IMAGE FORGERY DETECTION ON CUT-PASTE AND COPY-MOVE FORGERIES

[1]GARGI RATHOD, [2]SHRUTI CHODANKAR, [3]RUPALI DESHMUKH, [4]PRIYANKA SHINDE, [5]S. P. PATTANAIK

[1,2,3,4,5]Computer Engineering Department, RSCOE,
Rajarshi Shahu College of Engineering, Pune
E-mail: [1]rathodgargi14@gmail.com, [2]shrutichodankar06@gmail.com, [3]rupalivdeshmukh7@gmail.com,
[4]shindepriyanka425@gmail.com, [5]swapna_pattanaik1983@rediffmail.com.

Abstract— Now-a-days, it is possible to manipulate an image by removing or adding important features from it without leaving any clue of editing the original image. One can use advanced tools to digitally manipulate images to create non-existing situations which leads to discarding the originality of images. These modifications are not visible to the naked eye. Cut-paste and Copy-move forgeries are common image manipulations. In cut-paste forgery, a portion of another image is cut and pasted on another image. Whereas in copy-move, part/portion of the same image is copied and moved onto the same image. As devices like cameras are getting more and more digitized, there is an increase in the need for digital image authentication, validation and forgery detection. To detect cut-paste forgery, we use histogram equalization detection technique. It Detects whether the image's contrast has been enhanced or not. For copy-move detection, up till the useful techniques was block matching technique and key point based forgery detection method. These techniques are integrated along with adaptive over segmentation algorithm. The adaptive over segmentation algorithm divides the image into non-irregular blocks adaptively. Forgery regions are then detected using the forgery region extraction algorithm. Finally merged regions are detected by applying morphological operations.

Keywords— Copy-Move, Cut-Paste, Histogram Equalization, Adaptive Over-Segmentation, Block Matching, Forgery Region Extraction.

I. INTRODUCTION

The availability of powerful digital image processing programs, such as PhotoShop, makes it relatively easy to create digital forgeries from one or multiple images. Due to the development of computer technology and image processing software, digital image forgery has been increasingly easy to perform. However, digital images are a popular source of information, and the reliability of digital images is thus becoming an important issue. In recent years, more and more researchers have begun to focus on the problem of digital image tampering. Of the existing types of image tampering, a common manipulations of a digital image are cut-paste and copy-move forgeries, which is to paste one or several copied region of an image onto other parts of the same image or on another image. To hide cut-paste forgery, forgers usually enhance the contrast of the image. This enhancement can be detected by using Histogram Equalization Detection. In this technique, pf-cdf calculation is performed. In the end, plotting the cumulative graph shows whether the image is cut-paste forged or not. During the copy and move operations, some image processing methods such as rotation, scaling, blurring, compression, and noise addition are occasionally applied to make convincing forgeries. Because the copy and move parts are copied from the same image, the noise component, color character and other important properties are compatible with the remainder of the image some of the forgery detection methods that are based on the related image properties are not applicable in this case. In previous years, many forgery detection methods have been proposed for copy-move forgery detection. According to the existing methods, the copy-move forgery detection methods can be categorized into two main categories: block based algorithms and feature key point based algorithms. The existing block based forgery detection methods divide the input images into overlapping and regular image blocks then, the tampered region can be obtained by matching blocks of image pixels or transform coefficients. As an alternative to the block based methods, key point based forgery detection methods were proposed, where image key points are extracted and matched over the whole image to resist some image transformations while identifying duplicated regions. An adaptive over segmentation method is proposed to segment the host image into non overlapping and irregular blocks called Image Blocks (IB).Then, we apply the Scale Invariant Feature Transform (SIFT) in each block to extract the SIFT feature points as Block Features (BF). Subsequently, the block features are matched with one another, and the feature points that are successfully matched to one another are determined to be Labelled Feature Points (LFP), which can approximately indicate the suspected forgery regions. Finally, we propose the Forgery Region Extraction method to detect the forgery region from the host image according to the extracted LFP.

II. DETAILS EXPERIMENTAL

2.1. Cut-Paste Forgery

(a)

(b)

Fig.1. (a & b) Original image (a) has been contrastly enhanced to hide the pasted part (b).

