ISSN: 0067-2904

Multilevel Analysis to Recognize Original Voucher from Faked Voucher

Hasanen S. Abduallah*, Nesreen Waleed
Departmentof Computer Science, University of Technology, Baghdad, Iraq.

Abstract
Voucher documents have become a very important information carrier in daily lives to be used in many applications. A certain class of people could exploit the trust and indulge in forging or tampering for short or long term benefits unlawfully. This holds a serious threat to the economics and the system of a nation. The aim of this paper is to recognize original voucher document through its contents. Forgery of voucher document could have serious repercussions including financial losses, so the signature, logo and stamp that are used to determine being a genuine or not by using multilevel texture analysis. The proposed method consists of several operations. First, detection and extraction of signature, logo and stamp images from original voucher document by using auto crop method. Second, each image is processed in allotted level. Third, the voucher document is classified depending on a result of each level to determine being a genuine or not. Accuracy of 94% for identification process and 95% for verification process were achieved.

Keywords: Texture Feature, Naïve Bayesian, Faked, Voucher Document, Auto Crop process, Signature, Logo, Stamp.

تحليل متعدد المستويات لتميز المستند الاصلي من المستند المزيف

حسنين سمير عبد الله*، نسرين وليد عبد الواحد
قسم البرامجيات، علوم الحاسبات، الجامعة التكنلوجيا، بغداد، العراق.

الخلاصه
تعتبر مستندات الايصال ناقل معلومات مهم جدا في الحياة اليومية و ذلك لاستخدامها في العديد من التطبيقات. يوجد هنالك فئة معينة من الناس يمكن ان تستغل الثقة لغرض التزوير او العبث بشكل غير قانوني لتحقيق فوائد على المدى القصير أو الطويل. هذا التزوير يعتبر تهديدا خطيرا للاقتصاد ونظام الدولة. الهدف من هذا البحث هو لتميز المستند المزور بالاعتماد على محتوى المستند. تزوير المستندات يمكن أن يكون له انعكاسات خطيرة مثل الخسائر المالية، لذلك سوف يتم استخراج التوقيع ،الشعار والختم لتحديد كون المستند مزور أم لا باستخدام تحليل نسيج متعدد المستويات. تتكون الطريقة المقترحة من عدة عمليات. اولا، تحديد واستخراج التوقيع، شعار والختم من المستند الاصلي باستخدام طريقة القطع الالي. ثانيا، كل صورة يتم معالجتها في المستوى المخصص لها. ثالثا، يصنف المستند بالاعتماد على نتيجة كل مستوى لتحديد كونه اصلي أم لا. وكانت الدقة المحققة لعمليه تحديد الهوية هي 94٪ ولعملية التحقق هي 95٪.

1. Introduction

In spite of the enormous use of computer technology in various areas of our lives, voucher documents are still widely used all over the world for commercial forms, financial transactions, Bank cheques (checks) and government records etc. they still play a big role in the non-cash transactions in the world even after the arrival of debit cards, credit cards and other electronic means of payment [1].

*Email: nesreen.waleed@gmail.com

All vouchers issued by formal authorities are printed on a paper and contained a stamp or a signature to guarantee the originality of the content [2]. Even though voucher documents are secured, but nevertheless they suffer from a lack of security [3]. Document security does not play an important role in specific fields such as checks, passports and degrees and also in everyday documents like bills and voucher documents. Using high-security features for everyday documents are not useful due to the complexity and the cost of these methods [4]. Currently, the effort, money and much time can be saved, if the purpose of the process for recognition and verification of the voucher document based on image processing. In general, the need of the society has motivated to propose method depending on different approaches to detect fabrication in an original voucher.

2. Related Works

The work proposed in this paper includes significant previous works related to texture analysis methods. Such works are almost not available for ensuring the authenticity of the voucher and the summarized related work as follows:

In 2016, Kumar et al. [5] presented an approach for a logo classification system into three classes. The proposed system used color, texture, and shape features. The K-Nearest Neighbor (KNN) classifier was used for classification of logo. The experimental results showed that the most promising results were obtained.

In 2014, Zouari et al. [6] proposed the system for identification and verification processes of the signature. In this research, the fractal approach was used as a method of feature extraction. The identification phase had used K-Nearest Neighbor (KNN) classifier. The experimental results had shown Accuracy identification rate of 95% and verification rate of 83%.

In 2014, Ashok et al. [7] proposed a method to verification process of the signature. This method was based on GLCM as a method of texture analysis. The classification was used Feed Forward Back Propagation Neural Network (FFBPNN) to classify signature into two different classes either a genuine or a forged. The system gave an overall Accuracy of 92.08%.

In 2011, Barbora Micenkov [8] presented a thesis for verification process of stamp image. It was consisted from two stages: first, stamps were detected and extracted. Second, the stamp candidates were classified into two classes genuine and copied stamps. Feature extraction used in this stage was the standard deviation of hue and they proposed new features related to the quality of the print. Support Vector Machines (SVM) classifier was used for detection and verification. In this thesis, Precision and Recall of the identification were 90% and 89% respectively and Accuracy of the verification for copied 90% and for genuine 95%.

