TELKOMNIKA Telecommunication, Computing, Electronics and Control
Vol. 19, No. 3, June 2021, pp. 792~800
ISSN: 1693-6930, accredited First Grade by Kemenristekdikti, Decree No: 21/E/KPT/2018
DOI: 10.12928/TELKOMNIKA.v19i3.18763

An efficient method for stamps recognition using Haar wavelet sub-bands

Maha A. Rajab[1], Loay E. George[2]
[1]University of Babylon, Baghdad, Iraq
[2]University of Information Technology and Communication, Baghdad, Iraq

Article Info

Article history:

Received Sep 3, 2020
Revised Nov 13, 2020
Accepted Nov 25, 2020

Keywords:

Energy
Haar wavelet
Low order moment
Recognition rate
Stamp recognition

ABSTRACT

The problem facing certain organizations such as insurance companies and government institutions where a huge amount of documents is handled every day, hence an automated stamp recognition system is required. The image of the stamp may be on a different background, with different sizes, and suffers from rotating in different directions, also, the appearance of soft areas (patches) or small points as noise. Thus, the main objective of this paper is to extract and recognize the color stamp image. This paper proposed a method to recognize stamps, by using a technique named Haar wavelet sub-bands. The devised method has four stages: 1) extracts the stamp image; 2) preprocessing the image; 3) feature extraction; and 4) matching. This paper is implemented using C sharp (Microsoft Visual Studio 2012) programming language. The experiments conducted on a stamp dataset showed that the proposed method has a great capability to recognize stamps when using Haar wavelet transform with two sets of features (i.e., 100% recognition rate for energy features and 99.93% recognition rate for low order moment).

Corresponding Author:

Maha A. Rajab
College of Information Technology
University of Babylon
Baghdad, Iraq
Email: maha.a.r@ihcoedu.uobaghdad.edu.iq

1. INTRODUCTION

Pattern recognition is a branch of digital image processing and artificial intelligence. Thus, a pattern may be a fingerprint image, logo, stamp, a handwritten, a human face, and a speech signal [1]. Computer analysis of digital documents is now one of the most important areas of digital image processing and pattern recognition. Some of the very significant issues are the detection and recognition of stamps on digital documents, this issue remains open such remove background and extraction stamp area, despite the significant development in this field [2]. Thus, stamping is the process of locating the stamps on paper documents which hold a certain property such as (shape, complexity, background and typical patterns). The main objective of utilizing a stamp is to certify a document for various kinds of verification, such as authentication, and authorization. There is a need to organize and access digitized documents according to their contents in processing the image of the document. The legitimacy of the document is provided by a stamp [3]. The difficulty in extracting stamps is that: there is generally no template for stamps. The stamp can be a graphical and textual object which can be placed in the document at any position [4, 5]. Clustering is a method of grouping identical image pixels into one cluster according to the certain property. Clustering is an unsupervised classification of data points into groups or clusters [6, 7]. The most popular examples of clustering algorithms include K-means and Iterative Self-Organizing Data Analysis Technique Algorithm (ISODATA) clustering algorithms [8, 9]. Principal component analysis (PCA) can be defined as a preprocessing transformation that produces new images from the interrelated values of existing images. It is done by a linear transformation of variables that leads to rotation and translation of the original coordinate system. PCA becomes a powerful tool to analyze the data and used to perform rotation compensation to rotate the stamp image [10, 11]. Haar wavelet transform is the simplest wavelet transform variant. It was utilized in many researches due its good performance in feature extraction, the Haar transform has been used as a basic tool for decomposed the image into four sub-bands (i.e., approximation and detail sub-bands) [12]. Haar wavelet sub-bands has also been widely used in medical imaging and pattering approaches such as optical coherence tomography (OCT) [13, 14].

There are many problems facing the process of extracting stamp from the document image such as isolated complex background, eliminating noise, contrast enhancement, rotated in various directions, extracted local features, and recognition of the stamp. This paper aims at introducing a new method for recognizing the color stamp image robust against noise and rotation.

This paper is structured as follows: section 2 overviews the related work. Section 3 introduces the layout of the proposed stamp recognition system. Section 4 describes the results and discussion of conduct tests. Finally, the derived conclusions of this paper are shown in Section 5.

2. RELATED WORK

There are few studies that deal with the stamp recognition issues. These experiments are distinct from the approach suggested in terms of recognition and efficiency. Pawel and Dariusz [15] in 2012 proposed a method for solving the problem of identifying the stamp types extracted from digital images. It is separated into two principal phases. In the first phase, the stamp is represented using point distance histogram method depending on the shape descriptor. Principal component analysis transform was utilized in the second phase, with an additional reduction based on linear discriminant analysis. The dimensionality reduction conducted using PDH-calculated descriptors have shown that for classification purposes only the four first performance components are useful. The test results indicated that the best recognition rate is 89% [15]. Sheraz Ahmed *et al*. [16] in 2013 proposed a new method of segmenting stamps from document images. The system can be considered a general way to segment the single/multicolored, monochrome, and even black stamps. It is also capable of segmenting invisible stamps and also stamps in some arbitrary shapes. Also, used part-based/local features merged with geometric features to extract features. The test results achieved for recall and precision of 73% and 83%, respectively [16]. Sarthak [17] in 2015 suggested a method for segmentation using K-means algorithm and threshold technique. Image can be segmented into k clusters using iterative K-means algorithm. Initially, pixels are grouped based on their color and spatial features, where the grouping process is achieved. Cluster blocks are then combined into a specified number of regions. In thresholding technique, compared every pixel in an image with this threshold. Experimental results showed that the accuracy of K-means algorithm is 78.57%, while in threshold technique is 28.57% [17].

