not, there are many methodologies present to detect the forgery of the image. As the up-to-dated methodologies are not accurate to identify the forgery in the soft and minor regions, we are implementing a new methodology known as “Similarity Checking of Keypoints for Identification of Copy-Move Forgery (SCK-ICMF)”. Here we are aiming to check the similar keypoints even in the soft and minor regions. We are using “Copy-Move-Forgery-Detection” D0-dataset to evaluate the results of our designed methodology known as “Similarity Checking of Keypoints for Identification of Copy-Move Forgery (SCK-ICMF)” and compare with the different state-of-art methodologies to show the efficiency of our implemented methodology.

1. Introduction:

In the present era of digital world, digital data images are the main cause of information and it is the best way of conveying knowledge. For the purpose of evidence in courtrooms, the digital image maybe helpful. The pictures that are shown on news are only accepted with the certificate of genuine. Digital data pictures is utilized in several fields such as medical checkup, photography, military, and art work, and so on. Thus, the digital data picture forensics declares as rapidly growing requirement of the world and therefore it is necessary to have genuine image. However, these days, it is very simple to alter the pictures by not leave the clues of alteration using PCs and cheap tools for software and hardware. Therefore, it is harder for people to find out those alterations. By which, the originality and integrity of the pictures are lost. These alterations of the pictures maybe performed for hiding few necessary clues from the picture or the alter the
information of the picture and so on. Such that, the altered information is sent. For finding the picture’s integrity, we first required to find any alterations made on the picture [1]. With this, it is proved that we are in era of visual picturing with significant arrays. Whereas we can have generally had guarantee in the picturing integrity, present methodology of altering pictures are breaking down those trust. Beginning from the sensational magazines to the industry of fashion and also in typical outlets of media, courtrooms, scientific journals and the pictorial frauds that comes on our inboxes of email, doctored pictures are seen with sophistication and increasing frequency. From last few years, the background of digital forensics has come up to bring back the trust to digital pictures. Here we see few of the state of the art techniques in picture alteration field. Watermarking for the pictures was introduced for the purpose of authentication of pictures. The disadvantage of this method is that, the watermarking should be done during capturing of picture that will limit this method to digital cameras that are specially equipped for watermarking. With respect to these methods, inactive methodologies for forensics of picture perform when there is no signature or watermark. These methodology operates on the prediction, which while digital forgeries may not be seen that shows altering, they may tamper picture’s statistics that are underlying. Forgery of the image represents the cruel picture alteration with the motive of misleading the viewer’s observation. This methodology has a big past and it is simplified using picture editing applications that are sophisticated in digital world. Doctored images are looking in our lives with increasing frequency and difficulty [2]. Therefore, detection of forgery of image is becoming a challenging work. Alteration by Post-Processing like scanning and photofinishing makes the detection of forgery even harder. The old authentication of image approaches are signature and digital watermarking. These approaches need extra information to be added into or removed from picture. Unwantedly, many systems of digital imaging do not put information in pictures. While, passive blind forensics of picture was introduced to recognize the manipulation of the picture or to find out the source of the picture without any additional of actual picture. Enhancement in forensics of picture has helped to different techniques for recognition of tampered pictures. Forgery recognition methodology consists many methodology for recognizing the artifacts generated while resampling [3][4], variations because of arrays of color filter [5], Splicing [6], Cloning [7, 8], lighting variations [9, 10, 11] and pattern of noise because of troubles in the sensor of camera [12, 13]. Presently, pictures can be manipulated without any traces left behind using algorithm of computer graphics that are sophisticated and picture manipulation applications such as Acorn, Photosh, etc. [14, 15]. Forgery recognition methodology field has increased necessarily to resolve the problem of picture forgery in many fields such as medical science, administrations of legitimate and legal sciences [16, 17].

Recognition of image forgery is divided into two approaches. One is Active approach and another is Passive approach as represented in Figure-1.
The Active approach includes steganography and watermarking. These are designed during acquisition of picture. A special hardware design such as encoding the picture into another form or digital signature is required to authenticate the originality of the picture. The approach of watermarking is utilized to make invisibility of messages or mark on the image to secure its copyright during acquisition of picture and to verify the authenticity of message is taken from the picture and compared with watermarks of original. Making invisible of confidential message as it should not be utilized for wrong purpose by some others is known as steganography. The Passive approach do not need any pre-information regarding the picture and it is based on the clues remaining on the picture by various processing steps at the time of manipulating the image. Using various image forgery recognition algorithms, the forged region, amount and location of the forgery can be found. The example for this approach are image splicing, copy-move forgery recognition, etc. and it is used for finding the processes made on the image such as blurring, rotation, scaling, etc.

