A New Method for Detecting Altered Text in Document Images

Lokesh Nandanwar[1], Palaiahnakote Shivakumara[1](✉), Umapada Pal[2], Tong Lu[3], Daniel Lopresti[4], Bhagesh Seraogi[2], and Bidyut B. Chaudhuri[2]

[1] Faculty of Computer Science and Information Technology, University of Malaya, Kuala Lumpur, Malaysia
lokeshnandanwar150@gmail.com, shiva@um.edu.my

[2] Computer Vision and Pattern Recognition Unit, Indian Statistical Institute, Kolkata, India
umapada@isical.ac.in, to.bhagesh.sr@gmail.com, bbcisical@gmail.com

[3] National Key Lab for Novel Software Technology, Nanjing University, Nanjing, China
lutong@nju.edu.cn

[4] Computer Science and Engineering, Lehigh University, Bethlehem, PA, USA
lopresti@cse.lehigh.edu

Abstract. As more and more office documents are captured, stored, and shared in digital format, and as image editing software becomes increasingly more powerful, there is a growing concern about document authenticity. For example, texts in property documents can be altered to make an illegal deal, or the date on an airline ticket can be altered to gain entry to airport terminals by breaching security. To prevent such illicit activities, this paper presents a new method for detecting altered text in a document. The proposed method explores the relationship between positive and negative coefficients of a DCT to extract the effect of distortions caused by tampering operations. Here we divide DCT coefficients into positive and negative classes, then reconstructs images from the inverse DCT of the respective positive and negative coefficients. Next, we perform Laplacian filtering over reconstructed images for widening the gap between the values of text and other pixels. Then filtered images of positive and negative coefficients are fused by an average operation. For a fused image, we generate Canny and Sobel edge images in order to investigate the effect of distortion through quality measures, namely, MSE, PSNR and SSIM used as features. In addition, for the fused image, the proposed method extracts features based on histograms over the residual images. The features are then passed on to a deep Convolutional Neural Network for classification. The proposed method is tested on our own dataset as well as two standard datasets, namely IMEI and the ICPR 2018 Fraud Contest dataset. The results show that the proposed method is effective and outperforms existing methods.

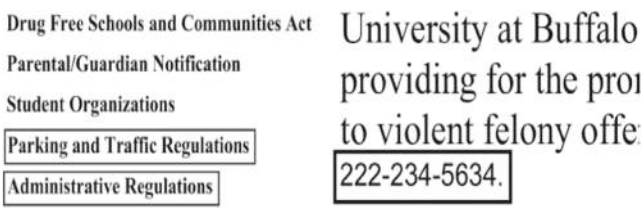
Keywords: Document digitization · DCT coefficients · Fused image · Altered text detection · Fraud document

Y. Lu et al. (Eds.): ICPRAI 2020, LNCS 12068, pp. 93–108, 2020.
https://doi.org/10.1007/978-3-030-59830-3_8

1 Introduction

Due to increased automated generation and processing of legal documents and records, the importance of document verification has received special attention from researchers [1, 2]. This is because professional and skilled fraudsters may take advantage of advanced tools such as Photoshop and Gimp to alter documents for various purposes at different levels. For instance, in airports, security checks require the verification of travel date and the name of the passenger an a boarding pass or identity card before allowing him/her to enter. There have been a number of cases where individuals have fooled security and gained access to an airport by alternates and names on documents [3]. To alter text in a document, in general, two common operations are used, namely, *copy-paste* and *insertion.* In copy-paste operation, the fraudster extracts desired text from a different document or a different location in the same document to be altered, while in the case of insertion, software tools are used to insert words to change the original text. In such a case, these tools can be used to create text in the same font, size and style as in the original (insertion or imitation) [4]. There are existing methods for forgery detection in document images based on printer identification [5–7]. Such work is based on studying characteristics of connected components in the document image. If the methods find abrupt changes in the shapes of characters or text of a document due to distortion introduced by a different printing device, the document is deemed to be forged. However, such approaches are not robust to altered text detection in document images and this is due to distortion introduced by forgery operations may not be prominent.

Hence, it is challenging to detect the altered texts in documents. It is evident from the illustration in Fig. 1, where the naked eye cannot find any obvious differences between the altered text (marked by a rectangle) and genuine text. However, when adversaries use such tools for altering text, there must be some disturbance in the content, which may be in the form of distortions, loss in pattern regularity, or misplacement of content at the pixel level. These clues lead us to propose several methods to detect altered document images.



Copy-paste operation Insertion (imitation) operation

Fig. 1. Illustration of sample forged PDF document images by copy-paste and insertion operations. Note: altered texts are enclosed by bounding boxes, which appear to be genuine text in terms of font, color and size.