3. PROPOSED STAMPS RECOGNITION SYSTEM

The general structure of the proposed stamp recognition system is presented in Figure 1. It is comprised of four essential stages: 1) stamp extraction, 2) preprocessing, 3) proposed feature extraction, and 4) matching. The steps involved in these stages are further described in the following stages.

3.1. Stamp extraction stage

In this stage, the stamp image may be put on the complicated background and surrounded by unnecessary data, such as word, text, patches, and other forms of noise produced as objects closer to the stamps. Thus, we need to separate background clusters from stamp clusters to remove noise. In this paper, a stamp extraction stage is similar to that mentioned in article [18]. It is consisting of four steps that are briefly described in the following:

- Step1: Improved K-means algorithm

Stamp image is the Bitmap image format (i.e., .bmp). The color resolution of the stamp image is 24 bit/pixel. Load image data (i.e. R, G, and B components) and clustering using the K-means clustering algorithm aimed at grouping a set of pixel values into k clusters based on the Euclidean distance determined between the data object and the chosen k centroids [19-21]. So, it will reduce the number of clusters by merging the same clusters using the ISODATA clustering algorithm.

- Step2: Compute minimum distance

The distance between all the remaining clusters must be determined. After that, the distance between the two clusters is conducted to find out the lowest distance, and they are combined according to the color value of the cluster pixels selected for merging. This method will be continued until combined all clusters have lowest-distance, so the number of clusters will be greatly decreased.

- Step3: Compute mean and standard deviation

This step aims to isolate background clusters from stamp clusters by computing the mean and standard deviation for the remaining clusters; then, the larger cluster will have less standard deviation than the rest of the other cluster. The largest cluster is the background cluster, which is not bound to central region and may have wide variation, then we have to determine which of the remaining clusters are part of the background. It is achieved by measuring the dominant color value for each cluster and keeping the value of the largest repetition because it is considered the background cluster. From experience, we note that the background cluster is known as the cluster, which contains the highest repetition whose value exceeds more than 10, but if less than 10 is known among the stamp clusters. As a result, we determined the number of background and stamp clusters, these clusters will combine to produce one image for both the background and the stamp.

- Step4: Extract stamp and remove object

The stamp clusters are isolated from the background and other unwanted information (noise patches come as isolated area), thus, we need to remove these information. This process is done by segmenting the stamp cluster using region growing to a number of segments and calculating the main diagonal and the secondary diagonal for each segment to produce four points of the two diagonals, so that, the segment contains four points is considered the stamp otherwise it is considered noise area. A binary mask is generated from the connected segment and multiplies with the original image after isolating the background to obtain the stamp image free from noise. Figure 2 shows the results of each step of the first stage (i.e., stamp extraction) of one image chosen randomly from the stamps dataset.