3. Theoretical Background

The proposed system has developed a number of different tools to be used for the automatic detection of the forgery as shown the following:

3.1 Region of Interest

A Region of Interest (ROI) means extraction of specific area within the image for investigation more closely. To do this, geometries operations are needed to modify the spatial coordinates of the image, the geometry operations include crop, shrink, translate, zoom, enlarge and rotate.The image crop process is selecting part of the image and then cutting it away from the rest of the image [9]. Auto Crop process (AC) is the fast procedure to extract the ROI with speed time, reducing the computational complexities and keeping the content information of the identified regions (ROI) without loss of any pixel and accurately [10].

3.2 Edge Detection

Edges are basic image features. They carry useful information about boundaries of an object. The edge detection is used for image analysis and object identification as well. The edge detection methods are used to find complex object boundaries when changes in brightness occur by marking potential edge points corresponding to places in an image, these edge points can be merged to form lines and object outlines. The Sobel operator is one of the best "sample" edge detection methods and used in this paper [9].

3.3 Morphological Operators

Morphological operators are used to distinguish objects and structure of objects in the image. It simplifies a segmented image to facilitate the search for objects of interest. This is done by smoothing out object outlines, filling small holes and eliminating small projections. The two principal morphological operations are dilation and erosion. Dilation allows objects in an image to be expanded, thus potentially filling in small holes and connecting disjoint objects in an image. Erosion shrinks objects in the image by etching away (eroding) their boundaries. The closing operator consists of dilation followed by erosion and can be used to fill in holes and small gaps in an image. The closing operator will be connecting small and adjacent objects in an image [9].

3.4 Feature Extraction

Feature extraction plays a very important role in the image processing. The main purpose of feature extraction is to extract the most important information from the original image [11]. In this paper, a set of features are used which are as follows:

3.4.1 Binary Object Features

The binary features are used to locate and classify the binary objects in image. In order to extract object features, all the binary objects in the image should be labeled, then, each object can be treated as a binary image. The binary object features can be computed from binary image including area, center of area, projection and aspect ratio [9].

3.4.2 Segmentation-Based Fractal Texture Analysis (SFTA)

SFTA is an efficient texture feature extraction method from binary images. This method is known to achieve accuracy and higher precision than Haralick and Gabor filter banks for image classification and content-based image retrieval. Additionally, SFTA is at least 1.6 times faster than that of Haralick and 3.7 times faster than that of Gabor with respect to feature extraction time. SFTA algorithm consists of two main stages which are: The first stage is decomposing the input gray level image into a set of binary images by using a new technique named Two-Threshold Binary Decomposition (TTBD). Then, the mean (M), size (pixel counting) and Fractal Dimension (FD) features are computed from resulted binary images [12].

3.4.3 Color Features

The color is an important feature that makes possible recognition of images. The color is used to show the difference between objects [13].Usually, color images are represented as RGB (Red, Green and Blue) images. These relationships are done by using the color transforms, the RGB color information is transformed into a mathematical space that decouples the color information from the brightness information. The Hue/Saturation/Lightness (HSL) color space decouples the image brightness from the color itself, where the hue is what we normally think of as “color” (like, green, or orange), the saturation is a measure of how much white is in the color and the lightness is the brightness of the color [9]. Color features are computed which are mean, variance and skewness [14].

3.4.4 Grey Level Co-occurrence Matrix (GLCM)

GLCM approach provides information about the relative position of the neighboring pixels in an image. It is one of the earliest techniques used for texture feature extraction and still popular so far [15]. It has become the most well-known method and widely used in many texture analysis applications.It is a tabulation of how often different combinations of gray level (pixel brightness values) occur in an image. To extract texture features from images GLCM is used, these features are important for image classification process. GLCM describes the probability of grey level (i) happening at a distance (d) in direction (θ) from grey level (j) in a gray level image. GLCM directions are: Horizontal ($\theta = 0°(0\ 1)$), right diagonal ($\theta = 45°(-1\ 1)$), vertical ($\theta = 90°(-1\ 0)$) and Left diagonal ($\theta = 135°(-1\ -1)$). These probabilities produce the co-occurrence matrix $M(I, J \mid d, \theta)$. Distance (d) to the next adjacent neighbor preferably is equal to one [16]. Feature extraction from GLCM can be extracted contrast, energy and homogeneity[17].

3.4.5 Law's Texture Energy Measures (LAWS)

Law's Texture Energy Measures (LAWS) method is used to extract texture features of the gray level image. It is based on masks (convolumeution masks) that target to filter the image. This method has a superior ability to extract texture of an image [18]. It is used for the best performance and ease of implementation based on providing several masked images that have a different view of the original image. Therefore, more useful features may be acquired from masked image [19].The LAWS has four vectors of one dimension (1-D) which are level= [1 4 6 4 1], edge= [-1 -2 0 3 1], spot = [-1 0 2 0 -1] and ripple = [1 -4 6 -4 1]. Each 1-D vector is used to determine the type of texture in the image. The masks (2D) are generated by computing outer products of pairs of 1-D vectors. Each mask has its own special properties. By convolumeuting the masks (2D) with a texture image and then calculating statistic features, the resulted features vector can be used for texture description. The statistical features of Mean and Variance are used for describing the filtered image [18].

