Spliced Detection Based on Illuminant Color Classification using HOG and IFLT Descriptor

Greeshma. B. K

Student of M.Tech (Applied Electronics)
Department of Electronics and Communication
Mahatma Gandhi University College of Engineering,
Thodupuzha, India

Mr. Anish. M. P

Lecturer
Department of Electronics and Communication
Mahatma Gandhi University College of Engineering,
Thodupuzha, India

Abstract—**The main function of image forensics is to assess the authenticity of images. Therefore, trustworthy digital image is a main concern for image forensics. Digital image editing software makes image modifications easy. In this paper, a method to detect splicing (inserting new person) is proposed. It makes use of difference in the color of the illumination of images. It is difficult to create the same illuminant condition while creating a composite image. Statistical-based illuminant estimator on image region is calculated, especially from face. Automated processing is required to avoid confusions. Hence, a pattern recognition scheme operating on illuminant maps is combined to increase the accuracy. From these illuminant estimates, edge based and texture -based features are obtained which is then provided to a SVM classifier for automatic decision-making. When images are captured using cameras, they are subjected to geometric distortions due to varying viewpoints; hence affine-invariant IFLT descriptor is combined for the analysis of real world texture images/patches. SVM classifier classified the combined feature vector with an accuracy of 84% on dataset containing 100 images.**

Keywords— Splicing,Illuminantion inconsistencies,texture and edge based descriptors,affine invariant descriptors component, IFLT.

I. INTRODUCTION

Nowadays, various digital image editing software are available, it makes image modifications easy. Splicing is one of the most common image modification operations. Spliced photographs are increasing with a high frequency, which is seen mainly in magazines, fashion industry, mainstream media outlets, scientific journals, political campaigns, courtrooms etc. Fig.1 shows a spliced image. The person with the helmet is actually inserted in the figure, but it is not easily detected. The main scope of spliced detection comes in insurance claim investigation, forensics or criminal investigation, legal proceeding and national intelligence analysis.

In the existing system for splicing detection, inconsistencies in the illuminant map obtained from images are used [1]. It is based on the fact that while creating a digital composite image, matching the lighting conditions from individual photograph is difficult . Illuminant color is estimated using the inverse intensity- chromaticity color space [3]. Each image region is recolored using the estimated illuminant to yield a so-called illuminant map. Implausible illuminant color estimates indicates a manipulated region [2]. From the face

regions, texture-based features and gradient-based features are extracted for machine learning. In order to describe the edge information, HOG descriptor [4] is proposed. Texture features are extracted using SASI descriptor [1]. These texture- and edge-based cues are combined using SVM classifier for automatic decision-making When images are captured using state of the art cameras, they get subjected to geometric distortions (e.g. translation, rotation, skew, and scale) due to varying viewpoints, and hence affine-invariant descriptors are required for the analysis of real world texture images/patches [5]. The vast majority of algorithms make an assumption that all images are captured under the same orientation (i.e., there is no inter-image rotation). Therefore IFLT descriptor along with HOG is implemented.

Fig.1. Example of Spliced Image

In Section II, related work in color constancy and illumination-based detection of image splicing is briefly reviewed. Section III describes system development and its implication. In Section IV experiment results of each stage of process is presented. Conclusions and potential future work are outlined in Section V.

II. STATE OF ART

M.Johnson and H.Farid [6] proposed a geometrical based method for determining the forged images. Specular highlights that appear on the eye are a powerful cue as to the shape, color and location of the light source. Inconsistencies in the properties of the light are used as an evidence of tampering. The 3-D direction to a light source

