Indian Statistical Institute

ISI Digital Commons

Journal Articles | Scholarly Publications

12-24-2020

Forged text detection in video, scene, and document images

Lokesh Nandanwar
Universiti Malaya

Palaiahnakote Shivakumara
Universiti Malaya

Prabir Mondal
Indian Statistical Institute, Kolkata

Karpuravalli Srinivas

Raghunandan
University of Mysore

See next page for additional authors

Follow this and additional works at: https://digitalcommons.isical.ac.in/journal-articles

Recommended Citation

Nandanwar, Lokesh; Shivakumara, Palaiahnakote; Mondal, Prabir; Srinivas, Karpuravalli; Raghunandan; Pal, Umapada; Lu, Tong; and Lopresti, Daniel, "Forged text detection in video, scene, and document images" (2020). *Journal Articles*. 2.
https://digitalcommons.isical.ac.in/journal-articles/2

This Research Article is brought to you for free and open access by the Scholarly Publications at ISI Digital Commons. It has been accepted for inclusion in Journal Articles by an authorized administrator of ISI Digital Commons. For more information, please contact ksatpathy@gmail.com.

Authors

Lokesh Nandanwar, Palaiahnakote Shivakumara, Prabir Mondal, Karpuravalli Srinivas, Raghunandan, Umapada Pal, Tong Lu, and Daniel Lopresti

This research article is available at ISI Digital Commons: https://digitalcommons.isical.ac.in/journal-articles/2

IET Image Processing

Research Article

IET Journals
The Institution of Engineering and Technology

Forged text detection in video, scene, and document images

ISSN 1751-9659
Received on 29th April 2020
Revised 4th December 2020
Accepted on 11th January 2021
E-First on 25th February 2021
doi: 10.1049/iet-ipr.2020.0590
www.ietdl.org

Lokesh Nandanwar[1] ✉, Palaiahnakote Shivakumara[1], Prabir Mondal[2], Karpuravalli Srinivas Raghunandan[3], Umapada Pal[2], Tong Lu[4], Daniel Lopresti[5]

[1]*Faculty of Computer Science and Information Technology, University of Malaya, Kuala Lumpur, Malaysia*
[2]*Indian Statistical Institute, Kolkata, India*
[3]*Department of Studies in Computer Science, University of Mysore, Karnataka, India*
[4]*National Key Lab for Novel Software Technology, Nanjing University, Nanjing, People's Republic of China*
[5]*Computer Science and Engineering, Lehigh University, Bethlehem, PA, USA*
✉ *E-mail: lokeshnandanwar150@gmail.com*

Abstract: Rapid advances in artificial intelligence have made it possible to produce forgeries good enough to fool an average user. As a result, there is growing interest in developing robust methods to counter such forgeries. This study presents a new Fourier spectrum-based method for detecting forged text in video images. The authors' premise is that brightness distribution and the spectrum shape exhibit irregular patterns (inconsistencies) for forged text, while appearing more regular for original text. The method divides the spectrum of an input image into sectors and tracks to highlight these effects. Specifically, positive and negative coefficients for sectors and tracks are extracted to quantify the brightness distribution. Variations in the shape of the spectrum are analysed by determining the angular relationship between the principal axes and the sectors/tracks of the spectrum. Next, it combines these two features to detect forged text in the images of IMEI (International Mobile Equipment Identity) numbers and document. For evaluation, the following datasets are used: own video dataset and standard datasets, namely, IMEI number, ICPR 2018 Fraud Document Contest, and a natural scene text dataset. Experimental results show that the proposed method outperforms existing methods in terms of average classification rate and *F*-score.

1 Introduction

Most popular social media services, such as Twitter, Facebook, Instagram, provide first-hand news provided by traditional journalists and ordinary users [1, 2]. At the same time, with the rapid development of multimedia technology and user-friendly image editing software, such as Adobe Photoshop and Premiere, and Mokey by Imagineer systems, manipulating video news by altering the content is becoming significantly easier. As a result, Facebook, Twitter, and Instagram report many cases of altered content in images and videos [1, 2]. Elsewhere, there are other sensitive applications involving traditional news media, military matters, and law enforcement, for which forged data detection is vital. Hence, there is an urgent need for new methods capable of detecting forged information in images and videos.

Video forensics can be broadly classified into two categories: active forensic and passive forensic [1]. In the case of the former, while generating a video, the system can include validation information such as a digital watermark or signature, and/or fingerprinting which can be used for authentication. In the case of the latter, the veracity and integrity of video information is authenticated without requiring any embedded validation information, making it more practical than active forensics [1]. This has made passive video forensics hot research topic.

Numerous methods have been developed for detecting forged images and videos created through copy-paste operations and splicing. For instance, Selvaraj and Karuppiah [3] proposed a technique based on SIFT features. Soni *et al.* [4] addressed forgery detection in images using density-based clustering and noise closeting algorithms. Fadi *et al.* [5] proposed an approach for inter-frame forgery detection in videos using spatio-temporal information. Chen *et al.* [1] explored automatic detection of object-based forgeries in videos. D'Amiano *et al.* [6] proposed a match-based dense field algorithm for copy-move forgery detection and localisation. Feng *et al.* [7] discussed motion adaptive frame deletion detection for digital video forensics. Pun *et al.* [8] proposed image forgery detection using adaptive over-segmentation and feature point matching. Yang *et al.* [9] explored copy-paste forgery detection based on hybrid features. Tian *et al.* [10] proposed robust independent elementary features and a similarity metric for detecting copy-move forgeries in images. We note that most of the above methods focus on general visual content, but not on text that appears in images. The scope of this method is limited in this regard. This makes forged text detection an interesting topic for research in forensic applications.

Methods [11, 12] developed for text detection in video and natural scene images work well irrespective of the text modality, i.e. caption/superimposed text which is edited into an image versus scene text which is an inherent part of the original image. Consider, for example, the results shown in Figs. 1*a* and *b* for video and scene images, respectively. Here the method in [12], which employs deep learning, detects the text correctly, including the indicated forged text. Based on this observation, our work uses the output of text detection as the input for our approach to forged text identification. Throughout our work we assume that the altered text in question has been created using standard image editing operations, namely, copy-paste and insertion. Figs. 1*a* and *b* show that using Photoshop software, one can copy desired words from a source image and paste them into a target image, and also insert words to create a fake message. Furthermore, when we look at the original and forged text, it is not easy to differentiate them visually without some assistance. This makes the problem interesting and challenging.

2 Related work

Since forged text detection in video images can be seen as related to forgery detection in document images, we review methods for identifying fraud in handwritten and printed documents. Likewise, since caption text in videos is super-imposed, this can be considered a similar alteration of the original video image, with

© The Institution of Engineering and Technology 2021

Source image Target image

a

Target image

b

Fig. 1 *Example of copy-paste and insertion operations for creating forged text, along with the results of the text detection method*

(a) Word 'MirrorNow' in the source image is copied and pasted into the target image. The text detection method segments the words in source and target images well as indicated by the bounding boxes, ***(b)*** Word 'LIVE' is inserted using Adobe Photoshop software. The text detection method segments the original and forged text in the natural scene image well as indicated by the bounding boxes

naturally occurring scene text viewed as a non-altered image. For this reason, we also review related past work on caption text disambiguation from natural scene text.

Barbosa *et al.* [13] proposed forged text detection in documents written using ballpoint pens. This method uses the ink as a feature. Elkasrawi and Shafait [14] proposed laser printer identification using supervised learning for document forgery detection. This approach explores using noise produced by different printers for fraudulent document classification. Ahmed and Shafait [15] proposed forgery detection based on intrinsic document contents. This method uses the similarity between blocks of an image to identify forged documents. Shafait *et al.* [16] proposed automatic ink mismatch detection for forensic document analysis. This method analyses inks from different pens to detect fraudulent documents. Luo *et al.* [17] proposed localised forgery detection in hyperspectral document images. This is an improved version of the above method, which uses ink quality in the hyperspectral space. Bertrand *et al.* [18] proposed a system based on intrinsic features for fraudulent document detection. This approach extracts features or characters to match with the ground truth for fraud estimation. Barboza *et al.* [19] proposed a colour-based model to estimate the age of a document for forensic purposes. This approach uses the ink quality of handwritten documents captured over different time periods. The method [19] helps us to identify a given image is old or new.

Halder and Garain [20] proposed a colour-based approach for determining ink age in printed documents. This method uses the same colour features as above for printed text images. Kumar *et al.* [21] proposed detecting alternations in ball point pen strokes. This approach employs colour and texture features for handwritten documents, and in particular bank cheques. Raghunandan *et al.* [22] proposed Fourier coefficients for fraudulent handwritten document classification using age analysis. This approach studies positive and negative coefficients for analysing the quality of images, which mark them as old or new. Wang *et al.* [23] proposed a method for printer identification by analysing the output of different types of laser printers. This is based on the premise that each printer introduces its own unique noise during the printing process. Shivakumara *et al.* [24] proposed a method for detecting forged IMEI (International Mobile Equipment Identity) numbers using a fusion approach based on colour space. The performance of this method depends on Canny and Sobel edge information. If a forged image contains sufficient distortion as a result of the editing, the method works well.

