Forgery Detection in Digital Images using Statistical Analysis

Thariq Shanavas, Theja Voora, Bandaru Sai Varun

***Abstract*—The rise of digital image processing techniques has led to new challenges in establishing the authenticity of digital images. We attempt to tackle one of the most commonly reported image manipulation techniques: copy-move forgery, wherein one part of the image is copied and pasted within itself to hide people, vehicles or other entities. We propose to use statistical properties of the image to highlight the parts of the image which has been possibly altered.**

***Index Terms*—Image Forgery, statistical moments, Forgery detection**

I. Introduction

The authenticity of a digital image can be used as an evidence or as an important resource of information in image processing. With the advent of various powerful editing software as well as sophisticated digital cameras, digital image forgery has become more and more popular in image manipulation, where a part of the image is copied and pasted on another part of the same image to hide unwanted information.
Digital watermarking has been proposed as a means by which an image can be authenticated. However, this requires that the images be taken with specially equipped digital cameras. Since this is not viable in many cases, we rely on statistical methods to detect forgery.

A. Proposed Method

The images are first converted to grayscale for ease of processing, albeit some loss of information for preliminary analysis. We propose to divide the image into smaller sub-parts, each centered around a pixel and suitably indexed. There will be a sub-image centered around all suspicious pixels, and hence there could be overlap between sub-images. The sub images will be $p * p$ pixel squares, where p is a suitable threshold. Let us call them cells.
The cells were fourier transformed. We then extract statistical features to form an $4 * 1$ matrix corresponding to each cell, with the four features being mean, variance, skewness and kurtosis.
If there were m cells in the suspicious region, we now form an $m * 4$ matrix with each row consisting of the extracted features of the corresponding cell. The rows are now lexicographically sorted. We now compare the statistical features in nearby rows, and if they are sufficiently similar (i.e. difference in terms are all less than a certain threshold), we claim that there is a 'link' between the two corresponding cells, and hence corresponding pixels in the original image are correlated. If a large number of adjacent cells link to another large number of adjacent cells elsewhere in the image, we can safely ascertain that the image has been manipulated digitally.
The features we have used, for an image of size $M * N$, cell size p and centered at (x, y)-

$$M = \frac{1}{p^2} \sum_{i=-(p-1)/2}^{(p-1)/2} \sum_{j=-(p-1)/2}^{(p-1)/2} I(x+i, y+j)$$

$$Var = \frac{1}{p^2} \sum_{i=-(p-1)/2}^{(p-1)/2} \sum_{j=-(p-1)/2}^{(p-1)/2} (I(x+i, y+j) - M)^2$$

$$Sk = \frac{1}{p^2} \sum_{i=-(p-1)/2}^{(p-1)/2} \sum_{j=-(p-1)/2}^{(p-1)/2} (I(x+i, y+j) - M)^3$$

$$Kr = \frac{1}{p^2} \sum_{i=-(p-1)/2}^{(p-1)/2} \sum_{j=-(p-1)/2}^{(p-1)/2} (I(x+i, y+j) - M)^4$$

II. Implementation

We set the subcell threshold to be 9 pixels, i.e. we consider a $9 * 9$ sub-image around each pixel. (We, of course, ignore the pixels which are less than 4 units from the edges. This should hardly be a concern as the image is at least hundreds of pixels long and wide)
We use FFT to obtain the Fourier transform of the sub-images (henceforth referred to as cells). There are $(m-p-1)*(n-p-1)$ cells, and the features of the cells are stored into a column matrix $(m-p-1)*(n-p-1)$ long. Each element in this 'Feature Vector' contains the mean, variance, skewness, kurtosis and the position coordinates of the corresponding pixel.

