CS 365 Project Report
Digital Image Forensics

Abhijit Sharang (10007)
Pankaj Jindal (Y9399)
Advisor: Prof. Amitabha Mukherjee

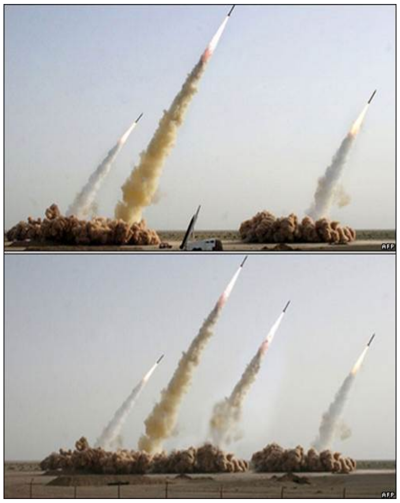
Abstract

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 in [1] and [2] with modifications of our own. We have then implemented two techniques discussed in [5] and [6] to 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 Introduction

By looking at the two images of Monalisa, the image on the left [1] and image on the right [2] by naked eyes, one can deduce that image on the right is fake but the same is not an easy task for image 2 [3].With the advancement in the technology in the field of computer graphics, it is very easy to manipulate digital images that are impossible to differentiate from authentic photographic images. This undermines the use of images as a credible source of some event .Moreover, digitally manipulated images can trigger off major controversies. Hence, there is a need for more sophisticated and mathematically sound techniques that can perform this task of classification of the image into real ones and digitally manipulated ones.

Presently, there are methods that can ensure the authenticity of an image like by inserting watermarks in it. But, such methods are not always feasible. Hence passive techniques for classifying the images are more popular and desired. One such passive technique exploits the fact that all camera clicked real images share a set of statistical features that are different from the statistical features that are possessed by the images that are manipulated by any software[7]. We exploit these features to classify the images and then detect the region of tampering in the manipulated images.Much of our work is based on the ideas presented in [1],[2],[5] and [6].



[1] source:wikipedia.org

[2] source:weheartit.com

[3] source:theiranianhome.com

2 Preliminaries

In this section, we discuss the application of wavelets in Digital Image Forensics.

The discrete wavelet transform of an image results in four sub-bands of varying orientation, namely the Diagonal sub-band(D_i), the horizontal sub-band(H_i), the Vertical Sub-band(V_i) and the low frequency sub-band which contains the approximation coefficients for decomposition to scale i+1.This transform is achieved by passing the image through a low pass filter and a high pass filter, which are orthogonal to each other.

The three high frequency sub-bands exhibit a statistical regularity which can be exploited in Image Forensics. Their distribution for clean images is non-Gaussian[1]. More specifically, the distribution is Laplacian with sharp peak at zero and large symmetrical tail[2]. Moreover it has been observed that a strong correlation exists between the sub-bands at scale i and scale i+1.[2]Hence, if there is a large coefficient for a sub-band at scale i, it is highly likely that the coefficient is also high at scale i+1 [3].

