Department of Electrical Engineering, IIT Jodhpur
BTP Report
August 2020

Project Title	**Image Tampering Detection**
Name of the Student/s and Roll Numbers	**Akash Jain (B18EE002)** **Ashish Ledalla (B18EE008)**
Name of the Supervisor	**Dr. Rajlaxmi Chouhan**

A. Abstract

Image splicing is a common image tampering method, where a particular area of an image is introduced into some other image with the objective of manipulating its content. Most existing image tampering detection methods can be grouped into statistical and physics-based approaches. Statistical methods can oftentimes be fully automated and achieve impressive results. Physics-based techniques try to detect image inconsistencies with the help of analytic models, and hence are more robust to general image processing operations like recompression or resizing. In this project, we have studied and implemented two statistical methods and a physics-based method for image splicing detection. These techniques are based on different features of digital images.

B. Motivation

In a media environment saturated with deceiving news, the threat of fake and altered images in our lives has become increasingly apparent. Further, today‟s smartphones are capable of digitally manipulating even ordinary photographs with minimal effort. With keeping in mind the dangers that have occurred in the past, we see great value in a tool capable of identifying fake images and reporting to users the nature of an image‟s alterations.
Image Tampering is a special type of image forgery that alters a part or multiple parts of the graphic content of a given image. Image Tampering can be classified into three categories[1]:

1. Copy-Move
2. Cut-Paste
3. Erase-Fill (Exemplar-based and Diffusion-based)

We will concentrate on the identification of the Cut-Paste type of Image Tampering.
Cut-Paste type of image tampering is a simple and typical image tampering operation, where a particular region from an image is pasted into another image with the aim to change its content. It is also called Image Splicing.

C. Methodology

To identify Cut-Paste type of image forgeries, there are different strategies from which three most effective methods are:

1. Double JPEG Compression detection.

2. Noise Variance Inconsistency detection.
3. Gradient-Based Illumination Description

There are various algorithms[2] to implement these methods. Among those, we intend to implement methods with high accuracy and ease of implementation.

C1. *Double JPEG Compression Detection*[2,3]

Due to the high compression ratio and good quality, the JPEG image format has been widely used in cameras and image processing software. Double JPEG compression is the result of decompressing a JPEG image to the spatial domain and then resaving using a different (secondary) quantization matrix. We consider that double JPEG compression occurs while tampering JPEG images, then it is of great significance to us in detecting image forgeries.

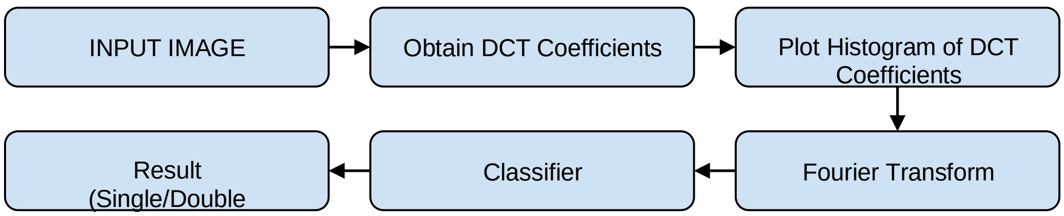

Figure 1

Algorithm : (Figure 1)

1. The input image in RGB format is first converted to YCrCb format.
2. Then the image is divided into blocks of 8 x 8 pixels.
3. DCT coefficients of each block are calculated at low frequencies.
4. The Zero Mean histogram corresponding to the Discrete Cosine Transform(DCT) coefficients is obtained.
5. Magnitudes of Fast Fourier Transform(FFT) of the DCT histogram are obtained.

