IMAGE FORGERY DETECTION USING CLUSTERING METHOD

Associate Prof.ANAND M[1], Krupa Gowda K[2], Kusuma B M[3], Meghana U[4], Mohita V[5]

[1]Associate Professor, Department of Electronics and Communication Engineering, East west Institute of Technology, Karnataka
[2-5]BE.Student, Department of Electronics and Communication Engineering, East west Institute of Technology, Karnataka

Abstract - *Nowadays duplicating an image has been easier with the advancement of editing tools. The act of forgery leads to loss of some important data required in the field of criminal investigation, forensic, documents etc. The process of duplicating or tampering images has become easier with the usage of powerful computer graphics and editing software such as Adobe photoshop, GIMP, Corel Paint Shop. Copy move forgery is one the most commonly used forgery technique which can be detected by decomposing the image and extracting its feature vector.*

Key Words: Forgery, Copy move forgery, Feature vector, k-means cluster, Confusion matrix.

1. INTRODUCTION

The rapid growth in digital technology and availability of many image processing softwares has made it easier to tamper the digital image. For tampering or altering the images several methods have been developed. The methods are: Active method and Passive method. Active method requires prior information like watermarking or signature generated at the time of creating an image. Passive method does not require any prior information of an image. It works based on the image statistics. Section II explains about the methodology used, Section III narrates the implementation of proposed methodology, Section IV illustrates the achieved experimental results and Section V concludes the idea.

2. METHODOLOGY

Priorly the given image is converted to grey scale image. Clustering algorithm is used in order to form clusters. One of the best clustering algorithms is k-means. The thirteen parameters like contrast, correlation, energy, homogeneity, mean, standard deviation, entropy, root mean square, variance, smoothness, kurtosis, skewness, inverse difference moments are obtained as feature extraction. The mean and standard deviation is given by equation (1) and (2):

(1) $$\text{Mean } (\bar{x}) = \frac{\sum x}{n}$$

(2) $$s = \sqrt{\frac{\sum (x - \bar{x})^2}{n - 1}}$$

2.1 Clustering Algorithm

Clustering algorithm is of grouping a set of objects or data points into clusters based on the similarities. There are various algorithms that bear up to this concept like, based on distance between two different clusters, density of the data space or statistical distributions.

K-means algorithm is used to resolve the clustering problems. It aims at partitioning set of observations or data points in optimal manner into number of clusters (k). It is an effective method to find out which group the object certainly belongs to. Priorly random k clusters centers are selected, then distance between each data point and cluster centers are calculated using Euclidean distance. The data point is selected such that it is the center of a cluster which has a mean distance from all the other cluster centers. Again, find a new cluster center and recalculate the distance between the data point and newly obtained cluster center. If there is no further more groupings then k means is terminated or the process continues until mean distance is obtained. The ease of implementation and high-speed performance makes k means algorithm efficient.

K-nearest neighbors algorithm is a simple supervised learning algorithm that is used for both classification and regression. The output depends based on the usage of k-NN as classification or regression. Based on the majority votes of its neighbor the objects are grouped in classification. In regression, based on the average values of its neighbor the objects are grouped.

The Fig-1 shows the block diagram of the clustering process. Input images is subjected to forgery detection. Then the input image is converted to grey scale to reduce unwanted information of the image. After the conversion, the parameters like mean, standard deviation and feature vector is extracted. The image is clustered using k-means clustering algorithm it groups the data which falls under similar criteria. After all the above process the tampered image is detected as forgery.