3.5 Image Classification

Image classification is a subject of pattern recognition in computer vision. It analyzes the numerical properties of various image features and organizes data into classes [20, 21]. In this paper, Naïve Bayesian (NB) is used in identification process. Naïve Bayesian (NB) is a classification technique built on Bayes' Theorem with independence assumption among predictors. In sample terms, a Naïve Bayesian classifier supposes that the presence of a particular feature in a class is unrelated to the presence of any other feature. This supposition is called class conditional independence. The NB classifier represents a statistical method for classification as well as a supervised learning method. The NB is easy to build and particularly useful for very large datasets [22, 23].

Studies comparing classification algorithms have found sample Naïve Bayesian classifier known as the Bayesian classifier to be comparable in performance with Neural Network classifiers and decision tree [23].

4. Proposed Multilevel Analysis System to Recognize Original Voucher

The proposed system is used in order to authenticate a voucher document depending on multilevel texture analysis methods. It processes an incoming voucher document by using a number of different methods to allow automatic detection of forgery. The proposed system consists of two phases which are the training phase that aims to train the classifier and the testing phase that aims to determine the voucher image being a genuine or not. These two phases consist of main processes which are preprocessing & feature extraction. Figure- 1 shows general block diagram of the proposed system.

5. Dataset

The dataset is a significant part to test the system for proving originality of the voucher image. As far as we know, there is neither a dataset of fraudulent voucher documents available nor a dataset of genuine voucher documents. Thus, to overcome this obstacle, fraudulent and genuine voucher documents are proposed (created and collected)**.** The total number of voucher images that are used in this paper is 200 under different light conditions and noise.

6. Preprocessing

Preprocessing stage contains several sub-processes that aim to make the input voucher image as uniform as possible to facilitate the Auto Crop process. The processes that are used in this paper are: gray level conversion, noise removal and binarization. Then, the Auto Crop process is applied on voucher image as follows:

a. Auto Crop Process

The Auto Crop process is a fast operation to select the ROI. A signature, logo and stamp (or any other important images) are extracted from voucher image in this paper through the Auto Crop process. It reduces the time cost and computational complexities and keeps the content information of the ROI. In this process, the signature, logo and stamp are detected and extracted from original voucher image and then passed to allotted level of the proposed system. After, the voucher image is converted into binary image, then, the Auto Crop process includes: edge detection, morphological closing operation, connect component labeling, binary feature extraction and classification.