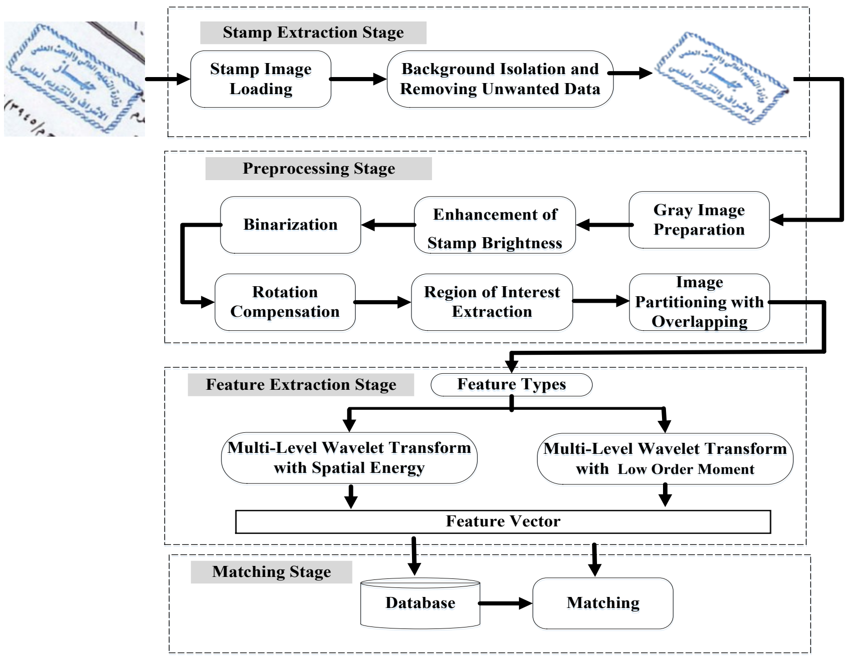

Figure 1. The layout of the proposed system

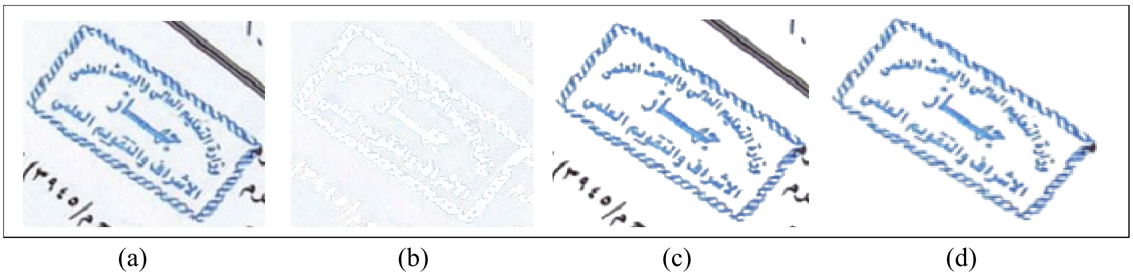

Figure 2. Result of applying stamp extraction stage; (a) original image, (b) background cluster, (c) stamp with noise, and (d) extraction stamp

3.2. Preprocessing stage

For every recognition system the main stage is the preprocessing. It was developed to modify the image data to make it more efficient for the feature extraction stage [22]. A set of tasks has been performed in the preprocessing stage to enhance the image quality. The following steps included these tasks:

- Step1: Gray image conversion

The RGB color stamp image format is converted to grayscale image as Figure 3 (a).

- Step2: Stamp image enhancement

Often the stamp image may be ambiguous, and this can influence the process of extracting good features for the image, and we need to improve the image brightness by applying nonlinear gamma correction. Gamma correction is done to enhance the degree of whiteness in light regions or the degree of blackness in dark regions, as Figure 3 (b).

- Step3: Image binarization

The most popular approach used in the segmentation of images is global thresholding. It is achieved by comparing image values with a threshold value T, the pixels with an intensity value of 1 belong to objects, while the pixels with an intensity value of 0 belong to the background, as Figure 3 (c).

- Step4: Rotation compensation

As we mentioned in the stamps dataset below, the images suffer from rotation at various angles and this is because the stamps are not necessarily positioned with the horizon line due to the stamping process. Consequently, these reasons have become one of the most important challenges facing the recognition process is to estimate the angle value that the image needs to return to the original point or to estimate the angle value which is approximately close to the point of origin. The dataset also contains stamps that are in different classes, thus the problem facing rotation compensation is how to estimate the angle of rotation. In this work, we utilized a statistical tool name Principal Component Analysis (PCA) to rotate the object. PCA is employed on white pixels in the binary image to determine the object is rotated around the mean of the object pixels, as Figure 3 (d).

- Step5: Image cropping

This step is useful to reduce the computational complexity and speed up the processing time when searching for stamp components regions. The clipping is the process of allocating the stamp in the image. This step can be achieved by doing four different scans (i.e., scanned each side of the image boundary even reached the row or column holding the stamp pixel), as Figure 3 (e).

- Step6: Image partitioning with overlapping

The main aim of this step is to divide the image into overlapping blocks to describe the local feature for each block. The block dimensions can be determined by dividing the image dimensions into the number of blocks. The overlapping ratio is the parameter that controlled the degree of overlapping between blocks which can be considered as the ratio between the extended block length and the original block, as Figure 3 (f).

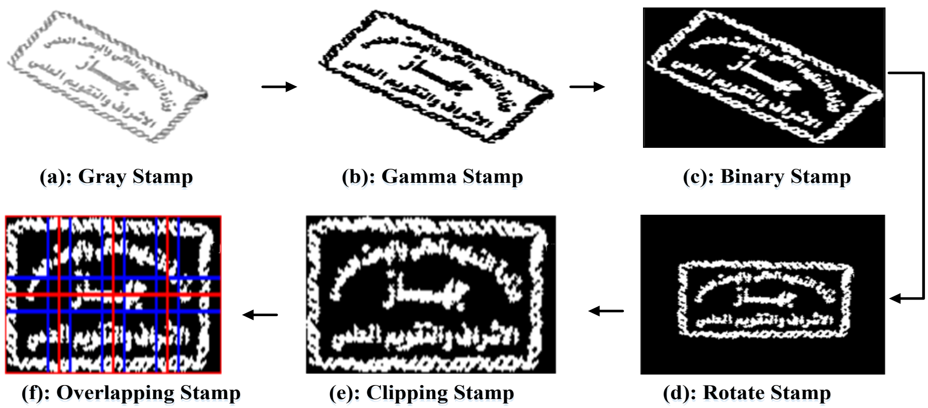

Figure 3. Results of applying preprocessing stage

3.3. Proposed features extraction stage

The image feature is defined as an image's distinctive significant feature or attribute, and, also describe the image properties that are extracted from several different information domains. Some of the discriminatory features are denoted as the visual appearance of an image such as luminance of pixel region and grayscale textural regions, while other features are obtained from specific manipulations of an image such as histograms of image amplitude and spatial frequency spectra. The most critical stage of any recognition system is an extraction of features. The main purpose of feature extraction is to extract significant information from the image. The information extracted must represent desirable features vector in order to discriminate between stamps. For the matching stage each extracted feature vector is stored in the database.