In our paper, we are implementing new methodology for identifying the forgery within the picture known as “Similarity Checking of Keypoints for Identification of Copy-Move Forgery” (SCK-ICMF). In here, we are performing following operations to identify the forgery of the picture. Feature Keypoint Collection and Similarity Checking of Keypoints are the operations performed on the given picture to identify the copy-move forgery. The rest of the paper is organized as following:

In section-2, we are describing our new designed methodology, in section-3, we are evaluating the result of our proposed methodology and compared with the state-of-art methodologies. Finally, in section-4, we are writing the conclusion of our designed methodology.
2. Similarity Checking of Keypoints for Identification of Copy-Move Forgery (SCK-ICMF):

In this section of our work, we are designing a new methodology for identification of forgery by copy-move method of the image for huge feature groups of an image using SIFT methodology for collecting the feature keypoints. There are two stages for identifying the forgery of the image is performed or not; i) Feature Keypoint Collection, 2) Similarity Checking of Keypoints. Both are described below:

2.1 Feature Keypoint Collection:

Because of awesome efficient based on noise distortion and transformation in geometrical way, SIFT technique is used in this research for collecting the feature keypoints. One main issue for methods based on keypoint is that they cannot produce enough keypoint features in soft and minor regions, which low-grades the efficiency. Luckily, after deep research on SIFT methodology, we observe that, we can produce enough SIFT keypoint features including in soft and minor regions. This operation can be performed using two techniques, which are simple but also efficient. 1) Minimization of Threshold of Contrast; 2) Given Picture Rescaling.

2.1.1 Minimization of Threshold of Contrast:

In SIFT methodology; the threshold of contrast is given as $T_c$, is already defined to decline unusable extrema whose values of contrast are lower than $T_c$. Anyways, in soft regions, the values of extrema’s contrast are assumed to be very less, showing that some extrema are qualifying the process of contrast filtering and at last be selected as SIFT feature keypoints. To confirm that enough feature keypoints can be produced in soft regions, our methodology is to minimize the threshold of contrast $T_c$, hence many of extrema having less value of contrast will be qualifying the process of contrast filtering. Anyways, as more the $T_c$ is low, so much the unusable features are collected, which cause in false similarities. So this section is used to minimize the $T_c$.

2.1.2 Given picture rescaling:

At the same time, only minimizing the threshold of contrast would not completely resolve the issue of not enough feature keypoints when the process of copy-move is done on minor regions. Our additional method is to rescale the size of the given picture by factor size, before evaluating the SIFT feature keypoints. It is observed that, big sized picture will select more feature keypoints. As the size value is increased, the amount of feature keypoints also rise that anyways are closely assigned in the plane of image, degrading the issue of similarity checking of keypoints. Thus via above stated two techniques, huge keypoint can be produced, even at minor or soft regions. Anyways, so huge feature keypoints can also make an issue of critical similarity check and also cause more wrong similarities.

2.2 Similarity Checking of Keypoints:
Let \( \text{img} \) be the given input image, representing the related feature keypoints and the alternative descriptors as \( \{\text{point}_1, \text{point}_2, \ldots, \text{point}_n\} \) and \( \{d\text{scptr}_1, d\text{scptr}_2, \ldots, d\text{scptr}_n\} \) individually, where the amount of feature keypoints are denoted as \( n \). Let point be a universal SIFT feature keypoints, and vector = \( \{\text{vector}_1, \text{vector}_2, \ldots, \text{vector}_{n-1}\} \) holds the Euclidean’s distance vector value from feature keypoint point to the rest all \( (n-1) \) feature keypoints in ascending order. point is said to be similar one and only when

\[
\mathbb{C} > \frac{d\text{scptr}_1}{d\text{scptr}_2}
\]  

(1)

Here, \( \mathbb{C} \in (0,1) \) is a constant that is already defined (mostly as 0.6). Unluckily, we can use the methodology of similarity checking because of the Issues of similarity checking of keypoints that is described below.