is measured from the position of the highlight on the eye. It is difficult to create the same illuminant condition while creating a composite image. Based on this concept Riess and Angelopolou [2] followed a physics based algorithm to indicate the original and tampered image. This proposed method operates on partially specular pixels. The image is segmented into superpixel of uniform chromaticity. Illuminant estimation is done for every superpixel. Difference in the illuminant color estimates indicates a tampered region. Tiago jose de carvalho [1] presented a machine-learning-based approach for forgery detection that requires minimum user interaction. To achieve this, two separate illuminant color estimators: the statistical generalized gray world estimates and the physics-based inverse-intensity chromaticity space are implemented. The illuminant color is used to obtain the illuminant map. From these illuminant estimates, using SASI and HOG descriptor, features are extracted which are then provided to a SVM Classifier for automatic decision-making. When an image is spliced, the statistics of these edges may differ from original images. To characterize such edge discontinuities, HOG descriptor is used. Muhammad Hussain et al. [7] combined Local Binary Pattern (LBP) and Discrete Cosine Transform (DCT) to detect spliced forgeries. In this technique, the features are extracted from the chromatic channel, which has been shown to capture the tampering artifacts better than other color channels. Human vision can perceive the luminance component in a better way than the chrominance component. When tampering is done, the original texture of the image is distorted. LBP can capture the texture differences. The chromatic component is divided into overlapping blocks and then LBP followed by 2D DCT are applied to each block. Standard deviations calculated from the corresponding DCT coefficients of all blocks are used as input features to SVM classifier to make the decision about the input image. Gholap and Bora [8] introduced a physics-based illumination method to image forensics. They examined inconsistencies in specularities based on the dichromatic reflectance model. The dichromatic lines for each specular region is found out using PCA and the distance between marginal median and intersection is estimated. The image is tampered if distance is greater than threshold. Specularities have to be present on all regions of interest, which limits the method's applicability.

III.   SYSTEM DEVELOPMENT

The proposed method consists of five steps which is shown in Fig.2.

A. Illuminant Estimation (IE):

   The illuminant color is estimated using physics-based inverse-intensity chromaticity space [3]. The illuminant color so obtained is used to recolor the image. The resulting representation is called illuminant map (IM) [1].

Inverse Intensity Chromaticity Estimator

   This method can estimate illumination chromaticity from any colored surfaces [3]. The main reason of choosing the physics- based method is that this algorithm is based on understanding the physical process of reflected light.
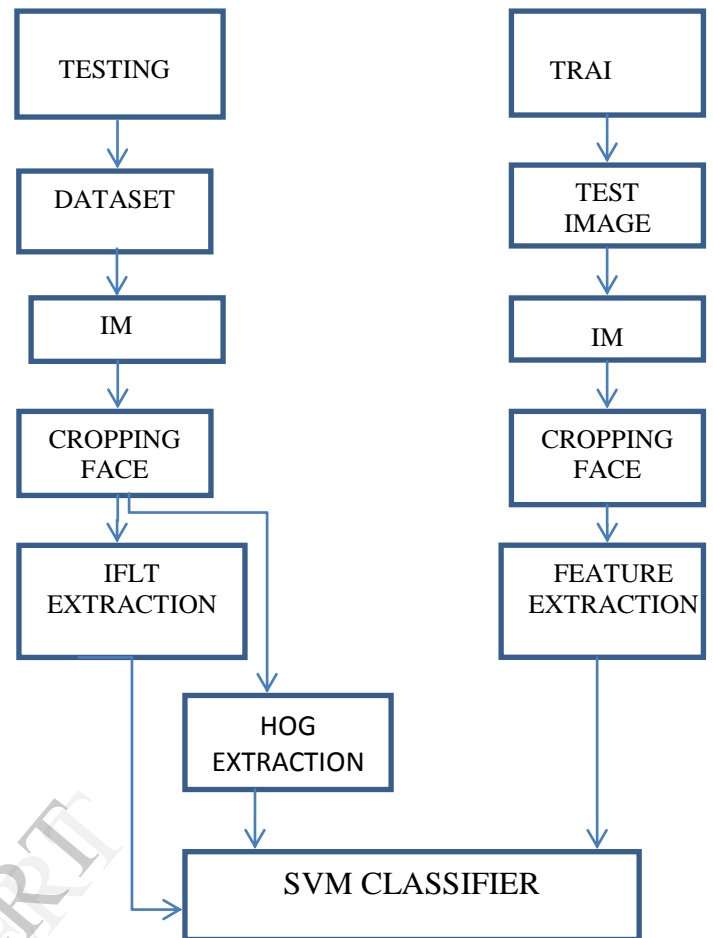