Bibi *et al.* [25] proposed a text-independent approach for document forgery detection through printer source identification. The method employs convolutional neural networks (CNNs). Kalbitz and Vielhauer [26] proposed forensic ink evaluation in handwritten documents using a clustering approach. This method exploits the idea of detecting the use of multiple inks used in a document as an indicator of forgery. Mukhtar and Malhotra [27] proposed a similar idea for legal handwritten documents. The method analyses the colour of the ink used in the document. Rahiche and Cheriet [28] proposed a technique for forgery detection in hyperspectral document images based on graph orthogonal non-negative matrix factorisation. Their method graphs regularised terms to exploit the geometric information lost in the matricisation of the images. Chen and Gao [29] proposed a method for forged handwriting number detection based on a CNN. This method pools different layers for forgery detection in the images. Then review of existing work shows that the methods focus on one type of image, whereas the proposed work considers different types of images in this paper. Nandanwar *et al.* [30] proposed method for detecting altered text in document images. Their approach splits the Discrete Cosine Transform (DCT) into positive and negative coefficients and reconstructs images for the respective positive and negative coefficients. However, since the method was developed specifically for document images, it may not work well for videos and natural scene images.

As indicated in our review above, there are methods that assume printed text on a plain background, while for forged IMEI number detection, there are methods that take advantage of a stable background. In addition, there are methods that use colour or character stroke components for forgery detection. However, these features may not work well for video images, where one cannot expect a plain or fixed background; rather these images can contain complex backgrounds with a large variety of variation. Finally, none of the previous methods we have surveyed here aims at detecting forged text across videos, natural scenes, and document images with a single unified approach.

To overcome the problems of methods that assume a plain background, a few approaches have been developed to detect caption text in videos. Shivakumara *et al.* [31] proposed a technique to separate graphics and scene text in videos. This method works based on the fact that caption text has high contrast and clarity, while scene text does not. Xu *et al.* [32] also studied graphics and scene text classification in videos. Their method extracts distinct features through distribution of eigenvalues. Roy *et al.* [33] proposed new features for scene and caption text classification in video frames. This method explores DCT coefficients to differentiate caption text from scene text. Bhardwaj and Pankjakshan [34] proposed image overlay text detection based on JEPG truncation error analysis. This method extracts error-based features through truncation given by a colour filter array for detecting caption text in videos. However, the above methods are not adequate for forged text created through copy-paste and insertion operations because the forged text might be classified as caption text and vice versa.

Hence, in this paper, we present a method which exploits brightness distribution (BD) and the shape of the Fourier spectrum through sectors which divide the spectrum into eight divisions based on eight directions and tracks, which are defined by drawing circle over sectors. The main basis for this work is that when a person performs forgery operations like copy-paste and insertion, the forged text will not perfectly align with the other text in the image and there will be irregularities in its background compared to regions around non-tampered text [8, 9]. Therefore, we expect the Fourier spectrum computed from original text will exhibit a regular pattern, while forged text will exhibit an irregular pattern. The main contributions of this paper are three-fold: (i) exploring the Fourier spectrum and its shape for text forgery detection in videos, natural scenes, and document images. (ii) Combining the spatial information and angular information of the spectrum in a new way for addressing this challenging task, and (iii) the way the

© The Institution of Engineering and Technology 2021

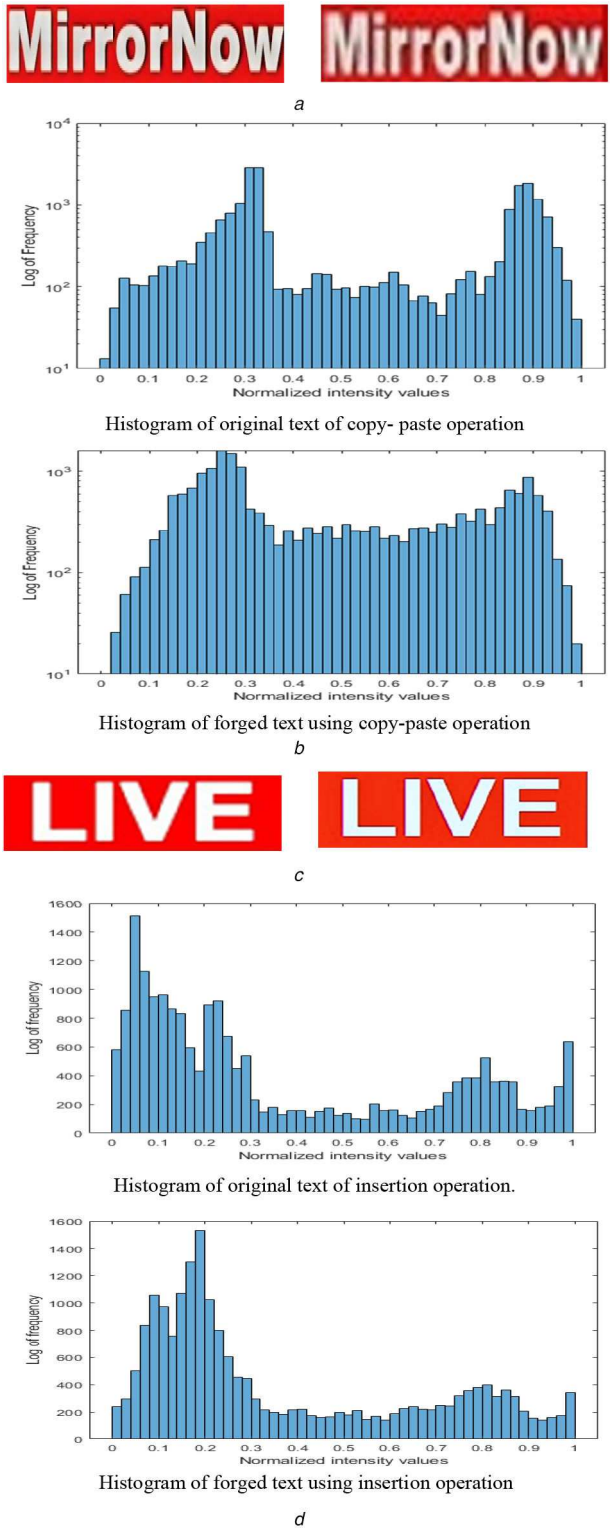

Fig. 2 *Pixel distribution of original and forged text using copy-paste and insertion operations*

(a) Original and forged text of copy-paste operation, ***(b)*** Intensity distribution of original and forged text using copy-paste operation, ***(c)*** Original and forged text of insertion operation, ***(d)*** Intensity distribution of original and forged text using insertion operation

proposed work combines features and neural networks (NNs) is new compared to the existing methods.

3 Proposed methodology

As mentioned in the previous section, the output of a text detection method is the input for our proposed forged text detection. We assume the forgery is created through copy-paste and insertion

Fig. 3 *Examples of BD and shape of the spectrum for the original and forged text using different operations*

(a) Original images for copy-paste and insertion operations, ***(b)*** Fourier spectrum of the images in (a), ***(c)*** Binary form for Fourier spectrum in (b), ***(d)*** Forged text corresponding to the above images in (a), ***(e)*** Fourier spectrum of the forged images in (d), ***(f)*** Binary Fourier spectrum of forged images shown in (e)

operations. Inspired by observations from the literature [8, 9], while forged words may appear authentic at first glance, when we look more closely at the pixel level, the text alignment and background exhibit irregular patterns compared to original text pixels.

This can be seen in the examples displayed in Figs. 2*a–d*, where for the original and forged images, the distribution of pixel values (visualised as histograms) exhibit clear differences between original and forged. This shows that even though the shapes of characters may look the same, there is a significant difference at the pixel level. As can be seen in the figures, text forged via copy-paste appears burred in comparison to original text, while text created through an insertion operation appears bright relative to original text.

This observation leads us to explore the distribution of brightness and the shape of the Fourier spectrum for achieving our goal in this work. The proposed method obtains spectrum for the original and forged texts of copy-paste and insertion operations as shown in Figs. 3*a–f*, where we can see clear differences in BDs and shapes of spectrum between the original text and the forged texts by both the two operations. This is the main basis for feature extraction to detect forged text in this work. At the same time, since the basis is the same for different types of text images (videos, natural scenes, and documents), we believe that the

