DETECTION OF FAKE COLORIZED IMAGES USING DENSE CONVOLUTION NETWORK

S Aashmi Miranda1 , Dr K Thulasimani2

1,2Department of Computer Science and Engineering,
Government College of Engineering, Tirunelveli

Abstract- ***Image forgery implies altering the digital image to some meaningful or valuable data. Image forensics is a well developed field that analyzes the images of specific conditions to build up trust and genuineness. Although image editing techniques can provide significant appreciation of the image or entertainment value, they may also be used with malicious intent. An emerging image editing technique is colorization, in which gray scale images are colorized with realistic colors. But this technique may also be intentionally applied to certain images to confound object recognition algorithms. In this work, it is observed that, colorized images, usually change the images using a variety of mechanisms. The digital image developed from the colorization Method possess statistical differences in their RGB channels, hue and saturation channels and also need to observe statistical inconsistencies in the dark and bright channels, because the colorization process will mainly affect the dark and bright channel values. Based on the experiments in the hue, saturation, dark and bright channels, two simple yet effective detection method Histogram based and Feature Encoding based Fake Colorized Image Detection along with the dense convolution network are proposed for detecting the fake colorized images***

Index Terms- ***image forgery, image detection, histogram, feature encoding, neural network.***

I. INTRODUCTION

Image forensics is a well developed field that observes the images of specific conditions to build up trust and honesty. Numerous image forensic technologies have been developed in the previous years. Image Forgery techniques classified into two classes, active and passive techniques. The active techniques usually called the hiding techniques.When the integrities of these images demand verification, watermark extraction procedures The original images are compared to the watermark images to detect forgeries. Since the active techniques require the watermark to be embedded prior to detection are limited in their applications.

Our proposed methods belong Passive image forgery detection approaches, to which, usually detect the manipulations to the input images directly. If these images are examined by humans, the cost increases peakly as the number of to-be-examined images increases. Obviously, detection based on human eyes is insufficient for the big data evolution On the other hand, conventional image forgery detection techniques are developed with different assumptions that may not be appropriate for generative fake colorized image detection. Therefore a detailed study of the fake colorized images is to be done. Among all the traditional techniques the colorization produces better performances and they difficult to identify the forgery by human eyes.

II. BACKGROUND

Forgery detection has been investigated for a period of ten years. In general, forgery detection possess different characteristics of images and attempts to get traces to analyze. As the observation shows, most of the traditional forgery detection techniques mainly classified into three types, copy-move , photomontage and image retouching detection.

Copy-move detection relies on identifying duplicated regions in a tampered image. Intuitively, these techniques develop to get an appropriate feature in a certain domain, such that the detection can be performed via searching the most similar two units (such as patches). Different methods usually exploit different features. [1] explores features in the frequency domain by dividing the image into overlapping blocks and detects the copy-move forgery via matching the quantized discrete cosine transform (DCT) coefficients. [2] performs a rotation invariant detection based on the Fourier-Mellin transform. [3]localizes the duplicated regions based on the Zernike moments, which exhibit the rotation invariance property, of small image blocks. [3] reports decent results especially when the duplicated regions are smooth. [4] implements the famous SIFT feature [5] to detect multiple duplicated regions and observes the geometric transformation performed by the copy-move operation.

The major techniques used for image forgery is the copy move attack. Photomontage detection mainly detects the changed or affected regions which generate from different original images. By summarizing each techniques it has been known that there are more techniques to detect the traditional forgery techniques , and it has been decided to propose the technique to detect the image forgery based on the colorization.

III. METHODOLOGY

The rapid development in colorization techniques has formed colorized images to be visually not distinguishable from natural images. To identify the fake colorized images from the natural images, the detailed study of the statistics of the fake colorized images, which are generated by traditional methods has tto be performed and two effective detection schemes has been proposed, FCID-HIST and FCID-FE.

The Hue-Saturation Value (HSV) color space represents the chrominance information in the hue and saturation channel separately, hence we calculate the normalized histograms (each containing 250 bins) of the red green blue ,hue and saturation channel in both natural images and fake colorized images.