Several methods related to altered text detection, including forged document identification have been investigated in the literature. Most of those are based on printer identification, ink quality verification, character shape verification, and distortion identification. For example, Khan et al. [5] proposed automatic ink mismatch detection for forensic document analysis. This method analyzes the ink of different pens to find fraudulent documents, which is effective for handwritten documents but not for printed document images. Luo et al. [6] proposed localized forgery detection in hyperspectral document images. This is an improved version of the previous method, which explores ink quality in the hyperspectral domain for fraud document identification. The method is often not effective on printed texts since when digitized, the quality change becomes very low. Khan et al. [7] proposed automated forgery detection in multispectral document images using fuzzy clustering. The method explores ink matching based on fuzzy c-means clustering to partition the spectral responses of ink pixels in handwritten notes into different clusters. The scope of the method is forgery detection in 'handwritten documents'. Raghunandan et al. [8] proposed Fourier coefficients for fraudulent handwritten document classification through age analysis. This approach studies positive and negative coefficients for analyzing image quality and identifying an image as old or new. The method may not work at the text line or word levels and requires the full document.

Most of the above methods target handwritten documents but not the typeset documents. However, Wang et al. [9] proposed a Fourier-residual for printer identification from document images. This method extracts features from residuals given by the Fourier transform for printer identification. The primary goal of this method is to identify printers rather than altered document images. Shivakumara et al. [3] proposed a method for forged IMEI (International Mobile Equipment Identity) in mobile images. The method explores color spaces of the images and it performs fusion operation for combining the different color spaces, which results in fused image. The features based on connected components are extracted from the fused image. The forged IMEI number is detected by comparing the features of template and the input image. However, the performance the method depends on the fusion operation.

It is observed from the above review that most methods extract features at the block level or connected component level for forgery detection in handwritten and typeset images. These methods are good when there are clear differences between forged and genuine text. If there is a minute difference at the pixel level or at character levels, as expected in typeset documents, the methods may not perform well. Therefore, there is a gap between altered text detection in printed documents and other documents.

Hence, in this work, we aim at detecting altered text detection in printed document images. Inspired by the knowledge that Discrete Cosine Transform (DCT) Coefficients have the ability to identify minute changes in image content [4], in this work, we propose to explore DCT coefficients for detecting altered text in PDF document images. Motivated by the divide and conquer method presented in [8] for fraudulent document identification based on the Fourier transform, we explore the idea for extracting features using DCT coefficients to study the differences between altered and genuine texts. Therefore, the main contribution here is to explore DCT coefficients for extracting features, namely quality measures and histogram based features for detecting altered text in document images.

2 Proposed Method

This work considers altered text lines, words or a short paragraph as input for detection. It is true that when the alteration is done using a copy-paste or insertions, some amount of distortion is introduced compared to genuine. Since the text is typeset, the changes may not be visibly apparent. Therefore, inspired by the property of DCT coefficients mentioned in [4] that it has the ability to extract minute changes at the pixel level, we classify DCT coefficients of an input image into positive and negative coefficients. It is expected that if an image is genuine, there is no misclassification of 'positive' as 'negative' or 'negative' as 'positive'. If the input image is altered, one can expect misclassification at the positive and negative coefficient levels due to the distortion in the content. To extract such a distortion, in the proposed method we reconstruct images by applying the inverse DCT on the respective positive and negative coefficients, resulting in two different images.

For widening the gap between text and other pixels, we then perform a Laplacian filtering over each reconstructed image. Then we fuse these two filtered images by averaging operation. It is expected that for genuine texts, the reconstructed image of the original texts must have better quality compared to the reconstructed image of the altered texts. With this cue, the proposed method extracts three quality measures as features, namely, Mean Square Error (MSE), Peak Signal to Noise Ratio (PSNR), and the Structure Similarity Index Metric (SSIM). Note that, for estimating quality measures, we consider Canny and Sobel operator of input images as ground for respective Canny and Sobel image of fused images. Since the problem is complex due to less distortion during alteration, the proposed method extracts more features based on histogram operation over fused and residual images. The residual image is the difference between the fused and input images. When the fused image of altered text is affected by quality, it is reflected in the intensity values compared to intensity values in the input images. So, for extracting more features, we use histogram operation. In total, our method extracts 76 feature vectors including a vector containing 6 quality measures. Then the extracted features are passed to a deep convolutional neural network for detecting altered text from the genuine text. The pipeline of the proposed method can be seen in Fig. 2.