© The Institution of Engineering and Technology 2021

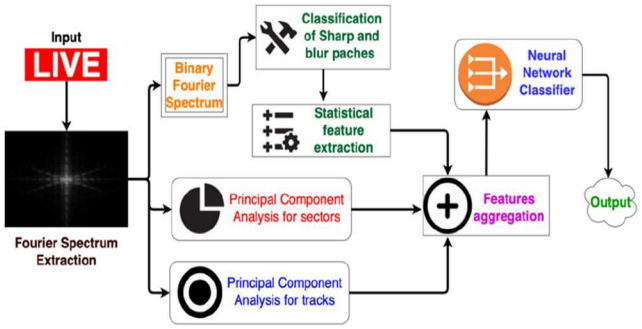

Fig. 4 *Block diagram of the proposed method*

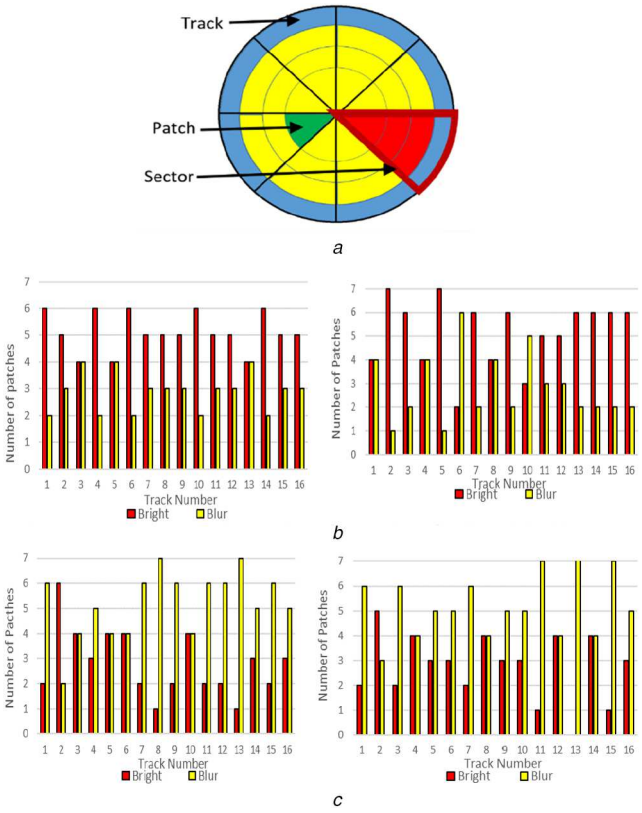

Fig. 5 *Classifying the number of SP and BP from the spectrum to study the BD*

(a) Constructing patches (green colour) for the sectors (red colour) and tracks (blue colour) of the spectrum, ***(b)*** Bar graphs to count the number of SP and BP for the original images shown in Fig. 3*a*, ***(c)*** Bar graphs to count the number of SP and BP for the forged images shown in Fig. 3*d*

aforementioned observations on BD and spectrum shape analysis should be the same. Therefore, the same set of features should work well without depending on the content of the images.

In Fig. 3, we can see the original and forged words created by copy-paste and insert operations, respectively, where we can see BD of spectrum shown in Figs. 3*b* and *e* is different for original and forged images. It is noted from the binary form of spectrum of the original and forged images shown in Figs. 3*c* and *f*, respectively, that spectrum shape of the original image appears different from that of the forged word. When we compare binary spectrum in Figs. 3*c* and *f*, the binary form of forged word loses pixels as, hence it affects the shape of spectrum (SoP). Note that the binary form for Fourier spectrum is obtained as described well-known steps in [35]. Our proposed method extracts statistical features by dividing the spectrum into sectors and tracks, which represent the BD. At the same time, to strengthen the statistical features, we extract angular-based features for the sectors and tracks through principal component analysis (PCA). The statistical and angular-based features are aggregated for the classification of forged text using a NN classifier. The block diagram of the proposed method can be seen in Fig. 4.

3.1 Brightness distribution for forged text identification

We divide the whole spectrum into eight equal divisions (sectors), and then construct rings with a fixed radius over the sectors, which are called tracks. The radius is determined experimentally and is presented in the experimental results. It is known that positive coefficients usually represent high contrast values (edges), while negative coefficients represent low contrast values (background). We propose to use positive and negative coefficients for classifying bright sectors and blur sectors to study the irregularity in BD.

Fig. 5*a* shows eight equal sectors according to angle information, where we can see red colour indicates sectors of the spectrum. The rings which are called tracks are formed with radius of 5 pixels over 8 sectors for the spectrum as shown in Fig. 5*a*, where we can see blue colour indicates tracks of the spectrum. This results in a number of patches denoted by the green colour in Fig. 5*a*. The number of tracks is decided based on the length of the major axis of the spectrum automatically. The radius of 5 pixels is determined empirically in this work.

The brightness of the spectrum depends on the number of positive and negative coefficients. Therefore, we calculate the percentage of positive coefficients (PPC) and negative coefficients (PNC) for each patch of each track as follows:

$$\begin{cases} \text{PPC} = \dfrac{c_{\text{pos}}}{c_{\text{total}}} \times 100 \\ \text{PNC} = \dfrac{c_{\text{neg}}}{c_{\text{total}}} \times 100 \end{cases} \tag{1}$$

where c_{pos} and c_{neg} represent the number of positive coefficients and negative coefficients in each patch of respective tracks, and C_{total} refers to the total number of coefficients in each patch of a track.

As expected, if a patch is bright, PPC should be larger than PNC. With this notion, we derive the condition that if PPC is larger than PNC, then the patch is classified as a sharp patch (SP), otherwise it is classified as a blurred patch (BP) as defined in the following equation:

$$\begin{cases} \text{SP} & \text{if PPC} > \text{PNC} \\ \text{BP} & \text{if PPC} \leq \text{PNC} \end{cases} \tag{2}$$

Our method uses (2) to classify each patch as sharp or blurred with respect to tracks. The proposed method counts how many SP and BP are there for each track, which is illustrated in Figs. 5*b* and *c* for original and forged images using copy-paste and insertion operations. It is observed from Figs. 5*b* and *c* that the count that satisfies (2) for SP is higher than that satisfies (2) for BP in case of the original image, while it is opposite for a forged image. In other words, the bar that represents bright marked by red colour is greater than that represents blur marked by yellow colour, and higher for most of the tracks in case of an original image while for a forged image it is vice versa. This shows that the number of SPs is greater for the original image compared to the number of BPs for a forged image. This makes sense because when an image is affected by a forgery operation, it degrades quality and hence it loses brightness. The proposed method records the total number of SP and BP for each track, which results in two sets, say, T_{SP} and T_{BP}, which are represented as follows:

$$T_{\text{SP}} = \{ S_{t1}, S_{t2}, \ldots, S_{t\alpha} \}$$
$$T_{\text{BP}} = \{ S_{t1}, S_{t2}, \ldots, S_{t\beta} \}$$

where α and β are the number of the members in SP and BP, respectively. The size of both T_{SP} and T_{BP} is the total number of the tracks of the spectrum, that is, $\alpha + \beta$.

To find the distance between the two sets, we calculate the common mean for both the sets, say, M, and then estimate the

© The Institution of Engineering and Technology 2021

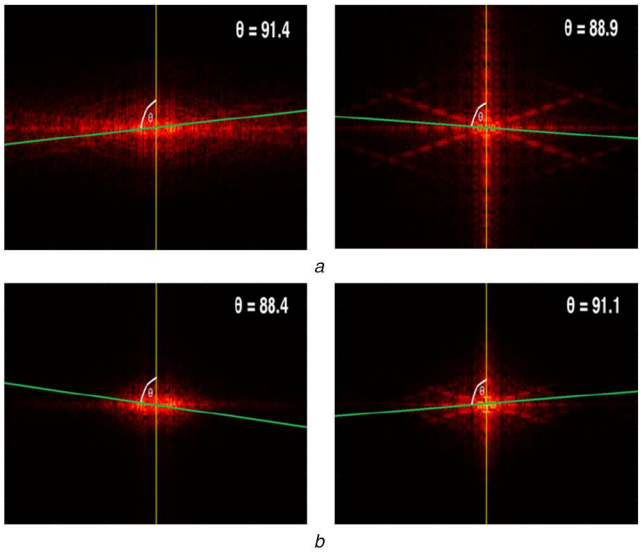

Fig. 6 *Principal axes (green colour line) for the original text and the forged text attacked by copy-paste and insertion operations*
(a) Principal axis (green colour line) for the original images of copy-paste and insertion operations, ***(b)*** Principal axis (green colour line) for the forged text attacked by copy-paste and insertion operations

standard deviation, say, S_M–T_{SP} and S_M–T_{BP}, for each set with respect to the common mean. Similarly, we calculate the standard deviation using the common median (Me), say, S_{Me}–T_{SP} and S_{Me}–T_{BP}, for each set. This procedure results in feature vector-1 (FV-1) which contains four standard deviation values as defined in the following equation:

$$\begin{cases} S_{MTSP} = \text{STD}(M_{TSP}) \\ S_{MTBP} = \text{STD}(M_{TBP}) \\ S_{MeTSP} = \text{STD}(\text{Me}_{TSP}) \\ S_{MeTBP} = \text{STD}(\text{Me}_{TBP}) \end{cases} \tag{3}$$

where STD() refers to the standard deviation, M_{TSP} denotes the common mean for the number of SPs, M_{TBP} denotes the common mean for the number of BPs, Me_{TSP} represents the common median for the number of SPs, and Me_{TBP} represents the common median for the number of BPs.

In the same way, to find the distance between SP and BP, our method employs K-means clustering with $K = 2$ on T_{SP} and T_{BP} sets. This outputs two clusters, namely, Max cluster which contains high values and Min cluster which contains low values. Again, we calculate the mean and standard deviation for Max and Min clusters, separately. This procedure gives four more values, which is labelled as feature vector-2 (FV-2). Note that both FV-1 and FV-2 represent the whole spectrum of the input image. The above steps outputs FV-1 with four features and FV-2 with four more features. In total, the proposed method extracts eight features from the spectrum distribution.

The motivation to use K-means clustering with $K = 2$ is as follows. The sets, namely, T_{SP} and T_{BP}, contain the number of SPs and BPs in each track. For the original image, the number of SPs varies from one track to another. However, the number of SPs is greater than the number of BPs for every track. This is due to the complexity of the background and multiple types images. Therefore, the set, T_{SP} contains high and low values. However, the difference in the values is not significant. In case of a forged image, the number of BPs is greater than the number of SPs for every track. The set, T_{BP} contains high and low values for the forged image. However, the difference between the values is significant in contrast to the original image. This is due to the distortion created by forgery operations. Therefore, we can conclude that both sets have high and low values with different degrees of variation in the values. To exploit this observation, we propose to use K-means with $K = 2$ to obtain two clusters, one called Max which contains the high values, and the other called Min, which contains the low values.