3.3.1. Proposed multi-level Haar wavelet transform

Wavelet transform is a signal analysis technique that can be used in many areas, including image processing and pattern recognition. The reason for using the discrete wavelet transform is to get more discriminating information by providing various resolutions at different parts of the image [23, 24]. In the proposed system, a multi-level 2D Haar wavelet transform (HWT) is accomplished. 2D haar wavelet transform applied to each partitioned block that decomposes it's into four sub-bands, where LL1 denotes the low frequency (approximation) sub-band, while the remaining sub-bands (i.e., LH1, HL1, and HH1) indicate to the high frequency component (detail). The approximation (LL1) sub-band is decomposed using two-level haar wavelet transform to created four sub-bands as (LL2, LH2, HL2, and HH2), each one of these sub-bands will be decomposed utilized three-level haar wavelet transform to create the output of new four sub-bands for each one of them arranged as (LL3, LH3, HL3, and HH3) and also refers to the output from all sub-bands named as (LL2, LH2, HL2, and HH2). After completing the decomposition of the (LL1) sub-band, the remaining sub-bands (i.e., LH1, HL1, and HH1) are then decomposed to created four sub-bands for each one of the remaining detail components from the one-level namely as (LL2, LH2, HL2, and HH2). Thus, the resulting from these decomposing is multi-level haar wavelet transform sub-bands (i.e. 28 sub-bands). The output of the multi decomposition level is illustrated in Figure 4. Two various sets of distinguished features will be used in the proposed system: (i) multi-level Haar wavelet transform with energy, (ii) multi-level Haar wavelet transform with low order moment.

<table>
<tr><td>LL_3</td><td>HL_3</td><td>LL_3</td><td>HL_3</td><td rowspan="2">LL_2</td><td rowspan="2">HL_2</td></tr>
<tr><td>LH_3</td><td>HH_3</td><td>LH_3</td><td>HH_3</td></tr>
<tr><td>LL_3</td><td>HL_3</td><td>LL_3</td><td>HL_3</td><td rowspan="2">LH_2</td><td rowspan="2">HH_2</td></tr>
<tr><td>LH_3</td><td>HH_3</td><td>LH_3</td><td>HH_3</td></tr>
<tr><td colspan="2">LL_2</td><td colspan="2">HL_2</td><td>LL_2</td><td>HL_2</td></tr>
<tr><td colspan="2">LH_2</td><td colspan="2">HH_2</td><td>LH_2</td><td>HH_2</td></tr>
</table>

Figure 4. The proposed multi-level wavelet decomposition

a. Multi-level Haar wavelet transform with energy

It should be taken into consideration, the feature vector cannot be extracted directly from these sub-bands, the energy of each wavelet sub-band (k) belong to each image block (b) is calculated to establish the feature vector, according to (1) [25].

$$Eng(k,b) = \sum_{x=x_s}^{x_e-1} \sum_{y=y_s}^{y_e-1} Wavelet(X,Y)^2 \qquad (1)$$

Where (x_s, y_s) and (x_e, y_e) are the range of coordinates of the sub-band (k) belong to block (b), wavelet (X, Y) represents the wavelet sub-bands.

b. Multi-level Haar wavelet transform with low order moment

Low order moment can be used to extract feature vector from the above sub-bands, where the moment should be utilized to describe the image segment and extract the properties. The values of low order moment can be considered an essential geometrical property of distribution function. Thus, the normalized of each wavelet sub-band (k) belong to each image block (b) is computed to establish the feature vector, according to (2) [26].

$$Norm(k,b) = \sum_{x=x_s}^{x_e-1} \sum_{y=y_s}^{y_e-1} |Wavelet(X,Y)|^{0.5} \qquad (2)$$

3.4. Matching stage

This stage is used to compute the degree of similarity between the feature vectors are extracted from the stamp image. The computed discriminating feature vectors for each class are used to generate the template. The particular template comprises the mean feature vector of each class, and the standard deviation feature vector is kept in a database. This stage is achieved by comparing the feature vector of the stamp image with all template vectors kept in the database.

The degree of similarity utilized in this paper is determined using Euclidean distance measure named as (normalized mean square distance (NMSD) as (3) [27].

$$NMSD(F_i\,,T_j) = \sum_{k=1}^{N} \left(\frac{F_i(k) - T_{\mu j}(k)}{T_{\sigma j}(k)} \right)^2 \quad (3)$$

Where, F_i represents the tested i^{th} feature vector extracted from stamp input image, T_j represents the j^{th} template feature vector registered in the database, while $T_{\mu j}$ and $T_{\sigma j}$ representing the mean and standard deviation template vectors respectively, which are loaded from the database.