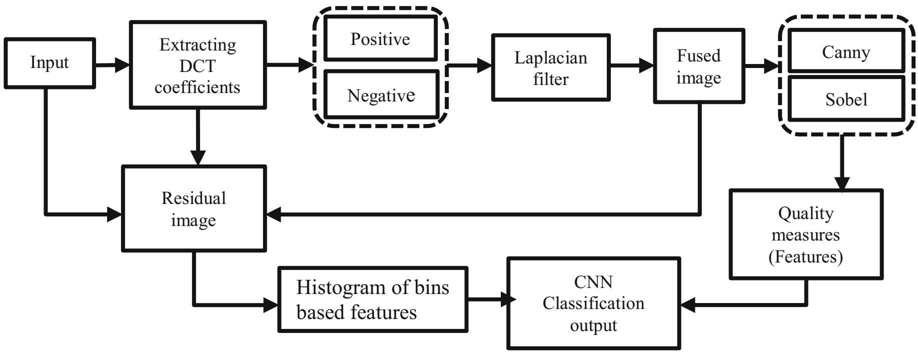


Fig. 2. Workflow of the proposed method

2.1 DCT Coefficient Analysis to Detect Altered Text

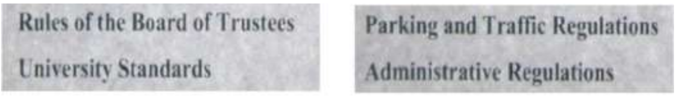
For the input original and altered images shown in Fig. 3(a), the proposed method applies the DCT, as defined in Eq. (1) and classifies the coefficients as positive, negative and zero as shown in Fig. 3(b), (c) and (d), respectively. It is noted from Fig. 3(b) that for the original image, there are more bright pixels at the top-left corner, while the brightness reduces gradually towards the right-bottom corner. On the other hand, for the altered one, bright pixels are scattered compared to the original. It is true that high frequency coefficients represent edge pixels, low frequency coefficients represent non-edge pixels, while zero coefficients represent background pixels. If there is no disturbance in the content (original) of the input image, according to DCT, the left-top corner should get high values that represent edge pixels, and the values get reduced gradually towards the right-bottom corner. When there is a disturbance in the content (i.e., altered), one cannot expect the same coefficient distribution as the original image. These observations holds for negative coefficients of the original and altered text images as shown in Fig. 3(c). However, the distribution of zero coefficients of the original and altered images is opposite as shown in Fig. 3(d), where we can see more pink colored pixels at the right-bottom corner for the original, and the pixels decrease gradually towards the top-left corner. Note that this is not true for altered document images. The distribution of zero coefficients is evident from the above observations made with respect to positive and negative coefficients.

$$F[u,v] = {}^{1}\!\big/_{N^2} \sum_{m=0}^{N-1}\sum_{n=0}^{N-1} f[m,n]\cos\left[\frac{(2m+1)u\pi}{2N}\right]\cos\left[\frac{(2n+1)v\pi}{2N}\right] \tag{1}$$

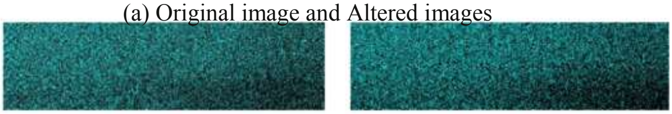
Here, u and v are discrete frequency variables (0, 1, …, N − 1), $f[m, n]$ represent gray value at (m, n) position of $N \times N$ image matrix (m, n = 0, 1, …, N − 1), and $F[u, v]$ is the (u, v) component of resultant DCT frequency matrix.

The proposed method reconstructs images by applying the inverse DCT on positive and negative coefficients of the original and altered images respectively, as in Fig. 4(a) and (b). It is observed from Fig. 4(a) and (b) that the original images reconstructed with respect to positive and negative coefficients appear brighter than the reconstructed altered image made in the same way. Since the alterations caused by copy-paste and insertion operations for printed texts do not create significant changes in content, it is hard to find the differences between the original and altered images. Therefore, to widen the gap between the original and altered text, the proposed method performs Laplacian filtering over the reconstructed images of positive and negative coefficients separately. This eliminates those pixels that represent unwanted noise as shown in Fig. 4(c) and (d), where it is noted that text pixels are enhanced and there are clear differences between the outputs of Laplacian filtering over positive and negative coefficients of the original and forged images.

To take advantage of this observation, we propose to combine the output of Laplacian filtering of positive and negative coefficients with a simple operation called averaging, which results in a fused image for the original and altered images as shown in Fig. 5(a).



(a) Original image and Altered images



(b) Positive DCT coefficients of the original and altered images

(c) Negative DCT coefficients of the original and altered images in (a).

(d) Zero coefficients of the original and altered images in (a)

Fig. 3. Positive, negative and zero coefficient distributions of the original and altered images.

(a) Reconstructed images for the positive coefficient of original and altered images

(b) Reconstructed images for the negative coefficient of original and altered images

(c) Laplacian filtering for the images in (a).