After comparing to the statistical differences of the color channels there are also some differences in their image priors because they exploit their differences in the images. In this dark channel priors and the bright channel priors are taken as the channel priors feature. In the dark channel prior, the dark channel of a natural image is minimum i.e(is close to zero), and in the bright channel prior, the bright channel of a natural image is maximum i.e.(is close to 255). The dark channel I_{dc} and bright channel I_{bc} of an image I are defined as shown below.

$$I_{dc} = \min_{y \epsilon \Omega(x)} \left(\min_{cp \epsilon (r,g,b)} \left(I_{cp}(y) \right) \right) \qquad (1)$$

$$I_{bc} = \max_{y \epsilon \Omega(x)} \left(\max_{cp \epsilon (r,g,b)} \left(I_{cp}(y) \right) \right) \qquad (2)$$

where x denotes the pixel location, Icp stands for a color channel of I and Ω(x) shows the local patch centered at the location x.

A. FCID-HIST

By observing the statistical differences the effective technique to detect the fake colorized images i.e. histogram based fake colorized image detection has been proposed.

In FCID-HIST four features are used to detect the forgeries they are the red feature F_r , green feature F_g , blue feature F_b hue feature F_h, the saturation feature F_s, feature of dark channel F_{dc} and the bright channel feature F_{bc}.

After calculating the index value the first order derivative i.e the first order hue features can be extracted.

$$F_h{}^{\alpha}(1) = \mathrm{Dist}_h{}^{\alpha}(v_h) \qquad (3)$$

The Distributions may vary according to their bins the second order derivative of the hue feature can be found by the $\mathrm{DistD}_{\alpha}{}^{h}(l) = \mathrm{Dist}_{\alpha}{}^{h}(l + 1) - \mathrm{Dist}_{\alpha}{}^{h}(l)$ to know the variation in the distribution of the histogram

$$F_h^{\alpha} = \sum_{l=1}^{k_h - 1} |DistD_h^{\alpha}\,(l)| \qquad (4)$$

The hue feature formed by combination of the first order and the second order derivative of the hue channel.

$$F_{\alpha}{}^{h} = [F_{\alpha}{}^{h}\,(1)\ F_{\alpha}{}^{h}\,(2)] \qquad (5)$$

. When all the features are calculated the final detected histogram feature $F_{HIST}{}^{\alpha}$ for the training images can be formed as

$$F^{\alpha}{}_{HIST} = [F^{\alpha}{}_{r}\ F^{\alpha}{}_{g}\ F^{\alpha}{}_{b}\ F^{\alpha}{}_{h}\ F^{\alpha}{}_{s}\ F^{\alpha}{}_{dc}\ F^{\alpha}{}_{bc}] \qquad (6)$$

After calculating the detected feature, FCID-HIST gives the input to the FCID-FE to get more better performance.

B. FCID-FE

A new technique has been proposed Feature Encoding based Fake Colorized Image Detection (FCID-FE), to better utilization of the statistical information by jointing the modeling of data distribution and shows the divergences inside different moments of the distribution.

The hue, saturation, dark and bright channels of a training image can be represented as $I^{\beta}{}_{r}$,$I^{\beta}{}_{g}$,$I^{\beta}{}_{b}$,$I^{\beta}{}_{h}$, $I^{\beta}{}_{s}$, $I^{\beta}{}_{dc}$ and $I^{\beta}{}_{bc}$ respectively, where β is the index of the image

In comparing to the histogram modeling FCID-FE models the Gaussian mixture model with the sample data distribution G created using the above equation

$$G\left(\frac{\phi}{\theta}\right) = \sum_{n=1}^{N} \log p\left(\frac{\phi_n}{\theta}\right) \qquad (7)$$

where N shows the number of samples in Φ, Θ represents for the parameter set of the GMM

Then, the likelihood function of Φ_n being formed by the GMM Θ can be modeled as given below

$$\rho\left(\frac{\phi_n}{\theta}\right) = \sum_{m=1}^{N_m} \log \omega_m \rho_m \left(\frac{\phi_m}{\theta}\right) \qquad (8)$$