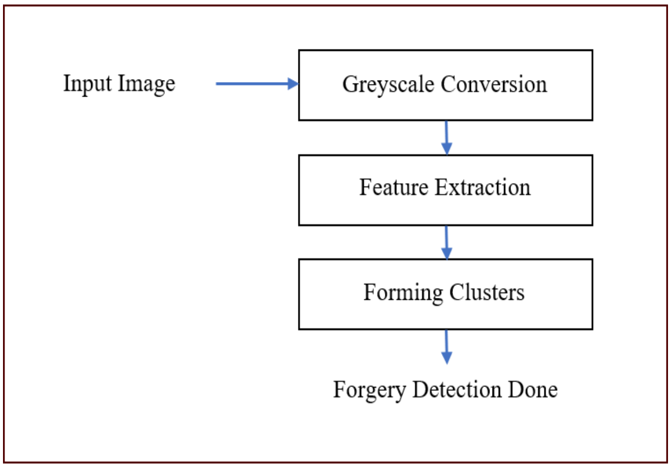


Fig-1: Block diagram of clustering process

2.2 Algorithm for Proposed Clustering Method

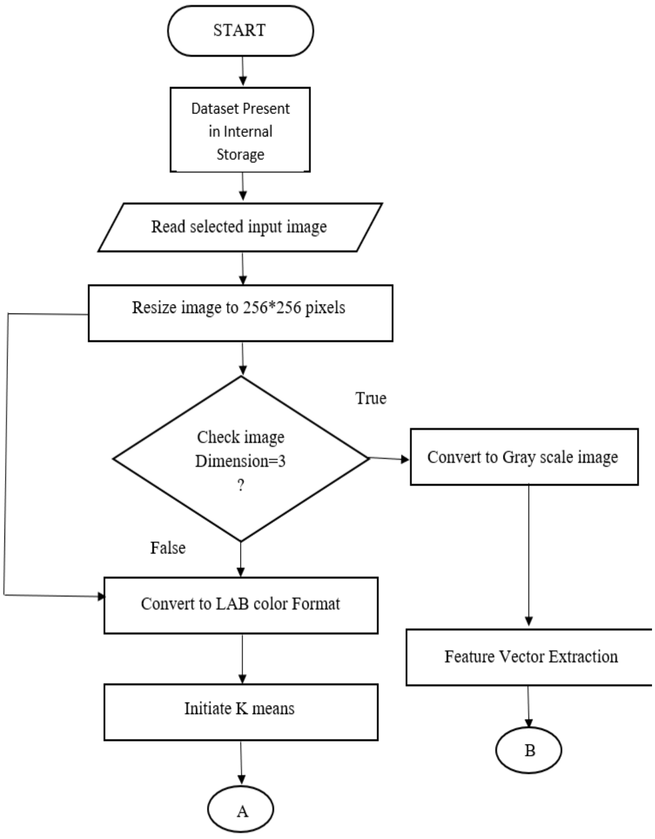
k-means uses an iterative method to group the clusters. Then the k-means classifier is used to do clustering analysis

Step 1: Consider the set of data points and set of centers.
Step 2: Define number of clusters K.
Step 3: Initialize centroids by reorganizing the data point and then randomly select the K data points for the centroids.
Step 4: calculate the distance between data point and centers by means of Euclidean distance.
Step 5: k points are assigned to the object data space representing initial group of centroids. Again, find the new cluster center
Step 6: After all the objects are assigned, the positions of the k centroids are recalculated.
Step 7: The above steps are repeated until the positions of the centroids no longer move.

3. IMPLEMENTATION

Firstly, input image is selected from the database. This image is tested and thirteen parameters are to be extracted. Then the selected image is resized to default format of 256*256 pixels. Based on the dimension value image is converted to grey scale to reduce unwanted information and noise. The RGB color image is a 3D image which is converted to l*a*b color space (2D image) to reduce processing time where 'l' is the luminosity layer, 'a' indicates the color falling in red-green axis and 'b' indicates the color falling in blue-yellow axis. This is done to distinguish the colors and obtain three main colors for cluster formation. The l*a*b output is given as input to k-means Next is initializing the k-means by using three color channels and measuring the distance using Euclidean distance based on the pixel value. Based on this distance we are segmenting the image. Here, we will group image pixel into three clusters. The clustered output is given as input to the Principal Component Analysis (PCA). PCA is used for dimension purpose. The output of PCA is given to Gray-Level Co-Occurrence Matrix (GLCM). Here, we are extracting parameters like contrast, homogeneity, correlation, mean, entropy, standard deviation, root mean square, energy, variance, smoothness, skewness, kurtosis, inverse difference moments.