(d) Laplacian filtering for the images in (b)

Fig. 4. Widening the difference between original and altered images using Laplacian filtering

2.2 Altered Text Detection in Document Images

As mentioned in the previous section, due to the operation for alternating text in the images, the image quality affects for the fused images as shown in Fig. 5(a) where we can see the clear difference between fused original and altered images in terms of brightness. It is evident from the Canny and Sobel edge images of original and altered fused images as shown in Fig. 5(b) and (c), respectively. It is observed from Fig. 5(b) and (c) that for the original fused images Canny and Sobel preserve structure of characters while for the altered fused image, the structure of the character lost compared to original fused image. This shows that the original fused image does not suffer from alteration of text while altered image suffer from the operation. To extract such difference, we estimate the quality measures, namely, Mean Square Error (MSE), Peak Signal Noise Ratio (PSNR) and the Structure Similarity Index (SSIM) for the fused and input images. Note: we consider Canny and Sobel of input images as ground truth for respective Canny and Sobel of fused images to estimate the above three quality measures. As a result, this process outputs 6 features, which includes 3 from Canny and 3 more from Sobel for the input image for detecting altered text in the document image. It is a vector containing 6 features.

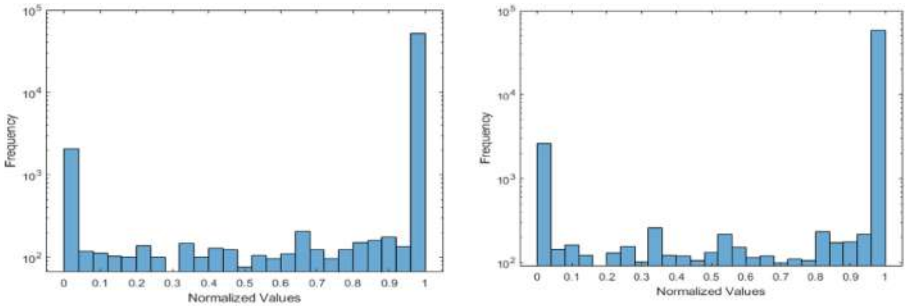
(a) Fused original and altered images

(b) Canny edge images of fused original and altered images in (a)

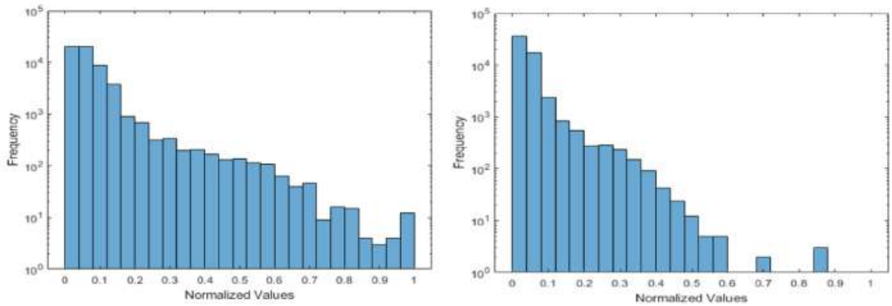
(c) Sobel edge images of fused and altered images in (a)

Fig. 5. The quality effect can be seen in Canny and Sobel edge images of input and fused images of original and altered.

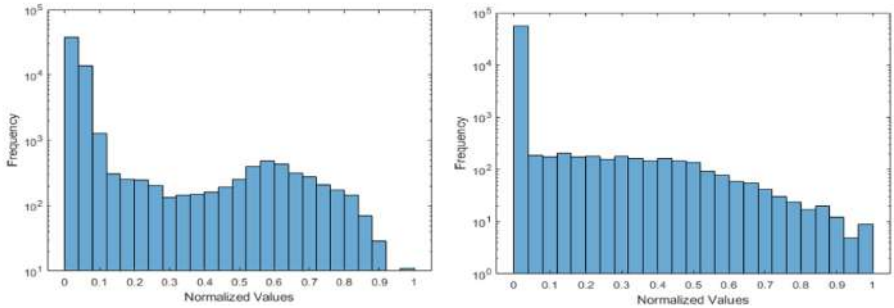
The problem we are considering here is challenging, so quality measures alone may not be sufficient for achieving better results for altered text detection in document images. The same clue of quality difference in fused images leads us to extract the difference in intensity values in the fused images of original and altered images. This is because when the quality changes, it affects intensity values in the images. It is evident from the histograms plotted over intensity values of input original and altered images with 25 bins as shown in Fig. 6(a), where it is noted that low values are increasing in case of altered image compared to the original images. In the same way, when we plot histogram with bin size for the fused original and altered images as shown in Fig. 6(b), it can be noticed that loss of high values in case of altered image compared to original image. Therefore, to extract the above observation, the proposed method finds difference between the fused and the input images, which results in Residual Image (RI) for respective original and altered images as defined in Eq. (2). The histogram with same the 25-bin size plotted for residual images of original and altered images show that there is a significant difference in distribution of intensity values. Since we consider each bin in the histogram as a feature vector (all the values that contributes to bin), the proposed method gets 25 + 25 = 50 feature vectors. To strengthen the above feature extraction, since there is difference between histograms of input and fused images of original, altered images, the proposed method further finds the difference between two histograms (CF) at frequency level as defined in Eq. (3), which results in 25 more feature vectors. In total, the proposed method extracts 50 + 25 + 1 = 76 feature vectors for detecting altered text in document images. Each vector contains number of features according to histogram bin operation.



