Real Time Image Processing and Forensic Verification of Documents

M. Ramamoorthy, N. Ayyanathan, M. Padma Usha

Received: 20 Jan 2018 • Revised: 25 March 2018 • Accepted: 18 April 2018

Abstract: Image Processing is a computer science field which involves the analysis and manipulation of digitized image, especially in order to improve its quality. The particular document is converted to an image format in order to achieve a common goal in image processing to get an enhanced image. Enhanced Image refers to the extraction of useful information which is used to detect the fabricated documents. We use Python inbuilt command for front-end and Machine Language (which can identify and execute algorithm) for Back-end. The documents which are converted into image format may not be original always. Even though some of current framework are available to find the fake documents the reliability is not assured. So there is the need to improve the performance of search results. Hence the effectual document verification will be provided using image processing in MATLAB software. The system is used to process the document verification without any Google assistant and to control the leakage of statistical information. Fabrication of documents can be found in an efficient way and can be protected from forgery activities. Thus the system helps the user to find out the fabricated documents using the original documents.

Keywords: Image Processing, Digitized Image, Verification Documents.

INTRODUCTION

Forensic verification of documents are necessary in all sorts of organizations, people and actions whose primary intent is to promote and maintain proper documents for the welfare of the people as well as the country. In older days, where people used some physical methods such as chemicals, inks, etc to be applied on the particular document to validate the forged one.

Currently, many document verification systems are emerging in forensic verification activities that uses computerized information systems to a certain extent such that the design and implementation of integrated documents helps the user to verify the document online through a web service of their preferences. Document verification systems lie at the intersection of efficiency and timely access to online services. Timely access is important for realizing good document outcomes and is also an important determinant of which is compared with the original document. Forged document with a verification system when they attempt to duplicate a particular document as same as the original document causes effect in many cases. Document verification system is mostly carried out by a person of the specific ganization and rarely the document is taken for verification of forged activities. Online verification system is by which an end user or directly, a user can approach the website and by the online software system, the user can conveniently verify their particular document. The system also deals with the problems such as, Leakage of statistical information. Sensitive to noise and it may generate error during detection. The document can be modified easily. Digital image processing, the manipulation of images by computer, is relatively recent development in terms of man’s ancient fascination with visual stimuli. In its short history, it has been applied to practically every type of images with varying degree of success. Several factor combine to indicate a lively future for digital image processing. A major factor is the declining cost of computer equipment.
Several new technological trends promise to further promote digital image processing. These include parallel processing mode practical by low cost microprocessors, and the use of charge coupled devices (CCDs) for digitizing, storage during processing and display and large low cost of image storage arrays. Image file formats are standardized means of organizing and storing images. This entry is about digital image formats used to store photographic and other images. Image files are composed of either pixel or vector (geometric) data that are converted to pixels when displayed (with few exceptions) in a vector graphic display. Including proprietary types, there are hundreds of image file types.

In most cases, forgers apply cut and paste techniques to manipulate figures, letters, or words that consequently alter the meaning of the document. Workers have reported the use of image processing techniques in forensic document examinations. The present study aims to explore the uses of image processing techniques for the examination of computer manipulated documents. The primary goal of this research is to study and characterize the various forms of alterations that have been found in a written document.

**RELATED WORKS**

A. Neural Network

It describes about compressing documents &images and they are managed respective to their gigabytes. Given the digit image obtained at the precedent step, this digit is compared to digits images in a dataset, and using the well-known Neural Network method, after interpolations, approximations and decisions algorithm, the OCR machine which is referred from [7] outputs the closest digit in the dataset to the digit image which was entered.

B. Full-Page Text Recognition

Text line detection and localization is a crucial step for full page document analysis, but still suffers from heterogeneity of real life documents. In this paper [3] is about a new approach for full page text recognition. Localization of the text lines is based on regressions with fully convolutional neural networks and multidimensional.

C. Optical Character Recognition

The optical character recognition (OCR) of the Telugu script has wide ranging applications including education, health-care, administration etc. The Telugu script however is very different from Germanic scripts like English and German. This makes the use of transfer learning of Germanic OCR solutions to Telugu a non-trivial task which have been discussed in [12].