4. RESULTS AND DISCUSSION

One dataset (i.e., Stamps Dataset) is used for evaluating the performance of the proposed stamp recognition system. The dataset was taken from Mendeley Data 2020 [28]. This dataset consists of different samples belonging to 173 classes; and for each class, a subset consists of nine samples of stamps. The stamp images are BMP 24 bit/pixel (bit depth). Table 1 shows the description of the summarized for the dataset.

Table 1. Show the description of the summarized for the dataset

No. of Classes	Samples per Class	Shapes Categories	Total Samples
173	9	6	1557

In multi-level 2D Haar Wavelet decomposition, our experiment produced three groups of features subsets as: a first subset depends on four features extracted from original sub-bands (i.e. F1(LL), F2(LH), F3(HL) and F4(HH)). A second subset consists of six features extracted from the combination of two sub-bands as (F5(LL-LH), F6(LL_HL), F7(LL_HH), F8(LH_HL), F9(LH_HH) and F10(HL_HH)). A final subset consists of fifteen features extracted from the combination of four sub-bands as (F11(LL_LH_LL_HL), F12(LL_LH_LL_HH), F13(LL_LH_LH_HL), F14(LL_LH_LH_HH), F15(LL_LH_HL_HH), F16(LL_HL_LL_HH), F17(LL_HL_LH_HL), F18(LL_HL_LH_HH), F19(LL_HL_HL_HH), F20(LL_HH_LH_HL), F21(LL_HH_LH_HH), F22(LL_HH_HL_HH), F23(LH_HL_LH_HH), F24(LH_HL _ HL_HH) and F25(LH_HH _ HL_HH)). Consequently, the number of all extracted features are twenty-five features. Thus, features are extracted from these sub-bands using two methods as described below.

Firstly, Table 2 demonstrates the effect of different blocks number and different overlapping ratio on the attained recognition results with (NMSD) similarity measure, also, shows the highest recognition rate is (100%) occur when the number of blocks (5×5), (7×7) and overlapping ratio 0.1, using features (F10(HL_HH), F13(LL_LH_LH_HL), F14(LL_LH_LH_HH), F15(LL_LH_HL_HH), F16(LL_HL_LL_HH), F17(LL_HL_LH_HL), F18(LL_HL_LH_HH), F19(LL_HL_HL_HH), F20(LL_HH_LH_HL), F21(LL_HH_LH_HH), F22(LL_HH_HL_HH), F23(LH_HL_LH_HH), F24(LH_HL_HL_HH) and F25(LH_HH_HL_HH)), consequently, when the overlapping rate 0.2 the highest recognition rate is (100%), occur when the number of blocks (7×7), using features (F11(LL_LH_LL_HL), F12(LL_LH_LL_HH), F13(LL_LH_LH_HL), F14(LL_LH_LH_HH), F15(LL_LH_HL_HH), F17(LL_HL_LH_HL), F18(LL_HL_LH_HH), F19(LL_HL_HL_HH), F20(LL_HH_LH_HL), F21(LL_HH_LH_HH) and F22(LL_HH_HL_HH)), Also shows when the overlapping rate 0.3 the highest recognition rate is (100%), occur when the number of blocks (5×5), (7×7), using features (F2(LH), F5(LL-LH), F7(LL_HH), F8(LH_HL), F9(LH_HH) and F10(HL_HH) F11(LL_LH_LL_HL), F12(LL_LH_LL_HH), F13(LL_LH_LH_HL), F14(LL_LH_LH_HH), F15(LL_LH_HL_HH), F16(LL_HL_LL_HH), F17(LL_HL_LH_HL), F18(LL_HL_LH_HH), F19(LL_HL_HL_HH), F20(LL_HH_LH_HL), F21(LL_HH_LH_HH), F22(LL_HH_HL_HH), F23(LH_HL_LH_HH), F24(LH_HL _ HL_HH) and F25(LH_HH_HL_HH)).

It can be notice from the results in the above table show that the best attained recognition rate (i.e. 100%) is obtained when taken a number of blocks is (5×5) and (7×7), overlapping ratio (0.1), (0.2) and (0.3), when using all three groups of extracted features with (NMSD) similarity measure. Secondly, Table 3 presents the effect of different blocks number and different overlapping ratio on the attained recognition results for the normalized of each wavelet sub-band explained in the above with (NMSD) similarity measure, also, shows the highest recognition rate is (99.93%) occur when the number of blocks (5×5), (7×7) and (5×5) respectively for overlapping ratio (0.1), (0.2) and (0.3) using features (F6(LL_HL), F7(LL_HH), F11(LL_LH_LL_HL), F12(LL_LH_LL_HH), F13(LL_LH_LH_HL), F15(LL_LH_HL_HH), F16(LL_HL_LL_HH),

F17(LL_HL_LH_HL), F18(LL_HL_LH_HH), F19(LL_HL_HL_HH), F20(LL_HH_LH_HL), F22(LL_HH_HL_HH)) from the second and final subset features.