(a) Histograms of input images original and altered images



(b) Histograms of fused images of original and altered images



(c) Histograms of residual images original and altered images

Fig. 6. Feature extraction using histogram of fused and residual images of original and altered images.

$$RI = \sum_{i=0}^{m}\sum_{j=0}^{m}|I(i,j) - FI(i,j)| \tag{2}$$

Where m denotes the size of the Input (I) and Fused Image (FI).

$$CF = \sum_{k=0}^{N}|HI(k) - HFI(k)| \tag{3}$$

where N denotes the bin size of the histogram which is 25 in this work. HI and HFI denote histograms of input image and fused images, respectively. The value of 25 is determined experimentally for choosing random samples from across datasets. It is justified in experimental section.

Motivated by the strength of discriminative power of deep convolutional neural networks, the extracted 76 feature vectors are passed to the neural network which is shown in Fig. 7 for detecting altered text in document images. In this architecture, we use 'ReLU' activation function for all the layers except the final layer where we use 'Sigmoid' [10] activation. With 'Adam' [11] as optimizer and learning rate of 0.01 and 'binary cross entropy' loss function, the proposed architecture is trained for 100 epochs with the batch size of 8. The loss function as defined in Eq. (4), which is binary cross entropy loss [12] used for classification of altered text in the PDF document images.

$$BCE = -\frac{1}{N}\sum_{i=1}^{N}(y_i.\log(p(y_i)) + (1 - y_i).\log(1 - p(y_i))) \tag{4}$$

where y is the label and $p(y)$ is the estimated probability for total number of N samples. The dropout rate, 0.2 is added in between the convolutional layers to reduce overfitting and more generalization of the results. For all the experiments, we use the system with Nvidea Quadro M5000 GPU for training and testing of the architecture and python framework Keras for this application. The dataset is divided into 80% and 20% for training and testing for all the experiments in this work.

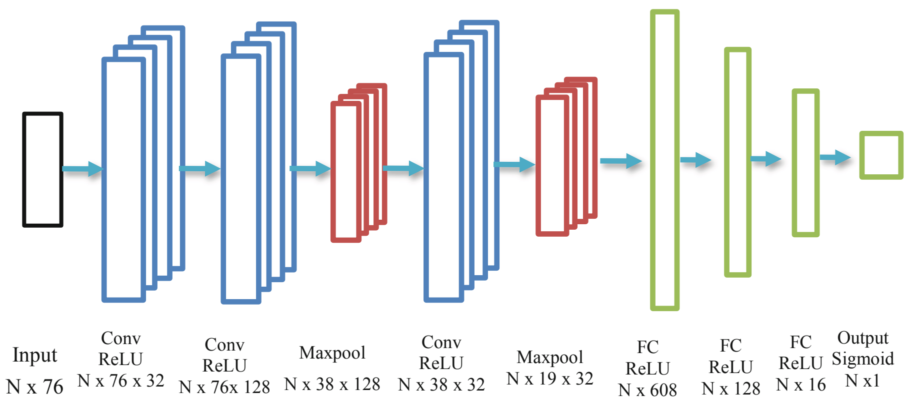
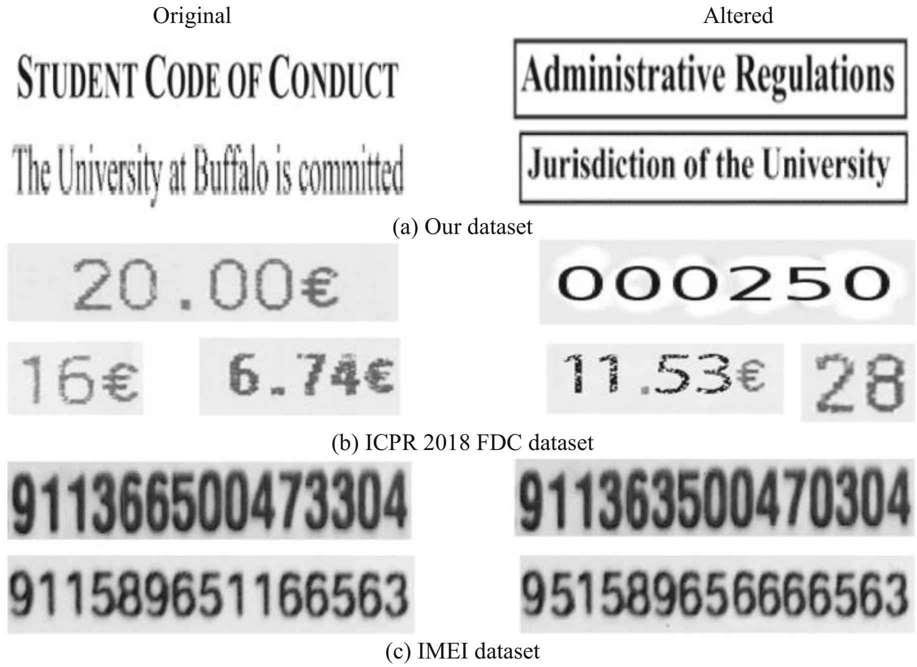