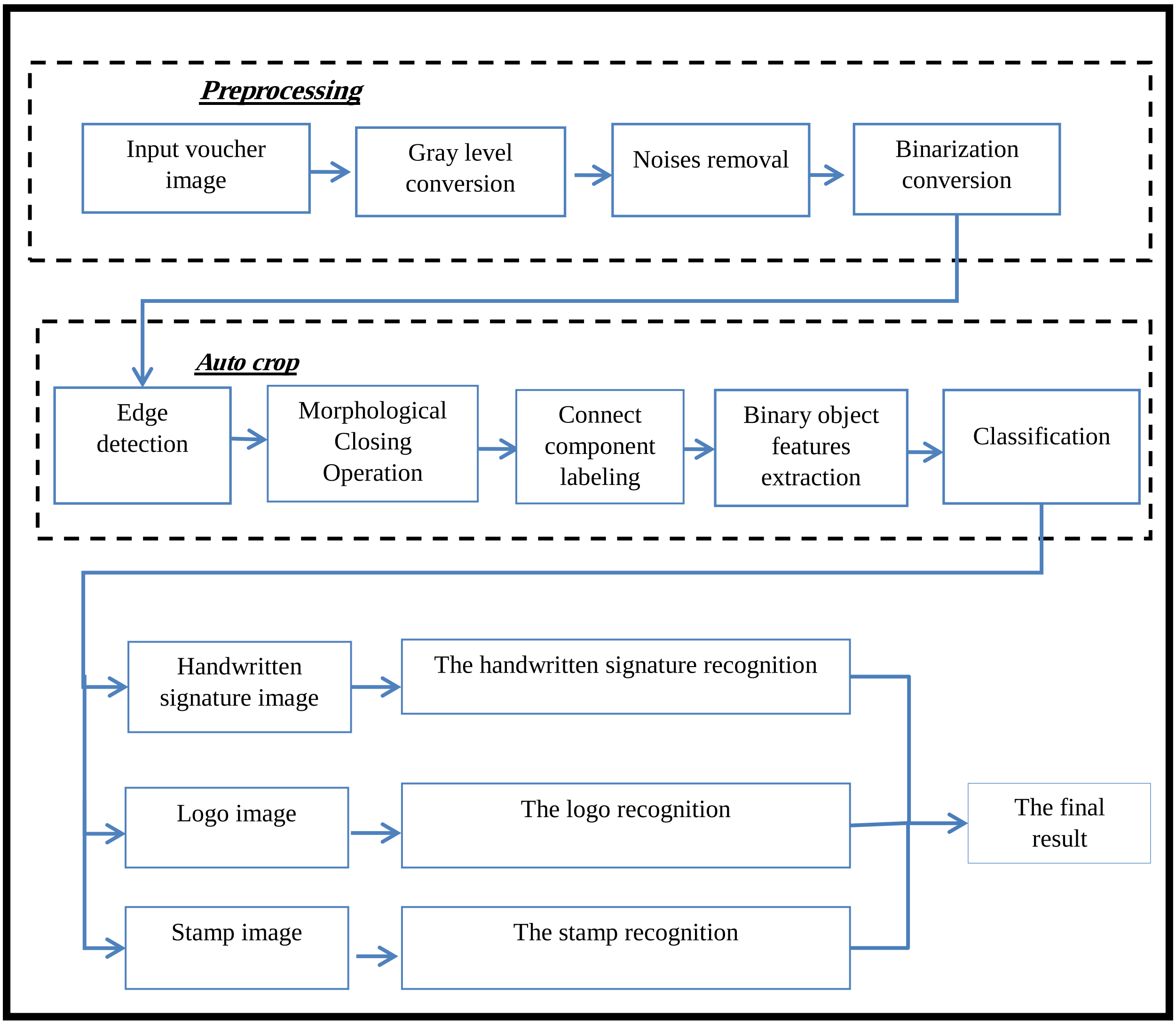

Figure 1- General block diagram for the proposed system.

7. the Training Phase

The training phase is a fundamental and significant phase to construct the classifier model**.** The proposed system request sample images of handwritten signatures, logo and stamps to be used in verifying the authenticity of the voucher documents. These samples consists 15 signature samples that are taken for each person, one sample for each logo image and 10 stamp samples for each stamp. The samples are used to train the classifier. The training phase consist preprocessing, feature extraction, training NB classifier and computing thresholds. Algorithm 1 describes the training phase of the proposed system.

7.1 Preprocessing

Preprocessing is an essential step due to its effectiveness on the identification and verification rates. Several steps have been taken in the preprocessing that make the proposed system obtains high accuracy. Different preprocessing steps were taken depending on type of sample (signature, logo or stamp) because texture feature is extracted from binary signature image while for the logo and stamp images preprocessing extract color features from color image and texture features from grey level image.

- Preprocessing of signature image consists signature body clipping, size normalization, converting a signature image to the gray level, noise removal by using a median filter, converting into a binary image by using Otsu threshold.
- Preprocessing of logo and stamp images consists image body clipping, size normalization, converting color image to gray level image to be used in the texture features and then noise removal by using a median filter.

7.2 Feature Extraction

Feature extraction is a significant step for image classification. Several features are extracted from samples, the SSE and improved SFTA texture features have been taken from signature image, the color and GLCM features from logo image and the color and LAWS features from stamp image.

i. Proposed Method for Signature Shape Extraction (SSE)

The proposed Signature Shape Extraction (SSE) method extracts the required features based on the signature shape in the input image. The input to the method is a gray level image and the output will be a number of features that will be saved as one dimension vector. The main steps of the proposed method are described as in the following:

- The first step of the proposed method is converting the input signature image into binary using Otsu threshold method. Then, dividing the obtained binary image into three vertical parts and save only the first part on the left side to be used for further processes.
- Applying thinning algorithm on the saved part of the signature in order to make the extracted features invariant to the characteristics of signature image such as quality of paper and pen type.
- Connected regions are extracted from the thinned image using the connected component labeling (CCL) process and each connected region is saved in an array of region pixels.
- Twelve pixels from the array of region pixels are detected and extracted based on their location in the signature image. Where, the first pixel of the array is taken and the next five pixels are ignored then the sixth one is taken and so on.
- The last step of the SSE method is creating the features vector which contains the number of the CCL in the thinned image and twelve pixels locations.

ii. Improved SFTA Feature Extraction Method

Segmentation-based Fractal Texture Analysis (SFTA) is an efficient texture feature extraction method and it achieves higher accuracy and precision. It is necessary for optimizing the feature extraction time and accuracy. As a result, the SFTA algorithm is improved in this research and it is suggested as a method for signature recognition.

The improvement of the SFTA algorithm as following:

- In the first step, signature image is converted to a binary image by Otsu threshold algorithm instead of multilevel Otsu threshold because Otsu threshold is based on a discriminant analysis which is a non-parametric, unsupervised method and reducing the computational cost to compute threshold. This method assumes that the image contains two classes of pixels: background pixels and foreground pixels. The optimal threshold is calculated for separating the background and the foreground information.
- In the second step, three types of features are extracted namely: computing Mean (M) feature the same as that used in SFTA algorithm, computing Area (A) feature instead of size (pixel count) feature because it takes shape into account and computing Fractal Dimension (FD) feature from edge detection image (edge detection image is found by applying Sobel operator to find edge detection instead of border image).

iii. Color Features

The color has considered as one of the most important characteristics of an image. The image (RGB color information) is transformed into HSL color model to decouple the color information from the brightness information. The HSL color space has three channels (H, S and L). For color feature extraction from logo image, the H and S channels will be taken because they contain the color information, the Lightness information (L) channel is ignored because it sensitive to less difference in lighting conditions of the input logo image and this is causing create color features differently. Each H and S channels are divided into four partitions because the logo image usually contains many colors. Then, each partition Mean, Color Variance and Color Skewness features are computed. For color feature extraction from stamp image, each H and S channels are computed Mean, Variance and Skewness features from stamp image without divided it because the stamp image contains mainly one color rarely it contains two color or more.

iv. Gray-Level Co-occurrence Matrix (GLCM)