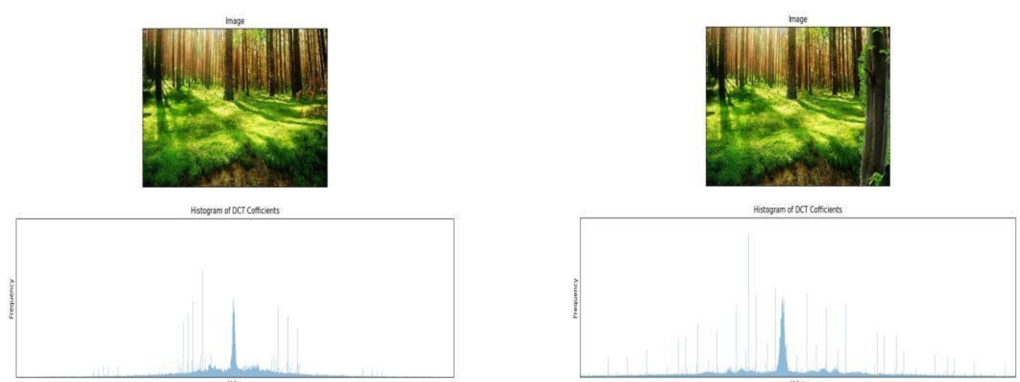

Figure 2

Figure 2 (Left) shows a single compressed JPEG image along with its zero-mean histogram of DCT coefficients which is continuous as expected. Whereas Figure 2 (Right) shows the tampered image with its zero-mean histogram of DCT coefficients which gives an indication that the image has tampered.

6. Based on the peaks (above a threshold say 0.5) of magnitudes of FFT obtained we use a classifier to classify the image as Single Compressed JPEG or Double Compressed JPEG. (an example is shown in Figure 2)

Classification

First Quantization Factor	Number of Peaks in FFT of Zero Mean Histogram DCT Coefficients.	
	Single Compressed Images	Double Compressed Images
LOW (< 50)	Less than equal to 19	Greater than equal to 21
MEDIUM	Between 11 and 19	Between 9 and 17
HIGH (>= 75)	Less than equal to 11	Less than equal to 11

C2. *Noise Variance Inconsistency*[2,5]

Typically, the amount of noise in an authentic image is uniform across the entire image. Inconsistencies in the image‟s noise are the result of adding random noise locally. Therefore, the detection of various noise levels in an image may signal the tampering of the image.

In natural images that are not tampered, the noise variances in different areas generally vary only slightly. In the Cut-Paste type of image tampering as there is a portion of the image which belongs originally to another image, there is significantly different intrinsic noise variance, the inconsistency of local noise variances becomes telltale evidence of tampering. Noise Variance Inconsistency method also helps to identify the Diffusion-based Erase-Fill type of Image tampering.

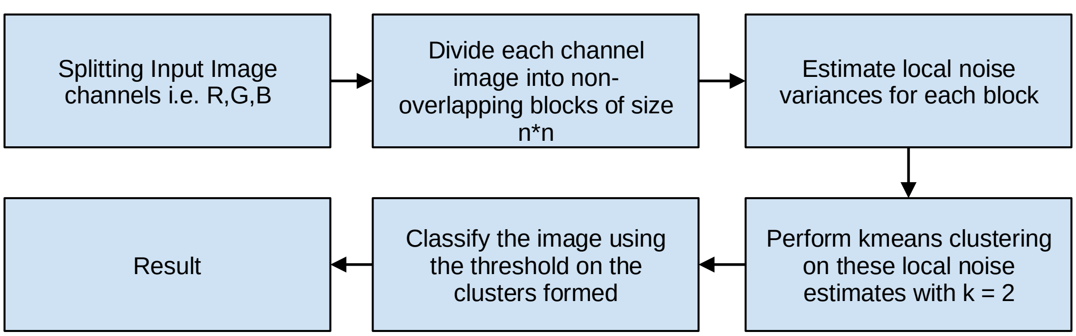

Figure 3

Algorithm: (Figure 3)

1. The input image is taken in the BGR color space.
2. Then decompose the image‟s each channel and get three single-channel images, one for each channel corresponding to B, G, R.

3. Each of the single-channel images is processed individually.
4. Each imaging channel is divided into non-overlapping blocks of size n*n. (We have done analysis on 16*16 and 32*32)
5. For each of the blocks, the noise is estimated.
6. All the estimates of noise per block are concatenated into a single vector to represent the local noise variance of the whole image.
7. We compute the mean of the corresponding values for each channel image to get a single noise variance vector for the input image.
8. Perform the K-means clustering on these estimated local noises variances with k = 2.
9. Using the clusters we choose the threshold that the spliced image has one cluster with fewer points with the varying or different centroid. But for the authentic image, the variances must follow a defined distribution. Hence, both the clusters might have an almost equal number of points. (one cluster might not have very less number of points)