As mentioned in the Introduction, the considered problem is complex, the features extracted from spectrum distribution may not be adequate for achieving better results and hence we propose to extract shape-based features for strengthening the above features, which will be presented in the next section.

3.2 Spectrum shape for forged text identification

When the BD is affected by forgery operations, one can expect the same effect on the shape of the spectrum as shown in Figs. 6*a* and *b* for the original and forged text, respectively. It is observed from Figs. 6*a* and *b* that the principal axis drawn for the original text is not the same as the principal axis drawn for the forged text. Therefore, the proposed method considers the binary form of spectrum as the input for studying shape changes. For the binary spectrum, the proposed method divides it into eight equally sized sectors and tracks as shown in Fig. 5*a*. It then finds the angle for the whole spectrum using PCA, which considers the first eigenvector for drawing principal axis, and labels it as the reference angle (RA). To study the effect of individual sectors and tracks with the whole spectrum shape, the proposed method estimates the angle for all the patches of sector-1 using PCA. Then it estimates the angle using PCA for all the patches of sector-1 and sector-2, and so on. As a result, this process gives eight cumulative angles $\left(\text{CA}_{\text{sec}}^{j}\right)$ for eight sectors. The proposed method computes the absolute difference of each of the eight cumulative angles with the RA, which gives eight difference values $\left(\text{DV}_{\text{sec}}^{j}\right)$ as defined in the following equation:

$$\text{DV}_{\text{sec}}^{j} = \left|\text{CA}_{\text{sec}}^{j} - \text{RA}\right|, \quad \forall j = 1, 2, \ldots, 8 \tag{4}$$

In the same way of obtaining cumulative angles for the sectors, the proposed method estimates the angles for all the patches of track-1, the angles for all the patches of track-1 and track-2, and so on. This process outputs cumulative angles of tracks, which is considered as feature vector-3 containing eight features. The number of cumulative angles $\left(\text{CA}_{\text{trac}}^{j}\right)$ is the number of the tracks of the spectrum. The cumulative angles are compared with the RA to find difference values $\left(\text{DV}_{\text{trac}}^{j}\right)$ w.r.t. tracks as defined in (5). This is considered as feature vector-4 containing 16 features

$$\text{DV}_{\text{trac}}^{j} = \left|\text{CA}_{\text{trac}}^{j} - \text{DV}_{\text{trac}}^{j}\right|, \quad \forall j = 1, 2, \ldots, 16 \tag{5}$$

where 16 is the number of the tracks of the spectrum.

The proposed method combines FV-1, FV-2, FV-3, and FV-4, namely, $4 + 4 + 8 + 16$ as one feature vector, which gives a vector of 32 features. The effect of the features can be seen in Figs. 7*a* and *b*, where it can be seen that the curves that represent the original and forged images of copy and insertion operations behave uniquely. Since the values in the feature vectors represent variation or deviation from the reference point, we expect high variations for the curve of forged images, and smooth variations for the curve of original images. This is valid because when a forged image is affected by forgery operation, the variance of the values increases compared to those in the original image. Therefore, based on the results shown in Fig. 7 it can be seen that the extracted features are good enough to differentiate the original and forged images. This leads to use a NN classifier for classification because when features are good, NN achieves the best results for classification.

We pass the extracted features though the NN architecture as shown in Fig. 8 for classifying forgery texts from the original ones. In this architecture, we use rectified linear unit as an activation function for all the layers except the final layer, where we use 'Sigmoid' [36] activation. With 'Adam' [37] as optimiser and learning rate of 0.01 and 'binary cross-entropy' loss (L_{BCE}) function, the proposed architecture is trained for 60 epochs with the batch size of 8. The loss function is defined in (6), which is binary

© The Institution of Engineering and Technology 2021

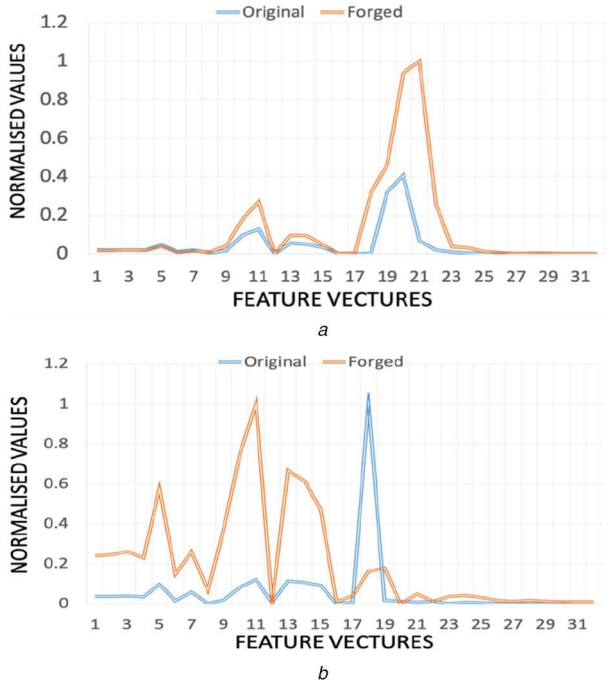

Fig. 7 *Effect of feature extraction for classifying original and forged text*
(a) Feature vectors of original and forged text image for copy-paste operation, ***(b)*** Feature vectors of original and forged text image for insertion operation

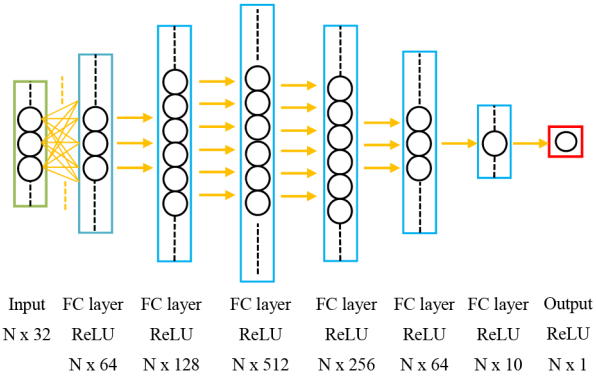

Fig. 8 *Architecture of NN classifier*

cross-entropy loss [38] used for the classification of altered text in PDF document images

$$L_{\text{BCE}} = -\frac{1}{n}\sum_{k=1}^{n}(y_k \times \log(P(y_k)) + (1-y_k) \times \log(1-P(y_k))) \quad (6)$$

where y is the label and $P(y)$ is the predicted probability for the total number of n samples. The dropout layer with rate of 0.2 is added between each layer to reduce overfitting. 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. Since the size of the dataset is small, overfitting and underfitting during training can be a serious issue. To overcome this, we combine testing and validation when evaluating our proposed method.

To alleviate overfitting and underfitting, we also use augmentation techniques, such as random flipping, zooming, rotations, and width and height adjustments to increase the number of samples for training. In addition, since the proposed approach extracts handcrafted features and uses the NN as a classifier and not as a feature extractor, it reduces dependency on the number of samples. For this reason, overfitting and underfitting may not be as significant here. The steps of the complete process are presented in the algorithm (see Fig. 9).

Initialize $Feature\_vectors$ = [] // empty array to store feature vectors of each image from each class
Initialize I = Input image
For each input image I **do**
  $Feature\_vector_I$ = []
  get Fourier Spectrum (F) from image I
  calculate the Binary Fourier Spectrum B from F
  T_{SP} = []; T_{BP} = []
  For each patch P_i **do**
    calculate PPC and PNC
    if $(PPC > PNC) \rightarrow S_{ti}$ = number of sharp patches
      Append S_{t1} to T_{SP}
    else $\rightarrow S_{ti}$ = number of blur patches; **Append** S_{t1} to T_{BP}
    End if
  End For
  calculation of $S_{M_{TSP}}, S_{M_{TBP}}, S_{Me_{TSP}}$ and $S_{Me_{TBP}}$
  Append $S_{M_{TSP}}$, $S_{M_{TBP}}$, $S_{Me_{TSP}}$ and $S_{Me_{TBP}}$ to $Feature\_vector_I$ // Feature Vector 1 (FV1)
  Apply k-means with $k = 2$ on T_{SP} and T_{BP} to get max_c and min_c
  Calculate Mean (M_1&M_2) and Standard deviation (S_1&S_2) of clusters max_c and min_c
  Append M_1, M_2, S_1&S_2 **to** $Feature\_vector_I$
  //Feature Vector 2 (FV2)
  $RA = PCA(F)$ //Principal Component Analysis of whole spectrum
  DV_{sec} = []
  For sectors in **range** $j = 1$ *to* $j = 8$ **do**
    $CA^j_{sec} = PCA(sector(j)) + PCA(sector(j-1)) + \cdots + PCA(sector(1))$
    $DV^j_{sec} = |CA^j_{sec} - RA|$
    Append DV^j_{sec} *to* DV_{sec}
  End For
  Append DV_{sec} *to* $Feature\_vector_I$ // Feature Vector 3 (FV3)
  DV_{trac} = []
  For tracks in **range** $j = 1$ *to* $j = 16$ **do**
    $CA^j_{trac} = PCA(track(j)) + PCA(track(j-1)) + \cdots + PCA(track(1))$
    $DV^j_{trac} = |CA^j_{trac} - RA|$
    Append DV^j_{trac} *to* DV_{trac}
  End For
  Append DV_{trac} *to* $Feature\_vector_I$ // Feature Vector 4 (FV4)
End For
Pre-process and Pass $Feature\_vectors$ along with class labels to the NN classifier to train and/or test it

Fig. 9 *Algorithm: the steps of the proposed method*

4 Experimental results

This section presents an experimental analysis of our method for detecting forged text in videos, natural scene images, and documents. This section consists of six sub-sections. Section 4.1 provides a description of our own dataset as well as standard datasets used for experiment purpose, our evaluation measures, and a list of existing methods used for a comparative study. In order to assess the contribution of the key steps in the proposed method, Section 4.2 examines the performance of each step on our dataset. Section 4.3 aims at validating the proposed method for forgery detection in video images. Experimental results and analysis for the IMEI number dataset are presented in Section 4.4. Section 4.5 tests the method for forgery detection in document images. A similar evaluation for the case of videos is provided in Section 4.6.