Table 2. The effect of different blocks number on the recognition rate of haar wavelet with energy features set using (NMSD) and overlapping rate set 0.1, 0.2 and 0.3

Feat.	Recognition Rate (%)-NMSD											
	Overlapping Ratio=0.1 No. of Blocks				Overlapping Ratio=0.2 No. of Blocks				Overlapping Ratio=0.3 No. of Blocks			
	3×3	3×5	5×5	7×7	3×3	3×5	5×5	7×7	3×3	3×5	5×5	7×7
F1	82.27	90.17	97.30	97.88	77.71	88.50	95.11	99.67	67.88	86.76	93.64	99.48
F2	91.52	98.13	99.42	99.16	80.86	96.78	99.03	99.61	88.50	94.09	98.07	100
F3	83.04	94.99	98.00	99.93	82.53	93.64	98.65	99.87	71.03	85.03	99.03	99.93
F4	90.43	93.77	98.13	99.35	85.87	95.82	98.65	99.48	85.29	92.74	97.81	99.87
F5	95.76	98.45	99.67	99.87	92.67	98.39	98.77	100	91.45	97.23	98.52	100
F6	94.21	97.81	99.48	99.93	93.57	96.27	98.97	99.93	87.98	96.14	99.03	99.93
F7	98.00	98.84	99.67	99.93	95.50	98.33	99.42	99.93	90.81	97.30	99.35	100
F8	97.43	99.48	99.80	100	95.11	98.84	99.74	99.93	94.79	98.84	99.87	100
F9	97.81	99.35	99.93	99.87	94.92	98.84	99.80	99.93	95.82	98.45	99.61	100
F10	97.10	99.29	99.61	100	94.15	98.33	99.48	99.93	93.89	97.94	99.87	100
F11	96.59	99.22	99.80	99.93	95.56	98.52	99.22	100	93.25	97.94	99.10	100
F12	98.52	99.16	99.87	99.93	96.08	98.77	99.61	100	93.77	98.13	99.48	100
F13	98.26	99.55	99.87	100	96.66	99.22	99.80	100	96.14	99.22	99.74	100
F14	99.35	99.55	99.93	100	96.91	99.35	99.74	100	96.98	98.90	99.74	100
F15	99.42	99.67	100	100	98.58	99.42	99.87	100	97.23	99.42	99.87	100
F16	98.39	99.03	99.87	100	96.33	98.26	99.74	99.93	92.93	97.68	99.48	100
F17	97.68	99.48	99.87	100	97.30	99.03	99.80	100	95.05	99.03	99.80	100
F18	99.42	99.67	100	100	98.58	99.42	99.87	100	97.23	99.42	99.87	100
F19	98.58	99.55	99.87	100	97.94	98.90	99.87	100	95.31	98.77	99.93	100
F20	99.42	99.67	100	100	98.58	99.42	99.87	100	97.23	99.42	99.87	100
F21	99.35	99.74	99.93	100	98.20	99.22	99.80	100	95.95	98.97	99.61	100
F22	98.90	99.48	99.93	100	98.39	99.22	99.80	100	95.11	98.65	99.87	100
F23	98.77	99.67	99.93	100	96.53	99.16	99.80	99.93	97.10	99.42	99.87	100
F24	98.65	99.67	100	100	96.66	99.16	99.87	99.93	97.23	99.29	100	100
F25	98.65	99.67	99.93	100	96.59	99.16	99.87	99.93	96.98	99.29	100	100

Table 3. The effect of different of blocks number on the recognition rate of haar wavelet with low order moment features set using (NMSD) and overlapping rate set 0.1, 0.2 and 0.3.