2.2.1 Issues of similarity checking of keypoints:

After getting huge SIFT feature keypoints, we usually think that the similarity pairs will also be more. However, we observe that the outcomes are opposite. By minimizing the threshold of contrast and rescaling the given picture into big size, the produced feature keypoints is maximized by 10 times. But, the similarity pair minimized by 5 times from before. This is caused because, the production of feature keypoints are selected from the close location or sometimes in the same place with various scale. By which, the alternative descriptors values will be mostly same, hence violating the Eqn-1 condition. We name is problem as Issues of similarity checking of keypoints-I. At the same time, the huge feature keypoints will simultaneously increase the burden of computation, as the similarity checking complexity is \( O(n^2) \). We name is problem as Issues of similarity checking of keypoints-II. In this paper, we are designing a new similarity checking of keypoints to resolve the above two issues.

2.2.2 Scale grouping based similarity checking in groups:

Notice that all feature keypoints are identified in the scale region at the place where pictures of Gaussian are clustered by octave. This will produce closely grouped feature keypoints at various scales while there is huge feature keypoints, which in turn causes the Issues of similarity checking of keypoints-I. Here, our motive is to highly classify the grouped feature keypoints identified at various scales. By this end, we design to perform similarity checking operation inside every single octave of minimum scales individually, whereas inside many octaves of maximum scales in integrative way. The motive is that for maximum scale octaves, the feature keypoints is very less when compared with minimum scale ones. Moreover, the feature keypoints in maximum octaves is also described to be highly efficient and of maximum perceptive ability.

Value of scale for feature keypoints point is denoted as \( S(\text{point}) \) that can be easily gained by evaluating the feature keypoints of SIFT. Assume \( \delta_x \) be the value of scale for 1\textsuperscript{st} DoG picture at \( x-\)
th octave. Every feature keypoints are grouped into 3-groups represented as $G = \{G_1, G_2, G_3\}$. Generally,

$$
G_1 = \{\text{point}_y | \delta_2 > S(\text{point}_y) \geq \delta_1, y = 1 \text{ to } n\}
$$

$$
G_2 = \{\text{point}_y | \delta_3 > S(\text{point}_y) \geq \delta_2, y = 1 \text{ to } n\}
$$

$$
G_3 = \{\text{point}_y | \delta_3 \leq S(\text{point}_y), y = 1 \text{ to } n\}
$$

(2)

After that, the similarity checking operation is done by $G_1$, $G_2$ and $G_3$ individually. Particularly, for $1^{st}$ and $2^{nd}$ octaves, the similarity checking operation is used only at every single octave individually. Whereas for maximum octaves, it is used in many octaves in combinative way. By grouping the scale, the features within various groups are classified, hence resolving the Issues of similarity checking of keypoints-I.

2.2.3 Gray level intersection based similarity checking in groups:

By seeing Eqn-1, it is clear that, the above similarity checking technique completely avoids the concept of finding the correct similarity pairs in two same local positions in copy-move methodology; also have same values of the pixel. To perform the similarity checking operation faster, our designed technique is to initially group the feature keypoints with respect to their values of gray, then perform the similarity checking technique given in section-3.2.2. on every single group, instead of complete identified feature keypoints. Anyways, violently separating feature keypoints into various groups may remove several correct similar pairs on various groups. In our work, we design to separate feature keypoints into intersecting groups. Mainly, the nearby groups will share some of their feature points. By this, we initially linearly separate the complete level of the gray (i.e. $[0, 255]$) into $m$ intersecting sub-levels, where step is the step factor and intersect represents the intersection factor. Remember that (intersect < step).

Consider $G_f = \{G_{f1}, G_{f2}, ..., G_{fm}\}$, here $f \in \{1,2,3\}$ representing the frame number and $G_{fx}$ holds all the feature keypoints in $G_f$ with value of gray belonging to $x$-th sub-level. After that, the similarity checking procedure can be performed on every single group $G_{fx}$ without depending on anything. Consider, $\text{SIMILAR}_{fx}$ be the set having the similar pairs of $G_{fx}$, and $\text{SIMILAR}$ be the set holding all the similar pairs. Scientifically,

$$
\text{SIMILAR} = \bigcup_{f \in \{1,2,3\}, x = 1 \text{ to } m} \text{SIMILAR}_{fx}
$$

(3)

To disregard any possible misunderstanding, we in our work use $(\text{point}, \text{point}')$ to represent a similar pair without checking similarity order, specifically, $(\text{point}, \text{point}') \neq (\text{point}', \text{point})$. Simultaneously, we utilize $(\text{point}, \text{point}')$ to represent similarity pair with directed similarity order that means point is similar to point$'$ by a particular affine transformation. It must be observed that complete similar pairs in SIMILAR are without similarity order, as the similarity checking operation does not include the position details.
3. Results and Analysis:

In this research work, the identifier of the copy-move forgery of the picture known as “Similarity-Checking-of-Keypoints for Identification-of-Copy-Move-Forgery (SCK-ICMF)” is performed on a software specification with Matlab v.2016 running on windows-10 operating system and with the hardware specification of 8GB of RAM along with 1GB of minimum hard-disk size. In our methodology, we are performing the execution of the forgery identification using Copy-Move-Forgery-Dataset (CMFD-D0). Here D0-dataset consists of tampered pictures only of simple copy and move. We perform this execution for different pictures from D0 dataset. After that, we note down the Precision and FPR (False-Positive-Rate) values for each picture and calculate the average of precision and FPR value.