Fig.2. Block diagram of proposed method

The product of illumination spectral energy distribution and surface spectral reflectance gives the spectral energy distribution of the light reflected from an object. The color of an object observed in an image will be a combination of multiple illuminants. The actual surface color is obtained by subtracting the color of illumination. Rough highlight regions are estimated by thresholding on brightness and saturation values. The estimated highlight regions are projected over inverse-intensity chromaticity space [1]. Highlight regions are found by thresholding on image intensity and saturation values.

$$\frac{I=I_r+I_g+I_b}{3} > T_a I^{max} \qquad 1$$

$$S=1 - \frac{\min(I_r,I_g,I_b)}{3} < T_b S^{max} \qquad 2$$

where I^{max} and S^{max} are the largest I and S in the image, respectively. T_a and T_b are the thresholds of image intensity and saturation, respectively.

   Tan et al. [3] derived a relation which shows the relation between illumination chromaticity and image chromaticity

$$\sigma_c (x)=p(x)\frac{1}{\sum_{i\in\{R,G,B\}}f_i(x)} + \gamma_c \qquad 3$$

where σ_c is chromaticity of the color channel, p(x) shows

geometrical influences, and γ_C is chromaticity of illuminant. This correlation is clearly described in "inverse intensity chromaticity space", a two-dimensional space. Based on this linear correlation, illumination chromaticity for both single and multi-colored surfaces is obtained without segmenting the color beneath the highlights. IIC space is a per-channel 2-D space, where the horizontal axis is $\frac{1}{\sum_i f_{i(x)}}$ and vertical

axis is the pixel chromaticity for that particular channel. Per color channel c, the pixels are projected onto inverse intensity-chromaticity (IIC) space.

B. Face Extraction

Illuminant maps are an important representation in determining the spliced forgery. To avoid the confusion while deciding the image is spliced or not, an automated machine learning is required. Hence, a pattern recognition scheme is combined with the features extracted from illuminant maps. Estimation on objects of similar material exhibits a lower relative error. The detection is limited to skin, and in particular to faces. The user draws a bounding box around each face in the image that should be investigated. Therefore from each illuminant map, face is cropped so that only the illuminant estimates of the face regions remain.

C. Compuation of Illuminant Features:

For all face regions texture-based and gradient-based features are computed on the IM values. Each one of them encodes complementary information for classification.

Hogedge Algorithm:

Statistics of edges differ from original image when the image is spliced. To characterize such edge discontinuities, HOGedge descriptor is used to characterize such variation in the edges. HOG descriptor is a feature descriptor which detects objects in image processing. The HOG descriptor technique is based on the distribution of gradient orientation in the image - detection window, or region of interest (ROI). The basic idea is that local object appearance and shape can often be characterized rather well by the count of local intensity gradients or edge directions. The main advantage of HOG descriptor is that it captures edge or gradient structure that is very characteristic of local shape. Implementation of the HOG descriptor algorithm is as follows [5]:

1. Divide the image into small cells, and for each cell compute a histogram of gradient directions for the pixel.

2. Discretize each cell into angular bins according to the gradient orientation.

3. Each cell's pixel contributes weighted gradient to its corresponding angular bin.

4. Groups of adjacent cells are considered as spatial regions called blocks.

5. Normalized group of histograms represents the block histogram. The set of these block histograms represents the descriptor.

The HOG method tiles the image with a dense grid of cells. For each pixel, the gradient vector from the image is computed. The angle of the vector is distributed into its corresponding bin and is weighted by the gradient magnitude. Gradient calculation is done from the pixels of each cell. From each cell 1-D histogram of gradient directions is

obtained using all cell pixels. To construct the feature vector, the histograms of all cells within a spatially larger region are combined. For gradient computation, the grayscale image is filtered to obtain x and y derivatives of pixels using simple derivative masks, $hx = [-1, 0, 1]$ and $hy = [-1\ 0\ 1]^T$. After calculating x, y derivatives the orientation and magnitude of the gradient is also computed:

$$Mag = \sqrt{{I_X}^2 + {I_Y}^2} \qquad 4$$

$$angles = \arctan \frac{I_X}{I_Y} \qquad 5$$

Each pixel calculates a weighted vote for an edge orientation histogram channel based on the orientation of the gradient element centered on it, and the votes are accumulated into orientation bins over local spatial regions called cells. Cells used are rectangular. The orientation bins are evenly spaced over -140-180. The vote is a function of the gradient magnitude at the pixel. Cells are computed with 9 orientation bins with 20 degree interval. For each pixel's orientation, the corresponding orientation bin is found and the orientation's magnitude $|G|$ is voted to this bin. Gradient strengths vary over a wide range owing to local variations in illumination and foreground-background contrast, so effective local contrast normalization turns out to be essential for good performance. Each R-HOG block has 3x3 cells and adjacent R-HOGs are overlapping each other for a magnitude of half-size of a block. This was achieved by means of 9 rectangular cells and 9 bin histogram per cell. The nine histograms with nine bins were then concatenated to make an 81-dimensional feature vector. The final descriptor is then the vector of all components of the normalized cell responses from all of the blocks in the detection window equations.

IFLTAlgorithm:

While capturing the image using cameras or sensors, they can get subjected to geometric distortions (e.g. translation, rotation, skew, and scale) due to varying viewpoints, and hence affine-invariant descriptors are required for the analysis of real world texture images/patches [5]. The vast majority of the algorithms make an explicit or implicit assumption that all images are captured under the same orientation (i.e., there is no inter-image rotation).Even when the image is rotated, it is always perceived as the same texture by a human observer. Therefore rotation invariant texture classification is highly desirable.The pixel intensities in a small image neighbourhood would provide us an approximate measure of the gradient in that specific neighbourhood of image pixels. This forms the basic theory behind the IFLT algorithm. This algorithm extracts features which are rotation invariant from a small texture patch around a centre pixel. With XC as the centre pixel, the gradient of intensity in all directions with reference to the centre pixel is calculated to obtain gradient components which would be scale invariant. The gradient intensities around a centre pixel can be rewritten as a 1-D vector as shown

$$I = [IC - I0 ... IC - I7]$$

Where I is a one-dimensional vector, IC is the intensity of the

centre pixel and I (0..7) are the intensities of the surrounding neighbourhood. Whenever there are rotational effects it would result in linear shifts in the one-dimensional vector. That is, rotations in cause linear shifts in the transformed space.

X1	X2	X3
X4	XC	X5
X6	X7	X8

Fig.3. Neighbourhood of pixel

Intensity vector derived from a small patch is normalized. Using Haar wavelet it is filtered to obtain rotation-invariant features. The gradient intensities around a centre pixel is estimated. It is normalized to further enhance scale invariance. The vector thus obtained represents the intensity gradient around the centre pixel and would also be (partially) illumination invariant. The DWT of signal I is calculated by passing it through filters. In this algorithm Pranam Janney and Zhenghua Yu used Haar wavelets because of their computational efficiency. The required filter coefficients are given

$$h=\left[\frac{1}{\sqrt{2}},\frac{-1}{\sqrt{2}}\right] \quad g=\left[\frac{1}{\sqrt{2}},\frac{1}{\sqrt{2}}\right]$$

The signal is passed through a highpass filter h and a low-pass filter g simultaneously. The filter outputs are then downsampled by 2 to make the calculation easier. The outputs from the highpass filter are known as the detail coefficients and those from the low-pass filter are known as the approximation coefficients. These detail and approximate coefficients so obtained have shift invariant energy distributions. Mean and standard deviation of the energy distributions of the high pass and the low pass filter outputs generated by the wavelet transform are used as the texture features. From the extracted local texture features a histogram is constructed. Mean and the standard deviation of the energy distributions of the high pass and the low pass wavelet bands are divided into a number of bins and calculating the count of local texture feature values in those bins. The wavelet transform is equivalent to a convolution followed by downsampling by 2.

$$ylow = (I*g)\downarrow 2 \qquad yhigh = (I*h)\downarrow 2$$

The histogram extracted serves as the texture descriptor of an image patch.

