**A**

**PROJECT REPORT**

**ON**

**Project Report Digital Image Forensics**

**IN PARTIAL FULFILLMENT FOR THE BACHELOR DEGREE COURSE**

**IN**

## COMPUTER ENGINEERING



**SUBMITTED BY**

**GIRAMKAR SHRUSHTI ANKUSH**

**BHOSALE SAIEE PRAMOD**

**SONAWANE PRATHAMESH NITIN**

**TUPE MANISHA NANA**

UNDER THE GUIDANCE OF

Dr. Suryawanshi.P.M

**DEPARTMENT OF COMPUTER ENGINEERING**

**HSBPVT’S PARIKRAMA COLLEGE OF ENGINEERING, KASHTI**

**HSBPVT’s GOI Faculty of Engineering, Kashti**

DEPARTMENT OF COMPUTER ENGINEERING, KASHTI



This is to certify that **Miss. Giramkar Shrushti Ankush** from **Fourth Year Computer Engineering** has successfully completed her seminar work titled **“Cyber Security Digital Forensic Mini Project”** HSBPVT College of Engineering, Kashti in the partialfulfilment of the Bachelor’s Degree in Engineering of Savitribai Phule Pune University.

Dr. Suryawanshi.P.M Proff. Sayyad. J.I Dr. Mahadik.S.N

**Guide Name HOD Principal**

**Abstract**

Determining the authenticity of an image is now an important area of research. In our work,we attempt to classify whether a digital image is a genuine image or is a manipulated version of some authentic image.We have applied three different algorithms discussed. with modifications of our own. We have then implemented two techniques discussed into detect the fake regions in the image. Our approach is invariant to the type of manipulation that has been done to the images. From our methods, we are able to classify nearly 80% manipulated images.

1. **Objective:**

In this project, I explored what is Digital Image Forensics (DIF), what are the type of those techniques. Then, I explored some techniques in the publications and write a small tool based on those techniques for DIF.

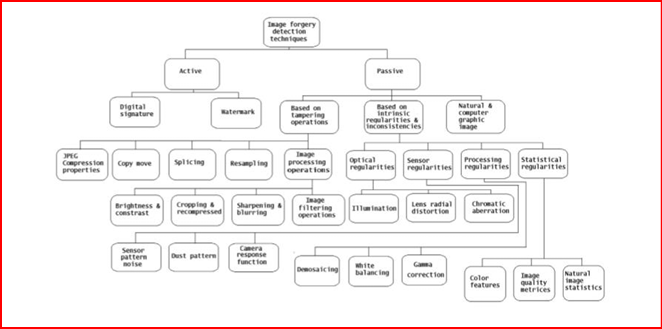
1. **Digital Image Forensics:**

Multimedia forensic aim at restoring some of the lost trustworthiness of digital media by developing tools to unveil conspicuous traces of previous manipulations, or to infer knowledge about the source device. Digital image forensics – a kind of multimedia forensics is a brand-new research field which aims at validating the authenticity of images by recovering information about their history. Two main problems are addressed: the identification of the imaging device that captured the image, and the detection of traces of forgeries. Digital Image Forensics is that branch of multimedia security that, together with Digital Watermarking, aims at contrasting and exposing malicious image manipulation. DIF aims at providing tools to support blind investigation. This brand-new discipline stems from existing multimedia security-related research domains (e.g. Watermarking and Steganography) and exploits image processing and analysis tools to recover information about the history of an image. Two principal research paths evolve under the name of Digital Image Forensics. The first one includes methods that attempt at answering question a), by performing some kind of ballistic analysis to identify the device that captured the image, or at least to determine which devices did not capture it. These methods will be collected in the following under the common name of image source device identification techniques. The second group of methods aims instead at exposing traces of semantic manipulation (i.e. forgeries) by studying inconsistencies in natural image statistics. We will refer to these methods as tampering detection techniques.