4.1 Dataset and evaluation

To evaluate the performance of the proposed method, we create a dataset from different sources, namely, YouTube and News channels and other internet sources. We use the standard operations, copy-paste and insertion stated in the literature for forgery detection in video and images [8, 9] for creating forged data. As discussed in the introduction section, the copy-paste operation is defined as copying from the source frame and then pasting it over a target frame. The insertion operation is defined as inserting words manually with the help of the Paint software/ Photoshop software. The words that we choose from source frames without any operations are considered as the original images and words created using the above two operations are considered as forged words. In this work, the copy-move operation is not considered for creating a forged dataset because the scope for this operation in the case of forging text is not much compared to the copy-paste and insertion operation. In the case of general image

© The Institution of Engineering and Technology 2021

forgery, the copy-move is the main operation to create duplicates in the same images. This challenge is considered a separate research issue for forged image detection. But for the text, the operation does not have much scope.

Our dataset includes 171 forged images created using copy-paste operations, 215 images created using insertion operations and 386original (unaltered) images, giving a total of 772 images for experimentation. Since the proposed NN classifier requires a larger number of samples for training, as mentioned earlier, we employ data augmentation operations, such as random flipping, zooming, rotating, and width and height adjustments to increase the amount of training data.

Sample original and forged images using copy-paste and insertion operations are shown in Fig. 10*a*, where we can see complexity of the classification of original and forged text images. It is noted from Fig. 10*a* that it is not so easy to differentiate original and forged images with naked eyes without any knowledge about the dataset. Note that the forgery is done at the word level in our dataset.

To test the effectiveness of the proposed method, we consider benchmark dataset [24], which consists of IMEI forged number images. In this dataset, for each IMEI number, 2–3 numerals are replaced either by copy-paste or insertion. Sample original and forged images are shown in Fig. 10*b*, where we can see difficulty in differentiating original and forged numbers. This dataset provides 500 original and 500 forged IMEI number images for experimentation. The images of this dataset do not have complex background as video and natural scene images considered in this work.

Fig. 10 *Examples of successful classification of the proposed method for our own and benchmark datasets. In case of Bhardwaj dataset, the images in (e) are only caption text (forged) images*

(a) Our dataset insertion operation, ***(b)*** IMEI number dataset, ***(c)*** ICPR 2018 Fraud Contest dataset, ***(d)*** Scene (original) and caption (forged) text chosen from Roy *et al.* dataset, ***(e)*** Caption texts chosen from Set-1 and Set-2 of Bhardwaj dataset

To test the robustness of the proposed method for detecting altered text detection in document images, we consider the benchmark dataset called 'ICPR 2018 Fraud Contest Data' (FCD) [39]. This provides forged text extracted from receipt images, where one can expect a plain background. In this dataset, forgery is done at the character level because there can be altered numerals in receipts to create fake prices. The main challenge of this dataset is the length of texts is too small due to strings or numerals with currency symbol. In addition, altering one number in the string of a few numerals makes the dataset more complex and challenging. The dataset provides 300 original and 302 altered samples, which gives total 602 images for experimentation. For our experimentation, since the proposed method requires text lines of the original and altered documents, we segment the altered part from respective document images. The ground truth for altered regions is marked by rectangles in all the documents in the dataset. We segment each rectangle region from the altered documents automatically. For the original documents, since almost all the documents are receipts and have prices in the same location, where a rectangle is drawn in the altered documents, we use the same segmentation step for extracting price information from the original documents. Sample images of the original and altered texts are shown in Fig. 10*c*, where difference is noticeable.

The scope of the proposed work is to use video images for forged text classification. If this is the case, it is noted from the methods [33, 34] that caption text in video image is considered as tampered because caption text is edited text over video. In the same video, it is expected text which is part of the image considered as scene text. Further, to test the ability of the proposed method, we consider two standard datasets, namely, Roy *et al.* [33] which consists of captions (forged) and scene text (original). These images are collected from different benchmark video sources, such as ICDAR 2013, ICDAR 2015, YVT, and authors' own sources. The dataset provides 650 forged images and 900 original images, which gives 1550 images for experimentation. Another standard dataset [34], which is similar to Roy *et al.* dataset, consists of two sets of images of different resolutions, namely, set-1 that contains 1280×720 pixels and set-2 that contains 1920×1080 pixels images. These two sets provide only forged text and therefore, we use original images of Roy *et al.* dataset for producing confusion matrix in this work. Set-1 provides 2233 and Set-2 a 2415, which gives total 4648 images for experimentation. Then sample images of Roy *et al.* and Bharadwaj datasets are shown in Figs. 10*d* and *e*, respectively, where we can see images with different complexities. These two sets help us to test the ability of the proposed method for the images of different resolution which can be independent of scaling. In summary, we consider 8574 images for evaluating the proposed and existing methods and it is reported in Table 1.

To measure the performance of the proposed and existing methods, we use standard metrics, namely, confusion matrix, average classification rate, and *F*-score. Classification rate is defined as the number of images classified correctly divided by the actual number of images. The average classification rate is defined as the mean of diagonal elements in the confusion matrix. The confusion matrix helps us to identify the misclassification error while the average classification rate is to measure the proposed method quantitatively. The *F*-score is the harmonic mean of recall

Table 1 Detail of different dataset used for evaluation

Datasets	Our dataset	IMEI dataset	ICPR 2018 FCD	Roy *et al.* dataset	Bhardwaj *et al.*	
					Set-1	Set-2
original	386	500	300	650	—	—
forged	386	500	302	900	2233	2415
total	772	1000	602	1550	4648	

© The Institution of Engineering and Technology 2021

and precision and helps us to evaluate whether the proposed method is accurate or not.

For comparison purposes, we tested a number of other methods for related problems. These include Shivakumara *et al.* [24] which was developed for forged IMEI number detection; Wang *et al.* [23], which was developed for printer identification; Nandanwar *et al.* [30], which was developed for altered text detection in document images. Since the objective of these three methods is the same as our proposed method, they make good choices for a comparative study. Similarly, we also implemented Roy *et al.* [33], which was developed for tampered text (caption) classification; Bhardwaj and Pankajakshan [34], which was developed for tampered text detection in video; Bibi *et al.* [25], which proposes a text-independent approach for printer source identification. Overall, to show that the above-mentioned methods are not effective to address the challenges of forgery detection in all three types of images, we consider the methods, which involve different approaches for forgery detection for our comparative study.

The proposed method draws sectors and tracks over the spectrum of the input image for feature extraction as in the methodology section. For deciding the number of sectors, we use angle information to determine 8, which divides the whole image into 8 equal sized sectors. In case of tracks, we fix 5-pixel radius between the tracks. To determine whether the value of 5 is optimal or not, we conduct experiments on samples chosen randomly from all the databases to calculate the average classification rate as shown in Fig. 11. It is noted from Fig. 11 that the average classification rate increases gradually as radius increases, but when radius value is 5, the average classification rate reaches the maximum and then drops. Therefore, the value of 5 is considered as the optimal value for determining the number of tracks. This shows that when radius value is small, the proposed method may lose vital information, and when it is large, the discriminative power reduces due to redundancy. Thus, 5 is a feasible value for achieving better results irrespective of different image types. Note that when the size of the images changes, the number of tracks also changes. This may affect the performance of the method and add complexity to the implementation, which will be discussed in Section 4.6. To overcome this issue, the size of the images is converted to a standard size of 160×160, which was determined based on the average size of the input images.

4.2 Ablation study

Our approach consists of three key steps for achieving the best results, namely, exploring BD, SoP, and the use of a NN classifier. Thus, we report on experiments to test the effectiveness of each of these key steps. We compute a confusion matrix, the average classification rate, and the *F*-score for using only features of BD, only features of SoP, and replacing NN with a conventional SVM classifier on the copy-base and insertion datasets, as reported in Table 2. From Table 2 we see that the features extracted from BD yield a better average classification rate and *F*-score than the features extracted from SoP for copy-paste forgeries, while of the opposite is true for insertion operations. This shows that for the latter, shape-based features are better, while for the former, BD-based features are better. This makes sense because copy-paste operation introduces more inconsistency during forgery compared to insertions. This is due to inserting a few characters in a word and hence one cannot expect high variation compared to copy-paste operations. However, we can expect some variations in shape and background colour compared to the original. This shows that BD and SoP are effective and contribute to achieve the best results as reported in Table 2.

