DS601: Assignment 3

A Multiscale and Multi-Perturbation Blind Forensic Technique For Median Detecting

INDIAN INSTITUTE OF TECHNOLOGY BHILAI

December 3, 2021

Abstract

This paper help us about detecting traces of median filtering in digital images, a problem of big importance in forensics given that filtering can be used to hide traces of image tampering such as re-sampling and light direction. To accomplish this objective, author have used a novel approach based on multiple and multi-scale(different region of interest) progressive perturbations on images able to capture different median filtering traces through using image quality metrics.The paper has used 8 IQM for each time filter is applied.After getting IQM we are able to get distinct feature space suitable for proper classification regarding whether or not a given image contains signs of filtering. Experiments using a real-world scenario with compressed and uncompressed images show the effectiveness of the proposed method

1 Introduction

There is huge flow of information in real words and most of them are also in form of images.So there should be some means to tell whether these images are real or fake. As image tampering has a lot of bad effects on our society and real world, therefore we want some way of detecting the image tampering.

One way of detecting the presence of image tampering is through the analysis of artifacts left by the re-sampling operations, but we also know that a non-linear filter such as the median filter can destroy these re-sampling artifacts by replacing each pixel with the median valued pixel within a neighborhood. So it would be difficult for detecting through re-sampling artifacts. Median filtering can also be used to fool some light direction-based forensic techniques and hiding edges.

Therefore there should be some way for detecting the use of median filtered on images. Most of these methods assume that the median filtering process leaves traces on application. These traces are called artifacts which can be detected in order to differentiate between a pristine and a median filtered one. In our paper, a median filtering detection algorithm that is based on the hypothesis that the median filtering streaking artifacts affect the image quality under multi-scale filtering (filtering with different regions of interest) and over progressive perturbations (henceforth perturbations are defined as cascade-wise successive image filtering). Then image quality metrics were evaluated upon perturbed images building a highly discriminated feature space for future classification. Experiments with compressed and uncompressed public datasets confirm the method's competitiveness without assuming anything about the underlying filtering process of the input images.

"There are certain terms which will be useful to know in order to understand the results. To compare the proposed method against the state of the art, we choose a set of standard metrics and conduct tests to identify if there is statistical significance in the reported results.

- **Accuracy -** The first metric used is the classification accuracy. It measures the ratio of the number of correct positive (in our case, median filtered images) and negative (pristine images) classifications and the total set of testing data. It is calculated as

$$Accuracy = \frac{TP + TN}{TP + FN + TN + FP}$$

where TP, TN, FP and FN are true positives, true negatives, false positives and false negatives respectively. We don't use the normalized accuracy here because the positive and negative examples in our data are always balanced.

- **Sensitivity -** The Sensitivity is the ratio of number of samples correctly classified as positive and the total number of positive samples in the testing data. It is also known as true positive rate and is calculated as

$$Sensitivity = \frac{TP}{TP + FN}$$

- **Specificity -** The Specificity is the ratio of number of samples correctly classified as negative and the total number of negative samples in the testing data. It is also known as true negative rate and is calculated as

$$Specificity = \frac{TN}{TN + FP}$$

- **Precision -** The precision tells the percentage of correct positive classifications given all the positive classifications given by the classifier. It is calculated as

$$Precision = \frac{TP}{TP + FP}$$

- **P-Value -** A p-value is used in hypothesis testing to help you support or reject the null hypothesis. The p value is the evidence against a null hypothesis. The smaller the p-value, the stronger the evidence that you should reject the null hypothesis"[3]

2 Materials & Methods

We have used the multi-perturbation approach which was inspired by the [1]. They proposed to perform a 'progressive insertion' of hidden message, in digital images. And they realised that pristine and the stigno images exhibit different behaviour whenever they are disturbed. So taking an inspiration from this observation, the authors of this paper found the same thing happening in the median filtered images. The already median filtered images suffered different degradation when compared to pristine images after a series of successive median filtering operations. The technique used in our paper is also inspired by [2]. In this 1986 paper the author E.L. Hauck a compression algorithm called Run Length Encoding. By observing that when a text file is compressed for the first time exhibits a behaviour differently than the file that has already been compressed." The authors used multi-scale perturbation and the rational behind this was that when an already filtered image is filtered once more using a different median filtering window, the streaking artifacts will get emphasised. When applied in succession, it tends to find groups of streaking pixels instead when done one time only finds a few of them" [3] . This paper contribute by using multiple and progressive perturbations allied with Image Quality Metrics (IQM) on image to detect median filtering. This process works effectively in distinguishing between the pristine and already filtered images.