D. Paired Face Features

To compare two faces, the same descriptors from each of the two faces are combined. Feature concatenation from two faces is different when one of the faces is an original and one is spliced. The IFLT descriptors and HOGedge descriptors capture two different properties of the face regions. Both descriptors are signatures with different behavior.

E. Classification

In this paper a machine learning approach to automatically classify the feature vectors is employed. An image is considered as spliced if at least one pair of faces in the image is classified as inconsistently illuminated. The illumination for each pair of faces in an image is classified as either consistent or inconsistent. Assuming all selected faces are illuminated by the same light source, an image is considered to be manipulated if one pair is classified as inconsistent. Individual feature vectors, i.e., IFLT or HOGedge features on IIC- based illuminant maps, are classified using a support vector machine (SVM) classifier with a radial basis function RBF kernel. The information provided by the IFLT descriptor features is complementary to the information from the HOGedge features. Each combination of illuminant map and feature type is classified using SVM classifier to obtain the distance between the image's feature vectors and the classifier decision boundary. The marginal distances provided by all individual classifiers are merged to build a new feature vector. SVM classifier then classifies the combined feature.

IV. RESULTS AND DISCUSSIONS

A. Dataset

To validate this approach, experiments were performed using images involving people. Spliced images were created by copying the object from one image and pasting on another image that is under different illuminants conditions. The editing was done using the photo editing software Adobe Photoshop. The text images created are in variable scales, illumination, situation and orientation. Data set consists of 100 images .Out of this 50 are original i.e. they have no adjustment, whatsoever and 50 are spliced. When necessary, image splicing operation has been done with postprocessing operation to increase photorealism.

B. Experimental Methodology

Illuminant map

The illuminant map is obtained using the physics-based method which exploits the IIC space. In order to easily estimate the illuminant color, a rough estimate of the specular region is obtained using threshold condition. This method is simple to implement and requires minimal user-interaction. Specular pixel is the one with the brightest pixels in the given image but its color is still not saturated. The threshold value varies the specular regions. Fig.4. is an original image. There is no variation in color inconsistencies in its corresponding IM. Fig.5. shows a spliced image where the person standing in the back is inserted. The inconsistencies in the IM show that it is a spliced image. IM provide a useful and a very important forensic tool for the analysis of color image.

Training

SVM Classifier is trained with HOG edge features along with IFLT features. HOG is implemented using 9 rectangular cells and 9 histogram per cell. The nine histogram with nine bins were then concatenated to make an 81-dimensional feature vector. ILFT yielded 200 –dimensional feature vector. SVM classifier is successfully trained with 40 original and 40 doctored images.

Evaluation on Data sets

Results are shown using classical ROC curves where sensitivity represents the number of composite images correctly classified and specificity represents the number of original images correctly classified. Fig.6. depicts a ROC curve of performance of the proposed method. The area under the curve (AUC) is computed to obtain a single numerical measure for each result. From the evaluated variants, it resulted in an AUC of 84%.
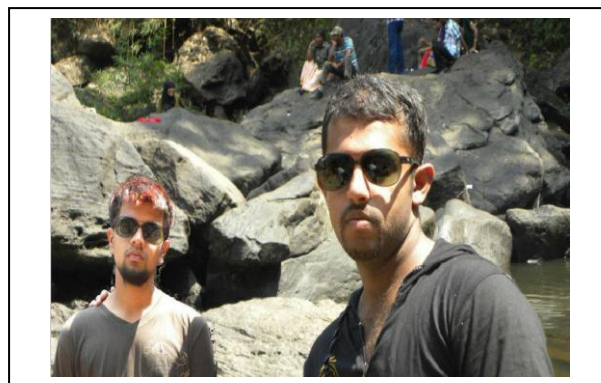

Fig.5. Spliced image with its IM

Fig.4. Original image with its IM

V. CONCLUSIONS AND FUTURE SCOPE

In this paper, a new method for detecting spliced images of people by the illuminant color classification using SVM classifier is presented. Estimation of illuminant color is done using a physics-based method which makes use of the inverse intensity- chromaticity color space. HOG edge descriptor and rotation invariant feature descriptor captures two different properties. These complementary cues (texture- and edge-based) which are computed on IM values are combined using SVM Classifier. Results are encouraging, yielding an accuracy of over 84% correct classification. The proposed method requires human interaction only for drawing rectangular box around face region.