Fig 2. shows flowchart of the implementation of proposed method. These parameters are represented in a single matrix and can be used for training purpose. The confusion matrix is used to describe the performance of the classifier. It summarizes the number of correct and incorrect predictions. It shows the way in which classifier model is confused when it makes prediction. It is table of four different combinations of predicted values and actual values. This matrix is useful for measuring recall, precision, specifity and accuracy. Next is to initialize the k-nn classifier. The input to this is tested data, trained data and prediction class label. After the classification the result is checked with trained data to ensure to which particular data it matches. Based on this the image is detected as forgery.


START
Dataset Present in Internal Storage
Read selected input image
Resize image to 256*256 pixels
Check image Dimension=3 ?
True
False
Convert to Gray scale image
Convert to LAB color Format
Feature Vector Extraction
Initiate K means
B
A

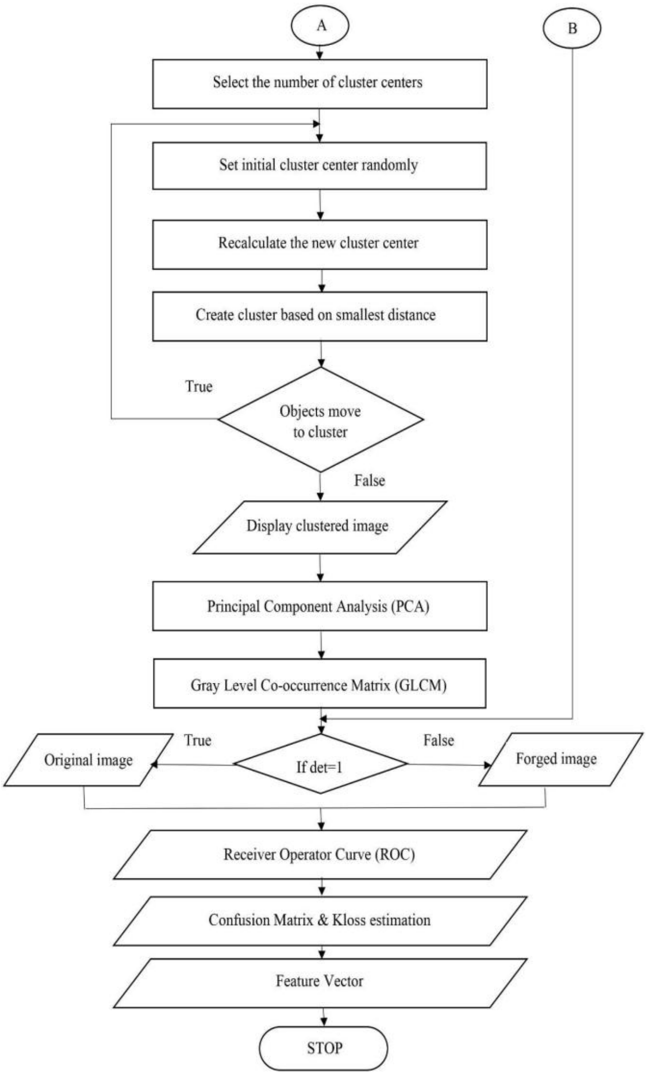

A
B
Select the number of cluster centers
Set initial cluster center randomly
Recalculate the new cluster center
Create cluster based on smallest distance
True
Objects move to cluster
False
Display clustered image
Principal Component Analysis (PCA)
Gray Level Co-occurrence Matrix (GLCM)
True
False
Original image
If det=1
Forged image
Receiver Operator Curve (ROC)
Confusion Matrix & Kloss estimation
Feature Vector
STOP

Fig-2: Flowchart for proposed method

The performance curve is true positive rate versus (TPR) false positive rate (FPR). The true positive rate states that the observation is correct and the prediction is also correct and it is also known as sensitivity. The false positive rate states that the observation is incorrect but the prediction is correct and is also known as probability of false alarm. This curve is known as receiver operating characteristic (ROC).