Where $\rho_m\left(\frac{\phi_m}{\theta}\right)$ is defined as given below

$$\rho_m\left(\frac{\phi_m}{\theta}\right) = \frac{exp\left[-\left(\frac{1}{2}\right)(\phi_m - \mu_a)^T \sigma_a^{-1}(\phi_m - \mu_a)\right]}{2\pi^{\frac{N_v}{2}} |\sigma_a|^{\frac{1}{2}}} \qquad (9)$$

where N_v shows the number of dimensions of each generated sample vector. Then, GMM can be developed by using the parameter set Θ. With the constructed GMM, FCID-FE uses different moments of the distribution and encodes each subset Φ^β of the sample vectors, of each training image, into training fisher vector values and the fisher vector can be expressed as follows

$$F_{FE}^{\beta} = \left[\frac{\lambda_1 \delta G\left(\frac{\phi^\beta}{\theta}\right)}{\delta \omega_a} \; \frac{\lambda_2 \delta G\left(\frac{\phi^\beta}{\theta}\right)}{\delta \mu_{a,v}} \frac{\lambda_3 \delta G\left(\frac{\phi^\beta}{\theta}\right)}{\delta \sigma_{a,v}} \right] \qquad (10)$$

where v = 1,2,...,N_v and λ_1, λ_2 and λ_3 are expressed in the following equations ,where λ_1, λ_2 and λ_3 are fisher vector values that are encoded from the inputs given.

C. DenseNet Neural Network

An artificial neural network is an algorithm that models computations using graphs of artificial neurons, mimicking how neurons work in the brain. Artificial neural networks are well-suited to solving complex nonlinear problems. Unlike traditional machine learning algorithms such as support vector machines, artificial neural networks have flexible structures that can be adapted according to the problem that is to be solved. This work uses an artificial neural network to differentiate natural images from fake colorized images.

The artificial neural network employed for detecting fake colorized images is based on the dense convolutional network (DenseNet) model [10]. DenseNet has a relatively simple structure, in which every layer of the network is connected to every other layer in a feed-forward manner. Compared with other neural network models, DenseNet strengthens feature propagation while reducing the number of parameters. Figure 2 shows the structure of the neural network used for fake colorized image detection. The neural network has six layers – an input layer, an output layer and four hidden layers. Each hidden layer is fully connected to the previous layers. For each hidden layer, the input of the layer is the sum of the outputs of the other hidden layers. The relationships of the hidden layers are given by:

$$Xi = Y1 + \cdots + Yi-1, i \geq 2$$

where Xi and Yi are the input and output of layer i, respectively. The selection of an appropriate activation function is an important aspect when designing a neural network. The proposed technique employs a parametric rectified linear unit (PReLU), an activation function with parameters that can be trained. This activation function is used in the hidden layers of the network. Table 4 shows the details of the neural network. Hidden layers 1 through 3 have 32 neurons each whereas hidden layer 4 has 128 neurons. The joint supervision of the softmax loss function and center loss function [21] was used to train the neural network. The softmax loss function is one of the most widely used loss functions. The center loss function has been demonstrated to minimize intra-class variations while keeping the features of different classes separable.

Dense convolution network is used as the training classifier. For testing, in FCID-FE the training test sample will be constructed then the Gaussian mixture model is used to encode the image into the fisher vector. Finally the features are trained with the neural network classifier to detect the fake images.

The softmax loss function L_S is:

$$L_S = -\sum_{i=1}^{m} \log \frac{e^{W_{y_i}^T x_i + b_{y_i}}}{\sum_j^n e^{W_j^T x_i + b_j}} \qquad (11)$$

where x_i is the ith deep feature, which belongs to the class y_i; m is the mini-batch; and n is the number of classes.

The center loss function is:

$$L_C = \frac{\lambda}{2} \sum_{i=1}^{m} \left|\left| x_i - c_{y_i} \right|\right|_2^2 \qquad (12)$$

where c_{Y_i} is the center of y_i of the deep feature and is updated as the deep feature changes. The joint supervision of the softmax loss function and center loss function are used to train the neural network.

The final loss function used to train the neural network is

$$L = -\sum_{i=1}^{m} \log \frac{e^{W_{y_i}^T x_i + b_{y_i}}}{\sum_j^n e^{W_j^T x_i + b_j}} + \frac{\lambda}{2} \sum_{i=1}^{m} \left|\left| x_i - c_{y_i} \right|\right|_2^2 \quad (13)$$

IV. EXPERIMENTAL RESULTS

This area shows about the databases and the measurements ,experiment results are shown detailed accordingly.

A. Setups and Measurements

In this work, dense convolution network , is implemented for classification and to get the predicted results. The VLFeat software is to be used for the GMM modeling and Fisher vector encoding. In the work two performance measure is to used to evaluate the performances, one is the half total error rate (HTER) measurement HTER is defined in as follows.

$$\text{HTER} = \frac{FPR+FNR}{2}$$

$$= \frac{\frac{FP}{(TN+FP)} + \frac{FN}{(TP+FN)}}{2} \qquad (14)$$

Where P represents the positive samples, N represents the negative samples, TP shows the true positive samples and the TN denotes the true positive samples. The natural images and the fake images are denoted as the positive and the negative samples.