D. Automatic Classification of Handwritten and Printed Text

Machine printed and handwritten texts intermixed appear in the ICR cells of variety of documents referred from [1] Recognition techniques for machine printed and handwritten text in these document images are significantly different.

**DOCUMENT VERIFICATION SYSTEM**

The ultimate aim of this paper is to create document verification system that will help users to be aware of the forged documents and can differentiate the duplicated documents from the original one. The system allows users to manage and verify their documents online. Users are allowed to verify their documents online and those documents are compared with the original one. The system manages the verification of documents for multiple purposes of various type of documents. Each time a user can check or verify his/her document with the original document updated to the database. The verification includes character recognition, preprocessing, etc which is used to verify the particular document whether forged or not. At the same time the user must also know to manage and identify the verification details of the duplicated documents. The original document which is to be compared is uploaded to the database prior to the duplicated document. Finally the particular document to be verified is validated through online with the preprocessing steps included and compared with the original document in the database.

Different computer monitors may use different sized pixels. The pixels that constitute an image are ordered as a grid (columns and rows); each pixel consists of numbers representing magnitudes of brightness and color. Image file size is expressed as the number of bytes that increases with the number of pixels composing an image, and the color depth of the pixels. The greater the number of rows and columns, the greater the image resolution, and the larger the file. Image file formats are standardized means of organizing and storing images.
This entry is about digital image formats used to store photographic and other images. Image files are composed of either pixel or vector (geometric) data that are converted to pixels when displayed (with few exceptions) in a vector graphic display. Including proprietary types, there are hundreds of image file types. The PNG, JPEG, and GIF formats are most often used to display images on the Internet.

In forensic verification of documents the particular document is given as an image format which is processed further with the preprocessing steps and so on which is described in the figure.

![Figure 1: Outline of document verification system](image1)

Image is a two-dimensional, such as a photograph, screen display, and as well as a three-dimensional, such as a statue. They may be captured by optical devices—such as cameras, mirrors, lenses, telescopes, microscopes, etc and natural objects and phenomena, such as the human eye or water surfaces. The modules of the proposed system are:

1. Preprocessing
2. Segmentation
3. Morphological Function
4. Character Recognition

![Figure 2: Verification System](image2)

**A. Preprocessing**

![Figure 3: Preprocessing steps](image3)
Noise Removal: An image may be "dirty" (with dots, speckles, stains). Noise removal is to remove speckles/dots on an image which can be referred in fig.3. Dots can be modeled as impulses (salt-and-pepper or speckle) or continuously varying (Gaussian noise) and can be removed by taking mean or median values of neighboring pixels (e.g. 3x3 window) which is equivalent to low-pass filtering.

Sharpening: To enhance line structures or other details in an image referred in fig.3. Enhanced image = original image + scaled version of the line structures and edges in the image. Line structures and edges can be obtained by applying a difference operator (high pass filter) on the image.

B. Segmentation

The goal of image segmentation is to cluster pixels into salient image regions, i.e., regions corresponding to individual surfaces, objects, or natural parts of objects. In computer vision segmentation refers to the process of partitioning a digital image to multiple segment referred in fig.4. The goal of segmentation is to simplify and/or change the representation of an image into something that is more meaningful and easier to analyze Image segmentation is typically used to locate objects and boundaries (lines, curves, etc.) in images. More precisely, image segmentation is the process of assigning a label to every pixel in an image such that pixels with the same label share certain visual characteristics.

The result of image segmentation is a set of segments that collectively cover the entire image, or a set of contours extracted from the image. Each of the pixels in a region are similar with respect to some characteristic or computed property, such as color, intensity, or texture. Adjacent regions are significantly different with respect to the same characteristics.

C. Morphological Function

A shape (in blue) and its morphological dilation (in green) and erosion (in yellow) by a diamond-shape structuring element. Mathematical morphology (MM) is a theory and technique for the analysis and processing of geometrical structures, based on set theory, lattice theory, topology, and random functions referred in fig.5. Mathematical morphology is most commonly applied to digital images, but it can be employed as well on graphs, surface meshes, solids, and many other spatial structures.