2.1.a. Cut-Paste Forgery Detection

Fig.2. Detection of Cut-Paste Forgery

Grey Scale Conversion

In photography and computing, a grayscale or greyscale digital image is an image in which the value of each pixel is a single sample, that is, it carries only intensity information. Images of this sort, also known as black and white, are composed exclusively of shades of grey, varying from black at the weakest intensity to white at the strongest. Grayscale images are distinct from one-bit bi-tonal black-and-white images, which in the context of computer imaging are images with only the two colours, black and white. Grayscale images are often the result of measuring the intensity of light at each pixel in a single band of the electromagnetic spectrum (e.g. Infrared, visible light, ultraviolet, etc.), and in such cases they are monochromatic proper when only a given frequency is captured. But also they can be synthesized from a full color image.

Display Histogram A histogram is a graphical representation of the distribution of numerical data. It is an estimate of the probability distribution of a continuous variables (quantitative variable) and was first introduced by Karl Pearson. To construct a histogram, the first step is to "bin" the range of values—that is, divide the entire range of values into a series of intervals—and then count how many values fall into each interval. The bins are usually specified as consecutive, non-overlapping intervals of a variable. The bins (intervals) must be adjacent, and are usually equal size.

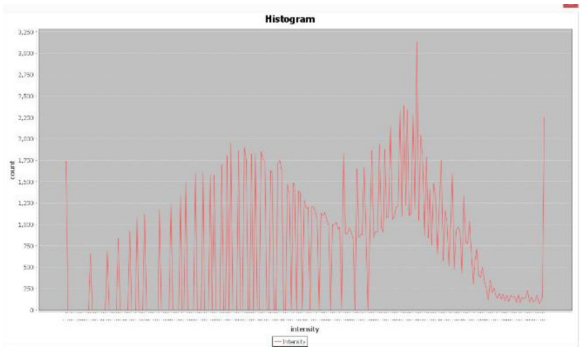

Fig.3. Histogram of fig.1.b

Display Cumulative Graph A cumulative frequency graph is a visual representation of ranked categorical data. An example is grouping people by height categories: under five feet, five feet to six feet, and above six feet. The number of people in each category is the frequency. To find the cumulative frequency, add to each successive category the totals of the lower categories. Graphing a normally distributed variable such as height will result in an s shape, with x as the height and y as the frequency. The cumulative frequency graph is useful in calculating the median and quartiles. The median is the x value when y is halfway up the y axis, or half of the total frequency. The quartiles are at 25% and 75% of the y axis.

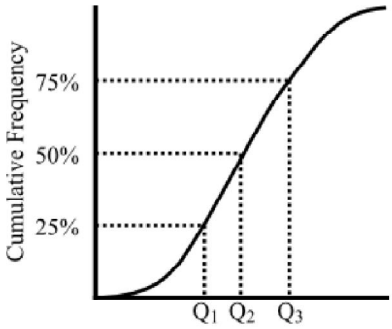

Fig.4.Example

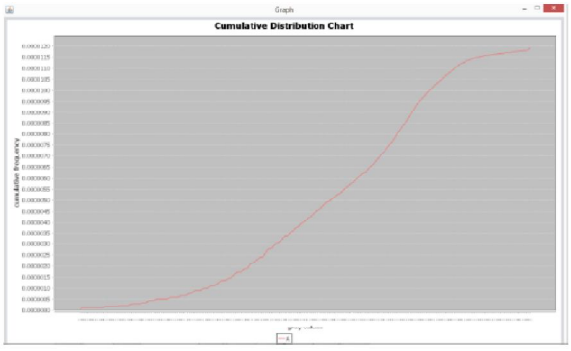

Fig.5.Cumulative Graph of fig.1.b

A smooth curve indicates that the digital image has undergone cut-paste forgery.

2.2. Copy-Move Forgery Detection

Fig.6. (a & b) Original image (a) Forged image (b).

Copy-move forgery is common image manipulating where a part of the image is copied and pasted on another parts. Copy-move forgery detection is used to search the copied regions and their pasted ones. Up to now the useful way to detect copy-move forgeries is block matching technique and key point based forgery detection method. The project integrates both these techniques along with adaptive over segmentation algorithm. The adaptive over segmentation algorithm divides the image in irregular blocks adaptively. Forgery regions are then detected using the forgery region extraction algorithm. Finally merged regions are detected by applying morphological operations. The method may successfully detect the forged part even when the forged area is enhanced to merge it with the background.