Texture plays a major role in recognition and gives us information about the spatial arrangement of the pixel in the image. The logo image is converted into gray level in order to produce Gray-Level Co-occurrence Matrix (GLCM) which is used to extract texture features from the logo image. The logo image is divided into four partitions. To extract the texture from the logo image, first the Gray-Level Co-occurrence Matrix (GLCM) will be computed for each partition in the diameter direction (0, 45) and distance d=1. Then, Contrast, Energy and Homogeneity features are computed for each partition.

v. Law's Texture Energy Measures (LAWS)

Texture feature is extracted to achieve important information from stamp image. After converting the stamp image into the gray level, Law's Texture Energy Measures (LAWS) is applied to extract texture features. The LAWS is suggested as a method for stamp identification and verification processes. Level (L), Edge (E) and Ripple (R) vectors are used because Level (L) vector is used to compute weight local average gray level for clear and unclear stamps while Edge (E) and Ripple (R) vectors are used to explain shapes and lines of the stamps because a fake stamp can be distinguish by their shapes and lines which are much sharper. To extract the texture features of the stamp image, the following steps should be applied:

- First step: Level (L), Edge (E) and Ripple (R) vectors are multiplied with themselves to produce a set of 9 masks each mask 5*5 dimension.
- Second step: Each mask is convolumeuted with the stamp image to produce 9 filtered images.
- Final step: For each filtered image Mean and Variance are computed.

7.3 Naïve Bayesian Classifier

The right choice of the classification algorithm to be used in identification process is highly dependent on the properties and the format of the features that represent the image. The NB classifier is used to train feature vectors.

7.4 Verification Process

Verification is the process of verifying whether an image is genuine or not. To do this, thresholds are proposed to be used in verification process as follows:

i. Proposed Method of Intra-Person Signature Thresholds (IPST)

In the handwritten signatures, there is a variation that appears even for the same person. An Intra-Person Signature Thresholds (IPST) method is proposed to compute the variation range between the signatures of the same person. The proposed method produces unique lower and upper thresholds for each person. The IPST method is proposed to be used in verification process of the testing phase. However, to calculate variation range of each person, the following steps are taken place:

- Step one, each person has fifteen feature vectors that are placed in Handwriting Signature Template Features database (HSTF).
- Step two, HSTF is divided into two parts A and B, first part A contains first five feature vectors and second part B contains ten remaining feature vectors.
- Step three, Euclidean distance is calculated between A and B. In which Euclidean distance is calculated between all feature vectors in A and the first feature vector in B, minimum distance is saved in Signature Template Ranges database (STR). The process is repeated for the remaining feature vectors in B.
- Step four, the lower and upper values are choosed from STR and called Lower Threshold (LT) and Upper Threshold (UT) respectively. Then, LT and UT values are saved in Handwriting Signature Template Thresholds database (HSTT).
- Step five, the obtained LT and UT for all persons in HSTT database are used to compute global thresholds for the all training signatures. Furthermore, the Global Lower Threshold (GLT) is computed from all LT in HSTT database by using Equation 2 and the Global Upper Threshold (GUT) is computed from all UT in HSTT database by using Equation 3.

$$GLT = LT1 + \cdots \ldots + LTn \,/\, N \tag{2}$$

$$GUT = UT1 + \cdots \ldots . + UTn \,/ N \tag{3}$$

Where N: number of persons.

ii. Accuracy Rate of Logo Image

The logo image has one situation at each time, but it is different in brightness when occurred during scanning, this causes different feature vectors. Therefore, the feature matching is applied to determine the Accuracy Rate (AR). AR is computed by using Euclidean distance between reference feature vector (training set) and feature vector of logo image (testing set).

iii. Proposed Method of Intra Stamp Thresholds (IST)

The stamps in different situations are not similar to each other. Therefore, it is important how to select stamps to determine thresholds to be used in verification process. The Intra Stamp Thresholds (IST) method is proposed to be used in verification process of the testing phase. Each stamp has 10 samples, first five samples contain clear stamp and the remaining five contain unclear stamps. The IST method produces four types of threshold for each stamp, where lower one and upper one thresholds were for clear stamp while lower two and upper two thresholds are for unclear stamp. However, in order to calculate threshold of each stamp the following steps are applied:

- Step one: Each stamp has ten feature vectors stored in stamp features database and it is placed in Stamp Template Feature database (STF).
- Step two: The STF method is divided into two parts A and B where each part contains five feature vectors.
- Step three, Euclidean distance is calculated between A and B. In which Euclidean distance is calculated between A and the first feature vector in B, where the minimum distance is saved in Template one database as well as for remaining feature vectors in B.
- Step four: Euclidean Distance is calculated between the first feature vector in part A and the second feature vector in part A and the distance is saved in Template two database. The process is repeated for the third, fourth and fifth feature vectors in A that are calculated Euclidean distance with the first feature vector in A.
- Step five: Lower Threshold1 (LT1) and Upper Threshold1 (UT1) values are determined from Template one database and saved in Stamp Template Threshold database (STT).
- Step six: Lower Threshold 2 (LT2) and Upper Threshold 2 (UT2) are determined from Template two and saved in Stamp Template Threshold database (STT).
- Step seven: The obtained LT1, UT1, LT2 and UT2 for all stamps in STT database are used to compute global thresholds (GLT1, GUT1, GLT2 and GUT2) for all the training stamps by using Equations 4, 5, 6 and 7 and then stored in stamp thresholds database.

$$\boldsymbol{GLT1} = LT1 + \ldots\ldots + LTN\,/N \quad (4)$$

$$\boldsymbol{GUT1} = UT1 + \ldots\ldots + UTN\,/N \quad (5)$$

$$\boldsymbol{GLT2} = LT2 + \ldots\ldots + LTN\,/N \quad (6)$$

$$\boldsymbol{GUT2} = UT2 + \ldots\ldots + UTN\,/N \quad (7)$$

Where N: number of stamps.