A. Lexicographic Sorting

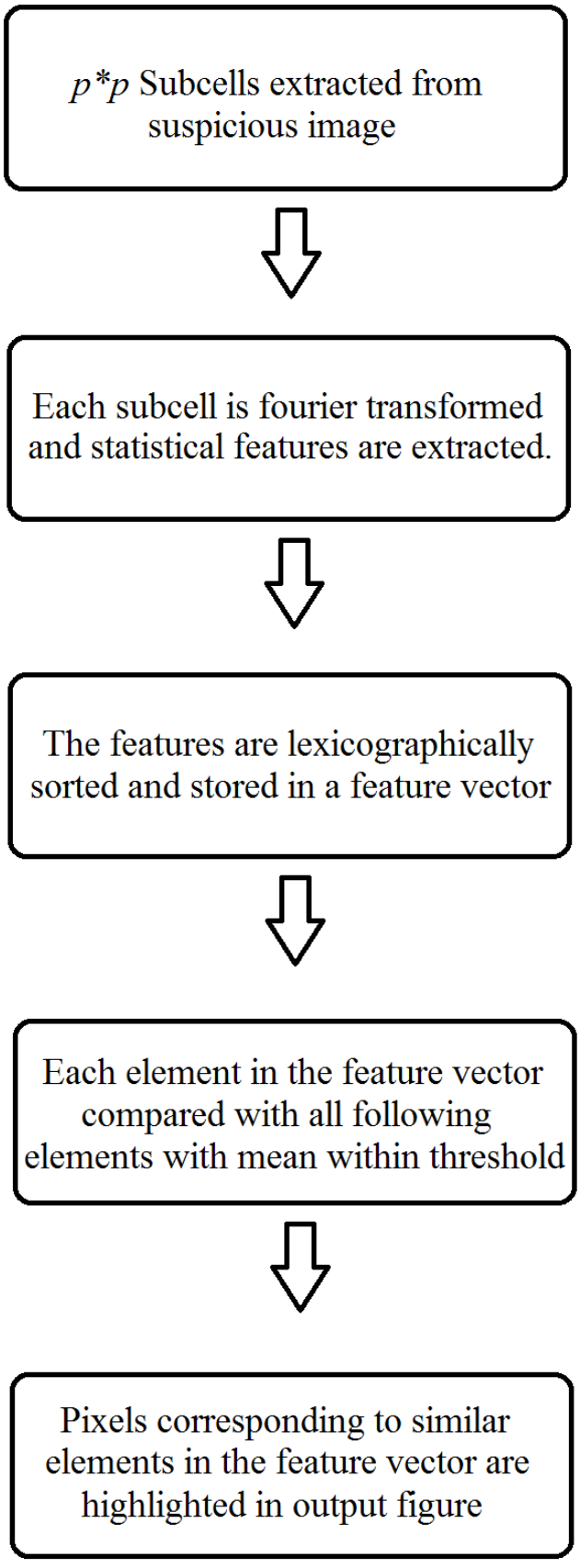
The feature vector is rearranged in the increasing order of mean, and within cells of same mean, they are sorted in the order of variance, followed by skewness and kurtosis.

Mr. Shanavas, Mr. Voora and Mr. Bandaru are undergraduates from the Department of Electrical Engineering, IIT Bombay

B. Identification of corresponding cells in the Feature Vector

The cells within the Feature Vector are processed sequentially. *Each element in the Feature Vector is compared with all following elements whose mean is within a preset threshold of the current element's mean*. This approach minimises the number of comparisons to be performed since the vector was initially sorted in the order of increasing mean. Within these, those elements which have a variance, skewness and kurtosis each within a set of predetermined thresholds are marked as having strong correspondence. In the output figure, pixels corresponding to these elements are highlighted.

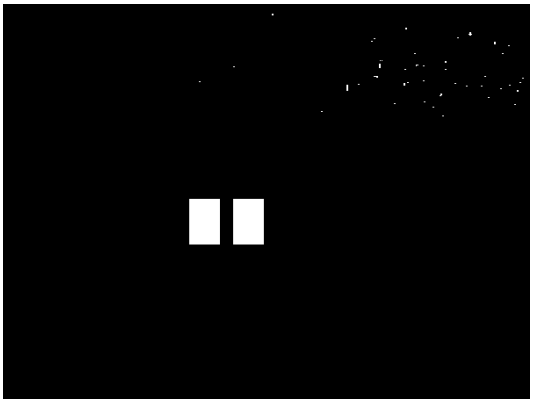
III. FLOWCHART



IV. Results

Example 1: The results of forgery detection. One person has been hidden in the forged image by copy-moving an adjacent patch of canopy over the person.

Forged image which has been fed into the algorithm

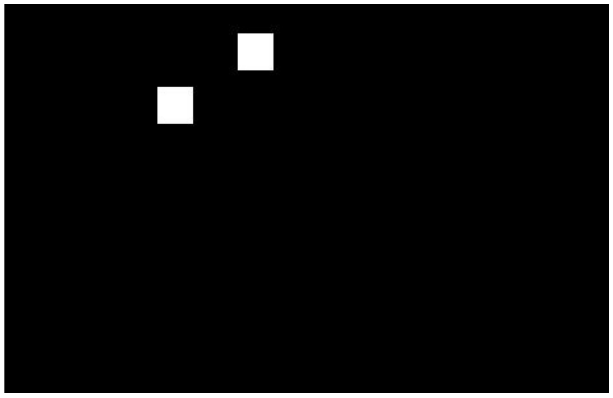


Output of the forgery detection algorithm

Original Image

Example 2: The number of missiles fired has been faked in the suspicious image.

Forged image which has been fed into the algorithm



Output of the forgery detection algorithm

Original Image

REFERENCES

[1] Sevinc Bayram, Ismail Avcibas, Bulent Sankur, Nasir D. Memon *Image manipulation detection with binary similarity measures*, Researchgate Jan 2005, 228937279

[2] Hany Farid *Image Forery Detection*, IEEE Xplore, IEEE Signal Processing Magazine, March 2009, 10.1109/MSP.2008.931079

[3] Rajeev Kaushika, Rakesh Kumar Bajajb, Jimson Mathewc, *On Image Forgery Detection Using Two Dimensional Discrete Cosine Transform and Statistical Moments*, 4th International Conference on Eco-friendly Computing and Communication Systems, ICECCS 2015.