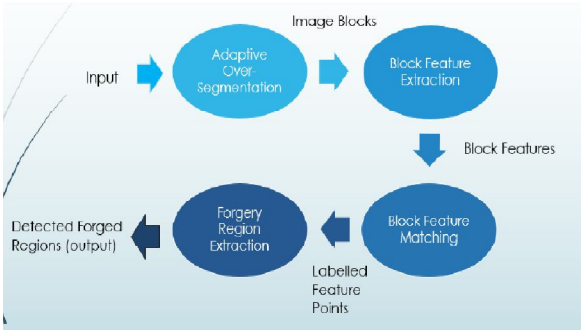

Fig.7.System Working of Copy-Move.

1. Adaptive Over-Segmentation Algorithm

In our copy-move forgery detection scheme, we first propose the Adaptive Over-Segmentation algorithm, which is similar to the traditional block-based forgery detection methods and can divide the host image into blocks. In previous years, a large amount of block-based forgery detection algorithms have been proposed. Of the existing block-based forgery detection schemes, the host image was usually divided into overlapping regular blocks, with the block size being defined and fixed beforehand. Then, the forgery regions were detected by matching those blocks. In this way, the detected regions are always composed of regular blocks, which cannot represent the accurate forgery region well; as a consequence, the recall rate of the block-based methods is always very low. Moreover, when the size of the host images increases, the matching computation of the overlapping blocks will be much more expensive. To address these problems, we proposed the Adaptive Over-segmentation method, which can segment the host image into non-overlapping regions of irregular shape as image blocks; afterward, the forgery regions can be detected by matching those non-overlapping and irregular regions. Because we must divide the host image into non overlapping regions of irregular shape and because the superpixels are perceptually meaningful atomic regions that can be obtained by over-segmentation, we employed the simple linear iterative clustering (SLIC) algorithm to segment the host image into meaningful irregular superpixels, as individual blocks. The SLIC algorithm adapts a k-means clustering approach to efficiently generate the superpixels, and it adheres to the boundaries very well. The different blocking/segmentation methods, where the overlapping and rectangular blocking, shows the overlapping and circular blocking, and the non-overlapping and irregular blocking with the SLIC segmentation method. Using the SLIC segmentation method, the non-overlapping segmentation can decrease the computational expenses compared with the overlapping blocking; furthermore, in most cases, the irregular and meaningful regions can represent the forgery region better than the regular blocks. However, the initial size of the superpixels in SLIC is difficult to decide. In practical applications of copy-move forgery detection, the host images and the copy-move regions are of different sizes and have different content, and in our forgery detection method, different initial sizes of the superpixels can produce different forgery detection results; consequently, different host images should be blocked into superpixels of different initial sizes, which is highly related to the forgery detection results. In general, when the initial size of the superpixels is too small, the result will be a large computational expense; otherwise, when it is too large, the result will be that the forgery detection results are not sufficiently accurate. Therefore, a balance between the computational expense and the

detection accuracy must be obtained when employing the SLIC segmentation method for image blocking. In general, the proper initial size of the superpixels is very important to obtain good forgery detection results for different types of forgery regions. However, currently, there is no good solution to determine the initial size of the superpixels in the existing over-segmentation algorithms. In this paper, we propose a novel Adaptive Over-Segmentation method that can determine the initial size of the superpixels adaptively based on the texture of the host image. When the texture of the host image is smooth, the initial size of the superpixels can be set to be relatively large, which can ensure not only that the superpixels can get close to the edges but also that the superpixels will contain sufficient feature points to be used for forgery detection; furthermore, larger superpixels imply a smaller number of blocks, which can reduce the computational expense when the blocks are matched with one another. In contrast, when the texture of the host image has more detail, then the initial size of the superpixels can be set to be relatively small, to ensure good forgery detection results. In the proposed method, the Discrete Wavelet Transform (DWT) is employed to analyze the frequency distribution of the host image. Roughly, when the low-frequency energy accounts for the majority of the frequency energy, the host image will appear to be a smooth image; otherwise, if the low-frequency energy accounts for only a minority of the frequency energy, the host image appears to be a detailed image. We have performed a large number of experiments to seek the relationship between the frequency distribution of the host images and the initial size of the superpixels to obtain good forgery detection results. We performed a four-level DWT, using the 'Haar' wavelet, on the host image; then, the low-frequency energy E_{LF} and high-frequency energy E_{HF} can be calculated using (1) and (2), respectively. With the low-frequency energy E_{LF} and high-frequency energy E_{HF}, we can calculate the percentage of the low-frequency distribution P_{LF} using (3), according to which the initial size S of the superpixels can be defined as in (4).

$$E_{LF} = \_|CA4| \quad (1)$$

$$E_{HF} = \_ \_i (\_|CDi| + \_|CHi| + \_|CVi|), \quad i = 1, 2, \ldots, 4 \quad (2)$$

where $CA4$ indicates the approximation coefficients at the 4th level of DWT; and CDi, CHi and CVi indicate the detailed

2. Block Feature Extraction Algorithm

In this section, we extract block features from the image blocks (IB). The traditional block-based forgery detection methods extracted features of the same length as the block features or directly used the pixels of the image block as the block features; however, those features mainly reflect the content of the image blocks, leaving out the location information. In addition, the features are not resistant to various image transformations. Therefore, in this paper, we extract feature points from each image block as block features, and the feature points should be robust to various distortions, such as image scaling, rotation, and JPEG compression. In recent years, the feature points extraction methods SIFT [20] and SURF [21] have been widely used in the field of computer vision. The feature points extracted by SIFT and SURF were proven to be robust against common image processing operations such as rotation, scale, blurring, and compression; consequently, SIFT and SURF were often used as feature point extraction methods in the existing key point-based copy-move forgery detection methods. Christlein et al. [22] showed that the SIFT possessed more constant and better performance compared with the other 13 image feature extraction methods in comparative experiments. As a result, in our proposed algorithm, we chose SIFT as the feature point extraction method to extract the feature points from each image block, and each block is characterized by the SIFT feature points that were extracted in the corresponding block. Therefore, each block feature contains irregular block region information and the extracted SIFT feature points.

3. Block Feature Matching Algorithm

After we have obtained the block features (BF), we must locate the matched blocks through the block features. In most of the existing block-based methods, the block matching process outputs a specific block pair only if there are many other matching pairs in the same mutual position, assuming that they have the same shift vector. When the shift vector exceeds a user-specified threshold, the matched blocks that contributed to that specific shift vector are identified as regions that might have been copied and moved. In our algorithm, because the block feature is composed of a set of feature points, we proposed a different method to locate the matched blocks. The block is calculated adaptively; with the result, the matched block pairs are located; and finally, the matched feature points in the matched block pairs are extracted and labeled to locate the position of the suspected forgery region. The detailed steps are explained as follows.

Algorithm Block Feature Matching Algorithm

Input: Block Features (BF);

Output: Labeled Feature Points (LFP).

STEP-1: Load the Block Features $BF = \{BF_1, BF_2, \ldots, BF_N\}$, where N means the number of image blocks; and calculate the correlation coefficients CC of the image blocks.

STEP-2: Calculate the block matching threshold $T R_B$ according to the distribution of correlation coefficients.
STEP-3: Locate the matched blocks *MB* according to the block matching threshold $T R_B$.
STEP-4: Label the matched feature points in the matched blocks *MB* to indicate the suspected forgery regions.

4. Detection of Copy-Move Forgery

Any Copy-Move forgery introduces a correlation between the original image segment and the pasted one. This correlation can be used as a basis for a successful detection of this type of forgery. Because the forgery will likely be saved in the lossy JPEG format and because of a possible use of the retouch tool or other localized image processing tools, the segments may not match exactly but only approximately. Thus, we can formulate the following requirements for the detection algorithm:
1. The detection algorithm must allow for an approximate match of small image segments
2. It must work in a reasonable time while introducing few false positives (i.e., detecting incorrect matching areas).
3. Another natural assumption that should be accepted is that the forged segment will likely be a connected component rather than a collection of very small patches or individual pixels.
In this section, two algorithms for detection of the Copy-Move forgery are developed – one that uses an exact match for detection and one that is based on an approximate match. Before describing the best approach based on approximate block matching that produced the best balance between performance and complexity, two other approaches were investigated – Exhaustive search and Autocorrelation.