1. **Digital Image Forensics:**

Image forgery detection aims to verify the authenticity of a digital image. Image authentication solution is classified into two types. Active and Blind or passive. An active forgery detection techniques, such as digital watermarking or digital signatures uses a known authentication code embedded into the image content before the images are sent through an unreliable public channel. By verifying the presence of such authentication code authentication may be proved by comparing with the original inserted code. However, this method requires special hardware or software to insert the authentication code inside the image before the image is being distributed.

Passive or blind forgery detection technique uses the received image only for assessing its authenticity or integrity, without any signature or watermark of the original image from the sender. It is based on the assumption that although digital forgeries may leave no visual clues of having been tampered with, they may highly likely disturb the underlying statistics property or image consistency of a natural scene image which introduces new artifact resulting in various forms of inconsistencies. These inconsistencies can be used to detect the forgery. This technique is popular as it does not need any prior information about the image. Existing techniques identify various traces of tampering and detect them separately with localization of tampered region.



1. **Digital Image Forensics is not Computer Forensics:**

Multimedia Forensics in general and Digital Image Forensics in particular is not Computer Forensics. Even though both computer forensics and multimedia forensics explore digital evidence, we believe that they form two distinct sub-categories of digital forensics. This may seem counter-intuitive at first sight, since in any case, the domain of evidence is limited to the set of discrete symbols found on a particular device. In multimedia forensics, however, it is assumed that these discrete symbols were captured with some type of a sensor and therefore the symbols are a digital representation of an incognizable reality.

The existence of a sensor that transforms natural phenomena to discrete projections, which are then subject to investigation, implies that multimedia forensics has to be seen as empirical science. This resembles the epistemological argument brought forward in the context of steganography in digitized covers. Literally, a forensic investigator can never gain ultimate knowledge about whether a piece of digital media reflects reality or not. Neither can a sophisticated perpetrator be sure whether his manipulation really has not left any detectable traces. Unlike computer forensics, digital evidence in multimedia forensics is linked to the outside world and cannot be reproduced with machines. Thus, while the principle of transfer does not necessarily apply to computer forensics, it does have a place in multimedia forensics.

1. **Small tool for Digital Image Forensics:**

I gathered some techniques in some paper and composed the small tools that implement those methods in DIF. The tool does not tell exactly this image is forged/photoshoped/tampered or not (even if some tool can tell exactly forged or not, you can go directly to the top of the world). This tools just give some warning, some weird “information” of the image, visual some strange region of the image to the users. Then, based on that information the users can make the decisions. Nowadays, the tampered image is more difficult to detect. Many techniques that counter the detection. My tool implements some techniques are a little bit old, but still useful in some image. Some tampered images are easy to detect by these methods but hard to detect by the others. Therefore, we need to combine many techniques to analyze one images.

1. **Exposing digital forgeries by EXIF metadata:**

Every image has the header (EXIF) which contains many information about the image are source device identification, date-time, thumbnail, etc. Analyzing that information and checking the consistency is the one of the methods to check the image is modified or not. That technique is simple and look like ‘stupid’ but this is the strongest method. We know that, the image with the right header can tell that the image is right or wrong. But, the image with the wrong header, we can absolutely be sure that image is wrong.

In my tool, I extracted the EXIF information and then compare it:

• If the modified date is not the same with the original date: the image has been modified.

• The header contains the tag of some software edited image like Adobe Photoshop, etc.: the image has been modified by those software.

• The image’s header is stripped => that’s weird.

• The thumbnail and the image are not the same.

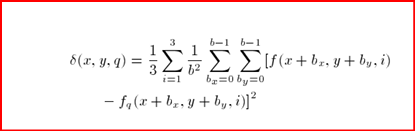
• The image has a strange resolution (this does not match with any camera resolution) => the image can be resized or crop.

• Etc. Gathers that information, the tool gives the WARNING for the users.

1. **Exposing digital forgeries from JPEG Ghost:**

When creating a digital forgery, it is often necessary to combine several images, for example, when compositing one person’s head onto another person’s body. If these images were originally of different JPEG compression quality, then the digital composite may contain a trace of the original compression qualities. Based on this idea, we need to compute the different map between 2 different quality images.

The difference image is first averaged across a bxb pixel region:

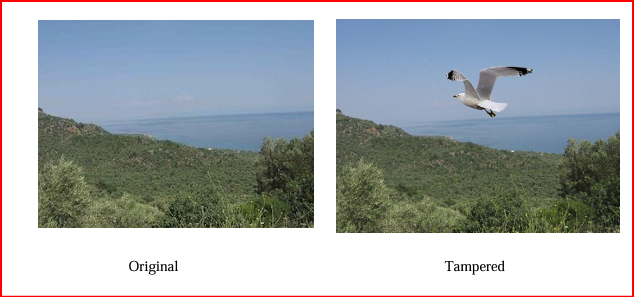


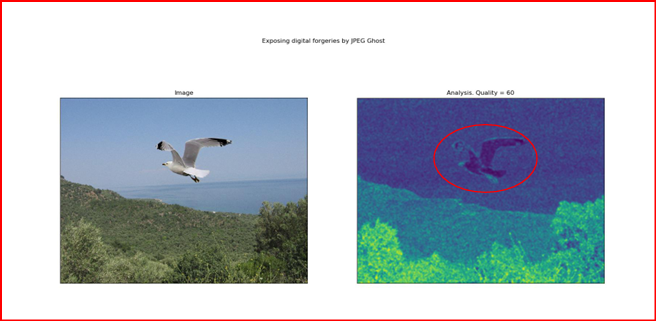
and then normalized so that the averaged difference at each location (x,y) is scaled into the range.

This is a simple and yet potentially powerful technique for detecting tampering in low-quality JPEG images. This approach explicitly detects whether part of an image was compressed at a lower quality than the saved JPEG quality of the entire image. Such a region is detected by simply resaving the image at a multitude of JPEG qualities and detecting spatially localized local minima in the difference between the image and its JPEG-compressed counterpart. Under many situations, these minima, called JPEG ghosts, are highly salient and easily detected.

For more detail of this method check it in the publication: Exposing digital forgeries from JPEG ghosts. Here is my actual result testing the implementation of this method on my image:

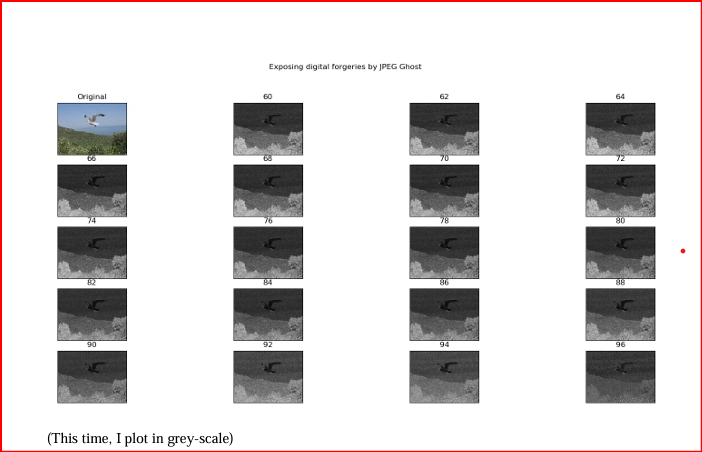
Original Image for all demo:





**Red: Tampered region:**

The tampered region in this method is the region with strong blue (or black if plot in greyscale). This happens when we choose the correct quality of the tampered region. Then, the different between tampered region and the resaved image with this quality is minimum => small value => dark colour. This method is very hard to choose the best quality, so I proposed the multiple map, that plot many smaller qualities with step of 2 from 60 of the resaved image, then we can see the different more easily.