Fig. 7. Architecture of Convolutional Neural Network for Classification

3 Experimental Results

To evaluate our approach, we created our own dataset with the help of copy-paste and insertion operation using Paint Brush Tool, tested on documents drawn from a variety of sources, including property documents, insurance documents and air tickets. Our dataset contains 110 altered text line images and the same number of original text line images, which gives a total of 220 text line images for experimentation. Sample images of our dataset are shown in Fig. 8(a), where it can be seen that it is hard to notice the difference between the altered text and original text line images. Our dataset consists of altered text at line level or words levels. In other words, our dataset does not include character level additions or deletions.



(a) Our dataset

(b) ICPR 2018 FDC dataset

(c) IMEI dataset

Fig. 8. Sample images of different datasets for original and altered text

In order to evaluate the effectiveness of the proposed method, we consider the benchmark dataset [13] called ICPR 2018 Fraud Detection Contest (FDC), which provide altered text at character level. Most of the documents considered in this dataset are receipts, where the price is altered and those are considered as fraudulent documents. The main challenge of this dataset is that the length of text is too small and is limited to strings of numerals with a currency symbol. In addition, altering only one character in a string of few numerals makes the data more complex and challenging. The dataset provides 300 samples for original and 302 samples for altered text, which provides a total of 602 images for experimentation. For our experiment, since the proposed method requires text line of original and altered documents, we segment the altered part from the documents and original documents. The ground truth for the altered region is marked by a rectangle in the dataset. We segment the rectangular region from the altered documents automatically. For original documents, since almost all of the documents are receipts and have the prices in the location where the rectangle is drawn in the altered documents, we use the same segmentation step for extracting price information from original documents. The sample images of altered text and original text are shown in Fig. 8(b), where the clear difference is noticeable. This makes problem easier compared to our own dataset and the IMEI dataset.

In the same way, to test the robustness of the proposed method, we also consider standard dataset of forged IMEI number detection [3], which provides 500 forged and 500 original images. This dataset is challenging because the forgery is done at the character level but not at the word level. In each forged image, we can expect that one or two characters, especially numerals, are forged using copy-paste and insertion operations. This dataset also includes images affected by blur and noise some sample images of which are shown in Fig. 8(c), which are poor quality images. When the image is of poor quality, it is more challenging for forgery detection, because the distortion created by alteration and the poor quality may accentuate each other. The background is complex for these images compared to ours and ICDAR 2018 FDC datasets. Overall, 1822 images are considered for our experiment.

To show the superiority of our proposed technique, we implement two recent methods for comparative study. Wang et al. [9] proposed a method for printer identification by studying the print of different printers. The main basis of the method is that the print of different printers are affected by unique noise/distortion introduced by the printers. The same basis is used for the proposed method. In addition, the text in the print document is printed type as the proposed work. In the same way, we implement Shivakumara et al. [3] method which explores color spaces obtaining the fused image and computing connected based features are extracted for forged IMEI number detection. The objective of the method is same as the proposed method and idea of fusion concept is similar to the proposed method. Furthermore, we run Convolutional Neural Network Inception V3 network [14] (CNN) by passing directly images as input for classification with transfer learning on pertained weights. This is to show that the combination of features and CNN is better than CNN alone especially for two classes with small sized dataset, In order to show that the above methods are not adequate for achieving the results for altered text detection in the document images, we use the above methods for comparative study. For measuring the performance of the proposed and existing methods, we use standard measure average classification rate calculated through confusion matrix. The classification rate is defined as the number of images detected correctly divided by the total number of images. The average classification rate is the mean of diagonal elements of confusion matrix.