We took an image and median filtered it through 's' number of different sized windows and then applied median filter on it 'n' number of times. Then now from one image we got (s x n) images. Following this IQMs are calculated for each of these images with respect to the original image. A total of 8 IQMs were calculated. The same process is carried for each image and we end up getting feature vector for each image. These feature vectors are calculated for each image in the database. Database contains set of images that are annotated as pristine or filtered. The model is then trained by feeding it all the calculated feature vectors. Once the classifier is trained, it is then given an unlabelled image and the same process follows again. The fed image undergoes multi-scale perturbations and followed by IQM calculation and then the feature vector is obtained. This obtained feature vector is then classified as pristine or filtered by the trained model. We have performed two experiments on each Four Perturbation Multiple Windows (FPMW), Three Perturbation Multiple Windows, Three Perturbation One Window combinations. Experiment one consist of compressed dataset and experiment 2 consist of both compressed as well as uncompressed dataset. We have trained our classifier on TPOW as well, which was not mentioned earlier in experiments section in the paper. because of it's poor performance as compared to given two methods.For each such

experiment, we generate confusion matrix and calculate below model performance metrics based on that: -Accuracy -Sensitivity -Specificity -Precision -Recall After collecting all the feature vector added it as.txt file and also we have added the readme file for running code.

3 Results

The table shown below are comparison between research papers' tables and our model which was run on 2 experiment case:.

	TPMW	FPMW	SPAM	MFF	GLF
Accuracy	82.1%	84.5%	70.1%	70.1%	65.1%
Sensitivity	92%	91%	98%	88%	99%
Specificity	72%	77%	42%	52%	31%
Precision	76%	80%	62%	64%	59%
Significant?	yes	–	yes	yes	yes

Table 1: Compressed dataset experiments results presented in paper.

	TPMW	FPMW	TPOW	SPAM	MFF	GLF
Accuracy	81%	84%	69%	70.1%	70.1%	65.1%
Sensitivity	90%	91%	80%	98%	88%	99%
Specificity	67%	72%	50%	42%	52%	31%
Precision	82%	86%	73%	62%	64%	59%
Significant?	yes	–	yes	yes	yes	yes

Table 2: Compressed dataset experiments results replicated by our group.

	TPMW	FPMW	SPAM	MFF	GLF
Accuracy	82.2%	80.8%	77.9%	74.2%	79.9%
Sensitivity	78.2%	74.4%	68.3%	76.9%	90.9%
Specificity	90.2%	91.5%	90.6%	78.6%	76.8%
Precision	88.9%	89.8%	87.9%	78.3%	79.7%
Significant?	yes	–	yes	yes	yes

Table 3: Cross dataset experiment with compressed and uncompressed images presented in paper.

	TPMW	FPMW	TPOW	SPAM	MFF	GLF
Accuracy	82%	80%	71%	77.9%	74.2%	79.9%
Sensitivity	80%	79%	71%	68.3%	76.9%	90.9%
Specificity	90%	84%	69%	90.6%	78.6%	76.8%
Precision	96%	94%	86%	87.9%	78.3%	79.7%
Significant?	yes	–	yes	yes	yes	yes

Table 4: Cross dataset experiment with compressed and uncompressed images' results replicated by our group.

The four tables mentioned in this section are denoting percentage values of four parameters which are Accuracy, Sensitivity, Specificity and Precision. These four parameters are calculated for the methods TPMW, FPMW, and TPOW and are mentioned in Table 1 for compressed images and in Table 2 for both compressed as well as uncompressed images.

4 Discussion

The whole discussion of paper was related to the implementation of the proposed technique of multi-scale perturbation. We also tried to reproduce the results mentioned in the paper. Matlab code was converted into python and using svm classifier the classification was obtained. We obtained the results for Three Perturbation Multi Window (TPMW) and Four Perturbation Multi Window (FPMW) and found it to be similar to the results mentioned in the paper. Apart from that, we also calculated all the metrics for the third case which was not mentioned in the paper. It was Three Perturbation One Window (TPOW) and we have mentioned all the findings of TPOW in the results section.

5 Conclusions

We have studied the paper and gain knowledge on another method of detecting median filtering.In our paper, author used a novel approach to forensically detect median blurring traces on digital images. The technique used is that it progressively perturbs the image by blurring it multiple times with different window sizes (filtering intensities), building a discriminative feature for later decision making by using image quality metrics. The model was was trained on FPMW,TPMW and TPOW and discussed in result section After we got different features,the method was trained with features extracted as input to SVM classifier.

6 Acknowledgements

Thanks To : Dr. Nitin Khanna Sir for his support during the whole time and guiding us in positive direction.

7 References

[Rocha] A. Rocha and S. Goldenstein, Progressive randomization: Seeing the unseen, Elsevier Computer Vision and Image Understanding (CVIU) 114(3) (2010), 349–362.

[Hauck] E.L. Hauck, Data compression using run length encoding and statistical encoding, (2 December 1986), uS Patent 4, 626, 829.

[Ferreira] Ferreira, Anselmo, Dos Santos, Jefersson A., and Rocha, Anderson. 'Multi-directional and Multi-scale Perturbation Approaches for Blind Forensic Median Filtering Detection'. 1 Jan. 2016 : S17 – S36