4. RESULTS

Here, we are going to discuss about the results of the algorithm proposed by us. Initially dataset consisting of both forged and original images is selected and given as a primary dataset, which is further renamed and organized. Using an input image and its feature vectors the results are obtained. Fig 3. shows the input image selected by the user from the given dataset, which is further processed to detect the forgery. Once the image is selected it needs to be resized to default unsigned 8-bit integer format 256*256 pixels. Fig 4. represents the resized input image. The resized image is then subjected to grey scale conversion. This is done based on dimension value. Fig-5. shows the greyscale image and Fig-6. Shows the clustered grey scale image.

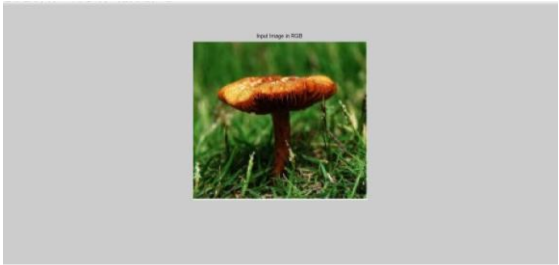


Fig-3: Selected input image

Fig-4: Resized input image

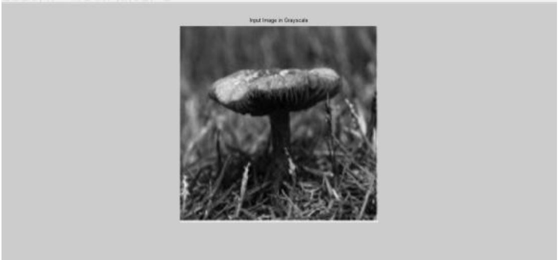


Fig-5: Greyscale image

Next the RGB image is converted to the l*a*b colour space, three most prominent colours are obtained. And k-means is initialized which forms cluster. Fig-7(a)., 7(b)., and7(c). represents the three clustered images.

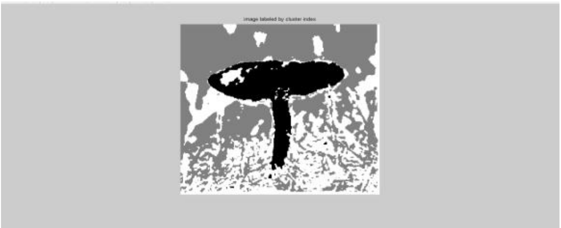


Fig-6: Clustered grey scale image

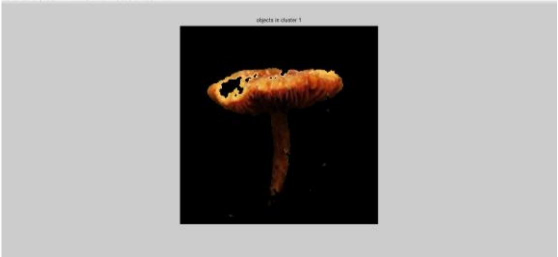
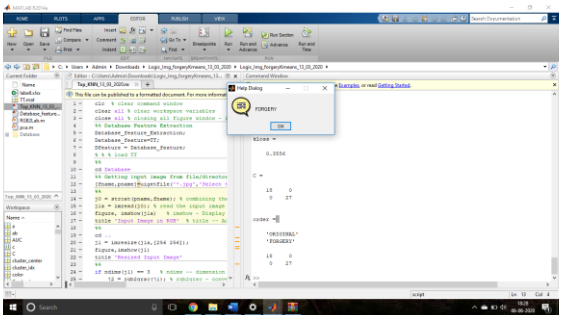


Fig-7(a): Clustered image 1

Fig-7(b): Clustered image 2

Fig-7(c): Clustered image 3

The k-nn classifier is initiated and based on it results the image is classified as forgery or original. After all the processing and calculations are done, the result indicating whether the image is forgery or not is shown with the help of pop-up dialog box 'Help Dialog' as shown in Fig-8.