As mentioned in the proposed methodology section, the size of the histogram bin is 25. To determine this value, we conduct experiments on random sample chosen from across datasets by varying the size of the bin vs average classification rate as shown in Fig. 9. It is observed from Fig. 9 that as bin size increases the average classification rate

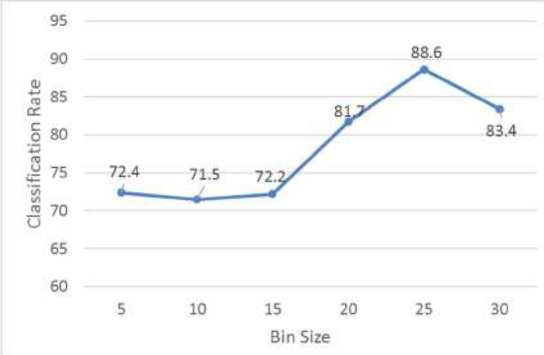


Fig. 9. Determining the value of 25 for achieving better results

increases gradually. This is valid because as the bin size increases, the proposed features acquire more local information and hence the average classification rate increases gradually up to 25 bin size. At the same time, if we continue in increasing bin size, at some point, the proposed features lose vital information in the images due to sparsity. However, when the bin size is 25, the average classification rate reaches highest score, 88.6% and then the average classification rate decreases after bin size 25. This shows that the value of bin size, 25 is the optimal value for achieving better results. The same value are used for all experiments in this work.

In the proposed method, we extract the features, namely, quality measures, histogram based features from fused images, residual images and from the histogram difference between input and fused images. In order to assess the contribution of each feature type, we calculate classification rate for each feature type and the results are reported in Table 1. It is observed from Table 1 that each feature type scores more than 65% average classification rate. Therefore, one can understand that each feature type contributes for achieving better results. At the same time, one can justify that feature type alone is not sufficient to achieve better results as the proposed method. The lowest average classification rate is reported by Quality measures and highest average classification rate is reported by features extracted from histogram of difference between input and fused images. This indicates that the step which classifies DCT coefficients into positive and negative to obtain fused image play a significant role in widening the gap between input and fused images. Since the proposed method uses Canny and Sobel edges of the input and fused images for estimating quality measures, the quality measures do not contribute much for altered text detection because Canny method introduces erratic edges and Sobel operator lose vital information for the fused images of original and altered images, as shown in Fig. 5(b). When we compare the results of fused and residual, the residual contributes more for detecting the altered text. This is because the both input and fused images contributes for the Residual. Overall, to achieve 88.6% average classification rate for the proposed method, four different feature types are employed.

Quantitative results of the proposed and existing methods for our dataset and standard dataset are reported in Table 2, where it is noted that the proposed method is the best in terms of average classification rate for all three datasets compared to the existing

Table 1. Confusion matrix and average classification rate of different features and the proposed method on our dataset (in %).

Features	Quality measures		Fused		Residual		Difference -Hist		Proposed	
Classes	Original	Altered	Original	Altered	Original	Altered	Original	Altered	Original	Altered
Original	61	39	62.8	37.2	75.2	24.8	82.6	17.4	87.4	12.6
Altered	27.5	72.5	29	71	18.4	81.6	20.2	79.8	10.2	89.8
Average	66.75		66.9		78.4		81.2		88.6	

methods. Interestingly, the proposed method scores its highest average classification rate for the ICPR 2018 FDC dataset compared to our and IMEI datasets. This is because the distortions introduced due to forgery operation are noticeable and visible compared to the original image. However, this is not true for our and IMEI datasets. It is evident from the sample image shown in Fig. 8(b), where we can see clearly the changes due to the forgery operation. It is confirmed from the results scored by two existing methods [9, 14] as the methods score high results for ICPR 2018 FDC dataset compared to our and IMEI dataset. Similarly, the proposed method returns its lowest results for the IMEI dataset compared to our own dataset and the ICPR 2018 FDC dataset. The reason is that the images of IMEI dataset are challenging because they were captured by mobile devices, while the images in the other two datasets are captured by scanning the documents. When we compare the results of the three existing methods, Wang et al. is better than the other two existing methods. This is because it extracts spatial, texture and gradient properties of text in the images while the method [13] extracts only the connected component based properties from the edge images and the method [14] does not consider advantage of handcrafted features. In other words, the features extracted in [13, 14] are not good enough to cope with the challenges of different datasets. It is noted from Table 2 that the method [13] achieves highest average classification rate for IMEI dataset compared to our and ICPR 2018 FDC dataset. This is understandable because the method [13] is developed for IMEI forged number detection. Table 2 shows that the CNN [14] reports lowest average classification rate compared to other existing methods for all the three datasets. This shows that CNN does not work well when the dataset size is small. Therefore, one can infer that, in this situation, the combination of handcrafted features with classifier is better than CNN alone. However, the accuracies of the existing methods are lower than the proposed method. Thus, we can conclude that the combination of frequency domain, spatial domain and deep convolutional neural network of the proposed method make difference compared to existing methods.

Although the proposed method achieves better results, there are some cases for which it does not work. One such case is alterations at the text line level, where the proposed method fails to detect the altered text. This is because when we copy and paste the whole text line, the effect of altered operation is minimal compared to character and word levels. One such sample is shown in Fig. 10, where it is hard to see the difference between the original and altered image. Similarly, when the images are degraded due to aging and distortion, the distortion introduced by alteration overlap with the distortion of the image, which is beyond the scope of the work. In this case, it is necessary to introduce context-based features for detecting altered information. This requires image processing and natural language processing to find a solution.

