SIGNATURE FORGERY DETECTION AND CLASSIFICATION USING MACHINE LEARNING

ABSTRACT

Signature verification and forgery detection are the process of verifying signatures automatically and instantly to determine whether the signature is real or not. There are two main kinds of signature verification: static and dynamic. Static, or offline verification is the process of verifying a document signature after it has been made, while dynamic or online verification takes place as a person creates his/her signature on a digital tablet or a similar device. The signature in question is then compared to previous samples of that person's signature, which set up the database. In the case handwritten signature on a document, the computer needs the samples to be scanned for investigation, whereas a digital signature which is already stored in a data format can be used for signature verification. Handwritten signature is one of the most generally accepted personal attributes for verification with identity whether it may for banking or business. While this method uses CNNs to learn the signatures, the structure of our fully connected layer is not optimal and GLCM is used to extract the texture features of the image. In the model we will create two classes for each user real and forgery.

INTRODUCTION

The Latest technological advancement has strengthened the growth of every field, Security is being one of them. Not only legal documents can be stolen and forged, criminal evidence such as photographs and security footage can be easily tampered with. One may feel it is enough for an institution to check IDs at the front gate, but they do not realize how criminal to get their hands on fake ID’s. As mentioned before, photo editing tools which on top being easily accessible are also extremely friendly. One can learn basic photo editing tips in a few hours, even if they have never seen an image editing software before. There is nothing too advanced about photo editing anymore, whereas forgery has become even more difficult to detect. Most of the techniques used to detect those manipulations employ machine learning and pattern recognition. Region duplication can be detected by calculating the scale invariant feature transform (SIFT) key points and then finding all the pixels within the duplicated region. Digital documents that have been rotated, scaled, or resized can also be detected easily using image processing tools. Since all the databases in a security system are digital, people mostly rely on the image features that can be extracted easily. For instance, gradient based texture features, with the help of a machine, can easily be calculated and compared. Another devised scheme is to divide the image into overlapping blocks, thinking of them as vectors and find the manipulated region through radix sorting. Signature forgery detection can also be done using only image processing and without any embedded security information. Copy-move forgery (CMF) can also be detected using based on Stationary Wavelet Transform (SWT), which is able to accurately detect the duplicated blocks. Another such method for digital forgery detection is proposed. This method uses a Graphical Interface (GUI), designed to detect whether an image has been morphed or not. The GUI is designed to load the images, do its pre-processing before determining whether it has been tampered with. This is a novel methodology that allows the user to simply load the image onto the interface. Then the image is preprocessed before being analysed. This includes global contrast enhancement. Then the image is partitioned into three segments using k-means clustering, which separates the data into three sections, each having a dissimilar set of data. Then the segment containing the most information is chosen for further analysis. Then the selected segment's GLCM features are calculated and cross-validated with those of the scan in the database to determine whether the image has been morphed or not. The accuracy of proposed method was initially calculated using CNN.

Convolutional Neural Network

- **Convolution Layer:** Convolutional layers are the layers the place filters are applied to the original image, or to different feature maps in a deep CNN. This is the place most of the user-specified parameters are in the network. The most important parameters are the wide variety of kernels and the dimension of the kernels.
- **Pooling:** Pooling layers operate a specific feature such as max pooling, which takes the maximum cost in a sure filter region, or average pooling, which takes the common fee in a filter region. These are typically used to limit the dimensionality of the network.
- **Fully Connected Layer:** Fully Connected Layer is just, feed forward neural networks. Fully Connected Layers shape the remaining few layers within the network. The enter to the thoroughly linked layer is that the output from the last word Pooling or Convolutional Layer, which is flattened so fed into the wholly related layer.

Gray Level Co-occurrence Matrix

- **Contrast:** Measures the local variations in the gray-level co-occurrence matrix.
- **Correlation:** Measures the joint probability occurrence of the specified pixel pairs.
- **Homogeneity:** Measures the closeness of the distribution of elements in the GLCM to the GLCM diagonal.

Literature Survey

“Forgery Detection of Spliced Images Using Machine Learning Classifiers and color Illumination” by Tamana Sharma, Er.Mandeep Kaur : In this paper described a technique for forgery detection of composite images using machine learning classifiers LSSVM & SVM and perceptron with the help of illuminating color. They proposed a method, that describes mismatches of color and that objects are considered for forgery detection of image based on the color illumination. In this technique they improved image forgery detection. The main advantage is totally preventing user interrelationship and gives a declaration on the actuality of the forged image. Comparison between the old SVM and modified SVM using LSSVM is based upon the accuracy, the result shows better and more accuracy.