Feat.	Recognition Rate (%)-NMSD											
	Overlapping Ratio=0.1 No. of Blocks				Overlapping Ratio=0.2 No. of Blocks				Overlapping Ratio=0.3 No. of Blocks			
	3×3	3×5	5×5	7×7	3×3	3×5	5×5	7×7	3×3	3×5	5×5	7×7
F1	92.93	97.23	99.29	99.55	86.38	95.50	98.71	99.67	80.02	94.15	98.65	99.29
F2	92.22	94.86	98.45	98.45	69.62	95.69	98.77	98.77	82.98	91.00	97.30	98.71
F3	91.32	98.00	99.42	99.22	87.21	97.68	99.16	99.67	85.09	94.60	98.52	99.16
F4	66.08	91.65	95.50	96.85	61.59	88.50	97.36	96.59	67.88	77.32	92.48	97.49
F5	97.81	99.22	99.67	99.55	92.35	98.39	99.48	99.87	91.71	97.88	99.67	99.74
F6	96.85	99.03	99.87	99.55	94.54	98.65	99.61	99.93	91.52	98.45	99.55	99.67
F7	97.75	99.16	99.67	99.42	93.19	98.65	99.42	99.93	91.65	97.36	99.74	99.87
F8	97.23	99.10	99.55	99.29	89.27	98.65	99.42	99.55	91.97	97.49	99.03	99.61
F9	93.83	96.85	98.45	98.45	72.83	95.56	99.16	98.39	86.89	91.58	97.62	99.03
F10	93.25	98.39	99.03	98.77	85.99	97.30	99.03	99.35	88.43	95.05	98.58	99.42
F11	98.26	99.10	99.74	99.61	95.24	98.97	99.55	99.93	93.12	98.45	99.74	99.74
F12	98.26	99.29	99.67	99.55	93.64	98.58	99.55	99.93	93.25	97.81	99.67	99.87
F13	98.65	99.29	99.80	99.48	94.09	99.10	99.67	99.93	93.96	98.39	99.87	99.87
F14	98.26	99.29	99.61	99.42	89.98	99.03	99.61	99.80	93.06	97.81	99.55	99.87
F15	98.77	99.35	99.80	99.42	94.21	99.35	99.74	99.93	95.24	98.58	99.87	99.87
F16	98.00	99.29	99.87	99.42	94.79	98.97	99.67	99.93	92.67	98.33	99.74	99.80
F17	98.39	99.48	99.93	99.29	95.24	99.29	99.80	99.93	94.79	98.90	99.93	99.87
F18	98.77	99.35	99.80	99.42	94.21	99.35	99.74	99.93	95.24	98.58	99.87	99.87
F19	98.00	99.16	99.93	99.42	94.34	99.42	99.80	99.93	93.44	98.65	99.87	99.87
F20	98.77	99.35	99.80	99.42	94.21	99.35	99.74	99.93	95.24	98.58	99.87	99.87
F21	98.39	99.16	99.42	99.35	90.04	98.77	99.67	99.74	93.38	97.68	99.67	99.87
F22	98.13	99.35	99.80	99.29	92.74	99.03	99.61	99.93	93.44	98.39	99.87	99.87
F23	96.85	98.52	99.10	98.90	84.20	97.75	99.55	99.55	91.52	96.40	98.71	99.42
F24	97.04	98.97	99.29	99.22	89.14	98.58	99.61	99.61	92.35	96.78	98.97	99.67
F25	95.37	98.39	98.84	98.71	82.46	97.49	99.29	99.22	90.81	95.56	98.39	99.42

To evaluate the efficiency of the new method, and to show that it has better results than other existing experiments. This section explains the results of our proposed method and comparison with many previously published studies. P. Forczma´nski and D. Frejlichowski [15] use a different dataset, but the same parameter for evaluation (i.e., recognition rate), while [22] use the same dataset and recognition rate. Table 4 presents the recognition rate achieved by some of the previously conducted experiments and the planned study. The results listed in Table 4 show that our approach leads to a high recognition rate when compared with other research.

Table 4. The recognition rate compared with previous studies

Reference	Number of images in Database	Recognition Rate (%)
[15]	140	89%
[22]	80	94%
Our Proposed with Features Set 1	1557	100%
Our Proposed with Features Set 2	1557	99.93%

5. CONCLUSION

In this paper, we suggested a new method leading to the recognition of color stamp image. In previous studies, this subject has not been extensively explored and addressed until now. The experimental results showed that partitioning into blocks with overlap helped to resolve the partial loss in the low-quality stamp image and increased accuracy of recognition, also, the recognition rate is extremely affected by the variation of number blocks and the overlapping ratio. The combination of more than two wavelet features gives the best recognition rate. Thus, the test results show that the multi-level Haar wavelet transform given a high recognition rate (i.e., 100%) when using energy features extracted from each sub-bands, and (i.e., 99.93%) when using the low order moment feature.

REFERENCES

[1] A. K. Jain, *et al*., " An Introduction to Biometric Recognition," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 14, no. 1, pp. 4-20, 2004.

[2] S. Gaur, *et al*., "Biometric Recognition Techniques: A Review," *International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering,* vol. 1, no. 4, pp. 282-290, 2012.

[3] A. V. Nandedkar, *et al*., "SPODS: A dataset of color-official documents and detection of logo, stamp, and signature," *Springer International Publishing,* vol. 10481, pp. 219-230, 2017.

[4] B. Micenkov´a and J. V. Beusekom, "Stamp detection in color document images," *IEEE International Conference on Document Analysis and Recognition, (ICDAR)*, 2011, pp. 1125-1129,

[5] S. Dey, *et al*., "Stamp and logo detection from document images by finding outliers," *IEEE 2015 Fifth National Conference on Computer Vision, Pattern Recognition, Image Processing and Graphics (NCVPRIPG)*, 2015, pp. 1-4.

[6] A. Gulhane, *et al.,* "A Review of Image Data Clustering Techniques," *International Journal of Soft Computing and Engineering (IJSCE)*, vol. 2, no. 1, pp. 212-215, 2012.

[7] S. Monalisa and F. Kurnia, "Analysis of DBSCAN and K-means Algorithm for Evaluating Outlier on RFM Model of Customer Behaviour," *TELKOMNIKA Telecommunication Computing Electronics and Control,* vol. 17, no.1, pp. 110-117, 2019.