Similarly, we also evaluate the proposed method by feeding the features to the SVM classifier on our dataset as reported in Table 2. It is evident from the results reported in Table 2 that the proposed method with NN is better than the proposed method with SVM. This is because NN involves more weights to learn the samples properly compared to SVM. In addition, the handcrafted features appear to work well for the SVM, which is more evidence that the feature selection process was a good choice. Overall, Table 2 shows that the key steps by themselves are effective, but not comparable to the results of the proposed method which combines all of them into one system.

4.3 Evaluation on our video image dataset

The quantitative results of the proposed and existing methods are reported in Table 3, where it is seen that the proposed method achieves the best average classification rate for both forgery datasets. When we compare the results among the existing methods, the method in [34] scores the best average classification rate and *F*-score for the copy-paste dataset. This is because the method is developed for studying compression errors caused by

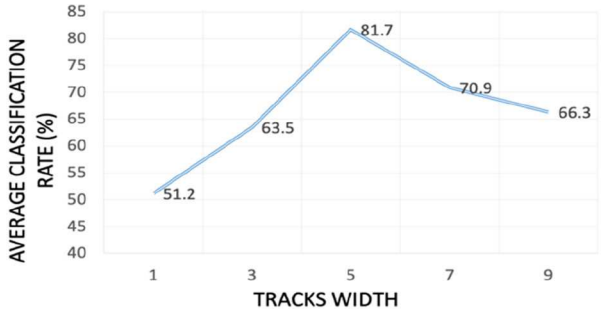

Fig. 11 *Illustrations to determine the feasible value for the radius of the tracks*

Table 2 Classification rate and confusion matrix of BD and SoP steps for forgery detection on our dataset (in %)

Operation		BD		SoP		Proposed + SVM		Proposed + NN	
O	F	O	F	O	F	O	F		
copy-paste	O	95.9	4.1	41.69	58.31	65.9	34.1	77.64	22.36
	F	64.3	35.7	1.75	98.25	25.7	74.3	16.3	83.7
	average	65.80		69.97		70.1		**80.67**	
	F-score	73.71		58.13		68.79		**80.08**	
insertion	O	96.06	3.94	79.53	20.47	73.4	26.6	80.6	19.4
	F	71.62	28.38	16.74	83.26	23.8	76.2	15.7	85.3
	average	62.22		81.4		74.8		**82.9**	
	F-score	71.77		81.04		74.44		**82.12**	

O: original and F: forged.
Bold indicates the best results achieved by the method

© The Institution of Engineering and Technology 2021

blur and distortion. However, it is worst for the insertion data compared to [23, 25, 30, 33]. This is because insertion operations do not introduce much blur, rather they change character shapes. The method in [30] achieves the best *F*-score compared to the proposed method and other existing methods for the insertion data because the features proposed in [30] are robust, which work based on the quality of the images but not extracting specific features. The method [25] has the best average classification rate for the insertion data. The reason is that this method uses a powerful deep learning model, which is a text-independent approach to forgery detection. Overall, the proposed method is better than all the existing methods for both the copy-paste and insertion data. This is not surprising as the existing methods were developed with one particular input type in mind. On the other hand, since the proposed method combines the advantage of BD and spectrum shape analysis, the proposed method detects forgery irrespective of the content of images and image types.

4.4 Evaluation on IMEI number dataset

The IMEI dataset is considered scene text images because the text is embedded on different mobile cases and covers (pack). The mobile cover and cases usually have a different background like scene images according to mobile. The main issue with this dataset is that 2–3 characters are used for creating forged images, which introduces little distortion. The results of the proposed and existing methods are reported in Table 4, where it can be seen that the proposed method is slightly better than the existing methods in terms of average classification rate and *F*-score. When we look at the results of the existing methods, the method [30] scores the highest results compared to the other existing methods. This is because the features extracted are based on the quality of the images, while the other existing methods use specific features based on text information. The methods in [33, 34] are developed for video and natural scene text and hence report poor results compared to the proposed method and [30]. Although, the method [24] is developed for forged IMEI number detection, it produces lower quality results compared to the methods [25, 30] and the proposed method. This is because the performance of the method depends on Canny and Sobel edge images, which are sensitive to complex backgrounds. On the other hand, the method [25] uses CNN for classification, which is more powerful for discriminating while the method [30] works based on the quality of the images, which are robust features.

4.5 Evaluation on ICPR 2018 fraud detection contest dataset

This dataset consists of forged receipts, where characters are altered to change the prices of items. The forgeries take place at both the character and word levels. The results of the proposed and existing methods are reported in Table 5, where it can be observed that the proposed method is the best for average classification rate and *F*-score. Among the existing methods, the method in [25] achieves the best average classification rate and *F*-score compared to other existing methods. The reason is that this method is a text-independent and it has the advantage of the deep learning model. In the same way, the method [30] also achieves better results than [23, 24, 33, 34]. This is because the method is developed to detect altered text detection in document images similar to the images of FCD. However, for the other existing methods, the scope does not match with the FCD dataset. When we compare the results of the proposed method to ours, IMEI number and FCD datasets, our approach achieves better results for the FCD dataset compared to the other datasets. This makes sense because images of the FCD dataset do not have much background variations compared to images of the other two datasets.

Table 3 Classification rate and confusion matrix of the proposed and the existing methods on our dataset (in %)

Operation	Classes	Proposed		[24]		[33]		[34]		[23]		[30]		[25]	
		O	F	O	F	O	F	O	F	O	F	O	F	O	F
copy-paste	O	77.6	22.3	75.6	24.4	76.0	24.0	90.0	10.0	72.2	27.8	72.1	27.9	67.5	32.5
	F	16.3	83.7	51.1	48.9	27.0	73.0	29.0	71.0	84.4	15.6	38.8	61.2	41.6	58.3
	average	**80.67**		62.25		74.5		80.5		43.9		66.65		62.9	
	F-score	**80.08**		66.70		74.88		82.19		56.27		68.37		64.5	
insertion	O	80.6	19.4	69.4	20.6	68.0	32.0	68.0	32.0	81.8	8.2	78.4	11.6	83.3	16.6
	F	15.7	85.3	47.2	52.8	30.0	70.0	38.0	62.0	29.1	70.9	11.3	79.7	22.5	77.5
	average	**82.9**		61.1		60.9		65.0		76.35		79.05		80.4	
	F-score	**82.12**		67.18		68.69		66.02		81.43		87.26		80.9	

Bold indicates the best results achieved by the method

Table 4 Classification rate and confusion matrix of the proposed and the existing methods on IMEI number [24] dataset (in %)

Classes	Proposed		[24]		[33]		[34]		[23]		[30]		[25]	
	O	F	O	F	O	F	O	F	O	F	O	F	O	F
O	83.6	16.4	82.2	17.8	22.4	77.6	34.0	66.0	83.2	16.8	84.4	15.6	83.6	16.4
F	13.2	86.8	18	82	5.6	94.4	29.6	70.4	25.6	74.4	14.2	85.8	18.1	81.9
average	**85.2**		82.1		58.4		52.2		78.8		85.1		82.75	
F-score	**84.99**		82.12		35.00		41,46		79.69		84.9		82.90	

Bold indicates the best results achieved by the method

Table 5 Classification rate and confusion matrix of the proposed and the existing methods on ICPR 2018 Fraud Contest Dataset (in %)

Classes	Proposed		[24]		[33]		[34]		[23]		[30]		[25]	
	O	F	O	F	O	F	O	F	O	F	O	F	O	F
O	93.6	6.4	92.0	8.0	39.7	60.2	27.6	72.4	84.6	15.4	84.0	16.0	92.1	7.9
F	3.1	96.9	49.4	50.6	5.1	94.9	6.07	93.9	10.7	89.3	4.5	95.5	5.5	94.5
average	**95.25**		71.33		67.32		60.77		86.99		89.1		93.33	
F-score	**95.17**		76.22		54.87		29.31		86.9		89.12		93.22	

Bold indicates the best results achieved by the method

© The Institution of Engineering and Technology 2021

4.6 Evaluation on Roy et al. and Bhardwaj et al. video

In these two datasets, caption which is edited text in the video is considered as forged text. The text which is a part of an image is considered as the original text. As a result, we can expect forgies at the word level and variations in the background. To test whether the proposed method is invariant to resolution and scaling, the Bhardwaj dataset provides two sets of datasets with different resolutions. Set-1 contains images of low resolution, while Set-2 dataset contains images of high resolution.

Quantitative results of the proposed and the existing methods for all the three datasets are reported in Tables 6–8, respectively. It is observed from Tables 6–8 that the proposed method is the best at average classification rate and *F*-score for all three datasets. Interestingly, the methods in [33, 34] score low results compared to the method [23, 25, 30] for all the datasets. This is due to the use of the heuristic rule in the methods [33, 34], while the methods [23, 25] use the combination of features and classifiers. Since the classifier is better than fixing rigid conditions, the methods [33, 34] report poor results. On the other hand, the proposed method achieves the best for all the three datasets compared to the existing methods in terms of average classification rate and *F*-score. It is noted from Tables 6–8 that the proposed method scores better results for the Set-2 dataset compared to Roy *et al.*'s and Set-1 datasets. This is understandable because the Set-2 dataset provides high-resolution images, which have high contrast and clarity. In the case of the other two datasets, the images suffer from low resolution and low contrast.

In summary, the experimental analysis on the various datasets shows that our proposed method is content independent, image type independent, and applications independent.