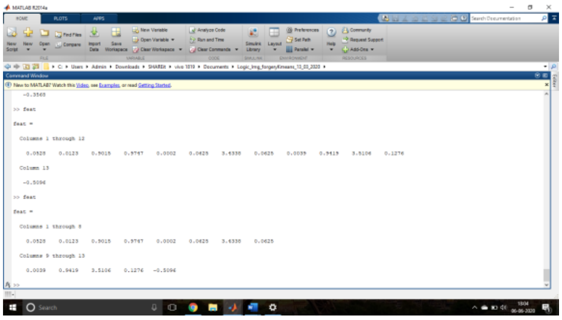
```
clc  % clear command window
clear all % clear workspace variables
close all % closing all figure window - i
%% Database Feature Extraction
Database_feature_Extraction;
Database_feature=TT;
Dfeature = Database_feature;
% % % load TT
%%
cd Database
%% Getting input image from file/director
[fname,pname]=uigetfile('*.jpg','Select t
%%
j0 = strcat(pname,fname); % combining the
j1a = imread(j0); % read the input image
figure, imshow(j1a)   % imshow - Display
title 'Input Image in RGB'  % title -- Ad
%%
cd ..
j1 = imresize(j1a,[256 256]);
figure,imshow(j1)
title 'Resized Input Image'
%%
if ndims(j1) == 3   % ndims -- dimension
   j2 = rgb2gray(j1); % rgb2gray - conve
```

FORGERY

```
kloss =

    0.3556


C =

    18     0
     0    27


order =

    'ORIGINAL'
    'FORGERY'

    18     0
     0    27

>>
```

Fig-8: Result of proposed algorithm

The thirteen parameters of the input image are extracted and stored as shown in Fig-9. These parameters are calculated for all the images provided in the dataset and stored. Based on these features the input image features are checked, in order to get the desired results.



```
  -0.3568

>> feat

feat =

  Columns 1 through 12

    0.0528    0.0123    0.9015    0.9747    0.0002    0.0625    3.4338    0.0625    0.0039    0.9419    3.5106    0.1276

  Column 13

   -0.5096

>> feat

feat =

  Columns 1 through 8

    0.0528    0.0123    0.9015    0.9747    0.0002    0.0625    3.4338    0.0625

  Columns 9 through 13

    0.0039    0.9419    3.5106    0.1276   -0.5096

>>
```

Fig-9: Feature vector of the input image

The overall performance of the proposed algorithm is graphically represented using Receiver Operating Characteristic (ROC) curve as shown in the Fig-10.

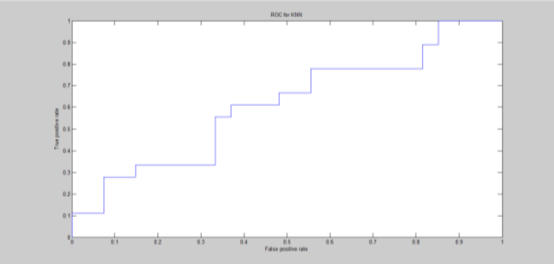


Fig-10: ROC curve

5. CONCLUSIONS

The proposed method when compared to the existing methods has higher success ratios. Thus, we believe that our technique can give a little contribution to the area of digital image forensics.

REFERENCES

[1] G. Nirmala and K. K. Thyagharajan, "A Modern Approach for Image Forgery Detection using BRICH Clustering based on Normalised Mean and Standard Deviation" 2019.

[2] Babak Mahdian, Stainslaysaicy, "Image Tampering Detection Using Methods Based on JPEG Compression Articrafts: A Real-Life Experiment" 2011.

[3] Cheng-Shian and Jyh-Jong Tsay," Passive Forgery Detection for JPEG Compress Image based on Block Size Estimation and Consistency Analysis" 2014 Applied Mathematics and Information Sciences, Vol:9, No:2 pp 1015-1028.

[4] Bo Liu, Chi-Man Pun, and Xiao-Chen Yuan, "Digital Image Forgery Detection using JPEG Features and Local Noise Discrepancies", 2014, Hindawi Publishing Corporation, doi:http//dx.doi.org/10.1155/2014/230425.

[5] Khaled W. Mahmoud, Arwa Husien Abu Al-Rukav," Moment Based Copy Move Forgery Detection Methods"2016, Vol.14 No.7, International Journal of Computer science and Information Security.

[6] Adam Novozamsky, Michal Sorel, "Detection of copy-move image modification using JPEG Compression Model", 2017, Forensic Science International, pp 47-57.