D. Character Recognition

In order to segment the characters in the binary license plate image the method named peak-to-valley is used. The methods first segments the picture in digit images getting the two bounds of the each digit segment according to the statistical parameter referred in fig.6.
Firstly the fabricated document is converted to image format to proceed the upcoming process. So, the fabricated document is converted to a RGB image format. Then the particular document image is processed with pre-processing steps. A web application for making document verification with ease. The system develops an online document verification system to verify the particular document is duplicated or not by comparing with the original document.

The system will also include various methods of preprocessing steps to be applied to the particular document image and the process is carried out with the image processing steps. The pre-processing includes grey scale conversion, filtering etc. After pre-processing the next step includes threshold segmentation in which the threshold value is given according to the pixel values of the image. After threshold segmentation it is processed by morphological function. In morphological function the particular document image is traced and forgery or morphological activities are found. The next process is character recognition in which character is recognized.

According to the font size, spacing, etc. Finally the image is matched with the original document and the fake document is identified efficiently. The working principle is described in figure 7. Neural network is the best tool in recognition and discrimination between different sets of signals. To get best results using the neural network, it is necessary to choose a suitable architecture and learning algorithm. Unfortunately there is no guaranteed method to do that. The best way to do that is to choose what is expected to be suitable according to our previous experience and then to expand or shrink the neural network size until a reasonable output is obtained. In this work we tried different sizes for the neural network using MATLAB.
The implementation of Real time image processing and forensic verification of documents is developed by using Python inbuilt commands (which is defined in inbuilt library) for front-end and Machine Language (which can identify and execute algorithm) for Back-end. The process of this developed system also includes a major coding activity which has been used to enhance the performance for performing better verification processes. It is carried out in MATLAB software which is an interactive system whose basic data element is an array that does not require dimensioning. This allows you to solve many technical computing problems, especially those with matrix and vector formulations, in a fraction of the time it would take to write a program in a scalar non-interactive language such as C or FORTRAN.

The first step includes the input of the particular document in the image format via image acquisition and further analyzing and manipulating the image. The altered image is then processed for preprocessing steps which includes image adjusting, image smoothening and filtering process. Texture is that innate property of all surfaces that describes visual patterns, each having properties of homogeneity. There are three principal approaches used to describe texture; statistical, structural and spectral. Statistical techniques characterize textures using the statistical properties of the grey levels of the points/pixels comprising a surface image. Typically, these properties are computed using; the grey level co-occurrence matrix of the surface, or the wavelet transformation of the surface. Structural techniques characterize textures as being composed of simple primitive structures called “textures” (or texture elements). These are arranged regularly on a surface according to some surface arrangement rules. Spectral techniques are based on properties of the Fourier spectrum and describe global periodicity of the grey levels of a surface by identifying high-energy peaks in the Fourier spectrum.

EXPERIMENTS AND RESULTS

The revised model of the project describes the image processing steps with the input as the image of the particular document. It is processed further with the preprocessing steps such as image sharpening, image smoothening and image filtering. The preprocessing methods are followed by the segmentation process which includes the threshold value.

Morphological function process are performed using Mathematical morphology. It is then followed by Character segmentation which includes the process of segmenting each character throughout the document in order to process the differentiate between duplicated and original documents. If a document is identified as a duplicate or a fake one then respected actions are ought to be taken. This paper determined a system to identify the fake documents with the above mentioned process and implementation.

In this era of cybercrime, computers are used most frequently to commit certain crimes, such as altering documents to change their meaning. In this paper, we studied alternations in system-generated documents, including documents with printed backgrounds as well as plain backgrounds. The manipulation of the documents was done with the help of image processing software applications including Adobe Photoshop and Paint. After manipulating the documents, they were examined for alterations; encouraging results were obtained. The results demonstrated that a number of features were associated with image manipulation and could easily be detected using standard image processing applications.

The future enhancement is focused on improving the system in a better way by providing the verification system in an efficient manner with a clear prediction of the document forgery. The document to be matched with is uploaded to the database prior to the document to be checked. Thus the database must be maintained without any loss of information.

REFERENCES