8. Testing Phase

The testing phase uses voucher images that are different from the images used in training phase. The testing phase consists of three steps which are preprocessing, feature extraction and classification**.** After the signature, logo and stamp are extracted from the voucher image by using Auto Crop process. Each image is processed in allotted level. Algorithm 2 describes the testing phase of the proposed system.

8.1 Preprocessing

This step is similar to the preprocessing step in training phase of the proposed system.

8.2 Feature Extraction

This step is similar to the feature extraction step in training phase of the proposed system.

8.3 Classification

Classification process is designed for identification and verification processes of a signature, logo and stamp images. Therefore, there are two different purposes for classification: (1 identification process and (2 verification process. Since the number of signature, logo and stamp images available for comparison is huge, therefore, the classification is designed to determine the class of each images that it belongs to and then verifying it. Each signature, logo and stamp is processed in allotted level.

Classification process of the signature image as follow:

- ✓ In identification process, the NB classifier should determine the class number that the test signature image belongs to.

- ✓ In verification process, the test signature is examined to determine whether it is a genuine or not. Therefore, the input for the verification process (1) feature vector for the test signature image (2) feature vectors for the class number (3) GLT and GUT. To verify the authenticity of the test signature image, the following steps should be applied:
- Step one, Euclidean distance is calculated between the feature vector of the test signature image and feature vectors of the class number.
- Step two, Minimum Euclidean Distance (MED) is determined.
- Step three, GLT and GUT are taken from signature thresholds database.
- Step four, compare MED with GLT and GUT, if GUT > MED > GLT, then the test signature image will be accepted; otherwise it will be marked as a forgery and will be rejected.

Classification process of the logo image as follow:

- ✓ In identification process, the NB classifier determines the class that a test logo image belongs to. The input of the NB classifier is feature vector (color, texture that have been extracted from the test logo image) and output from the NB classier is a class number that determine the class that it belongs to.

  After identification process, the feature vector is taken from logo features database based on class number.
- ✓ In verification process, the test logo image is examined to determine whether it is a genuine or not. Therefore, the input of verification step (1) feature vector for the test logo image (2) feature vector for the class number. The process of feature matching is applied as in the following steps:
- First step, Accuracy Rate (AR) of the test logo is computed between features vector of the test logo and feature vector of the class number.
- Second step, If AR of the test logo > = AR of the training logo, then the test logo image will be accepted; otherwise it will be marked as a forgery and will be rejected.

Classification process of the stamp image is as follow:

- ✓ In identification process, the NB classifier determines the class of a test stamp image. The input of the NB classifier is features vector (color, texture that have been extracted from the test stamp image) while the output from the classifier is the class number that it belongs to.

  After identification step, the feature vectors are called from stamp features database based on class number and GLT1, GUT1, GLT2 and GUT2 are taken from stamp thresholds database to be used in verification step.
- ✓ In verification step, test stamp is examined to determine whether it is a genuine or not. Therefore, the input of verification step will be (1) feature vector for the test stamp (2) feature vectors for the class number (3) GLT1, GUT1, GLT2 and GUT2.

To verify the authenticity of test stamp, the following steps are applied:

- Step one: Euclidean distance between the feature vector of the test stamp image and feature vectors for class number is calculated.
- Step two: Minimum Euclidean Distance (MED) is determined.
- Step three: GLT1, GUT1, GLT2 and GUT2 are taken from stamp thresholds database.
- Step four: Compare MED with GLT1, GUT1, GLT2 and GUT2 and if GUT1 > MED > GLT1 OR GUT2 > MED > GLT2, the test stamp will be accepted. Otherwise it will be marked as a forgery and will be rejected.
-

Algorithm 1: The Training Phase of the Proposed System
Input: Sample images of the signature, logo and stamp.
Output: Trained NB classifier and computed thresholds.
Begin **Step1:** The signature sample images are entered to handwritten signature recognition and processed as follows: **1.1:** Preprocessing of the signature sample images. **1.2:** The SSE and improved SFTA features are extracted. **1.3:** A training NB classifier of the handwritten signature recognition level. **1.4:** A Intra-Person Signature Thresholds (IPST) method are applied to Compute GLT and GUT (applying equations 2 and 3).

Step2: The logo sample images are entered to logo and processed as follows:
2.1: Preprocessing of the logo sample images.
2.2: The color and GLCM features are extracted.
2.3: A training NB classifier of the logo recognition level.
2.4: Feature matching is applied to compute Accuracy Rate.
Step3: The stamp samples images are entered to stamp recognition and processed as follows:
3.1: Preprocessing of the stamp samples images.
3.2: The color and LAWS features are extracted.
3.3: A training NB classifier of the stamp recognition level.
3.4: A Intra-Person Signature Thresholds (IPST) method are applied to Compute GLT1, GUT1, GLT2 and GUT2 (applying equations 4, 5, 6 and 7).
End

Algorithm 2: The Testing Phase of the Proposed System
Input: Image of voucher document.
Output: Result faked voucher (type of faked) or not.