[8] S. A. Elavarasi, *et al*., "A survey on partition clustering algorithms," *International Journal of Enterprise Computing and Business Systems,* vol. 1, no. 1, pp. 1-14, 2011.

[9] T. Velmurugan and T. Santhanam, "A Survey of partition based clustering algorithms in data mining: An experimental approach," *Information. Technology Journal*, vol. 10, no. 3, pp. 478-484, 2011.

[10] J. Yang, *et al*., "Two-Dimensional PCA: A New Approach to Appearance-Based Face Representation and Recognition," *IEEE Transaction Pattern Analysis and Maching Intelligence*, vol. 26, no. 1, pp. 131-137, 2004,

[11] H. Z. U. Rehman and S. Lee, "Automatic Image Alignment Using Principal Component Analysis," *IEEE Access*, vol. 6, pp. 72063-72072, 2018.

[12] K. H. Talukder and K. Harada. "Haar Wavelet Based Approach for Image Compression and Quality Assessment of Compressed Image," *IAENG International Journal of Applied Mathematics* (IJAM*),* vol. 37, no. 1, pp. 1-8, 2007.

[13] R. K. Meleppat, *et al*., "An Efficient Phase Analysis-Based Wavenumber Linearization Scheme for Swept Source Optical Coherence Tomography Systems," *Laser Physics Letters*, vol. 12, pp. 1-7, 2015.

[14] K. M. Ratheesh, *et al*., "Spectral Phase-Based Automatic Calibration Scheme for Swept Source-Based Optical Coherence Tomography Systems," Institute of Physics and Engineering in Medicine, vol. 61, pp. 7652-7663, 2016.

[15] P. Forczma´nski and D. Frejlichowski, "Classification of Elementary Stamp Shapes by Means of Reduced Point Distance Histogram Representation," *Springer-Verlag Berlin Heidelberg*, vol.7376, pp. 603-616, 2012.

[16] Sheraz Ahmed*, et al*., "A generic Method for Stamp Segmentation Using Part-based Features," *IEEE International Conference on Document Analysis and Recognition (ICDAR)*, 2013, pp. 708-712.

[17] S. Panda, "Color Image Segmentation Using K-means Clustering and Thresholding Technique," *IJESC,* pp. 1132-1136, 2015.

[18] M. A. Rajab, and Loay. E. George, "Stamps Extraction Using Local Adaptive K- Means and ISODATA Algorithms," *Indonesian Journal of Electrical Engineering and Computer Science.*, vol. 21. no. 6, pp. 137-145, 2021.
[19] M. N. Qureshi and M. V. Ahamad "An Improved Method for Image Segmentation Using K-Means Clustering with Neutrosophic Logic," *International Conference on Computational Intelligence and Data Science (ICCIDS)*, vol. 132, pp. 534-540, 2018.
[20] Y. Hamzaoui, *et al.,* "Enhancenig OLSR Routing Protocol Using K-Means Clustering in MANETs," *International Journal of Electrical and Computer Engineering (IJECE),* vol. 10, no. 4, pp. 3715-3724, 2020.
[21] I. Q. Abduljaleel and A. H. Khaleel, "Hiding Text in Speech Signal Using K-Means, LSB Techniques and Chaotic Maps," *International Journal of Electrical and Computer Engineering (IJECE),* vol. 10, no. 6, pp. 5726-5735, 2020.
[22] Z. J. Ahmed and L. E. George, "Fingerprints Recognition Using the Local Energy Distribution over Haar Wavelet Subbands," *International Journal of Science and Research (IJSR)*, vol. 6, no. 9, pp. 979- 986, 2017.
[23] P. Porwik and A. Lisowska. "The Haar-wavelet Transform in Digital Image Processing: Its Status and Achievements," *Mathematics. Machine graphics & vision,* vol. 13, no. 1-2, pp. 79-98, 2004.
[24] A. W. Altaher and S. K. Abbas, "Image Processing Analysis of Sigmoidal Hadamard Wavelet with PCA to Detect Hidden Object," *TELKOMNIKA Telecommunication Computing Electronics and Control,* vol. 18, no. 3, pp. 1216-1223, 2020.
[25] S. Ekici, et al., "Energy and Entropy-based Feature Extraction for Locating Fault on TransmissionLlines by Using Neural Network and Wavelet Packet Decomposition," *Elsevier. Expert Systems with Applications,* vol. 34, pp. 2937-2944, 2008.
[26] Z. J. Ahmed, "Fingerprints Matching Using the Energy and Low Order Moment of Haar Wavelet Subbands," *Journal of Theoretical and Applied Information Technology,* vol. 96, no. 18, pp. 6191-6202, 2018.
[27] Z. J. Ahmed and L. E. George, "Fingerprints Identification and Verification Based on Local Density Distribution With Rotation Compensation," *Journal of Theoretical and Applied Information Technology* , vol. 96, no. 24, pp. 8313-8323, 2018.
[28] M. A. Rajab and L. E. George, "Loay & Maha Stamps Dataset, *Mendeley Data*," Published: 2020-02-26, 2020.