It is difficult to decide whether the image is spliced or not from IM in some of the images. Thus, further improvements can be achieved by improving the learning. The fusion algorithm is a classifier that receives the likelihoods from the others single-classifiers and decides the class [9] .Two or more different classifiers can be combined for spliced detection.
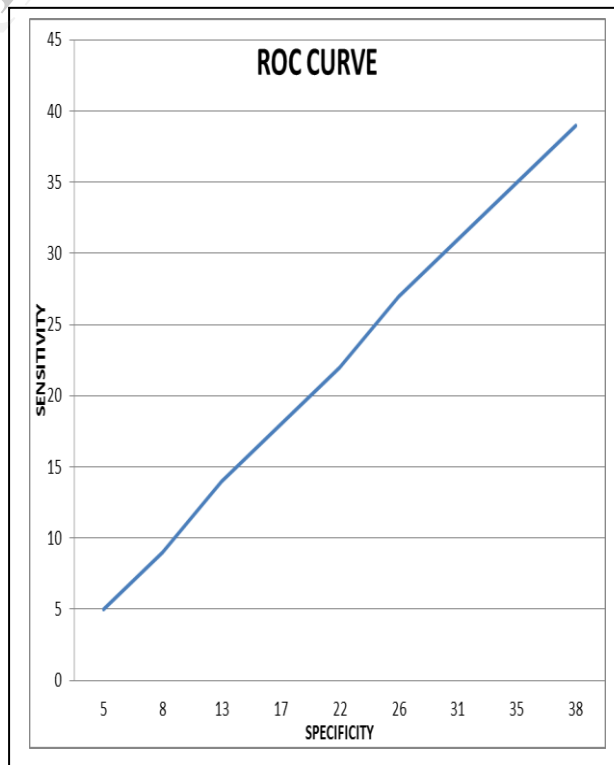

Fig.6. ROC CURVE

REFERENCES

[1] Tiago José de Carvalho,C.Riess and Angelopolou ,"Exposing Digital Image Forgeries by Illumination Color Classification" IEEE Transactions on information forensics and security, vol8 july 2013

[2] C.Riess and Angelopolou ,"Scene manupilation as an indicator of image manupilation", Inf. Hiding, vol. 6387, pp. 66–80, 2010

[3] R. Tan, K. Nishino, and K. Ikeuchi, "Color constancy through inverse intensity chromaticity space," J. Opt. Soc. Amer. A, vol. 21, pp. 321–334, 2004

[4] N.Dalal and B.Triggs,"Histograms of oriented gradients for human detection," in Proc. IEEE Conf. Comput. Vision and Pattern Recognition, 2005, pp. 886–893

[5] Pranam Janney and Zhenghua Yu, "Invariant Features of Local Textures—a rotation invariant local texture descriptor,"IEEE Conference on computer vision and pattern recognition 2007

[6] M. Johnson and H. Farid, "Exposing digital forgeries through specular highlights on the eye," in Proc. Int.Workshop on Inform. Hiding, 2007, pp. 311–325

[7] Amani A. Alahmadi, Muhammad Hussain, Hatim Aboalsamh, Ghulam Muhammad, "Splicing Image Forgery Detection Based on DCT and Local Binary Pattern" GlobalSIP 2013 IEEE.

[8] S. Gholap and P. K. Bora, "Illuminant colour based image forensics,"in Proc. IEEE Region 10 Conf., 2008, pp. 1–5

[9] O. Ludwig, D. Delgado, V. Goncalves, and U. Nunes, "Trainable classifier-fusion schemes: An application to pedestrian detection," in *Proc.* IEEE Int. Conf. Intell. Transportation Syst., 2009, pp. 1–6.