4.1 Exhaustive search

This is the simplest and most obvious approach. In this method, the image and its circularly shifted version (see Figure 5) are overlaid looking for closely matching image segments. Let us assume that *xij* is the pixel value of a grayscale image of size M×N at the position *i, j*. In the exhaustive search, the following differences are examined:

$$|xij - xi+k\ mod(M)\ j+l\ mod(N)|,\ k = 0, 1, \ldots, M-1,\ l = 0, 1, \ldots, N-1 \text{ for all } i \text{ and } j.$$

It is easy to see that comparing *xij* with its cyclical shift $[k,l]$ is the same as comparing *xij* with its cyclical shift $[k',l']$, where $k'=M-k$ and $l'=N-l$. Thus, it suffices to inspect only those shifts $[k,l]$ with $1 \le k \le M/2$, $1 \le l \le N/2$, thus cutting the computational complexity by a factor of 4. or each shift $[k,l]$, the differences $\Delta xij = |xij - xi+k\ mod(M)\ j+l mod(N)|$, are calculated and threshold with a small threshold *t*. The threshold selection is problematic, because in natural images, a large amount of pixel pairs will produce differences below the threshold *t*. However, according to our requirements we are only interested in connected segments of certain minimal size. Thus, the threshold difference Δxij is further processed using the morphological opening operation. The image is first eroded and then dilated with the neighborhood size corresponding to the minimal size of the copy-moved area (in experiments, the 10×10 neighborhood was used). The opening operation successfully removes isolated points. Although this simple exhaustive search approach is effective, it is also quite computationally expensive. In fact, the computational complexity of the exhaustive search makes it impractical for practical use even for medium-sized images. An estimate of the computational complexity of the algorithm is given below.
During the detection, all possible shifts $[k,l]$ with $1 \le k, l \le M/2$ need to be inspected. For each shift, every pixel pair must be compared, thresholded, and then the whole image must be eroded and dilated. The comparison and image processing require the order of *MN* operations for one shift. Thus, the total computational requirements are proportional to $(MN)2$. For example, the computational requirements for an image that is twice as big are 16 times larger. This makes the exhaustive search a viable option only for small images.

4.2 Autocorrelation

The autocorrelation of the image x of the size M×N is defined by the formula:

$$, , 0, \ldots, 1. , 0, \ldots, 1$$
$$1\ 1$$
$$, = \Sigma\Sigma , , = - = -$$
$$= =$$
$$r\ x\ x + + i\ k\ M\ j\ l\ N$$
$$M$$
$$i$$
$$N$$
$$j$$
$$k\ l\ i\ j\ i\ k\ j\ l\ .$$

The autocorrelation can be efficiently implemented using the Fourier transform utilizing the fact that $r = x * x^{\wedge}$, where $x^{\wedge}ij = xM+1-i,N+1-j$, $i = 0, \ldots, M-1$, $j = 0, \ldots, N-1$. Thus we have $r = F-1\{F\ (x)\ F\ (x^{\wedge})\}$, where F denotes the Fourier transform.
The logic behind the detection based on autocorrelation is that the original and copied segments will introduce peaks in the autocorrelation for the shifts that correspond to the copied-moved segments. However, because natural images contain most of their power in low-frequencies, if the autocorrelation *r* is computed directly for the image itself, *r* would have very large peaks at the image corners and their neighborhoods. Thus, we compute the autocorrelation not from the image directly, but from its high-pass filtered version. Several high-pass filters were tested: Marr edge detector, Laplacian

edge detector, Sobel edge detector, and noise extracted using the 3×3 Wiener filter (see, for example, [ImgProcBook]). The best performance was obtained using the 3×3 Marr filter. Assuming the minimal size of a copied-moved segment is *B*, the autocorrelation copy-move detection method consists of the following steps:

1. Apply the Marr high-pass filter to the tested image.
2. Compute the autocorrelation r of the filtered image.
3. Remove half of the autocorrelation (Autocorrelation is symmetric.).
4. Set $r = 0$ in the neighborhood of two remaining corners of the entire autocorrelation.
5. Find the maximum of r, identify the shift vector, and examine the shift using the exhaustive method (this is now computationally efficient because we do not have to perform the exhaustive search for many different shift vectors).
6. If the detected area is larger than B, finish, else repeat Step 5 with the next maximum of r. Although, this method is simple and does not have a large computational complexity, it often fails to detect the forgery unless the size of the forged area is at least ¼ of linear image dimensions (according to our experiments).

Both the exhaustive search and the autocorrelation method were abandoned in favor of the third approach that worked significantly better and faster than previous approaches.

V. DETECTION OF COPY-MOVE FORGERY BY BLOCK MATCHING

Exact match

The first algorithm described in this section is for identifying those segments in the image that match exactly. Even though the applicability of this tool is limited, it may still be useful for forensic analysis. It also forms the basis of the robust match detailed in the next section.

In the beginning, the user specifies the minimal size of the segment that should be considered for match. Let us suppose that this segment is a square with $B \times B$ pixels. The square is slid by one pixel along the image from the upper left corner right and down to the lower right corner. For each position of the $B \times B$ block, the pixel values from the block are extracted by columns into a row of a two-dimensional array A with $B2$ columns and $(M–B+1)(N–B+1)$ rows. Each row corresponds to one position of the sliding block.

Two identical rows in the matrix A correspond to two identical $B \times B$ blocks. To identify the identical rows, the rows of the matrix A are lexicographically ordered (as $B \times B$ integer tuples).

This can be done in $MN\log2(MN)$ steps. The matching rows are easily searched by going through all MN rows of the ordered matrix A and looking for two consecutive rows that are identical.

The blocks form an irregular pattern that closely matches the copied-and-moved foliage. The fact that the blocks from several disconnected pieces instead of one connected segment indicates that the person who did the forgery has probably used a retouch tool on the pasted segment to cover the traces of the forgery. Note that if the forged image had been saved as JPEG, vast majority of identical blocks would have disappeared because the match would become only approximate and not exact (compare the detection results with the robust match in Figure 8). This also why the exact match analysis of images from Figures 2 and 4 did not show any exactly matching blocks. In the next section, the algorithm for the robust match is given and its performance evaluated.

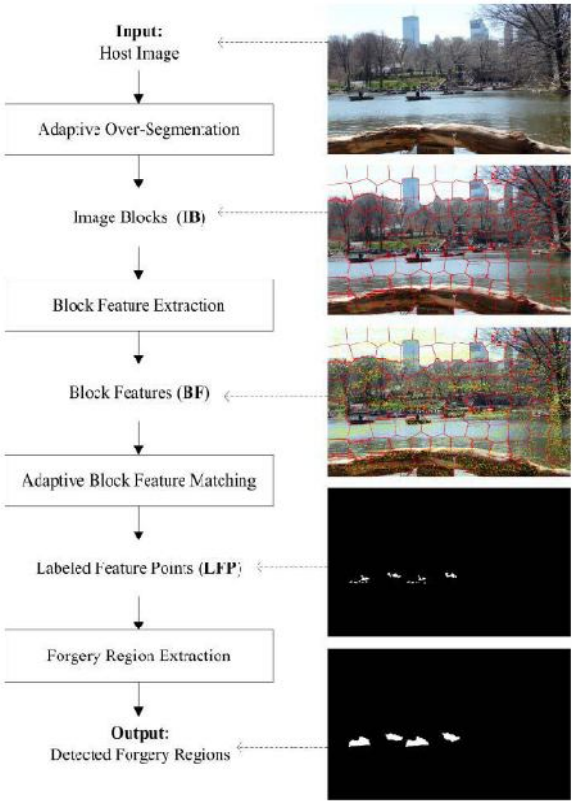

Fig.8.Working of Copy-Move Forgery Detection System.

CONCLUSION

With the increase in digital image forgery, the need of forgery detection algorithms has increased. The above mentioned algorithms help in detecting forged image regions from an image using Cut-Paste and Copy-Move Forgery Detection System.

REFERENCES

[1] J. Fridrich, “Methods for "Methods for Tamper Detection in Digital Images", Proc. ACM Workshop on Multimedia and Security, Orlando, FL, October 30−31, 1999, pp. 19−23.

[2] S. Saic, J. Flusser, B. Zitová, and J. Lukáš, “Methods for Detection of Additional Manipulations with Digital Images”, Research Report, Project RN19992001003 "Detection of Deliberate Changes in Digital Images", ÚTIAAV ČR, Prague, December 1999 (partially in Czech).

[3] J. Lukáš, "Digital Image Authentication", Workshop of Czech Technical University 2001, Prague, Czech Republic, February 2001.

[4] Chi Man Pun, Senior Member, IEEE, Xiao Chen Yuan, Member, IEEE, and XiuLiBi "Image Forgery Detection Using Adaptive Over segmentation and Feature Point Matching" IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY, VOL. 10, NO. 8, AUGUST 2015 1705.

★★★