For creating forged text in our dataset, we randomly replace a single character, two characters, and so on, and sometimes, all the characters in the original images using copy-paste and insertion operations. Therefore, to analyse the performance of the proposed method on a different sized forgery regions, we calculate the average classification rate for varying degrees of forgery. For example, if the insertion operation alters one character in the original images of four characters, we consider it to be a 25% forgery region. In the same way, if all the original characters are replaced by the forged characters, it can be 100% forgery region as shown in Fig. 12, where we can see sample forged images for different percentage of forged regions. We plot a graph for average classification rate versus varying percentage of forgery regions as shown in Fig. 13*a*, where it is noted that as the percentage of forgery region increases, the performance of the proposed method

Table 6 Classification rate and confusion matrix of the proposed and the existing methods on Roy *et al.* dataset (in %)

Classes	Proposed		[24]		[33]		[34]		[23]		[30]		[25]	
	O	F	O	F	O	F	O	F	O	F	O	F		
O	91.1	8.9	40.7	59.3	75.0	25.0	56.0	44.0	86.3	13.6	83.5	16.5	86.6	13.3
F	14.4	85.5	10.0	90.0	37.0	63.0	60.0	40.0	10.9	89.1	10.1	89.9	11.3	88.7
average	**88.32**		65.85		69.0		48.0		87.73		86.7		87.71	
F-score	**88.64**		54.01		70.75		51.85		92.14		86.26		87.53	

Bold indicates the best results achieved by the method

Table 7 Classification rate and confusion matrix of the proposed and the existing methods on Bhardwaj *et al.*-Set-1 dataset (in %)

Classes	Proposed		[24]		[33]		[34]		[23]		[30]		[25]	
	O	F	O	F	O	F	O	F	O	F	O	F	O	F
O	93.8	6.2	95.2	4.8	66.0	34.0	91.3	8.7	91.9	8.1	84.8	15.2	78.3	21.7
F	10.8	89.2	44.6	55.4	36.0	64.0	81.6	18.3	9.1	90.9	17.2	82.8	18.3	81.6
average	**91.59**		69.95		65		54.81		91.4		83.8		79.98	
F-score	**91.69**		79.40		65.35		66.89		91.44		83.96		79.64	

Bold indicates the best results achieved by the method

Table 8 Classification rate and confusion matrix of the proposed and the existing methods on Bhardwaj *et al.*-Set-2 dataset (in %)

Classes	Proposed		[24]		[33]		[34]		[23]		[30]		[25]	
	O	F	O	F	O	F	O	F	O	F	O	F	O	F
O	95.9	4.1	95.7	4.3	62.0	38.0	93.3	6.7	92.4	7.6	89.6	10.4	88.2	11.8
F	2.2	97.8	61.3	38.7	33.0	67.0	72.7	27.3	10.7	89.3	12.1	87.9	9.6	90.4
average	**96.8**		67.2		64.5		60.3		90.85		88.75		89.3	
F-score	**96.82**		74.47		63.59		70.15		90.99		88.84		89.18	

Bold indicates the best results achieved by the method

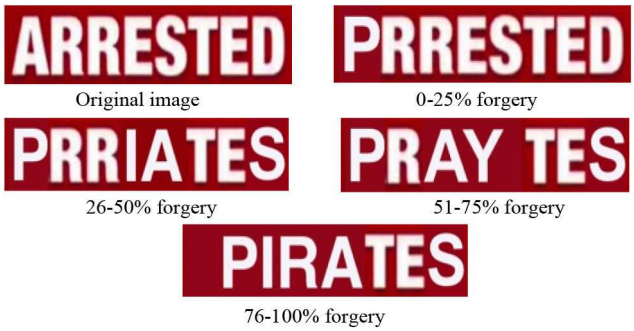

Fig. 12 *Sample images with different percentage of forgery using for insertion operation*

© The Institution of Engineering and Technology 2021

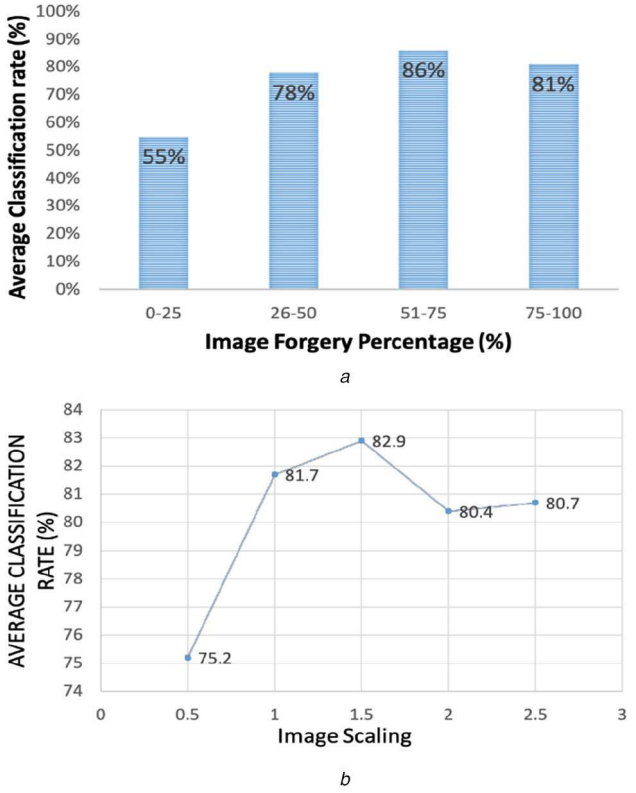

Fig. 13 *Testing the robustness of the proposed method on different scaling and percentage of image forgery*

(a) Average classification rate of the proposed method for different percentage of image forgery, ***(b)*** Average classification rate of the proposed method for different scaling

also increases (up to a 75% forgery region). However, the performance of the proposed method decreases after 75%. The reason is that when a whole word is forged, the feature loses discriminative power because the word does not provide inconsistency pattern for the feature extraction in contrast to the word having both original and forged characters. Furthermore, when the words contain a very small portion of the forgery region (<25%), the performance of the proposed method degrades. Therefore, with this analysis, one can conclude that the forged image must have a 51–75% forged region to achieve the best results. However, in this work, the forged dataset includes the images with the forged region from the range of 25–100% for experimentation.

To study the impact of image size variations on determining the number of tracks for extracting feature, we calculate average classification rate for our dataset. This is because the number of tracks depends on the size of the images. It is illustrated for varying scale of the input images in Fig. 13*b*. It is observed from Fig. 13*b* that as scale increases, the average classification rate also increases up to the 1.5 scaling ratio. After a scaling of 1.5, the average classification rate decreases. This shows that when the image size decreases to 0.5 and increases to more than 1.5, the image loses quality, and it results in a degraded image. Therefore, the performance of the proposed method degrades. However, the scaling up and down of the images does not affect for overall the performance of the proposed method because in this work, the size of the input images is converted to a standard size of 160 × 160.

In summary, the above experimental analysis shows that the proposed method is robust to different types of images, and it works well for forgery detection irrespective of content of images. Hence, the method is generic in nature and can be extendable for other application and datasets. The scope of the work is limited to forgery detection created using copy-paste and insertion operations. However, there are counter-measures [40–43] that use anti-forensic operations to mislead or confuse forgery detection methods through image processing operations, such as sharpening, blurring, and so on. In this case, detecting forgeries becomes more challenging, which is a topic beyond the scope of this work. However, since our method considers inconsistencies and irregular patterns for forgery detection, we believe it may be capable of detecting forgeries even in the presence of anti-forensic attacks. This is because the anti-forensic attack is also one kind of forgery operation and hence inconsistency is inevitable. It is true that as the number of operations increases, the quality of the text image decreases, and hence small abrupt changes are detectable. Therefore, our method can be extended to counter anti-forensic attack. This is a new direction of the proposed work and this will be our future target.

5 Conclusion and future work

We have proposed a method based on analysing Fourier coefficients and shapes of Fourier spectrum for identifying forgery in images of different types, including video, scene, and document types. The proposed method explores characteristics of spectrum distribution and shapes of the spectrum for extracting features. To achieve the best results for different type images, the proposed method integrates the advantage of BD of spectrum and shape of the spectrum for forgery identification. The extracted features are fed to a NN for classification. To validate the proposed method, we conduct experiments on our own dataset created by copy-paste and insertion, the IMEI number dataset, the ICPR 2018 Fraud Detection Contest dataset, and two more natural scene image datasets to show that the proposed method is effective and generic.

6 Acknowledgment

Tong Lu, Palaiahankote Shivakumara, and Umapada Pal received support for this work from the National Natural Science Foundation of China under grant no. 61672273. Palaiahnakote Shivakumara also received partial support from Faculty Grant (GPF096A-2020, GPF096B-2020 and GPF096C-2020), University of Malaya, Malaysia. They are thankful to D. Bhardwaj and V. Pankajakshan, Department of Electronics and Computer Engineering, Indian Institute of Technology, Roorkee, Utarakhanda, India and S. Roy, University of Malaya, Malaysia for sharing datasets and codes for experimentation and comparison.