Begin
Step1: Read voucher image.
Step2: Convert color voucher image into gray level image.
Step3: Convert into binary image by Otsu threshold method.
Step4: Apply Edge Detection by using Sobel operator.
Step5: Apply Morphological closing operation.
Step6: Each connected components is labeled.
Step7: Binary object features are extracted for all Connected Components Labeling (CCL).
Step8: Apply k-Nearest Neighbors (KNN) to detect signature, logo and stamp.
Step9: The coordinate points of the CCL (signature, logo and stamp) in the binary image are used to extract signature, logo and stamp from original image.
Step10: The signature image is processed to be identification and verification in handwritten signature recognition level as follows:
10.1: Preprocessing of the signature image.
10.2: The SSE and improved SFTA features are extracted.
10.3: The NB classifier is applied to classify features to determine which class that the signature image belongs to.
10.4: Verification process is performed to determine the signature image being a faked or not.
Step11: The logo image is processed to be identification and verification in logo recognition level as follows:
11.1: Preprocessing of the logo image.
11.2: The Color and GLCM features are extracted.
11.3: The NB classifier is applied to classify features to determine which class that the logo image belongs to.
11.4: Verification process is performed to determine the logo image being a genuine or not.
Step12: The stamp images is processed to be identification and verification in stamp recognition level as follows:
12.1: Preprocessing of the stamp image.
12.2: The color and LAWS features are extracted.
12.3: The NB classifier is applied to classify features to determine which class that the stamp image belongs to.
12.4: Verification process is performed to determine the stamp image being a genuine or not.
End

9. Results of Voucher Recognition System

Evaluation of the proposed voucher recognition system needs system performance measurements. The effectiveness of the proposed system is tested by using Precision and Recall measurements as well as the accuracy to offer a clear overall indicator of system performance. Table- 1 shows the accuracy measurement of identification process based on improved Segmentation-based Fractal Texture Analysis (SFTA) texture features and the proposed Signature Shape Extraction (SSE) methods. From the Table- 1, it is clear that the obtained results were enhanced in comparison with the using of Improved SFTA only due to the effectiveness of SSE that can serve as attractive features of identification process with improved SFAT texture features.

Table- 2 describes the precision, recall and accuracy measurements of the identification and verification processes for signature, logo and stamp respectively. In the Table- 2, the signature, logo and stamp are tested with various datasets in order to prove their ability of the proposed recognition system and help in producing more statistical reliable experimental results. Through experiments, when the number of the training dataset is increased, the accuracy of identification and verification processes is also increased.

After testing each level with its related datasets as shown in Table- 2. Table- 3 and Table- 4 the results of identification and verification processes are illustrated respectively of all stages of the proposed system (including signature, logo and stamp recognition). The dataset contains 200 vouchers document images which are proposed and collected through the work. The proposed system achieves accuracy rate of 94% for identification process and 95% for verification process, these rates are supported with the features extraction, the proposed threshold and NB classifier.

Table 1- The signature identification process results of proposed system

Method	Accuracy
Improved SFTA	91.03%
SSE and Improved SFTA	97.94%

Table 2-Results of the identification and verification process for signature, logo and stamp images

	Identification			Verification		
	Precision	Recall	Accuracy	Precision	Recall	Accuracy
Signature	99.58%	98.34%	97.94%	99.17%	99%	98.19%
Logo	99.48%	97.23%	96.75%	99.49%	98.98%	98%
Stamp	97.35%	96.84%	94.35%	96.32%	97.87%	96.41%

Table 3- Results of the identification process of the overall proposed system

	The testing set	No. of correct	No. of wrong	Precision	Recall	Accuracy
Proposed system	200	189	11	97.42%	96.92 %	94%

Table 4- Results of the verification process of the overall proposed system

	The testing set	No. of correct	No. of wrong	Precision	Recall	Accuracy
Proposed system	200	190	10	96.44%	98.44%	95%

As a comparison between the proposed system with other related works, the proposed method processed the voucher documents that contain all of signature, logo and stamp and as far as our knowledge there is no other method that can do that, in other words, other methods (related works) processed either only the signature, logo or stamp. Therefore, the signature, logo and stamp recognition is compared separately with other works based on the available information of each work.

Table- 5 describes that the highest accuracy obtained from the proposed system, this is due to the selected features, proposed threshold with NB classifier.

Table 5-Comparison of the Proposed System with Related Works

	Kumar et al.[5]	Zouari et al.[6]	Ashok et al.[7]	Barbora M.[8]	Proposed system
Signature	-	83%	92.08%	-	98.19%
Logo	High accuracy	-	-	-	98%
Stamp	-	-	-	95%	96.41%

10. Conclusions

Forgeries of voucher documents are a high- risk criminal activity that adversely affects society. In this paper, the proposed signature, logo and stamp recognition system decided an efficient accuracy rate of automatic authentication method of the voucher document. Several factors lead to reach other quality characteristics about the proposed system, the choice of efficient Auto Crop process to extract signature, logo and stamp from original voucher image consume less run time process and computational complexity. Also the efficient feature extraction methods (improved SFTA, proposed SSE, GLCM, LAWS and color feature) resulted to precise feature vector. NB classifier proved that is a good choice for identification process and verification process with the proposed threshold. Experimental results showed that the proposed system gave better accuracy rate than the existing systems.