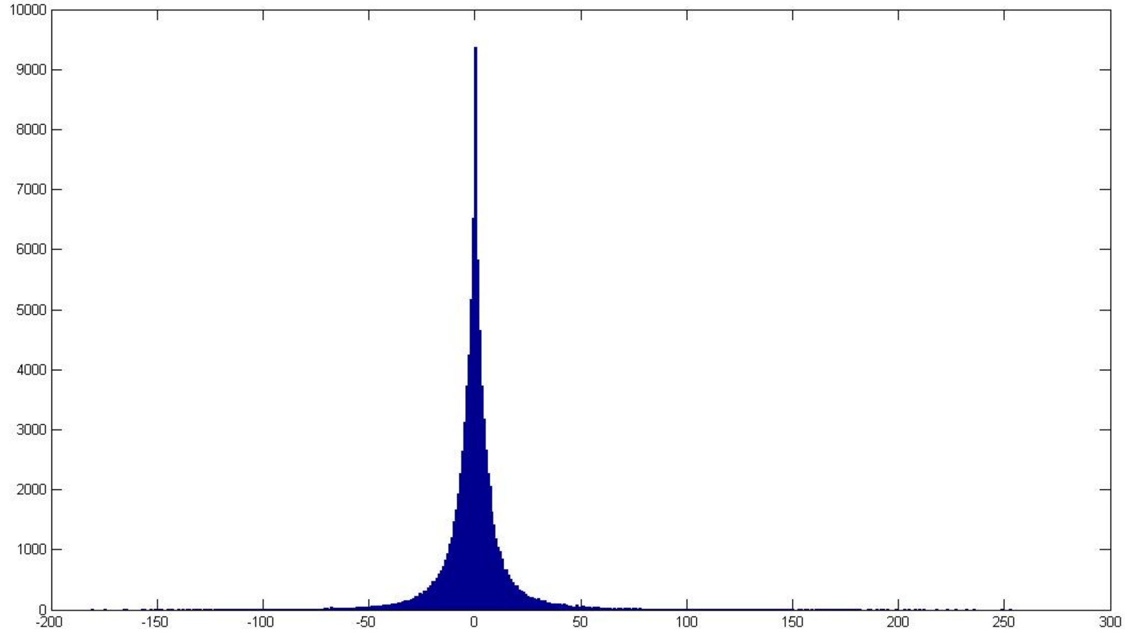


Fig: A distribution of Diagonal Sub-band

These statistical properties hold good for natural images which have not undergone any tampering. When a tampering operation is done on the image,it is likely that the wavelet properties are modified to conceal the tampering.

3 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[1] and the third was based on the method used by paper[2]. The authors of [2] had restricted their domain to classifying an image as photographic or photorealistic, 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, $6 \times 3 = 18$ feature vectors were obtained.
    Relevant Code: denois.m
  - In the second classifier, the RGB image was first converted to its gray-scale equivalent(Reason explained below). For each image, a single level discrete wavelet transform was performed to obtain the high frequency H_i, D_i and V_i 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 gray-scale 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 distribution of the sub-band coefficients.
    Relevant Code: wavelet.m
  - In the third classifier, the RGB image was again converted to its gray-scale 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:
    $$O_i(x, y) = w_1 O_i(x-1, y) + w_2 O_i(x, y-1) + w_3 O_i(x, y+1) + w_4 O_i(x+1, y) + w_5 O_{i+1}(x/2, y/2)$$
    where O_i is the i_{th} 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 $6 \times 2 = 12$ features vectors were obtained.
    The conversion to gray-scale was done to reduce the computational complexity in calculating the residues.
    Relevant Code: neigh.m

- Detection of manipulated regions

  Two methods were employed for manipulated region detection. The first was based on the idea used in [6] but with a different set of features, and the second used the algorithm described in [5].

  - 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

4 Results and discussions

- Digital Image manipulation detection

  We used 61 digitally manipulated images and 71 authentic images obtained from the Dresden Image Database[4] 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

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 [5]. 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.

[4]

[5]

[6]

[7]

[8]

[9]

Some manipulated images correctly classified by all 3 classifiers

[4] Source: Wikipedia

[5] Source: Dresden Image Database

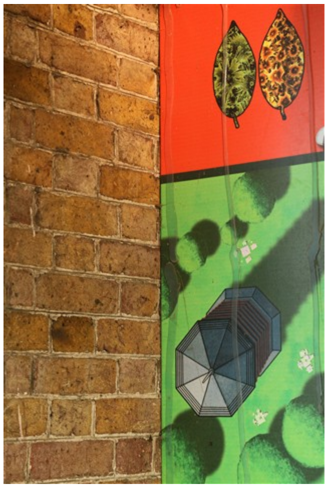
[6] Source: Wikipedia

[7] Source: Own Image

[8] Source: Wikipedia

[9] Source: Wikipedia

[10]


[11]

[12]

Some authentic images wrongly classified as manipulated images

[13]

[14]