7 References

[1] Chen, S., Tan, S., Li, B., *et al.*: 'Automatic detection of object based forgery in advanced video', *IEEE Trans. CSVT*, 2016, **26**, (11), pp. 2138–2151

[2] Su, L., Li, C., Lai, Y., *et al.*: 'A fast forgery detection algorithm based on exponential Fourier moments for video region duplication', *IEEE Trans. Multimed.*, 2018, **20**, (4), pp. 825–840

[3] Selvaraj, P., Karuppiah, M.: 'Enhanced copy-paste forgery detection in digital images using scale invariant feature transform', *IET-Image Process.*, 2020, **14**, pp. 462–471

[4] Soni, B., Das, P. K., Thounaojam, D. M.: 'Keypoints based enhanced multiple copy-move forgeries detection system using density spatial clustering if application with noise clustering algorithm', *IET-Image Process.*, 2018, **12**, pp. 2082–2099

[5] Fadi, S.M., Han, Q., Li, Q.: 'Inter-frame forgery detection based on differential energy of residue', *IET-Image Process.*, 2019, **13**, pp. 522–528

[6] D'Amiano, L., Cozzolino, D., Poggi, G., *et al.*: 'A PatchMatch-based dense-field algorithm for video copy-move detection and localization', *IEEE Trans. Circuits Syst. Video Technol.*, 2019, **29**, (3), pp. 669–682

[7] Feng, C., Xu, Z., Jia, S., *et al.*: 'Motion adaptive frame deletion detection for digital video forensic', *IEEE Trans. CSVT*, 2017, **27**, (12), pp. 2543–2554

[8] Pun, C.M., Yuan, X.C., Bi, X.L.: 'Image forgery detection using adaptive over segmentation and feature point matching', *IEEE Trans. IFS*, 2015, **10**, (8), pp. 1705–1716

[9] Yang, F., Li, J., Lu, W., *et al.*: 'Copy-move forgery detection based on hybrid features', *Eng. Appl. Artif. Intell.*, 2017, **59**, pp. 73–83

[10] Tian, X., Zhou, G., Xu, M.: 'Image copy-move forgery detection algorithm based on ORB and novel similarity metric', *IET- Image Process.*, 2020, **14**, pp. 2092–2100

[11] Yin, X.-C., Zuo, Z.-Y., Tian, S., *et al.*: 'Text detection, tracking and recognition in video: A comprehensive survey', *IEEE Trans. Image Process.*, 2016, **25**, pp. 2752–2773

[12] Baek, Y., Lee, B., Han, D., *et al.*: 'Character region awareness for text detection'. Proc. Computer Vision and Pattern Recognition, Long Beach, CA, USA, 2019, pp. 9365–9374

[13] Barboza, R.D.S., Lins, R.D., Lira, E.D.F.D., *et al.*: 'Later added strokes of text fraud detection in documents written with ballpoint pens'. Proc. 14th Int. Conf. on Frontiers in Handwriting Recognition, Crete, Greece, 2014, pp. 517–522

[14] Elkasrawi, S., Shafait, F.: 'Printer identification using supervisedlearning for document forgery detection'. Proc. Document Analysis Systems, Tours, France, 2014, pp. 146–150

© The Institution of Engineering and Technology 2021

[15] Ahmed, A., Shafait, F.: 'Forgery detection based on intrinsic document features'. Proc. Document Analysis Systems, Tours, France, 2014, pp. 252–256
[16] Shafait, K.F., Mian, A.: 'Automatic ink mismatch detection for forensic document analysis', *Pattern Recognit.*, 2015, **48**, pp. 3615–3626
[17] Luo, Z., Shafait, F., Mian, A.: 'Localized forgery detection in hyperspectral document images'. Proc. 13th Int. Conf. on Document Analysis and Recognition (ICDAR), Nancy, France, 2015, pp. 496–500
[18] Bertrand, R., Kramer, P.G., Terrades, O.R., *et al.*: 'A system based on intrinsic features for fraudulent document detection'. Proc. 13th Int. Conf. on Document Analysis and Recognition (ICDAR), Washington, DC, USA, 2013, pp. 106–110
[19] Barboza, R.D. S., Lins, R.D., Jesus, D.M.D.: 'A color based model to determine the age of documents for forensic purpose'. Proc. 13th Int. Conf. on Document Analysis and Recognition (ICDAR), Washington, DC, USA, 2013, pp. 1350–1353
[20] Halder, B., Garain, U.: 'Color feature based approach for determining ink age in printed documents'. Proc. 20th Int. Conf. on Pattern Recognition, Istanbul, Turkey, 2010, pp. 3212–3215
[21] Kumar, R., Pal, N.R., Chanda, B., *et al.*: 'Forensic detection of fraudulent alterations in ball point pen strokes', *IEEE Trans. IFS*, 2012, **7**, (2), pp. 809–820
[22] Raghunandan, K. S., Shivakumara, P., Navya, B. J., *et al.*: 'Fourier coefficients for fraud handwritten document classification through age analysis'. Proc. 15th Int. Conf. on Frontiers in Handwriting Recognition (ICFHR), Shenzhen, People's Republic of China, 2016, pp. 25–30
[23] Wang, Z., Shivakumara, P., Lu, T., *et al.*: 'Fourier-residual for printer identification'. Proc. Int. Conf. on Document Analysis and Recognition (ICDAR), Kyoto, Japan, 2017, pp. 1114–1119
[24] Shivakumara, P., Basavaraja, V., Gowda, H. S., *et al.*: 'A new RGB based fusion for forged IMEI number detection in Mobile images'. Proc. Int. Conf. on Frontiers in Handwriting Recognition (ICFHR), Niagara Falls, NY, USA, 2018, pp. 386–391
[25] Bibi, M., Hamid, A., Moetesum, M., *et al.*: 'Document forgery detection using printer source identification: A text independent approach'. Proc. Int. Conf. on Document Analysis and Recognition Workshops (ICDARW), Sydney, Australia, 2019, pp. 7–12
[26] Kalbitz, M., Vielhauer, C.: 'Automated forensic ink determination in handwritten document by clustering'. Proc. EUSIPCO, A Coruna, Spain, 2019
[27] Mukhtar, M., Malhotra, D. D.: '$S_iS_bd_p$-the technique to identify forgery in legal handwritten documents'. Proc. ICSSIT, Tamil Nadu, India, 2020, pp. 1103–1108
[28] Rahiche, A., Cheriet, M.: 'Forgery detection in hyperspectral document images using graph orthogonal nonnegative matrix factorization'. Proc. CVPRW, Seattle, WA, USA, 2020, pp. 2823–2831
[29] Chen, Y., Gao, S.: 'Forgery numeral handwriting detection based on convolutional neural networks'. Proc. ITOEC, Chongqing, People's Republic of China, 2020, pp. 201–205
[30] Nandanwar, L., Shivakumara, P., Pal, U., *et al.*: 'A new method for detecting altered text in document images'. Proc. Int. Conf. on Pattern Recognition and Artificial Intelligence, Milan, Italy, 2020
[31] Shivakumara, P., Kumar, N.V., Guru, D.S., *et al.*: 'Separation of graphics (superimposed) and scene text in videos'. Proc. Document Analysis Systems, Lyon, France, 2014, pp. 344–348
[32] Xu, J., Shivakumara, P., Lu, T., *et al.*: 'Graphics and scene text classification in video'. Proc. Int. Conf. on Pattern Recognition, Stockholm, Sweden, 2014, pp. 4714–4719
[33] Roy, S., Shivakumara, P., Pal, U., *et al.*: 'New tampered features for scene and caption text classification in video frame'. Proc. Int. Conf. on Frontiers in Handwriting Recognition (ICFHR), Kolkata, India, 2016, pp. 36–41
[34] Bhardwaj, D., Pankajakshan, V.: 'Image overlay text detection based on JPEG truncation error analysis', *IEEE Signal Process. Lett.*, 2016, **23**, (8), pp. 1027–1031
[35] Ko, J., Kim, C.: 'Low cost blur image detection and estimation for mobile devices'. Proc. 11th Int. Conf. on Advanced Communication Technology, Gangwon-Do, Republic of Korea, 2009, pp. 1605–1610
[36] Narayan, S.: 'The generalized sigmoid activation function: competitive supervised learning', *Inf. Sci.*, 1997, **99**, pp. 69–82
[37] Kingma, P. D., Bai, J. L.: 'Adam: A method for stochastic optimization'. Proc. ICLR, San Diego, CA, USA, 2015, pp. 1–15
[38] Nasr, G.E., Badr, E.A., Joun, C.: 'Cross entropy error function in neural networks: forecasting gasoline demand'. Proc. of the Fifteenth Int. Florida Artificial Intelligence Research Society Conf., Pensacola Beach, FL, USA, 2002, pp. 381–384
[39] Artaud, C., Sidère, N., Doucet, A., *et al.*: 'Find it! fraud detection contest report'. Proc. 24th Int. Conf. on Pattern Recognition (ICPR), Beijing, People's Republic of China, 2018, pp. 13–18
[40] Li, X., Yan, D., Dong, L., *et al.*: 'Anti-forensics of audio source identification using generative adversarial network', *IEEE Access*, 2019, **7**, pp. 184332–184339
[41] Wu, J., Wang, Z., Zeng, H., *et al.*: 'Multiple operation image anti-forensics with WGAN-GP framework'. Proc. Asia-Pacific Signal and Information Processing Association Annual Summit and Conf. (APSIPA ASC), Lanzhou, People's Republic, 2019, pp. 1303–1307
[42] Singh, G., Singh, K.: 'Counter JPEG anti-forensics approach based on the second-order statistical analysis', *IEEE Trans. IFS*, 2019, **14**, pp. 1194–1209
[43] Chen, C., Xiong, Z., Liu, X., *et al.*: 'Camera trace erasing'. Proc. Computer Vision and Pattern Recognition, 2020, pp. 2947–2956

© The Institution of Engineering and Technology 2021