B. Databases

Four benchmark datasets based on the ImageNet LSVRC 2012 Validation Set were employed in the experiments. The datasets, which are widely used in image colorization and fake image detection research, contain many categories of images, including images of people, animals, buildings and landscapes. The D1 dataset corresponds to the ctest10k dataset, which has 10,000 fake colorized images and their corresponding 10,000 natural images from the ImageNet LSVRC 2012 Validation Set. Datasets D2 and D3 each contain the 5000 natural images as well as 5000 fake colorized images in both databases. The images are in the equal dimesion of 256 *256

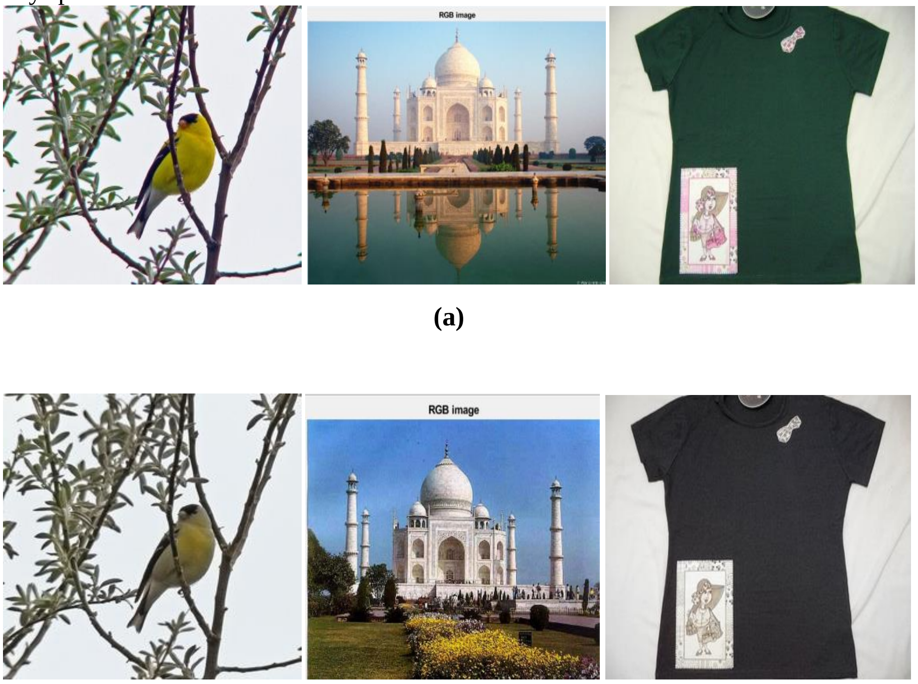
D. Results

The dataset which contains the real and their fake images are given as input and they will undergo both the training and the testing .The images are loaded and they have been to transformed to rgb and hsv image to find the hue saturation and the value images and their respective histograms. Then the dark channel and the bright channel of the images and their respective histograms are found. Using these four channels the histogram features can be extracted . Then the features are given to the GMM to produce the samples then they process with the GMM and produce the fisher vector values and it undergoes for the neural network classifier and classify whether the image is real or fake. Finally analysis of the project will be done. Analysis process makes the project more efficient. The performance and accuracy of the algorithm are analyzed using the above measures.

V. CONCLUSION AND DISCUSSION

In this project, histogram based fake colorized image detection and the feature encoding based fake colorized image detection has been proposed to identify the fake colorized images. Most of the algorithms are developed only for the detection of the traditional image forgery techniques. So these algorithms are used to detect the fake image. The observation shows that that fake colorized images and their corresponding natural images shows a major differences in their hue, saturation, dark and bright channels. In comparing both the algorithms feature encoding based algorithm produces better results than the histogram based algorithm. The work produces an accuracy of 90%.So these algorithms are used in the field of image forgery detection to get better performance for finding the fake images.