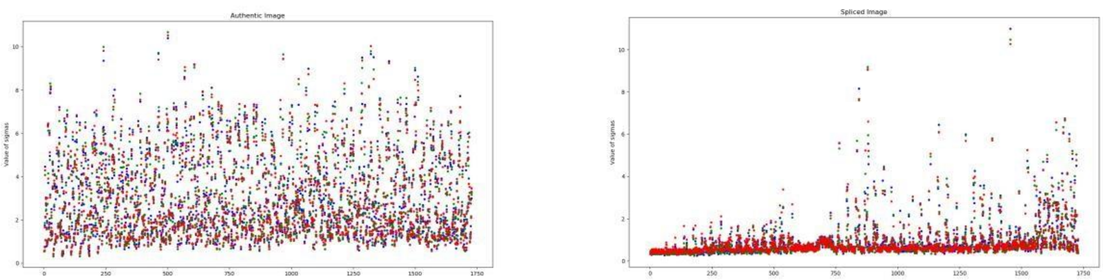

Figure 4
The left image shows the plot of estimated noise variance for an authentic image, where the whole local noise follows some distribution and the two clusters formed will have equal no of points, whereas the right image shows the local noise variances for the spliced image and the plot shows there is no such distribution and the two clusters formed will be highly imbalanced.

10. Thresholds (t) have been chosen using the grid search, that number of points in small clusters is less than t% of the total number of points indicates that the image is spliced.
11. Using the computed threshold, we show the output of the classification result on a dataset[6].

C3. *Gradient-Based Illumination Description*[7]

The above two methods for image splicing detection are statistical. Physics-based techniques are generally more adaptable to downsampling and recompression. These techniques characterise the image by searching for lighting based inconsistencies in images to expose forgeries, for example in the direction of incident light and the resulting shadows. The illumination information is used as a physics-based cue to detect spliced images.

These techniques analise intensity distributions of whole image areas and are thus are less affected by downsampling and compression. We tried to implement a physics-based method which analyses the inconsistency of incident light on pairs of objects in the 2-D plane of the image. This method works on intensity gradients on the surfaces of objects. The sum of all the intensity gradients on spherical objects points to the direction of the light source. By means of observation, this condition can be applied to non-spherical objects. For this implementation, we particularly worked on very practical scenarios with people.

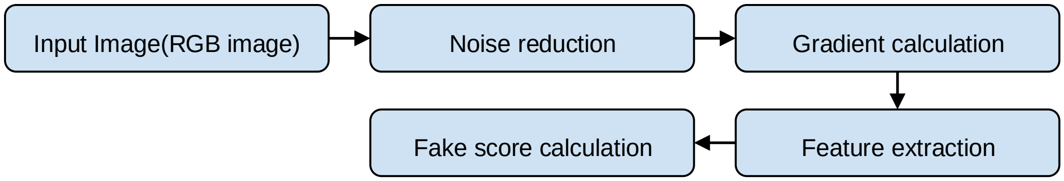

Figure 5

Algorithm: (Figure 5)

1. An RGB input image is taken.
2. Using a bilateral filter on the RGB input image we try to reduce the effect of noise.
3. We convert the image grayscale by taking the mean of the RGB color channels.
4. The gradients are obtained in vertical and horizontal directions separately using corresponding Sobel operators.

$$\mathrm{d}I(x,y) = (\mathrm{d}I_x(x,y), \mathrm{d}I_y(x,y))$$

5. We filter the gradients that are above a threshold „t" which is defined as the sum of mean and standard deviation of the gradients.
6. We obtain the mean of the gradient vectors after normalization as follows, this will be the dominant lighting direction

$$\overline{dI} = \left(\frac{1}{N} \sum_{x,y} d\hat{I}_x(x,y), \frac{1}{N} \sum_{x,y} d\hat{I}_y(x,y) \right)^{\mathrm{T}}$$

$$d\hat{I}(x,y) = \left(\frac{dI_x(x,y)}{\|dI(x,y)\|}, \frac{dI_y(x,y)}{\|dI(x,y)\|} \right)^{\mathrm{T}}$$

7. Using step 6, the approximate incident lighting directions, a and b of two segmented images A and B are obtained. Cosine dissimilarity LD, is used to differentiate between them as follows.

$$L_D(\boldsymbol{a}, \boldsymbol{b}) = 1 - \frac{(c(\boldsymbol{a}, \boldsymbol{b}) + 1.0)}{2}$$

$$c(\boldsymbol{a}, \boldsymbol{b}) = \frac{\boldsymbol{a}^{\mathrm{T}} \cdot \boldsymbol{b}}{\|\boldsymbol{a}\| \cdot \|\boldsymbol{b}\|}$$