The future works need efficient segmentation method that can segment each text or word without any overlapping or missing parts. Support Vector Machine (SVM) can be used for the segmentation and classification which may improve the results. In addition, it can be planed for applying the proposed system in future for verifying the handwritten in voucher documents.

References

1. Jayadevan, R., Kolhe, S.R., Patil, P.M. and Pal, U. **2012**. Automatic processing of handwritten bank cheque images: a survey. *International Journal on Document Analysis and Recognition (IJDAR)*, **15**(4): 267-296.
2. Barbora, Micenkov, and Jost, van Beusekom. **2011**. Stamp detection in color document images. In International Conference on Document Analysis and Recognition (ICDAR), pp. 1125-1129.
3. Bertrand, R., Gomez-Krämer, P., Terrades, O.R., Franco, P. and Ogier, J.M. **2013**. A system based on intrinsic features for fraudulent document detection. In International Conference on Document Analysis and Recognition (ICDAR), pp. 106-110.
4. Van Beusekom, J., Shafait, F. and Breuel, T.M. **2008**. *Document signature using intrinsic features for counterfeit detection*, *Springer*, Volume 5158, ISSN 0302-9743, pp. 47-57.
5. Kumar, N.V., Kantha, V.V., Govindaraju, K.N. and Guru, D.S. **2016**. Features fusion for classification of logos. International Conference on Computational Modeling and Security (CMS 2016), pp.370-379.
6. Zouari, R., Mokni, R. and Kherallah, M. **2014**. Identification and verification system of offline handwritten signature using fractal approach. In Image Processing, Applications and Systems Conference (IPAS), pp. 1-4.
7. Kumar, A.D. and Dhandapani, S. 2014. A bank check signature verification system using FFBP neural network architecture and feature extraction based on GLCM. *The International Journal of Emerging Trends & Technology in Computer Science* (*IJETTCS*), **3**(3): 46-52.
8. Barbora Micenková . **2011**. Verification of Authenticity of Stamps in Documents. MSc. Thesis, Brno University of Technology, Germany.
9. Scott, E. Umbaugh. **2010**. *Digital image processing and analysis: human and computer vision applications with CVIPtools.* Second Edition. CRC Press, USA.
10. Gautam, C.M., Sharma, S. and Verma, J. S**.** **2012**. A GUI for Automatic Extraction of Signature from Image Document. *International Journal of Computer Applications*, 54(15): 13-19.

11. Kumar, G. and Bhatia, P.K. **2014**. A detailed review of feature extraction in image processing systems. International Conference on Advanced Computing and Communication Technologies, pp. 5-12.

12. Costa, A. F., Humpire-Mamani, G. and Traina, A. J. M**. 2012**. An efficient algorithm for fractal analysis of textures. *Institute of Electrical and Electronics Engineers (IEEE),* pp: 39 -46.

13. Tripathi and Garima. **2014**. Review on color and texture feature extraction techniques. *International Journal of Enhanced Research in Management and Computer Applications*, **3**(5): 77-81.

14. Mandloi, G. **2014**. A Survey on Feature Extraction Techniques for Color Images. *International Journal of Computer Science and Information Technologies*, **5**(3): 4615-4620.

15. Banday, S.A., Mir, A.H. and Khursheed, F. **2013**. Multi-unit iris biometric fusion using gray level co-occurrence matrix features. International Conference on Advanced Electronic Systems (ICAES), 21-23 September, pp. 225-229.

16. Pathak, B. and Barooah, D.**2013**. Texture analysis based on the gray-level co-occurrence matrix considering possible orientations. *International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering,* **2**(9): 4206-4212.

17. Rao, C. N., Sastry, S. S., Mallika, K., Tiong, H. S. and Mahalakshmi, K. B. **2013**. Co-occurrence matrix and its statistical features as an approach for identification of phase transitions of mesogens. *International Journal of Innovative Research in Science, Engineering and Technology (IJRSET)*, 2(9): 4531-4538.

18. Setiawan, A.S., Wesley, J. and Purnama, Y. **2015**. Mammogram classification using law's texture energy measure and neural networks. *Science Direct*, **59**, pp.92-97.

19. Lee, D. C. and Schenk, T. **1993**. Image segmentation from texture measurement. *International Archives of Photogrammetry and Remote Sensing*, pp.195-199.

20. Umamaheswari, J. and Radhamani, G. **2012**. An amalgam approach for DICOM image classification and recognition. *World Acad Sci Eng Technol,* **6**(2): 807-812.

21. Nandita Chasta and Manish Tiwari**. 2016**. Optimized Image Classification based on Universal Image Distance and Support Vector Machines. *International Journal of Advanced Research in Computer and Communication Engineering (IJARCCE),* **5**(2): 130- 139.

22. Sunita Joshi, Bhuwaneshwari Pandey, Nitin Joshi. **2015**. Comparative analysis of Naive Bayes and J48 Classification Algorithm***s****, International Journal of Advanced Research in Computer Science and Software Engineering (IJARCSSE),* **5**(12): 813- 817.

23. Han, J., Pei, J. and Kamber, M. **2011**. *Data mining: concepts and techniques*. Elsevier. Third Edition.