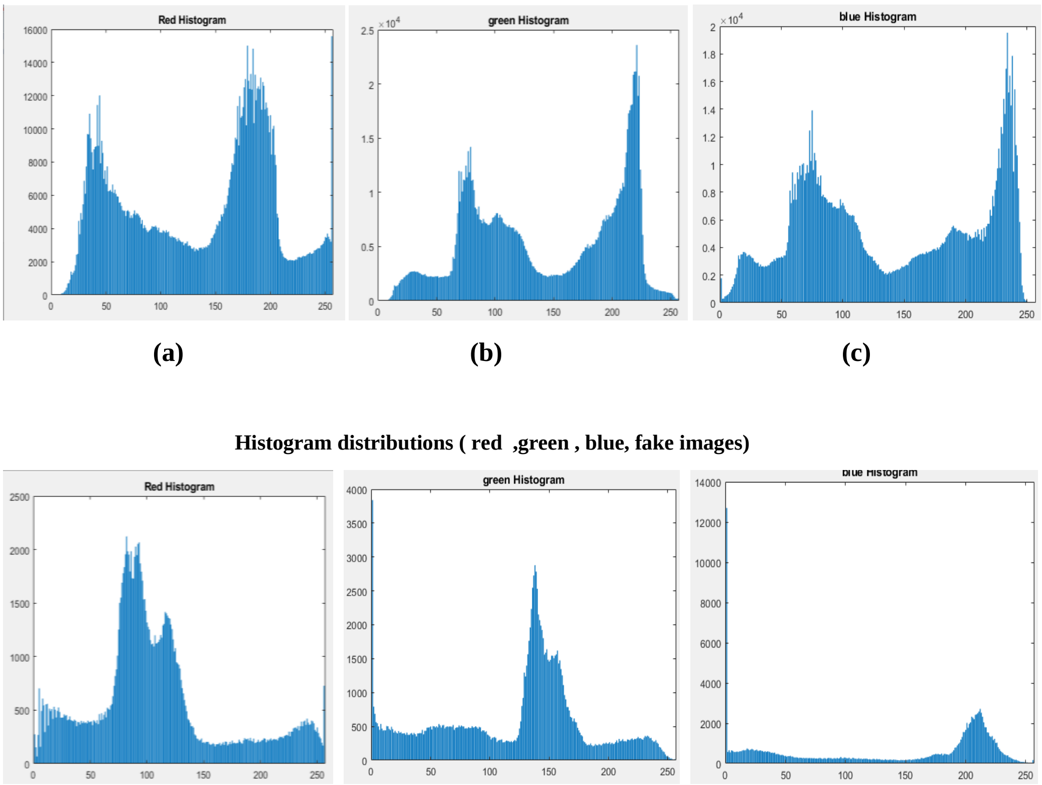
By using the performance analysis of the system it has been clearly known that the work produce the ROC curve of the range 0.90.When compared to the algorithms it has been shown that



(a)

(b)