8. The gradient field is then separated as four quadrants taking the centroid of each segmented object as origin and then divergence is computed at that origin.
9. We take the absolute difference of the divergences calculated for objects A and B as follows.

$$D_D(A, B) = |\mathrm{div}A - \mathrm{div}B|$$

10. The quadrants which we obtained, we use them to calculate an additional feature using the cosine dissimilarity of each quadrant as follows, with K = 4, as there are 4 quadrants.

$$T_D(A,B) = \frac{1}{K}\sum_i L_D(A_i, B_i)$$

11. For the fourth feature, we take an angular resolution of 5 degrees and discretize the gradient vectors into 72 bins. Now a histogram is obtained and then normalized.
12. The histograms of two objects A and B, h(A) and h(B) are taken and we compute ***H(A,B)*** via zero-normalized cross-correlation which serves as a comparison.

$$H_D(\boldsymbol{h}(A), \boldsymbol{h}(B)) = \frac{(\boldsymbol{h}(A) - \bar{h}(A))^{\mathrm{T}}(\boldsymbol{h}(B) - \bar{h}(B))}{\|\boldsymbol{h}(A) - \bar{h}(A)\| \cdot \|\boldsymbol{h}(B) - \bar{h}(B)\|}$$

13. We train a logistic regression model on a dataset of these features to calculate the fake score.
14. Since we have a classification problem, we visualize the performance of the classifier using AUC (Area Under The Curve) ROC (Receiver Operating Characteristics) curve.
15. ROC is a probability curve and AUC represents degree or measure of separability. It shows the capability of a model to distinguish between classes, in this case, if the image is tampered or not.

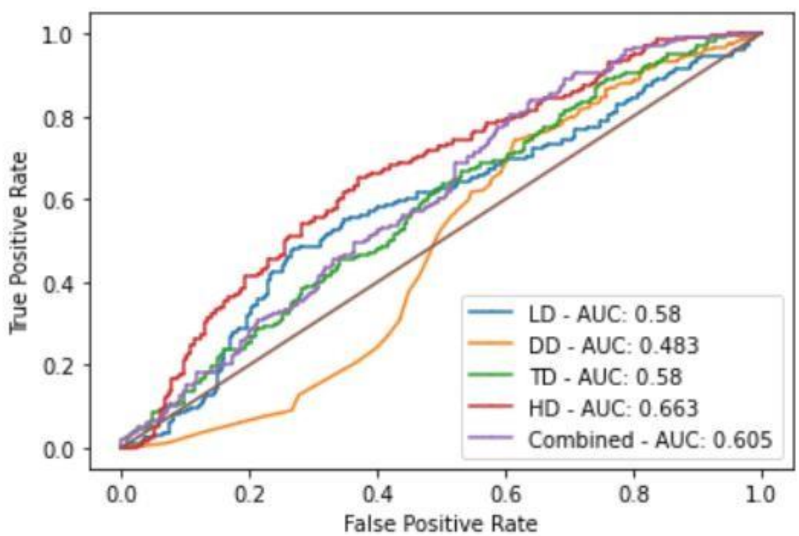

FIgure 6

ROC curves of individual features and combined feature set on Dataset[8]

D. Results and Discussion

In Double JPEG Compression detection, the test dataset is prepared by randomly picking images from UCID[4] and single quantization of those images with different compression ratios and then double compressing those images with different compression ratios.

Results from Double JPEG Compression Detection

Threshold (Tampered for number of peaks greater than =)	Accuracy	False Positive
11	0.876	0.55
13	0.77	0.45
15	0.66	0.33

For a data set of images, the upper bound of the set of the number of peaks in the FFT of the single compressed images overlaps with the lower bound of the set of the number of peaks in the FFT of the double compressed images. So it is difficult to distinguish between single compressed and double compressed images using a general classifier. Hence, there will be a trade-off between Accuracy and False Positives. We achieve high accuracy (more than 90%) when the first compression ratio is less and negligible accuray (around 1%) when the first compression ratio is high.

In Noise Variance Inconsistency detection, it may happen that the noise pattern of an image is concentrated at a small part of the image, which gives us false positives. Hence, there will be a trade-off between Accuracy and False Positives. This method has high accuracy for detecting the tampered regions in a tampered image. We have tested our algorithm on Columbia uncompressed image splicing detection evaluation dataset [6].