Framework for Image Forgery Detection and Classification Using Machine Learning” by Shruti Ranjan, Prayati Garhwal, and Anupama: This implemented on Graphical User Interface, based on this solution of the digitally morphed image or document; it is a foundation for investigating the digitally forged image and differentiating between the original images from digitally forged document. They can also use the advanced image editing tools in this project, it implements the required solution more instantly. Artificial neural network classifier gave a high accuracy of 96.4%as compared to linear support vector machine which gave less accuracy 87.6%

“Image forgery detection using support vector machine” by Dr Palanivel, Arthi.Z, Deepika.G and Latha.S: In this paper they have proposed the technique to find a forged region in the image or document by using the SVM classifier and PCA algorithm; it is based on the forgery detection technique of CMFD. They also did the copy-move and splicing forgery detection at the same time. They are using the SVM classifier for developing and detecting image forgery in which image include addition, replacement and removal part in the images, SVM is used to find out the similar part of an image by its matching image block. And finally, the result of output is highlighted the duplicate region in the forged image. The result shows accuracy 85.86%.

Data collection, picture processing, normalization, clustering, and evaluation are carried out in the paper "Offline Signature Recognition and Verification System utilizing Efficient Fuzzy Kohonen Clustering Network (EFKCN) Algorithm" by Dewi Suryani, Edy Irwansyah, and Ricki Chindra. RGB to Grayscale Format conversion, binary image conversion, binary image inversion, border removal, and bounding box extraction were the pre-processing techniques used. An accuracy of roughly 70% was attained utilizing this strategy.

In Tejas Jadhav's paper titled "Handwritten Signature Verification using Local Binary Pattern Features and KNN," the pre-processing methods used: RGB to Gray Scale conversion, Otsu Thresholding, and Boundary box cropping, and feature extraction methods used: LBP image generation, texture features, and name features are used as the feature extraction methods. KNN is used in this methodology together with Euclidian distance. The accuracy of this strategy is 73.34%.

The Deep Convolutional Neural Networks (DCNN) and Explainable Deep Learning are used to implement the model in the paper titled "An Offline Signature Verification and Forgery Detection Method Based on a Single Known Sample and an Explainable Deep Learning Approach" by Hsin-Hsiung Kao and Che-Yen Wen. The pre-processing techniques include rotation, block-based data augmentation, and RGB to grayscale conversion. The accuracy of this model is 94.37%.

PROBLEM STATEMENT

- The forger simply writes another person’s name, details with no basis or concern for what the other person’s signature looks like. This is common in check fraud, where a thief has stolen checks from a person and does not have a genuine signature of the account holder to use that check for accessing in the bank process.
- The forger may practice simulating the true signature which look like a hand-signature prior to signing the document.
- The forger makes duplicate certificates of driving license, passport, identification-card etc, which looks like the original certificates.
- The person attempting to trace the signature of the victim’s authentic signature on the document to fabricate.

- The currencies are always tended to get printed illegally.

OBJECTIVES

- To verify if a given signature is original or forged.
- To understand the characteristics of a signatures.
- Investigation of digitally manipulated signature and provides a solution to distinguish original signature from a digitally morphed signature.
- Examining the obtained results after transformation techniques on the morphed signature and original signature, Finally displaying the result to the user.

REQUIREMENT SPECIFICATIONS

HARDWARE REQUIREMENTS:

Processor	: Intel-Core i3, i5
Processor Speed	: 2.0. GHz
RAM: 1GB Hard Disk	: 40GB to 80GB

SOFTWARE REQUIREMENTS:

Jupyter Notebook	: Anaconda
Language	: Python

METHODOLOGY

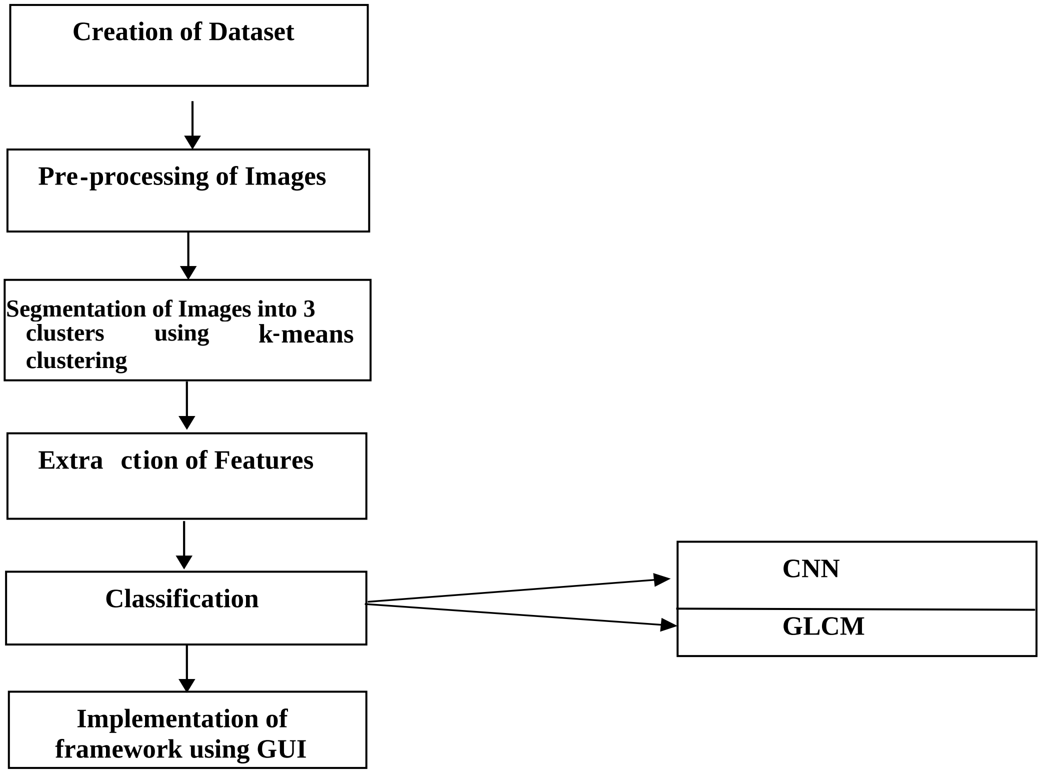
Creation of Dataset
Pre-processing of Images
Segmentation of Images into 3 clusters using k-means clustering
Extra ction of Features
Classification
CNN
GLCM
Implementation of framework using GUI

Fig: Flowchart showing cases in methodology adopted

- **Creation of Dataset:** The images used for this project were collected from various internet sources and morphed using photo editing tools. These images were edited using Adobe Photoshop CC 2017 to create a dataset with 120 pairs of images- one original and its edited version. These images were used in further analysis using PYTHON.
- **Pre- Processing of the Images:** To make the details of the images stand out more, the query image was enhanced using histogram equalization. It is a necessary step because sometimes minute forgeries go undetected through the entire process. It is important that the machine gets most of the details in one go. Histogram equalization, as the name suggests, is a method, where the intensities are adjusted using the histogram of the image. This technique is used here for contrast enhancement. Another essential stage in the pre-processing of an image is the removal of noise. De noising is again done so that the details

of the image are sharper and are not missed while extracting the features of the image. In this paper, de-noising is done using the median filter in PYTHON.

- **Segmentation:** The image is segmented into 3 clustering by using k-means clustering. K-means clustering is a technique for quantizing vectors. This method divides the image into segments, each containing mutually exclave data. This is a common method when it comes to pattern recognition and machine learning. One of the segmented images is chosen based on the information contained in it. To determine this, the GLCM features of each segment are calculated and the segment with the highest mean is chosen. The GLCM of the segmented image are then compared with the original image using cross-validation, which gives another array, which is studied to determine whether an image is morphed or not, and function for the result is added on the basis of that.
- **Extraction of GLCM Features:** Out of all the methods to analyses an image, extraction of GLCM features has proven to be efficient time and time again. The gray level co variance matrix is a tabulation that provides with statistical measures for texture analysis. This method considers the spatial relationship between the intensities of pixels in a gray-level image. In this paper, the GLCM features were calculated to study the differences in the original image and the digitally forged image. This gave 22 texture values (for each image) to work with, most of which were similar when it came to an image and its fraudulent counterpart. In practice, this would lead to redundancy and would also increase the time to run the algorithm. Also, the histogram of oriented gradient (HOG) features was calculated which gave another set of features for the original and the morphed image. The HOG values of the original and the morphed images were reasonably apart from each other, which meant that these values will be useful in differentiating the original document from the morphed one.
- **Classification:** Initially, the classifier used for classification of dataset into two parts as original. A CNN is the most suitable classifier for two-class classification problems. It finds an equivalent hyper-plane which separates the whole data by a specific criterion that depends on the algorithm applied. It tries to find out a hyper-plane which is far from the closest samples on the other side of the hyperplane while still classifying samples. It gives the best generalization techniques because of the larger margin. So, Convolutional Neural Network (CNN) classifier was applied on the dataset. CNN networks are basically a system

of interconnected neuron like layers. The interconnection of the network can be adjusted based on the number of available inputs and outputs making it ideal for a supervised learning. The CNN model was trained by providing 220 images.

Expected Outcome

Once the user uploads an image to the application it detects whether the image is an original signature and displays the result.

The results of the CNN and GLCM is showed, the final accuray is calculated separately for CNN and GLCM algorithms.

By comparing the accuracy value, the best method is stated for signature forgery detection.