Fig. 1: (a) Real images. (b) Fake colorized images

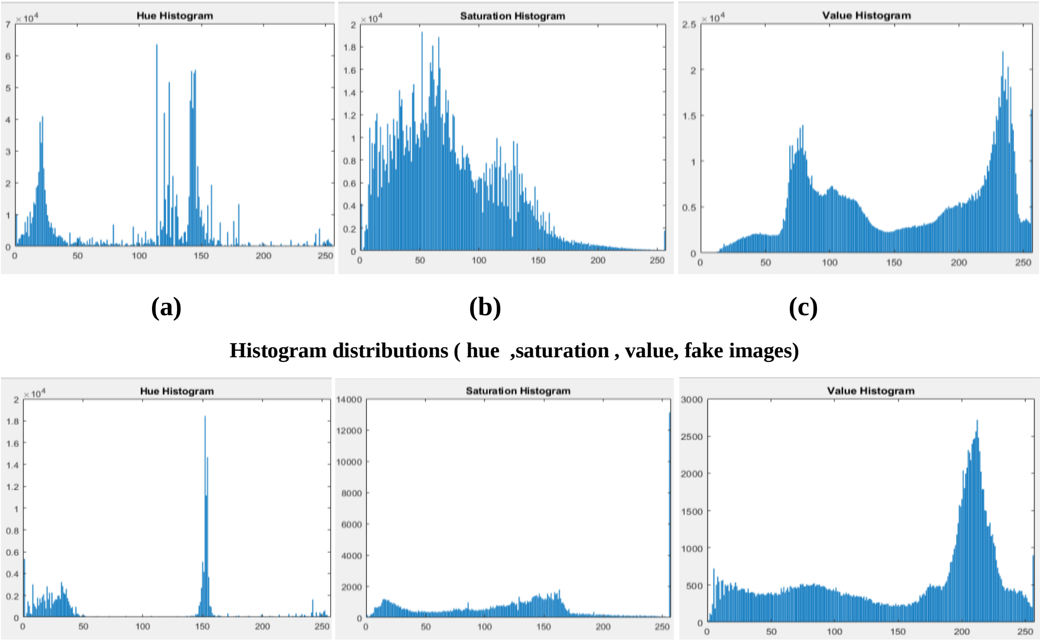

Histogram distributions (red ,green , blue, natural images)

(a) **(b)** **(c)**

Histogram distributions (red ,green , blue, fake images)

(d) **(e)** **(f)**

Fig. 2: (a)Red Histogram Distribution (natural images). (b) Green Histogram Distribution (natural images). (c) Blue Histogram Distribution (natural images). (d) Red Histogram Distribution (fake images). (e) Green Histogram Distribution (fake images). (f) Blue Histogram Distribution (fake images)

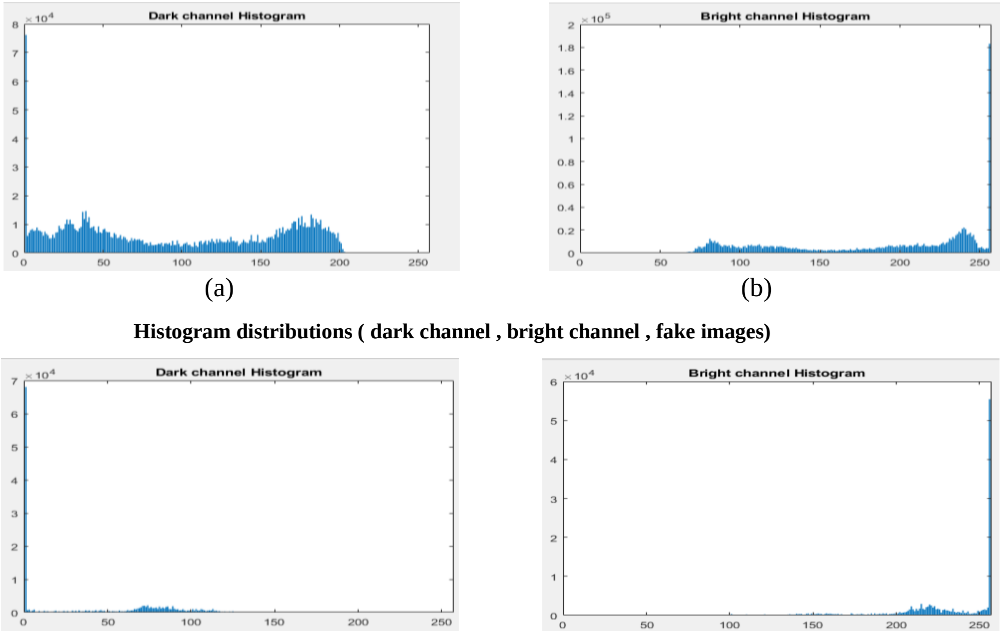

Histogram distributions (hue ,saturation , value, natural images)

(a) **(b)** **(c)**

Histogram distributions (hue ,saturation , value, fake images)

(d) **(e)** **(f)**

Fig. 3: (a)Hue Histogram Distribution (natural images). (b) Saturation Histogram Distribution (natural images). (c) Value Histogram Distribution (natural images). (d) Hue Histogram Distribution (fake images). (e) Saturation Histogram Distribution (fake images). (f) Value Histogram Distribution (fake images)



Histogram distributions (dark channel , bright channel , natural images)

(a) (b)

Histogram distributions (dark channel , bright channel , fake images)

(c) (d)

Fig. 4: (a) Dark channel Histogram distribution (natural images). (b) Bright channel Histogram distribution (natural images). (c) Dark channel Histogram distribution (fake images). (d) Bright channel Histogram distribution (fake images).