Results from Noise Variance Inconsistency Detection

Threshold	Accuracy	False Positive
15%	0.688	0.344
20%	0.811	0.448
25%	0.877	0.535

The statistical techniques that we used (Double JPEG Compression Detection and Noise Variance Inconsistency Detection) come under hypothesis testing scenarios where our main aim is to detect tampered images. Here, we have two types of errors. Type-1 error occurs when we claim a tampered image to be an untampered image and type-2 error occurs when we claim an untampered image to be a tampered image i.e. false positives. According to hypothesis testing, type-1 error is more dangerous and we need to minimize it. So, we try to choose thresholds with high accuracies and descent false positive errors.

In Gradient-Based Illumination Description, a logistic regression model is trained using 7000 segmented person pairs as training data, from the COCO 2017[8] dataset. For evaluating our model, we keep aside 500 pairs of images and use it as validation data. We also train the model separately with each of the four features, and also with all the four features taken together. If the image is spliced, then we assume it shows differences in the lighting conditions, whereas a natural image displays almost similar lighting conditions. But this assumption does not always hold.

AUC scores on Dataset[8]

Feature	AUC-Score
LD	0.580
DD	0.483
TD	0.580
HD	0.663
Combined	0.605

E. Conclusion

Detection of double JPEG compression in images plays a major role in crime detection and image forensics. The detection method used is based on histograms of DCT coefficients. We can reduce false positives and increase accuracy by using SVM (Support Vector Machine)[3].

In Noise Variance Inconsistency Detection, we consider that the spliced region and the original image have different intrinsic noise variances. Therefore, wherever their difference in noise variances is not significant, our method may fail to detect whether it is a spliced image. This method is mainly used for detecting spliced regions in an image. Better results can be obtained by using CNN to classify the images.

The physics-based method, which we implemented, is very robust for comparing 2D lighting environments. This technique is developed using the four features that are obtained using the gradients vector fields of segmented objects. Though the theoretical conditions are not strictly met, these gradient distributions of object surfaces are still very informative. Using a logistic regression model on the obtained four features we are able to differentiate if the objects are under the same lighting conditions or not and hence the image is spliced or not, this is tested and validated on the dataset extracted from the COCO dataset and we were able to achieve an AUC score of 0.605. One of the major strengths of this technique is that it is highly robust to JPEG compression or downsampling, which are generally applied after tampering.

References

[1] L. Zheng, Y. Zhang, and V. L. Thing, “A survey on image tampering and its detection in real-world photos,”Journal of Visual Communication and Image Representation, vol. 58, pp. 380 – 399, 2019. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S104732031830350X

[2] A. Levandoski and J. Lobo, “Image forgery detection : Developing a holistic detection tool,” 2019.

[3] B. Mahdian and S. Saic, "Detecting double compressed JPEG images," 3rd International Conference on Imaging for Crime Detection and Prevention (ICDP 2009), London, 2009, pp. 16, doi: 10.1049/ic.2009.0240.

[4] G. Schaefer and M. Stich, “UCID: an uncompressed color image database” in Storage and Retrieval Methods and Applications for Multimedia 2004, M. M. Yeung, R. W. Lienhart, and C.-S. Li, Eds., vol. 5307, International Society for Optics and Photonics.SPIE, 2003, pp. 472 – 480. [Online]. Available: https://doi.org/10.1117/12.525375

[5] X. Pan, X. Zhang and S. Lyu, "Exposing image splicing with inconsistent local noise variances," 2012 IEEE International Conference on Computational Photography (ICCP), Seattle, WA, 2012, pp. 1-10, doi: 10.1109/ICCPhot.2012.6215223.

[6] Y.-F. Hsu and S.-F. Chang, “Detecting image splicing using geometry invariants and camera characteristics consistency,” inInternational Conference on Multimedia and Expo, 2006.

[7] F. Matern, C. Riess and M. Stamminger, "Gradient-Based Illumination Description for Image Forgery Detection," in IEEE Transactions on Information Forensics and Security, vol. 15, pp. 1303-1317, 2020, doi: 10.1109/TIFS.2019.2935913.

[8] “COCO - Common Object in Context.” [Online]. Available: cocodataset.org

Signature of Supervisor

PChouhan
07/Aug/2020
Dr. Rajlaxmi Chouhan

Signature of Students

Akash Jain
Akash Jain (B18EE002)

Ashish L
Ashish Ledalla (B18EE008)