Some manipulated images classified as authentic

[10] Source: www.dpreview.com

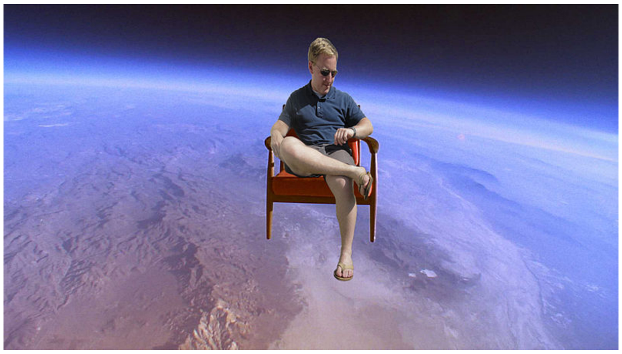
[11] Source: www.dpreview.com

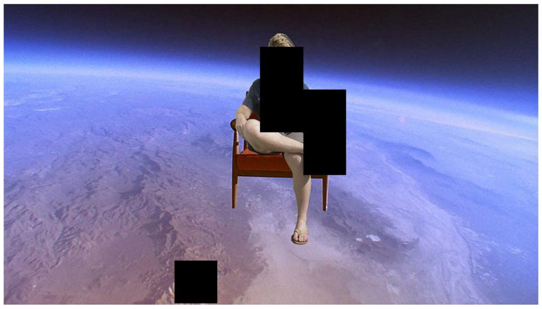
[12] Source: www.dpreview.com

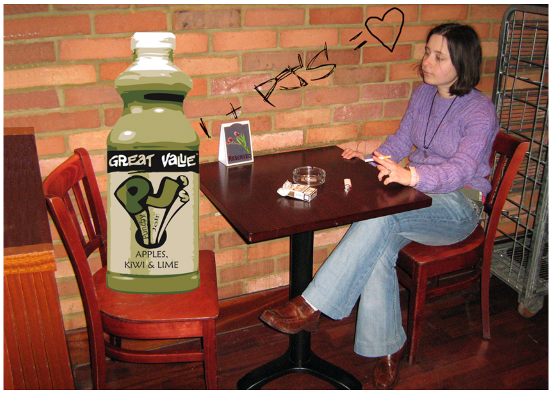
[13] Source: www.existingvisual.com

[14] Source: www.bollywoodhungama.com

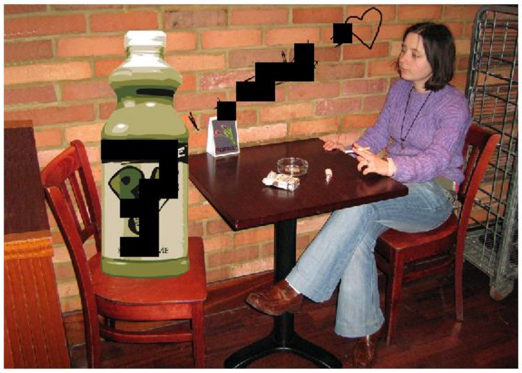
Some results by using our method for detection of manipulated regions

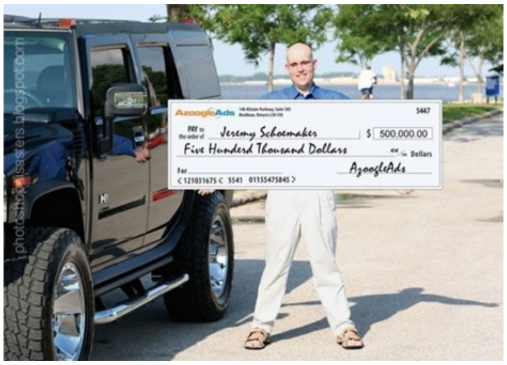




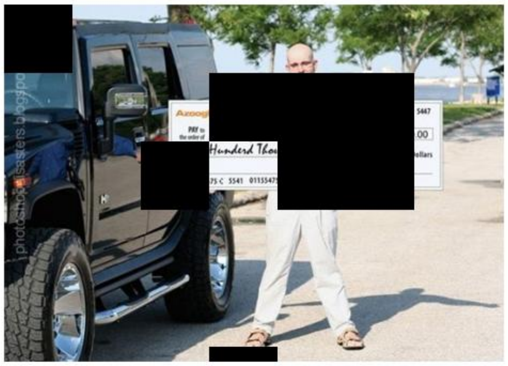


[15]