After performing the execution for various 10 pictures from D0 dataset, we obtained their Precision and FPR values for those 10 pictures. The following table-1 represents the precision and FPR values for those 10 pictures in percentage.

Table 1: Representation of Precision and FPR values for 10 test pictures.

<table>
<thead>
<tr>
<th>Precision (in %)</th>
<th>FPR (in %)</th>
</tr>
</thead>
<tbody>
<tr>
<td>99.52</td>
<td>5.69</td>
</tr>
<tr>
<td>87.87</td>
<td>0.25</td>
</tr>
<tr>
<td>97.29</td>
<td>0.22</td>
</tr>
<tr>
<td>100.0</td>
<td>2.61</td>
</tr>
<tr>
<td>100.0</td>
<td>1.12</td>
</tr>
<tr>
<td>99.50</td>
<td>3.25</td>
</tr>
<tr>
<td>98.42</td>
<td>1.34</td>
</tr>
<tr>
<td>99.35</td>
<td>5.43</td>
</tr>
<tr>
<td>99.90</td>
<td>11.75</td>
</tr>
<tr>
<td>98.74</td>
<td>4.66</td>
</tr>
</tbody>
</table>

We get the average precision value and average FPR value representing the Precision and FPR value of our proposed methodology. After computing the average Precision and FPR value of our methodology, we compare our values with the previously implemented methodologies so show the efficiency and accuracy of our methodology. The below table-2 represents the values of precision and FPR value of proposed along with state-of-art methodologies.

Table 2: Comparison between Proposed v/s state-of-art methodologies.

<table>
<thead>
<tr>
<th>Methodologies</th>
<th>Precision</th>
<th>FPR(False-Positive-Rate)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proposed</td>
<td>98.06</td>
<td>3.63</td>
</tr>
<tr>
<td>Huang et al. [18]</td>
<td>78.45</td>
<td>1.22</td>
</tr>
<tr>
<td>Yuan et al. [19]</td>
<td>69.69</td>
<td>2.00</td>
</tr>
</tbody>
</table>
We can clearly observe that, our proposed methodology is producing the best results compared with all other state-of-art methodologies. We also represent the above table into a graph representing to have the clean picture of the accuracy of our designed methodology compared with state-of-art methodologies.

The below figure-2 represents the pictorial representation of Precision value in percentage for various methodologies and guaranteeing that our methodology is better than other state-of-art methodologies.

![Precision Graph](image)

**Figure 2**: Graphical representation of Precision percentage for various methodologies.

We also represent the pictorial representation of FPR value in percentage in figure-3 for various methodologies and show that our methodology is better than state-of-art methodologies.
Figure 3: Graphical representation of FPR percentage for various methodologies.

Here, we can clearly observe that, the proposed methodology known as “Similarity-Checking-of-Keypoints for Identification-of-Copy-Move-Forgery (SCK-ICMF)” is performing the identification of forgery of the picture in better way. The above tables and graphs prove that our designed methodology shows good results when compared with the above mentioned state-of-art methodologies [18-22].

4. Conclusion:

In our designed methodology, we are implementing a new technique known as “Similarity-Checking-of-Keypoints for Identification-of-Copy-Move-Forgery (SCK-ICMF)”. In this methodology, we are identifying the given picture is forgery picture or not. The main aim of this methodology is to identify the forgery picture even in soft and minor regions of the picture. Here we are using Copy-Move-Forgery-Detection (CMFD-D0) dataset pictures for testing the outcome of our designed methodology known as “SCK-ICMF”. After performing the execution of the designed methodology, we are noting down the Precision and FPR values in percentage to compare with the state-of-art methodologies [18-22] and see whether our designed methodology “SCK-ICMF” is better than old methodologies or not. After comparing the results with state-of-art methodologies, we observe that our methodology known as “Similarity-Checking-of-Keypoints for Identification-of-Copy-Move-Forgery (SCK-ICMF)” is out performing compared to old methodologies.

References:


