ISSN: 0067-2904

Copy Move Forgery Detection Using Forensic Images

Ayat Fadhel Homady Sewan*, Mohammed Sahib Mahdi Altaei
Department of Computers Science, Collage of Science, AL-Nahrain University, Baghdad, Iraq

Received: 30/9/2020 Accepted: 20/11/2020

Abstract

Digital images are open to several manipulations and dropped cost of compact cameras and mobile phones due to the robust image editing tools. Image credibility is therefore become doubtful, particularly where photos have power, for instance, news reports and insurance claims in a criminal court. Images forensic methods therefore measure the integrity of image by apply different highly technical methods established in literatures. The present work deals with copy move forgery images of Media Integration and Communication Center Forgery (MICC-F2000) dataset for detecting and revealing the areas that have been tampered portion in the image, the image is sectioned into non overlapping blocks using Simple liner iterative clustering (SLIC) method. Then, Scale invariant feature transform (SIFT) descriptor is applied on the grey of the handled image to gives distinctive key points that classified by K-Nearest neighbor to detect and localize the forged portion in the tempered image. The forgery detection results gave a performance percent of about 98%, which reflects the ability of the KNN classifier that cooperated with SIFT descriptor to detect the forged portions even if the forged area is rotated or scaled or both of them.

Keywords: Copy move, Image forgery detection, SIFT, SLIC, Digital forensics.

كشف التزوير بالنسخ والتحريك باستخدام صور الطب الشرعي

آيات فاضل حمادي صيوان*، محمد صاحب مهدي الطائي
قسم علوم الحاسوب, كلية العلوم ,جامعة النهرين ,بغداد , العراق

الخلاصة

الصور الرقمية مفتوحة للعديد من التلاعبات وانخفاض تكلفة الكاميرات الرقمية والهواتف المحمولة بسبب أدوات تحرير الصور القوية. وبالتالي أصبحت مصداقية الصورة موضع شك ، لا سيما عندما تكون للصور قوة ، على سبيل المثال ، التقارير الإخبارية ومطالبات التأمين في محكمة جنائية. لذلك ، تقيس طرق الطب الشرعي للصور سلامة الصورة من خلال تطبيق أساليب تقنية عالية مختلفة تم وضعها في الأدبيات. ويتعامل العمل الحالي مع نسخ نقل الصور المزورة لمجموعة بيانات تكامل وسائل الاعلام وتزوير مركز الاتصال(MICC-F2000) للكشف عن المناطق التي تم العبث بها في الصورة ، تقسم الصورة إلى كتل غير متداخلة باستخدام طريقة التجميع التكراري البسيط (SLIC). بعد ذلك ، يتم تطبيق واصف تحويل الميزة الثابتة للمقياس (SIFT) على اللون الرمادي للصورة التي تمت معالجتها لإعطاء نقاط مفاتيح مميزة والتي ستصنف بواسطة K-Nearest المجاورة لاكتشاف وتوطين الجزء المزيف في

*Email: progayat@gmail.com

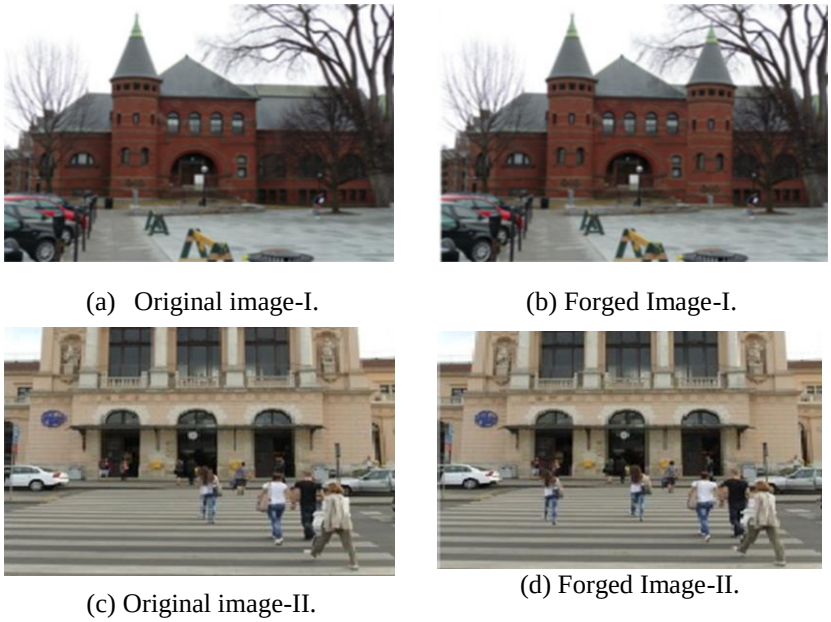
الصورة المخففة. أعطت نتائج كشف التزوير نسبة أداء بلغت حوالي 98٪ مما يعكس قدرة مصنف KNN الذي تعاون مع واصف SIFT على اكتشاف الأجزاء المزورة حتى لو تم تدوير المنطقة المزورة أو تحجيمها أو كليهما .

1. Introduction

Digital images are still representing a major component of the knowledge expand in the customary routine communication, such: social network, website, newspaper, TV, and journals. There is a great risk of modification and integrity attacks due to the saturation of the digital image published on a routine basis on the mob and the simple-to-arrival multimedia channel (e.g., net). In addition, the comfortable availableness of little value and handy strong images redaction code material (such as, Adobe Photoshop, Pixar), have place the legitimacy and safety of image[1]. Current software package permits users to form special effects that cannot be distinguished from real photos or maybe to get hybrid generated visual content[2]. Recently, digital images not only work as master loads to information, but they also take a great job both as a sort of certificate for a criminal court. Away, keeping the correctness and safety for digital images has become a main challenge of late[1].

2. Copy Move Forgery

The copy move (or called cloning) forgery (CMF) is common as one of the most widely used and difficult type to image manipulation techniques. In this technique, a part of the image must be covered in order to add or hide details [3]. A part of the image is copied and pasted elsewhere within the image in the copy move technique[4]. Since the copied portion come from the image itself, its essential properties, as noise, brightness and texture, will be consistent with the rest of the image, making it harder for experts to recognize and detect the modification. A sample of image copy move is illustrated in Figure-1 [5].



(a) Original image-I. (b) Forged Image-I.

(c) Original image-II. (d) Forged Image-II.

Figure 1-Copy move forgery of images [5].

Therefore, in digital images, forgery is linked to a changed or manipulated object. The primary goal of forgery analysis is therefore to decide if any adjustments have been made to alter the meaningful image content[6]. Detection of image forgery is conducted to examine whether the image considered displays the unaltered captured scene or has been forged to deceive the viewer [1]. Generally, they can be categorized into two separate approaches: block-based and key-point-based approaches. Pre-processing, extraction of features, matching as shown in the Figure-2 are the main steps in the detection of copy move forgery[7].

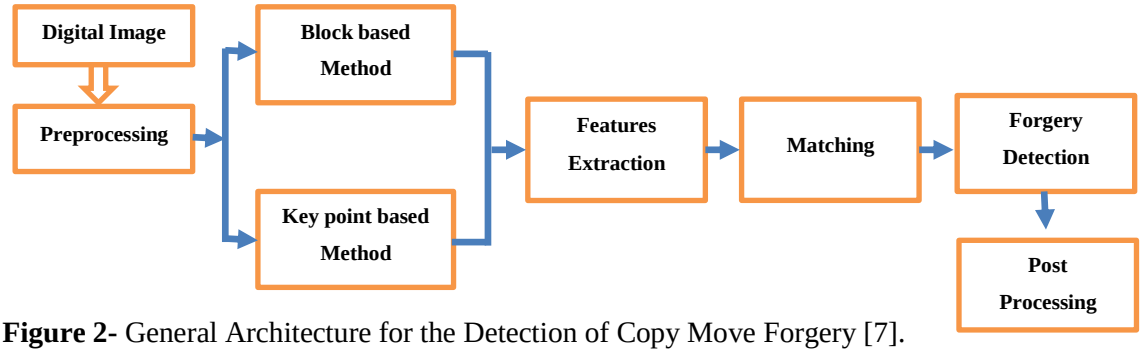


Figure 2- General Architecture for the Detection of Copy Move Forgery [7].

3. Related Works and Contribution

Several studies have been published in the field. They vary in many aspects, like material image, method used, or even limitations of applications. The feasibility that use forgery image detection based on key point is discussed in the following:

3.1 Related Works

Copy move forgery (CMF) detection has been given a great deal of attention. In order to obtain more efficient techniques for supporting particular applications, various approaches have been developed. The most important literature is listed with detail in the following:-

In 2015, Lee et al [8] the images were split into overlapping blocks and the histogram oriented gradient (HOG)of each block's was added. While this approach is capable of detecting multiple CMF instances, rotation and scaling over large areas are poor. The experimental results of this technique showed an accuracy of about 90%. In 2016, Parihar et al [9] proposed a method to copy move forgery detection (CMFD) using SIFT algorithm to extract the feature vectors. Afterward extract feature vectors are stored in a matrix and then matching process is performed. The duplicate areas can be identified via this method. But the big downside is making false matches in flat regions as false positives .The accuracy of the method used reached about 90%. In 2017, Warif, N. et al [10] suggested a robust CMF detection method, called SIFT-Symmetry, that innovatively combines Scale Invariant Feature Transform (SIFT) based Copy Move Forgery detection approach with correspondence dependent on symmetry. With three proven methods based on SIFT, multi-scale analysis, and patch matching, they evaluated the SIFT-Symmetry utilizing two additional data set that cover easy transformation and reflection-based attacks. The results show that the SIFT-Symmetry method's F-score exceeded the average value of 80% for all geometric transformation instances. Including quick transformation and reflective attack, with the exception of rotation case reflections with an overall F-score of 65.3 per cent. In 2017, Dixit, R., R. Naskar, and S. Mishra [11]proposed Blur-invariant copy move forgery detecting strategy with enhanced detection accuracy utilizing Stationary wavelet transform and singular value decomposition (SWT SVD), which demonstrates a copy move forgery recognition approach. This work is close to the results of our work because its accuracy also exceeded 98%, but we worked on adding a verification step as it compared the resulting image with the fake image mask. In 2018, Mahmood T. et al [12] a robust technique for CMF detection and recognition in digital image has been proposed. The method extracts features based on (SWT) for the disclosure of forgeries to digital image. More precisely approximation sub-band of the stationary wavelet transformations is used as this sub-band houses much of the information ideally suited to detecting forgery. By applying discrete cosine transform (DCT) the dimensions of the feature vectors are reduced. The experimental result shows that in terms of true and false detection rate, the proposed technique outperforms the current techniques, where he achieved results exceeding 93%.

4. CMF Detection Methodology

Copy move is the most common method of tamper in the images because of its effectiveness. It's simple to create the CMF by copying part(s) of the image to cover the element(s) in the same image, but it's very difficult to detect it by the naked eye when performed carefully.

4.1 Gray Scale Conversion

The color transformation is done if there is a necessity to transform the color image into a gray

image using the common formula (intensity, or luminance)[13]. Such that, the intensity grey image (*I*) is computed using the luminance formula from the three color bands as follows:

$$\mathbf{I} = \mathbf{0.299} \times \boldsymbol{R} + \mathbf{0.587} \times \boldsymbol{G} + \mathbf{0.114} \times \boldsymbol{B} \quad \textbf{(1)}$$

4.2 SLIC Image Segmentation

Image segmentation is significant for digital image processing. In computer vision and image processing, image segmentation is significant. For image segmentation, there are many existing methods. However, it is difficult to make the segmentation results fit human experience. The definition of super pixels was suggested. Which super pixel is a perceptually significant atomic area. After this, the simple linear iterative clustering (SLIC) super pixel was formulated. It is advanced version of the super pixel and can be created in a very efficient way [14]. With a lower handling times and costs of storage, the SLIC algorithm achieves good quality segments than another approaches. That approach is very easy and have a one variable k , Which is the appropriate numeral of super pixels of equal size to generate[15]. This forms super through grouping pixel using a 5-dimensionals (labxy) space depending on their colors likeness and nearness in picture level, Where its lab Color is a more accurate color space. It uses three values (L, a, and b) to specify colors. The a-axis (green to red), b-axis (blue to yellow) and L is Lightness axis.That approach is very easy and have a one variable k , Which is the appropriate number of super pixels of equal size to generate [5]. SLIC takes a desired number of approximately equally-sized super pixels Kslic as input. So each super pixels will have approximately Av=N/K average area of super pixel ,where N is number of pixels in the input image. Hence, for equally sized super pixels, there would be a super pixel center at every grid interval S=$\sqrt{N\ /\ K}$. Euclidean distances in CIELAB color space are meaningful for small distances[15]. Equations (2, 3, and 4) show how calculate spatial distance measure (Ds) as follows:

$$\Delta \boldsymbol{E} = \sqrt{(\boldsymbol{Li} - \boldsymbol{Lk})^2 + (\boldsymbol{ai} - \boldsymbol{ak})^2 + (\boldsymbol{bi} - \boldsymbol{bk})^2} \quad \textbf{(2)}$$

Where

$$\mathbf{L^*=116(Y/Yn))1/3-16} \quad \textbf{(3)}$$

$$\mathbf{a^*=500[(X/Xn)1/3\ -(Y/Yn)1/3]} \quad \textbf{(4)}$$

$$\mathbf{b^*=200[(Y/Yn)1/3\ -(Z/Zn)1/3]} \quad \textbf{(5)}$$

Where Xn, Yn, Zn being the XY Z values of the white point. Auxiliary definitions are:

$$\mathbf{chroma} = \mathbf{c} * = \sqrt{(\mathbf{a}^*)^2 + (\mathbf{b}^*)^2} \quad \textbf{(6)}$$

$$\mathbf{hue\ angle} = \mathbf{h} * = \mathbf{arctan} + \frac{\mathbf{a}^*}{\mathbf{b}^*} \quad \textbf{(7)}$$

Roughly, a* the maximum and minimum of value a correspond to red and green, while b* ranges from yellow to blue. Chroma is a scale of colorfulness, with more colorful (more saturated) colors occupying the outside of the CIELAB solid at each L brightness level, and more washed-out (de saturated) colors nearer the central achromatic axis. The hue angle expresses more or less what most people.Gradients for images are computed as follows:

$$\boldsymbol{G}(\boldsymbol{X}, \boldsymbol{Y}) = \left|\left|\boldsymbol{I}(\boldsymbol{x}\ +\ \mathbf{1}, \boldsymbol{y}) - \boldsymbol{I}(\boldsymbol{x} - \mathbf{1}, \boldsymbol{y})\right|\right|^2 + \left|\left|\boldsymbol{I}(\boldsymbol{x}, \boldsymbol{y} + \mathbf{1})\ -\ \boldsymbol{I}(\boldsymbol{x}, \boldsymbol{y} - \mathbf{1})\right|\right|^2 \quad \textbf{(8)}$$

where I(x, y) is the lab vector corresponding to the pixel at position (x, y) and || || is act distance computing[15].

4.3 Scale Invariant Feature Transform

Scale invariant feature transform (SIFT) SIFT is a process for finding or extracting important features from the images .The feature should have two main requirements: The repetition in the original image should be avoided and dimensionality of the data must be reduced. That purpose of the SIFT is used to locate the key points (features) in various size areas and to measure the predominant direction of the key point. That key-points detected via SIFT were several notable key-points, like angles, corners, high points in the black region, vice versa, that are not affected via brightness, transformations, and distortion [16]. constructing scale space (octaves) measure L (x , y, σ), that was generated using conversion of Gaussian variables scales, G(x, y, σ), by use a source images, I(x, y):

$$\mathbf{L}(\mathbf{x}, \mathbf{y}, \boldsymbol{\sigma})\ =\ \mathbf{G}(\mathbf{x}, \mathbf{y}, \boldsymbol{\sigma})\ *\ \mathbf{I}(\mathbf{x}, \mathbf{y}) \quad \textbf{(9)}$$

where * is the operations of convolution in (x, y) also , where σ is scale parameter

$$\mathbf{G}(\mathbf{x}, \mathbf{y}, \boldsymbol{\sigma}) = \frac{\mathbf{1}}{\mathbf{2\pi\sigma^2}}\ \mathbf{e}^{-\frac{\mathbf{x^2+y^2}}{\mathbf{2\sigma^2}}} \quad \textbf{(10)}$$

To generate Different of Gaussian (DOG) 2nd order derivative use scale spaces extreme in DOG converted as well as the image , Gf(x, y, σ), to effectively detect stable key point positions in scale space, that could be determined as per comparison of 2 nearest scales, isolated by a static multiplication operator k [**13**, **17**].

$$\boldsymbol{Gf(x,y,k\sigma) - G(x,y,\sigma)) * Img(x,y) \approx (k-1)a^2\nabla^2 Gf} \quad \textbf{(11)}$$

Also, Sift is eliminate the Edge Response ,when D(X) less than 0.03 .To illustrate this orientation computation , a HOG is computed in the neighborhoods of the key points. With the same position and scale, it produces key points, but distinct directions. In the situation of the image test L(x, y,σ) given to scales , an direction ∅(x,y) and gradients size m(x,y) are pre-calculated by pixels variance using the In the given formulas[18]. In the given formulas:

$$\boldsymbol{m(x,y) = ((L(x+1,y) - L(x-1,y) + \big(L(x,y+1) - L(x,y-1)\big)^2\ )\ ½} \quad \textbf{(12)}$$

$$\boldsymbol{\emptyset(x,y) = \tan^{-1}\frac{LG(x,y+1) - LG(x,y-1)}{LG(x+1,y) - LG(x-1,y)}} \quad \textbf{(13)}$$

where,(x , y, σ, f) In the given details (x , y) describes image plane coordinate,σ represent scales , and f contain the final descriptor**.**

4.4 Image Matching

Different parts are copied and moved to the same image during the copy move forgery process , so there is a robust correlation between these parts. This could be used for the detection of forgery as evidence. But identifying effective features and matching algorithms for identifying the associated regions is the main challenge. The matching of features is carried out to define the great similarities or matching between descriptors of features. If the similarities between the descriptors of the feature is found, it is interpreted as an indication for the duplicated regions[10]. Several method of identifying these similarities can be utilized like KNN method is supervise learning. . The first stage of K-NN is choose parameter (k), which is number of nearest neighbors. Next, calculation distance between the query (test sample) and all training samples by using equation:

$$\boldsymbol{ED = \sqrt{\sum(Xi - Yi)^2}} \quad \textbf{(14)}$$

Where, *ED* is the distance, *Xi* is training sample, and *Yi* is test sample .Later, must sorting distance and determine the nearest neighbors. final stage is apply simple majority to determine the predicate class[19,20].

5. Contribution

The motivation behind the present paper is to determine the original and copy place. Due to this matter was neglected by the previous literatures and was not touched upon, the present research focuses on this particular point. Also, the process of determining the copy move location in the image is more interest and requires a comprehensive study for the contents of the image. The contribution of the current work is the use of verification of the resulting image, and this step is more stringent to determine the places that have been manipulated by comparing the resulting image with Mask. The employed method will be compare and verify to existing state-of-art methods in terms of the effectiveness, robustness, matching time complexity, detection reliability, and forgery location accuracy, which is useful to verify the authenticity and integrity of digital images.

6. Proposed Forgery Image Detection (FID) Method

The general structure of the proposed forgery image detection is depicted in Figure-3, it contains two main stages: the first is the forged image detection (FID), which deals with the grey images that firstly goes to be segmented using SLIC method, and then applying the SIFT descriptor on the gray converted images to find the significant features. This prepare to use KNN classifier for matching features of multiple image segments and making a decision related to the existence of forged segment and its location. Algorithm (1) illustrate the main stages of the proposed method.

The proposed FID stage includes multiple sequential steps within: first, the input colored image is segmented by SLIC method into non-uniform several image parts, then the segmented image is converted from RGB colored bands into one gray scaled image that input into SIFT feature extraction to extract the dominant features for each part in the image. These features are achieved and then used to compared with each other that belong to another image segments. KNN classifier is used to detect image segments that shows same image features to declare them as similar or matched .The next parts illustrate further detail about the sequent steps of the propose images FID stage.

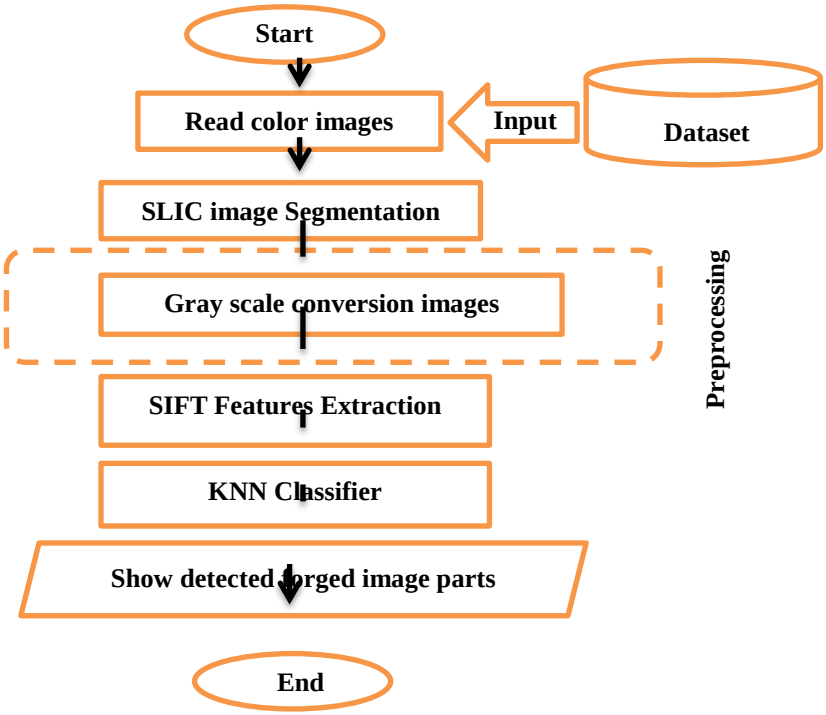


Figure 3-Block diagram of the proposed CMFD method.

Algorithm 1- Forged image part(s) detection by SIFT descriptor.

Input	Color image
Output	***D*** \\ Decision (forged or authentic image)

Procedure	**Step 1**: Read color image.
	Step 2: apply segmentation by using SLIC algorithm as Eq (2,3,4,5).
	Step 3 :gray conversion as Eq(1) // Convert the image to gray scale.
	Step 5: Feature Extraction as Eq(6,7,8,9,10) // Extract key-points feature and their descriptor by use SIFT algorithm.
	Step 6 : Matching by KNN as Eq(11)// Euclidean distance between only key-point for two similar cluster is computed
	Step 7 :Decision // if no match key point then the image is authentic
	Else ,the image is forged**.**
	Step 8 : Investigation by compare output image with mask using Intersection Over Union // IF IOU >= 0.5 then good detection
	Else ,bad detection.

End	

6.1 SLIC Image Segmentation

Simple Linear Iterative Clustering (SLIC) is an adjusted method of clustering by which images pixels are grouped into super pixels. With low processing time and memory expense, the SLIC algorithm provides better quality segments than other state-of-the-art methods. A single parameter (k) of similarly sized super pixels of (N*N) size is significantly present in the algorithm. Figure- 4 provides an example of the SLIC super pixel segmentation image in which case (a) provides the forgery image, whereas case (b) displays the output of the SLIC super pixel segmentation method being applied.

(a) Forged Image. (b) Segmented Image.

Figure 4- SLIC segmentation method applied on guns image.

6.2 Image Preprocessing

The pre - processing phase contain gray band computing and images spectral boost. The gray scale computing is depend on the belief that the three color gamut of images is described like a single gray bar, which reduces the efficiency of explaining pixel images from (24 bits) to (8 bits) per pixel, so the range of gray color intensity should range from 0-255 Values. The cause for recognizing such image from each another isolate of color image is that minimal information wants to be supplied for each pixel. In addition, grey scale image is quite enough for many jobs and thus there is no want to utilize much complicated and harder-to-process color image. The adopted method for converting the colored image into its grey scale, this is due to the SIFT applied only on gray images.

6.3 Features Extraction

SIFT method is used to extract many key-points from the images, which can be regarded as good image features for image description process. These features may be a piece of data that have relevancy for solving the computational task associated with description purpose. The most of the SIFT features are noticed aggregated with a specific structures within the image like points, edges or objects. Also, SIFT features are invariant to different factors and eminently special. For that, the probability of detecting a match between one feature to a data base of feature is highly possible. Figure-5 shows the block diagram of SIFT features extraction process.

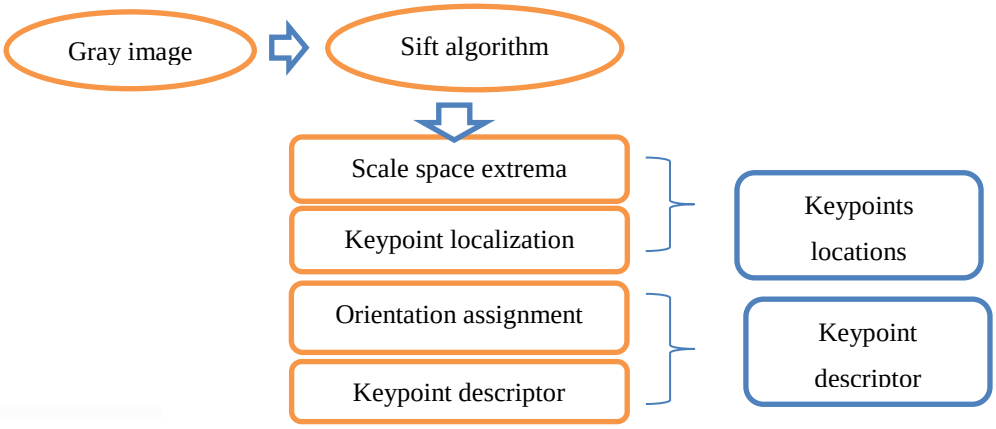


Figure 5- Block diagram of features extraction phase using SIFT algorithm.

6.4KNN Image Classification

Brute-Force (BF) matcher is straightforward. It gives the descriptors of one feature in first set and is matched with all another features in second set utilizing many distance computation, and the nigh one is returned. Whereas, BF-KNN uses the same idea of the BF for the match, but with return *k* best matches. Where, *k* is a number less than or equal the number of features in the feature vector. KNN is used to match the features vector of each image segment with the features set,

where the features set is composed of a number of features vector is equal to the number of segments of target image as shown in Figure-6.

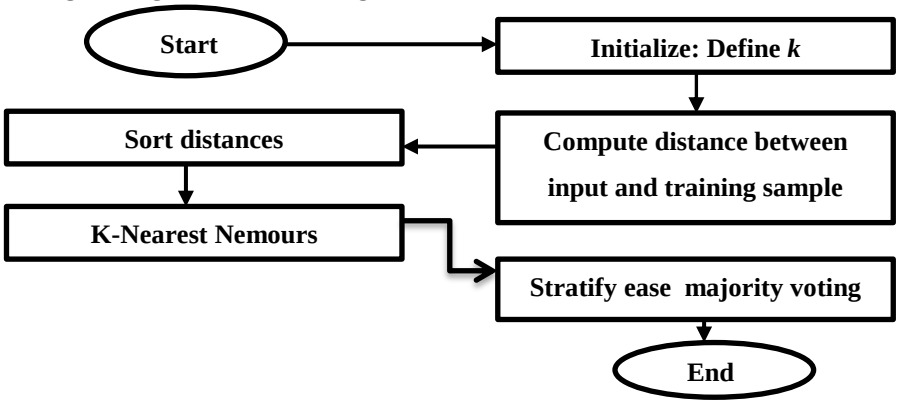


Figure 6- Block diagram of features matching by using KNN algorithm.

7. Results and Discussion

one well-known dataset is used to test and compare the performance of different forgery detection methods, this is: MICC-F2000 . These forgery images in this dataset is construct by cloning parts of the image and pasting them into the same image For the purpose of challenging the credibility of the image. Many types of transformation have been applied to fake images, such as rotation (90˚, 180˚ angle), translation, scaling or combination of them. this dataset is composed of images that have different sizes . There 248 image samples are selected from MICC-F2000 data set; in which 200 are tampered and 48 are originals. The MICC-F2000 data set consists of an images have different sizes, the Media Integration and Communication Center (MICC) dataset are JPEG images format. The MICC is a portion of the knowledge Engineering Division of Florence University. It is a multidisciplinary research facility for the processing, transmission and interpretation of videos and images , on synthetic vision and multimedia techniques applied to health, social applications, and cultural heritage. Figure-7 illustrates some dataset samples of the tampered images of MICC-F2000 dataset. The following subsections presents more explination about the results of each stage of the proposed FID meth.

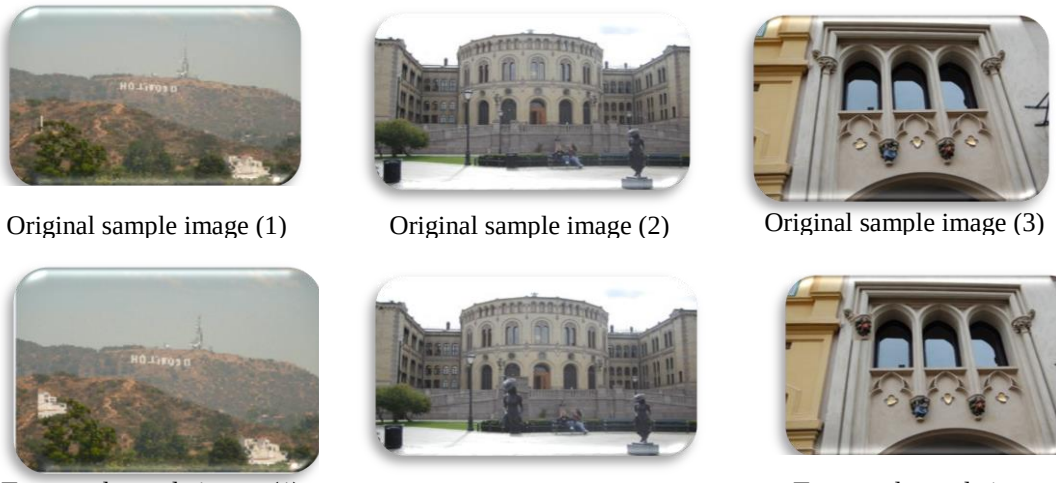


Figure 7-Sample images of MICC-f2000 dataset, images in upper row are authentic while images in lower row are its forged ones[3].

7.1 Image Segmentation Result

The medium resolution of material images used in the present work make the size of the resulted image segments is proper, the results of the image segmentation are shown in Figure-8, in which the use of proper number of segments (*n*) lead to make acceptable segmentation results, greater *n* values leads to confuse the classification results, while less *n* value leads to poor image

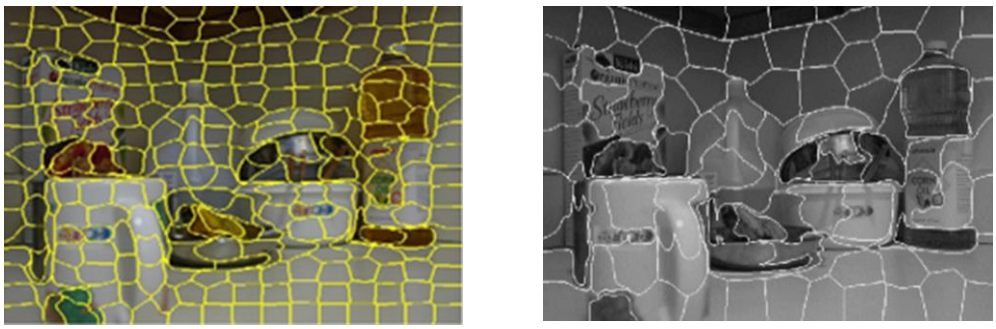
segments and no information may found. Such segmentation results enable to locate objects in terms of lines, curves, and boundaries of the input color images. The threshold (T=0.4) is value determines the number of matching segments in advance and also the size of image based on several experiments, it was found to be the best value. It is found that the change in image resolution leads to decrease in accuracy of the segmentation results, and this effect may leads to change the image dimensions and angles that directly affects the process of features extraction by SIFT, which mainly depends on the angles of the shapes.

(a) Input image (1). (b) Segmented Image (1).

Figure 8- SLIC image segmentation results.

7.2 Result of pre - processing

The pre-processing step is apply to the incoming image to transform the twenty four bits spectral resolution incoming color sample image through an eight bits spectral resolution gray scale images. The above step allows the incoming image to be well analyzed by machine learning (ML) because to its direct effects on time of calculation and detection rate. The gray image conversion technique is applied and checked with the features extraction methods to assess that ability for extracting beneficial descriptors with the gray scale images. As seen in Figure-9, the weighted contribution of color of three bands (R,G,B) reported a high contrast results. It is obvious that the precise details of the resulting gray image are clearly seen, where the image is still reserved in the same sense and did not lose any of its details, which indicates its ability to be input the next image description step.



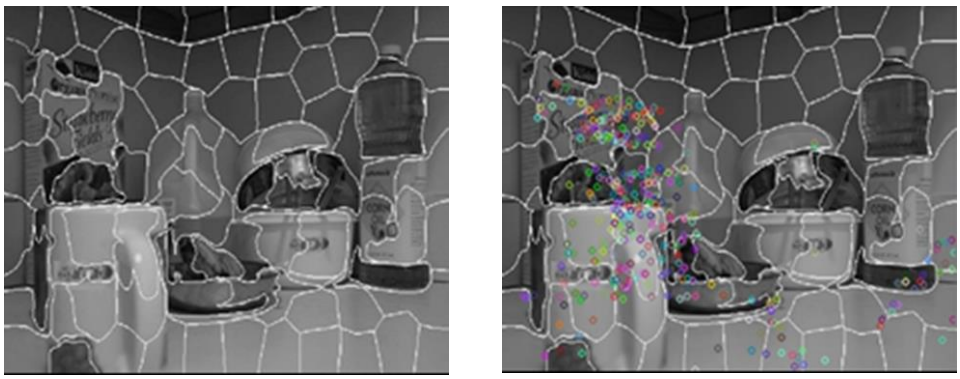
(a) Input image (1). (b) Segmented Image (1).

Figure 9- Grey image conversion results.

7.3 Feature Extraction Results

SIFT descriptor is a particular approach used in 2 steps for extracting features from gray images : 1) detection of Interest point :- In the detection stage, the Hessian matrix is used to detect blob as structure on the integral images 2) Description of the interest points :-several interest points are observed at various scales; the number of interest point is proportional to the number of spectral variance in a particular area. Three test images and whose corresponding interest points are shown in Figure-10. Outcome of the detection stage were applied to the localization of interest points. Probably depends on the spectral distribution of images intensity, it is seen that the numbers of interest points per test varies from image to image. The highest peak in the histogram is taken and

any peak above 80% of it is also considered to calculate the orientation. As a consequence, to every interest point, there are 64 features that are achieved. It indicates that for each image segment, n * 64 features are extracted. The localization of the points of interest was added to the outcomes of the detection stage. Based on the spectral distribution of that region, it is seen that the number of interest points per test varies from one image segment to another. The number of keypoints in a specific image segment is unlimited and depends on the number of resulted key points, which is differs from one image segment to another. Practically, these keypoints are stored in a two dimensional array represents the features array of that image segments in the database. It is noticeable that the key points of any image segments are found on the corners and edges of the objects found in that region, while there is no key points are shown in the empty regions; i.e., no objects in the region of interest.

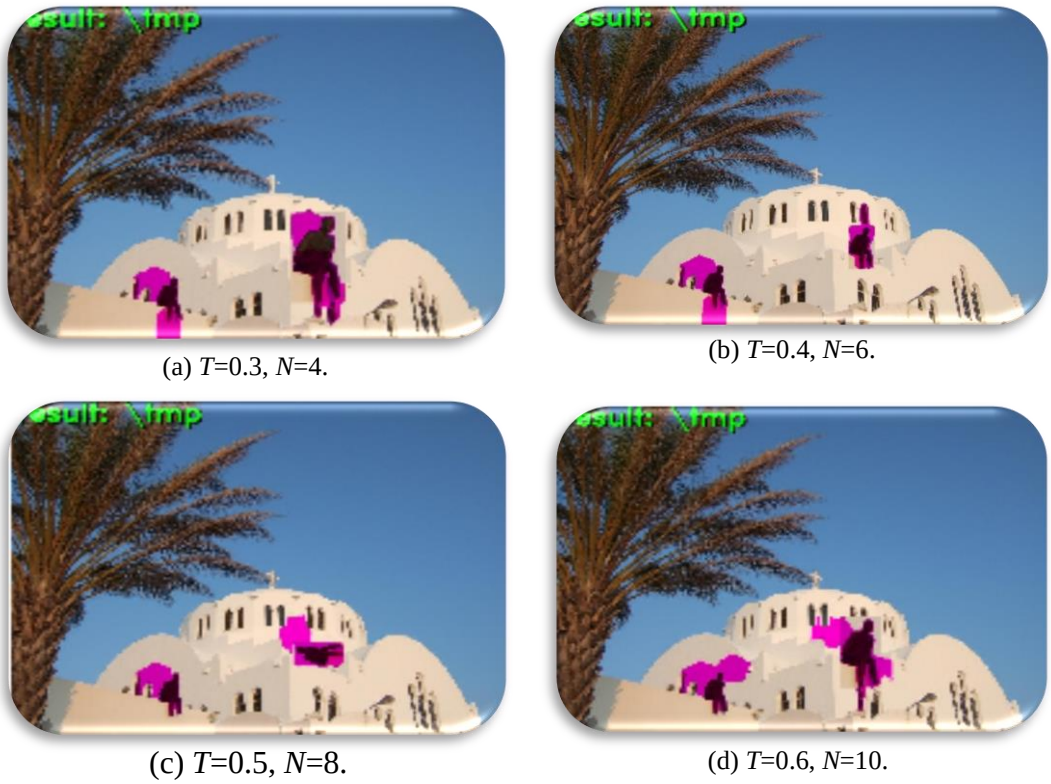


(a) Segmented image (1). (b) Key points in segmented Image (1).

Figure 10-Resulted SIFT keypoints in the grey segmented image sample.

7.4 Matching Result

The SIFT results refers to the results of the similarity matching of the proposed FID system. The increase of similarity fraction threshold (*T*) value of the match leads relatively to raise the number (*N*). The number of segments containing key point exceeds the threshold condition when it matches two segments and gives a true output in both the original and forged image parts, in which the mismatched is also shown connected incorrectly. On the contrary, when decrease the similarity threshold, the number of matched key points is relatively reduced, this associated with false connecting between dissimilar key points. Many tests have been implemented to check the accuracy of the proposed FID method, in which different threshold values and number of matches have been tried. The effect of different matching thresholds and number of matched key points is illustrated in Figure-11. The features vector of 128 dimension that belongs to either key points of a particular section in the test image is evaluated by the different values of similarity threshold and number of matches to verify the description ability of its features. Then it is aimed at reducing the features that lead mis-classification outcomes. Numerous experiments are performed in which the detection rate is calculated with an acceptable number of adjectives feature takes into account. The outcomes of the proposed FID are intended to be a detection region between two similar segments for just four match threshold values (T=0.4, 0.5, 0.6, and 0.7). The detected key points represents the most dominant features that show best discriminant behavior than others. It is found that the use of fixed threshold and fixed number of matches for all image parts does not always lead to acceptable results, this is due to the different texture is found in each segment. Thus, it is necessary to find out the best value of both T and N. The results show that the best FID performance is occurred when T=0.4 and N=6 are used. Such that, one can considered such values of T and N are useful for running the next KNN detection stage.



(a) T=0.3, N=4. (b) T=0.4, N=6.

(c) T=0.5, N=8. (d) T=0.6, N=10.

Figure 11- Effect of T and N on resulted matching when No. of segments=300.

7.5 KNN Detection Results

The comparison process of any key point is done by comparing them to the nearest neighbors using KNN. The process of determining the value of the k is based on specifications of the used sample image. The increase of the value of k leads to an increase in a number of comparisons and also increasing the processing time, and vice versa. To make the balance between the consumed processing time and the number of required comparisons, several values of k have been considered to access the best one that gives the best detection results. Table-1 illustrates the effect of increasing k values on the processing time applied on this image. Thus, The circular neighbor region radius is needed to be modified with various value to enhanced best for behavioral classification results. It is clear that the accuracy of detection of the various run at which the circular coverage area radius (k) ranges from 1-7 is detected differ as per the value of the radius . Compared to others, the three pixel radius value provided the highest detection results. It is also shown that accuracy of detection is increased by increasing the radius value until the optimum one is reached when the radius is Three pixels, which brings a mean detection score of approximately 98 % when using 80% of the used dataset that randomly chosen to be contributed in the test performance measurement. Then this score had been fluctuated about same achieved level and with value of the radius rising .The explanation behind these activity is that the rise in radius allows more important information to be included within the region considered, which contributes to a rise in the detection level. The noticed disadvantage of more increasing k value leads to late the detection decision making the system to consumed more additional time. Thus, one can decide that the best value that gave acceptable detection results is k=3. The use of such k value make the required comparisons were performed with the convenient amount of the processing time. It is improbable to find the matched key points after the 3rd neighbor, so any comparison after that may be considered a waste of time and without any usefulness.

Table 1- Effect of increasing *k* versus processing time .

Value of K	processing Time
K=3	4
K=5	6
K=7	8
K=9	10
K=11	12
K=13	14

The matching between each two image segments is carried out by comparing the current segment with all remaining ones. When the key points of current segment is same as that of other segment, then one can consider them as similar to each other. In such case, the classifier refers to the closest two similar image segments as forged image parts. The numerical comparison used the distance measure between the two features vectors belong to the two image segments under consideration to determine the convergence between them. To investigate the true results, the location of the detected two forged portions are compared with mask associated with the handled image in the dataset. In case of matching the locations of the two detected forged portions with that of the mask, then the detection percent is 100% for that image, while the detection percent is 0% when the mentioned locations are do not identical. In case of identifying one location with the mask, then the detection percent is 50% only. Table-2 seen the gained detection score of KNN algorithm depended on SIFT descriptor. Figure-12 pictures the behavior of the KNN classifier given in Table-2 that indicates the forgery detection for ten randomly chosen forged images from the used dataset. These results showed that the mean forgery detection scores for KNN algorithm that based on computing means (μ) for detection score for ten runs of the proposed FID method was about 98.514% with a standard deviation (σ) of 2.013. In fact, those encouraging results show that the SIFT descriptor utilized acts positively mostly with classifier to obtain the better forgery detection , the descriptors are establish helping the classifier to reach the high detection rates .

Table 2-Forgery detection result

Runs	Detection Score
1	98.01
2	98.01
3	98.02
4	98.1
5	98.65
6	98.8
7	98.85
8	98.9
9	98.9
10	98.9
μ	98.514
Σ	2.013

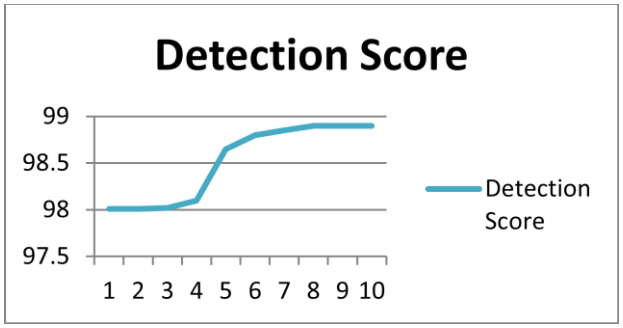
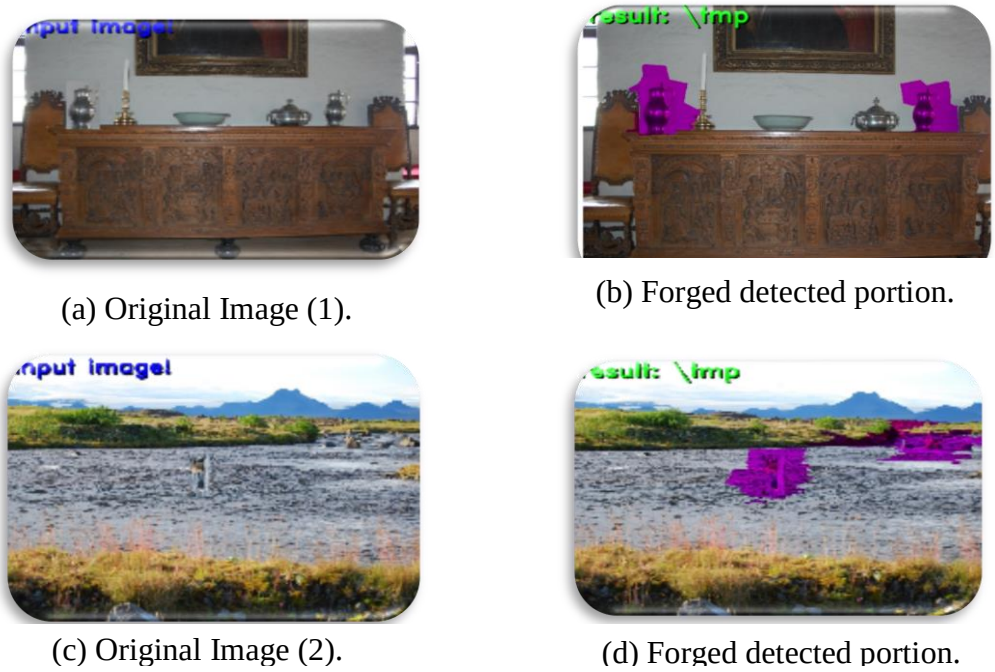


Figure 12-Detection results of the KNN classifier based on SIFT descriptor.

8. FID Results Evaluation

The proposed algorithm has been tested using 248 images determined from the used dataset. The FID results evaluation is an important test based on using the remaining 20% of the used dataset that are not previously used in the test. Both TP & TN are calculated in this test to assess the performance of the proposed approach, leading to the predicted FP and FN errors that determine the accuracy of the process. These parameters prepare to compute the three performance measures: Prescion, Recall, and F-score. Figures- (13,14) shows Some original and tampered images used in the testing and the achieved values of the evaluation parameters given in Table-3, in which the average processing time for each test was about 4*sec*.



(a) Original Image (1). (b) Forged detected portion.

(c) Original Image (2). (d) Forged detected portion.

Figure 13-Detection results of sample tampered images

Table 3- confusion matrix for two- class(authentic, forgery).

	Actually positive (1)	**Actually negative (0)**
Predicated positive(1)	TP=47	FP=0
Predicated negative(0)	FN=4	TN=196

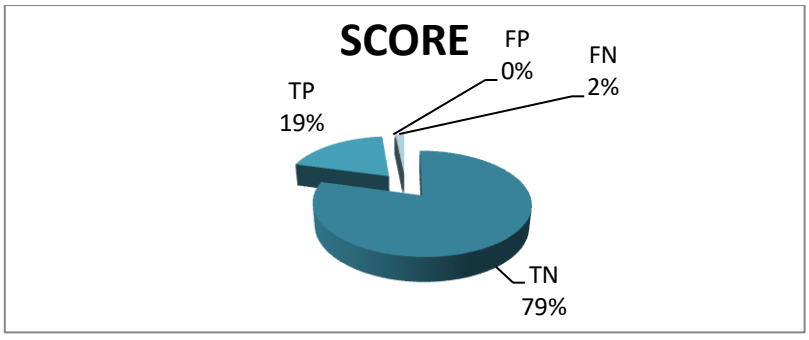


Figure 14-Forgery detection scores of test results.

9. FID Results Investigation

In order to check the efficiency of the forged portion detection in the handled image, the image segments of highest matching probability have been considered while the smallest probabilities

were eliminated. This procedure of non-maximum suppression (NMS) is usually used for object detection, where the location of each image segment of accepted probability is compared with that determined in the original mask to find the amount of the overlapping between them. The measure that describes the amount of closing these regions is the intersection over union (IOU). IOU is actually used to measure the overlap between two images regions belong to different image references. IOU is computed by dividing the intersection of two images on the union of them. Such that, when there is no intersection between two images the IOU is zero, while when the intersection is totally cover them then the IOU is one. The partial intersection between them refers to the percent of identifying the location of the compared image regions. The results of this procedure are to present a box enclose the original and tampered regions in the target image as shown in Figure15.

The application of the IOU on 100 tampered image sample containing only two similar regions within showed that there always was another identical portion in spite of little shift occurred in the location of the detected tampered region in comparison with its location in the mask image. In general, the overlap rate was about 70%, which indicates the assurance of existing a tampered region in the image in a shifted location. Several tests showed that the used method has the ability to detect the tampered regions in terms of its positions in the mask image. Less IOU percent refers does not underestimate the importance of such method, it is refers to existence a tempered portion bun not exactly fit its location in the mask image. The results of IOU pointed in Figure- (19-b and d) showed two green boxes in each image, each enclose the tampered and original portions in the image, which is refers to the effectively detection of both tampered and original portions in the image. It is usually of these results to possess IOU about 0.5 due to expanding the box to be greater than that found in the image mask. In such case, this method success to decide whether the image is forged or not, and also gave the approximate location of the tampered region. Many tests proved that this method was never wrong in detecting tampered, but it gave

approximate results for tampered locations. Such that, the FID results achieved 100% forgery detection.

10. Conclusions

Throughout the implementation, it is concluded that the segmentation by SLIC method is proper when using medium resolution of material images, in which the use of proper number of segments (*n*) lead to make acceptable segmentation results. The threshold (*T=0.4*) value in the SLIC method determines the number of segments in advance and also the size of image. The change in image resolution leads to decrease the accuracy of the segmentation results, which affects the features extraction by SIFT descriptor. The accuracy of detection is improved by raise the value of the radius until the optimum one is reached when the radius is Three pixels, giving a mean detection value of approximation 98 %. The mean forgery detection value of KNN-classifier that based on computing means (μ) of detection score for ten runs of the proposed FID method was about 98.514% with a standard deviation (σ) of 2.013. The measured average processing time for FID implementation was about (3-4) hours.

8. Refrences

1. Roy, A., et al. **2018**. *Digital Image Forensics: Theory and Implementation*. Vol. 755. Springer.
2. Meyer, G.W., et al. **1986**. An experimental evaluation of computer graphics imagery. ACM *Transactions on Graphics* (TOG), **5**(1): 30-50.
3. Sharma, V., S. Jha, and D.R.K. Bharti. **2016**. Image Forgery and it's Detection Technique: A Review. *International Research Journal of Engineering and Technology* (IRJET).
4. Ansari, M.D., S.P. Ghrera, and V. Tyagi. **2014**. Pixel-based image forgery detection: A review. *IETE journal of education*, **55**(1): 40-46.
5. Sridevi, M., C. Mala, and S. Sandeep. **2012**.Copy–move image forgery detection in a parallel (6)Singh, R., A. Oberoi, and N. Goel. **2014**. Copy move forgery detection on digital images. *International Journal of Computer Applications*, **98**(9).
6. environment. in Proceedings of Im age And Signal Processing, Cisp'09. 2nd International Congress.

7. Mahdi, M.S. and S.N. Alsaad. **2019**. Detection of Copy-Move Forgery in Digital Image Based on SIFT Features and Automatic Matching Thresholds. in International Conference on Applied Computing to Support Industry: Innovation and Technology. Springer.

8. Lee, J.-C., C.-P. Chang, and W.-K. Chen. **2015**. Detection of copy–move image forgery using histogram of orientated gradients. *Information Sciences*, **321**: 250-262.

9. Parihar, V. and B. Mehtre. **2016**. Copy move forgery detection using key-points structure. Sardar Patel University of Police, Security and Criminal, 2016.

10. Warif, N.B.A., et al. **2017**. SIFT-symmetry: a robust detection method for copy-move forgery with reflection attack. *Journal of Visual Communication and Image Representation*, **46**: 219-232.

11. Dixit, R., R. Naskar, and S. Mishra. **2017**. Blur-invariant copy-move forgery detection technique with improved detection accuracy utilising SWT-SVD. I*ET Image Processing*, **11**(5): 301-309.

12. Mahmood, T., et al. **2018**. A robust technique for copy-move forgery detection and localization in digital images via stationary wavelet and discrete cosine transform. *Journal of Visual Communication and Image Representation*, **53**: 202-214.

13. Lowe, G. **2004**. SIFT-the scale invariant feature transform. *Int. J*, **2**: 91-110.

14. Joglekar, N.P. and P. Chatur. **2015**. A compressive survey on active and passive methods for image forgery detection. *International Journal of Engineering and Computer Science*, **4**(1): 10187-10190.

15. Kavzoglu, T. and H. Tonbul. **2018**. *A*n experimental comparison of multi-resolution segmentation, SLIC and K-means clustering for object-based classification of VHR imagery. *International journal of remote sensing*, **39**(18): 6020-6036.

16. Li, G., et al. **2007**. A sorted neighborhood approach for detecting duplicated regions in image forgeries based on DWT and SVD. in 2007 IEEE international conference on multimedia and expo. 2007. IEEE.

17. Gupta, C.S. **2016**. A review on splicing image forgery detection techniques. IRACST-*International Journal of Computer Science and Information Technology & Security* (IJCSITS), **6**(2).

18. Prasad, S. and B. Ramkumar. **2016**. Passive copy-move forgery detection using SIFT, HOG and SURF features. in 2016 IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT). 2016. IEEE.

19. Amerini, I., et al. **2011**. *A sift-based forensic method for copy–move attack detection and transformation recovery.* IEEE transactions on information forensics and security, **6**(3): 1099-1110.

20. Chen, H., X. Yang, and Y. Lyu. **2020**. Copy-Move Forgery Detection Based on Keypoint Clustering and Similar Neighborhood Search Algorithm. *IEEE Access*, **8**: 36863-36875.