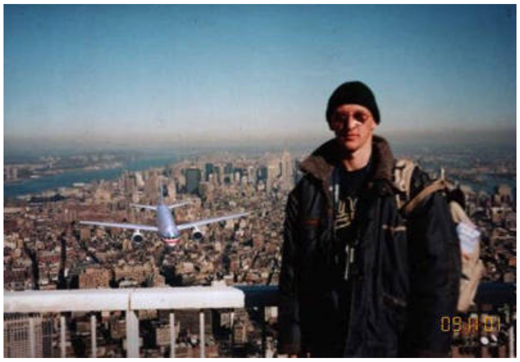


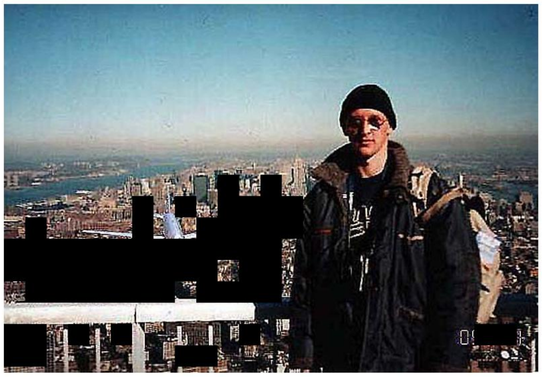

AzoogleAds
Jeremy Schoemaker
$ 500,000.00
Five Hundred Thousand Dollars
AzoogleAds
photoshopdisasters.blogspot.com

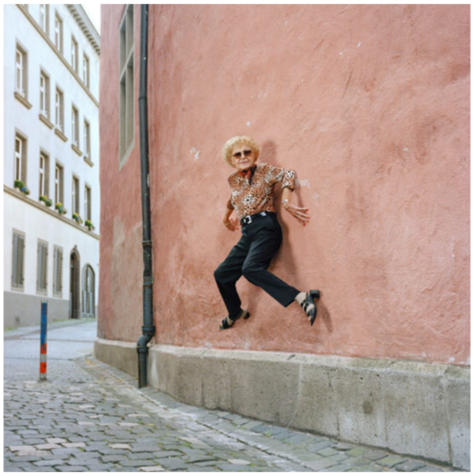
[16]

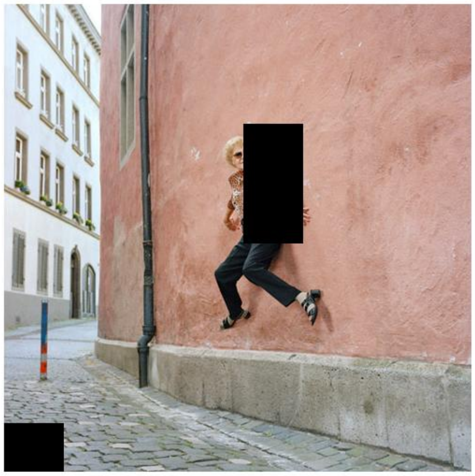


[15] Source: annanedelcheva.wordpress.com
[16] Source: www.umaxforum.com









[17]

[17] Source: www.ivomayr.com

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. SignalProcessing,* 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., Bhöme, 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).

5. Xunyu Pan, Xing Zhang, Siwei Lyu, "Exposing image forgery with blind noise estimation". *Proceedings of the thirteenth ACM multimedia workshop on Multimedia and security*, 2011

6. A. C. Popescu and H. Farid, *Exposing Digital Forgeries by Detecting Duplicated Image Regions*, 2004 :Dartmouth College

7. S.Dehnie,H.T. Sencar,and N.Memon,""Digital image forensics for identifying computer generated and digital camera images" in Proc. *IEEE Int. Conf.Image Processing (ICIP)*, Atlanta, GA, 2006, pp. 2313-2316.