TABLE I: The Histogram Based Features For the real and the fake image

Histogram Feature	Fh1	Fh2	Fs1	Fs2	Fdc1	Fdc2	Fbc1	Fbc2
Real image Values	32561	973392	1782	359263	76180	285206	183097	365965
Fake image Values	1642	95023	13123	32658	68045	99715	55444	86952

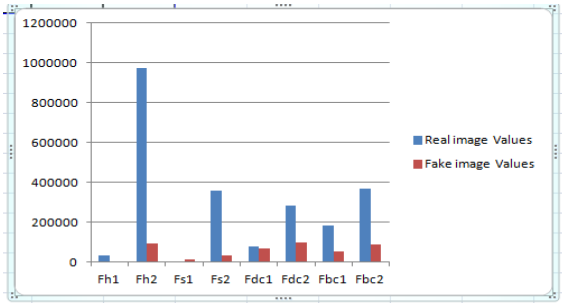


Fig 5: Results Of Histogram Features

the feature encoding algorithm produces a slightly better than the histogram algorithm. The performance of current methods sometimes degrades obviously when the training images and the testing images are generated from different colorization methods or different datasets, thus blind fake colorized image detection features and methods may be developed in the future by studying the common characteristics of other colorization techniques and the channels features and also the algorithm will be well tuned for the further process.

REFERENCES

[1] H. Farid, "Exposing digital forgeries from JPEG ghosts," *IEEE Trans.Inf. Forensics and Security*, vol. 4, no. 1, pp. 154-160, 2009.

[2] J. Li, X. Li, B. Yang and X. Sun, "Segmentation-Based Image Copy-Move Forgery

Detection Scheme," *IEEE Trans. Inf. Forensics and Security*, vol. 10, no. 3, pp. 507-518, 2015.

[3] G. Cao, Y. Zhao, R. Ni and X. Li, "Contrast Enhancement-Based Forensics in Digital Images," *IEEE Trans. Inf. Forensics and Security*, vol. 9, no. 3, pp. 515-525, 2014.

[4] G. Larsson, M. Maire and G. Shakhnarovich, "Learning representations for automatic colorization," *in Proc. European Conf. Comp. Vision (ECCV)*, pp. 577-593, 2016.

[5] I.J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville and Y. Bengio, "Generative adversarial nets," *in Procs. Advances in Neural Inf. Process. Systems (NIPS)*, pp. 2672-2680, 2014.

[6] F. Huang, X. Qu, H.J. Kim and J. Huang, "Reversible data hiding in JPEG images," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 26, no. 9, pp. 1610-1621, 2016.

[7] J. Yin, R. Wang, Y. Guo and F. Liu, "An adaptive reversible data hiding scheme for JPEG images," *in Proc. Int. Workshop on Digital-Forensics and Watermarking (IWDW)*, pp. 456-469, 2016

[8] J. Wang, S. Lian and Y.-Q. Shi, "Hybrid multiplicative multi watermarking in DWT domain", *Multidimensional Systems and Signal Process.*, vol. 28, no. 2, pp. 617C636, 2017.

[9] Y. Yang, W. Ren, Y. Guo, R. Wang and X. Cao, "Image deblurring via extreme channels prior," *in Procs. IEEE Int. Conf. Comp. Vision and Pattern Recognition (CVPR)*, 2017, Accepted.

[10] J. Farquhar, S. Szedmak, H. Meng and J. Shawe-Taylor, "Improving "bag-of-keypoints" image categorization," *Technical report, University of Southampton*, 2005.

[11] A. Levin, D. Lischinski and Y.Weiss, "Colorization using optimization," ACM Trans. Graphics, vol. 23, no. 3, pp. 689-694, 2004.

[12] J. Pang, O.C. Au, K. Tang and Y. Guo, "Image colorization using sparse representation," in Proc. IEEE Int. Conf. Acoustics, Speech and Signal Process. (ICASSP), pp. 1578-1582, 2013.

[13] G. Charpiat, M. Hofmann and B. Scholkopf, "Automatic image colorization via multimodal predictions," in Proc. European Conf. Comp. Vision (ECCV), pp. 126-139, 2008.

[14] X. Chen, J. Li, D. Zou and Q. Zhao, "Learn Sparse Dictionaries for Edit Propagation," IEEE Trans. Image Process., vol. 25, no. 4, pp. 1688-1698, 2016.

[15] Y. Li and J. Zhou, "Image copy-move forgery detection using hierarchical feature point matching," in Proc. Asia-Pacific Signal and Inf. Process. Association Annual Summit and Conf. (APSIPA ASC), pp. 1-4, 2016.