Table 2. Confusion matrix and average classification rate of the proposed and existing methods on our and benchmark datasets (in %)

Methods	Dataset	Own Dataset		IMEI dataset [3]		ICPR 2018 [13]	
		Original	Altered	Original	Altered	Original	Altered
Proposed	Original	**87.4**	12.6	**84.4**	15.6	**84.0**	16.0
	Altered	10.2	**89.8**	14.2	**85.8**	4.50	**95.50**
	Average	**88.6**		**85.1**		**89.41**	
Wang et al. [9]	Original	80	20	83.2	16.8	84.66	8.67
	Altered	13.4	86.6	25.6	74.4	7.95	89.33
	Average	83.3		78.8		86.99	
Shivakumara et al. [3]	Original	60	40	82.2	17.8	92	8
	Altered	35	65	18	82	49.44	50.66
	Average	62.5		82.1		71.33	
Szegedy et al. [14]	Original	34.8	65.2	53.8	46.2	81.8	18.2
	Altered	16.2	83.8	33.7	66.3	47.8	52.2
	Average	59.3		60.1		67.0	

(a) Altered text classified as Original (b) Original text classified as Altered

Fig. 10. Limitation of the proposed method

4 Conclusion and Future Work

In this work, we have proposed a new method for detecting altered text in document images. The proposed method explores applying DCT coefficients in different way for obtaining fused image for the input image. The proposed method extracts features from the fused images based on quality measures and histogram-based features. The extracted features are then passed to deep neural network classifier for detecting altered text in the document images. Experimental results on our own dataset and two standard datasets show that the proposed method outperforms existing methods in terms of average classification rate. In addition, the results show that the proposed method is robust to images which are altered at the character level. However, when the alteration is done at the text line level instead of at the word or character levels, the performance of the proposed method degrades. In addition, when the image is affected by degradations and poor quality, the proposed method may not perform well. This provides scope for future work where we plan to introduce context information through the use of natural language models to find solutions to these remaining problems.

References

1. Alaei, F., Alaei, A., Pal, U., Blumenstein, M.: A comparative study of different texture features for document mage retrieval. Expert Syst. Appl. **121,** 97–114 (2019)
2. Beusekom, J.V., Shafait, F., Breuel, T.: Text-line examination for document forgery detection. Int. J. Doc. Anal. Recogn. (IJDAR) **16**(2), 189–207 (2013)
3. Shivakumara, P., Basavaraja, V., Gowda, H.S., Guru, D.S., Pal, U., Lu, T.: A new RGB based fusion for forged IMEI number detection in mobile images. In: Proceedings of ICFHR, pp. 386–391 (2018)
4. Wang,W., Dong, J., Tan, T.: Exploring DCT coefficients quantization effect for image tampering localization. In: Proceedings of IWIFS, pp. 1–6 (2011)
5. Khan, Z., Shafait, F., Mian, A.: Automatic ink mismatch detection for forensic document analysis. Pattern Recogn. **48**, 3615–3626 (2015)
6. Luo, Z., Shafait, F., Mian, A.: Localized forgery detection in hyperspectral document images. In: Proceedings of ICDAR, pp. 496–500 (2015)
7. Khan, M.J., Yousaf, A., Khurshidi, K., Abbas, A., Shafait, F.: Automated forgery detection in multispectral document images using fuzzy clustering. In: Proceedings of DAS, pp. 393–398 (2018)
8. Raghunandan, K.S., et al.: Fourier coefficients for fraud handwritten document classification through age analysis. In: Proceedings of ICFHR, pp. 25–30 (2016)
9. Wang, Z., Shivakumara, P., Lu, T., Basavanna, M., Pal, U., Blumenstein, M.: Fourier-residual for printer identification. In: Proceedings of ACPR, pp. 1114–1119 (2017)
10. Narayan, S.: The generalized sigmoid activation function: competitive supervised learning. Inf. Sci., 69–82 (1997)
11. Kingma, P.D., Bai, J.L.: Adam: a method for stochastic optimization. In: Proceedings of ICLR, pp. 1–15 (2015)
12. Nasr, G.E., Badr, E.A., Joun, C.: Cross entropy error function in neural networks: forecasting gasoline demand. In: Proceedings of FLAIRS, pp. 381–384 (2002)
13. Artaud, C., Sidère, N., Doucet, A., Ogier, J., Yooz, V.P.D.: Find it! Fraud detection contest report. In: Proceedings of ICPR, pp. 13–18 (2018)
14. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z.: Rethinking the inception architecture for computer vision. In: Proceedings of CVPR, pp. 2818–2826 (2016)