1. **Exposing digital forgeries by noise inconsistencies:**

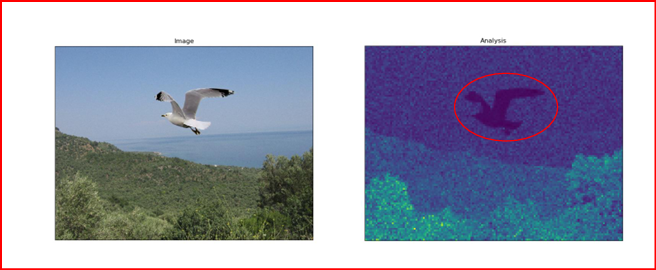
A commonly used tool to conceal the traces of tampering is the addition of locally random noise to the altered image regions. The noise degradation is the main cause of failure of many active or passive image forgery detection methods. Typically, the amount of noise is uniform across the entire authentic image. Adding locally random noise may cause inconsistencies in the image’s noise. Therefore, the detection of various noise levels in an image may signify tampering. This method is a segmentation method detecting changes in noise level. The main drawback of the method is that authentic images also can contain various isolated regions with totally different variances.

The proposed method is based on a few main steps:

• wavelet analysis,

• tiling sub-band HH1 with non-overlapping blocks,

• blocks noise variance estimation.



**Methods**

• Detection of digital image manipulation Three different classifiers were designed using feature vectors that exploited three distinct statistical properties of authentic images. Two of these were based on the method used by paper and the third was based on the method used by paper. The authors of had restricted their domain to classifying an image as photographic or photo realistic, but we found that the method works equally well with image manipulation .In the first classifier, the RGB image was filtered using a locally adaptive Weiner filter with a neighbourhood size of 5×5 to remove the high frequency noise. The image was separately filtered using a median filter with a neighbourhood size of 4 ×4 to remove the salt and pepper noise. For each of the two matrices of noise obtained, the mean, standard deviation and kurtosis were calculated. Hence, 6 ×3=18 feature vectors were obtained. Relevant Code: denois. m– In the second classifier, the RGB image was first converted to its grayscale equiv alent(Reason explained below). For each image, a single level discrete wavelet transform was performed to obtain the high frequency Hi, Di and Vi sub-bands. The distances of their distribution was calculated from the corresponding Gaussian Distribution. The Gaussian distribution was used as a benchmark in conformation with the Central Limit Theorem. These distances and the standard deviation of the sub-band coefficients were used as feature vectors. Hence, 6 feature vectors were obtained from each image.

The conversion to grayscale was done to prevent the feature vectors of the standard deviation from dominating the feature vectors of the distances. Authors of [1] had solved the problem by normalising the energy of each image, but we found that normalising tends to disturb the Laplacian dis tribution of the sub-band coefficients. Relevant Code: wavelet’s– In the third classifier, the RGB image was again converted to its grayscale equivalent (Details below). For each image, a multi-scale discrete wavelet transform was done to obtain the wavelet sub-band coefficients for the first three scales.

For the first two”parent” sub-bands, a neighbourhood prediction model was constructed for each coefficient as follows: Oi(x,y) = w1Oi(x − 1,y) + w2Oi(x,y − 1) + w3Oi(x,y + 1) + w4Oi(x + 1,y) + w5Oi+1(x/2,y/2) where Oi is the its scale diagonal, vertical or horizontal sub-band. Using the method of least squares, the residue matrix was obtained for each sub band for the first two scales. The standard deviation and kurtosis were used as feature vectors. Hence, a total of 6 × 2 = 12 features vectors was obtained. The conversion to grayscale was done to reduce the computational complexity in calculating the residues. Relevant Code: neigh.

**• Detection of manipulated regions:**

Two methods were employed for manipulated region detection. The first was based on the idea used in but with a different set of features, and the second used the algorithm described.

– In the first method, a wavelet based de-noising filter was applied on each colour channel of the RGB image. The standard deviation for each noise matrix was obtained. Moreover, a discrete wavelet transform was performed for each block and the standard deviation of the coefficients was obtained. A k-means classifier was used to obtain the suspicious regions based on these features. Then the cluster with the minimum number of blocks was chosen as the cluster containing the susupicious blocks. To reduce false positives, the value of k was different for different images.Moreover,clusters with size less than a minimum threshold were discarded since these could have been outliers. Relevant Code: patchy.m

– In the second method, we divide the image into non-overlapping blocks and then we lexicographically sort them. After sorting, we take lexicographically ”close” blocks and compute the distance between their positions. If the distance is less than a minimum threshold, then we mark both the blocks as manipulated blocks. Relevant Code: tro4.m

Results and discussions:

• Digital Image manipulation detection We used 61 digitally manipulated images and 71 authentic images obtained from the Dresden Image Database for training the classifier. The classifier used was a Support Vector Machine with a radial basis function as the kernel. The test dataset consisted of 61 authentic images obtained from www.dpreview.com and 59 manipulated images from various sources on the internet. The following results were obtained -

– Classifier from image de-noising

True positive: 50/59

False positive: 19/61

– Classifier from wavelet statistical properties

True positive: 48/59

False positive: 18/61

– Classifier from wavelet neighbourhood prediction

True positive: 44/59

False positive:13/61

– Voting from the three classifiers

True positive: 46/59

False positive: 13/61 5

We did not have any pre-assumption on the kind of manipulation that has been done on the image. The original paper we worked upon has assumptions about the kind of manipulation. Their result was detection of 90% true positive images and 5% false positive images. Their results seem to be better than our results because of the pre-assumption.

**• Fake region detection:**

We took a random sample of 30 images from the test and the training data-set which were digitally manipulated. Barring 3 images, the algorithm using the de-noising technique was able to detect the manipulated region in the images. However, blocks having high similarity with the tampered blocks were wrongly classified. This points at the need to improve the algorithm by using a better de-noising filter than the wavelet de-noising filter as this causes some amount of loss of information in the original image. This seems to blur the thin line of divide between manipulated regions and other regions. We did not proceed with the experiment on other images because of this handicap. The experiment was also conducted using the method as described in. In this method, we mostly obtained bad results. The reasons for such results are:

– The image need not be a of a size multiple of the block size, so we extended the incomplete blocks by filling zeros in them which affected the sorting and hence, the overall results:

Reasons why method is not a appropriate:

∗ It does a lexicographically sorting but this method is not useful if the manipulated region is copy pasted from some other image.

∗ The copy-pasted regions must completely fit into the blocks and the corresponding rows and columns should be same for this method to work.

∗ A simple rotation of the copied region while pasting in the image will render this method useless.



**Exposing digital forgeries based on demosaicing artifacts:**

This technique is a tamper detection technique based on artifacts created by Color Filter Array (CFA) processing in most digital cameras. The techniques are based on computing a single feature and a simple threshold-based classifier. My tool implements the first of two approaches in this paper. For more detail of this method check it in the publication: Image tamper detection based on demosaicing artifact. [8] Here is my actual result testing the implementation of this method on my image:



**Conclusion:**

Digital Image Forensics has become increasingly vital as digital images are widely used across personal, professional, and legal contexts. With the rapid advancements in digital editing tools, it has become easier to manipulate images, posing challenges for authenticity and credibility. Digital Image Forensics provides methods to detect tampering, verify authenticity, and establish the origins of digital images.

The field leverages techniques like metadata analysis, pixel-level examination, noise pattern analysis, and machine learning to detect forgery, splicing, and cloning, among other forms of manipulation. It also plays a key role in cybersecurity, intellectual property protection, and law enforcement, as it can uncover the authenticity of visual evidence and even aid in identifying sources of manipulated content.

**References:**

1. Hongmei Gou; Swaminathan, A.; Min Wu; ”Noise Features for Image Tampering Detection and Steganalysis,” Image Processing, 2007. ICIP 2007. IEEE International Conference on, vol.6, no., pp.VI-97-VI-100, Sept. 16 2007-Oct. 19 2007

2. S. Lyu and H. Farid,” How realistic is photorealistic?” IEEE Trans. Signal Processing, vol. 53, no. 2, pp. 845-850, 2005

3. Lu, W., et al. 2008. Digital image forensics using statistical features and neural network classifier, Proc. The 7th International Conference on Machine Learning and Cybernetics, pp. 2831–2834, July 2008.

4. Gloe, T., Bh¨ome, R. (2010). ” The Dresden Image Database for benchmarking digital image forensics”. In Proceedings of the 25th Symposium on Applied Computing (ACM SAC 2010) (Vol. 2, pp. 1585-1591).