SVM-BASED SEAL IMPRINT VERIFICATION USING EDGE DIFFERENCE

Yu-Chen Su† , *Yeong-Luh Ueng*†‡ *and Wei-Ho Chung* ‡

†Department of Electrical Engineering, National Tsing Hua University, Hsinchu, Taiwan
‡Institute of Communications Engineering, National Tsing Hua University, Hsinchu, Taiwan

***Abstract*—In Asian countries, seals are widely used for authenticating the identity of a person or organization. Therefore, the ability to efficiently verify whether a seal is either genuine or forged is important. We propose an effective method of verification based on Hough transformation to approximate the imprint borders and the four vertexes, and use geometric transformation to align the perspective of the detected imprint image with the genuine imprint. After the edge-difference images between the original image and the detected image are created, distance transformation and connected-component labeling are applied. Finally, the number of edges in the connected component and the distance to the closest point in the edge of original image are used to calculate the input vector for the SVM (support vector machine). The imprint is then determined to be either genuine or forged. The experimental results show the effectiveness of the proposed verification approach.**

***Index Terms*—** Forgery seal imprints, Image processing, Image recognition, Edge difference

I. Introduction

Verifying the seal imprint to authenticate the identity of an individual or organization is a general and important task in the financial industry. In many Asian countries, people often use their private seal and imprint to prove the authenticity of their identification. In addition, the person's seal imprint is usually attached to a document with legal status for non-repudiation purposes, which means that the owner cannot deny the legal effect that exists in the document. Consequently, the seal imprint plays an important role in the financial industry.In financial applications, bank personnel often need to recognize whether a seal is genuine or forged. However, in practical banking situations, imprint verification is generally performed manually by the bank staff [1]. Most of the manual verification methods are based on the process of overlapping two imprints and visually comparing the difference. By artificially folding the detected imprints on the screen, the bank staff then visually observe whether the patterns are from the same or from different seals based on their experience. Therefore, verifying a seal imprint may take several minutes and the process is obviously inaccurate. Some research has been conducted in order to solve this problem.

There have been a number of prior works that investigated the issue surrounding seal imprint identification, beginning with the extraction of the imprint and continuing to the subsequent verification. These steps include the detection and extraction of the seal imprint [2] [3] [4][5], registration or alignment of the imprint [6], and seal verification [1] [7] [8]. In [2] and [4], the seal imprint is transformed from the RGB to the HSV (hue, saturation, value) color space, which is tolerant to changing illumination in extraction process, while in [3] the seal imprints were extracted by projecting the pixels onto an appropriate axis in the RGB color space. In [5], geometric distortion caused by angle of view is considered and geometric transformation is applied to eliminate the effect of it. In [6], the authors proposed a method which exploits the outer contour of the frame to locate the four edges. Rotation and translation were then performed to align the seal imprints. However, perspective distortion was not considered in this study. In [1], a distance-weighted correlation between seals is defined and applied to the dataset which satisfies strict constraints on image quality. Note that, the correlation is only applied on the skeletonized images since distance transform was still costive to perform at that time. To reduce the computational complexity, skeletonization processing was performed and there may be useful information for discriminating the genuine and forgery seal abandoned in the processing. In [7], Horiuchi clarifies a specific characteristic of the seal classification issue. The proposed algorithm is robust and insensitive to variations in image quality, but it may be lacking in its ability to detect a forgery that has a highly similar pattern to that of a genuine seal, e.g., a clone seal. In [8], an algorithm that quantifies the edge difference between two seals was proposed. Only the number of connected components of non-overlapped edges is used to judge the authenticity of the seal in the final step, the distance between the edges is solely used to remove those edges which are considered to be noise. For a more efficient verification technique, we propose an effective method based on Hough transformation [9] to approximate the imprint borders, and use geometric transformation to adjust the perspective of the detected imprint image. The proposed features would then be calculated and used as an input vector for the SVM (support vector machine)[10]. In addition, in order to compare the performance of the proposed method with existing methods, we apply the methods to a dataset composed of images collected from both genuine seals and clone seals. The clone seals were manufactured by scanning the imprint pattern from genuine seals and then carving a new seal using a laser. The proposed method

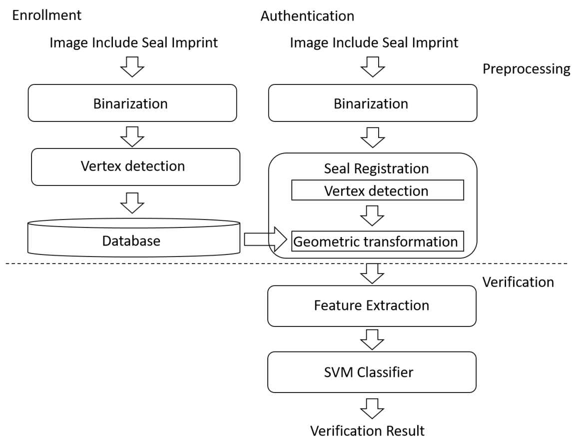

Fig. 1: Process flow for the proposed system

achieves a recognition rate of up to 92.5%, while the other methods are lower than 85%.

II. PROPOSED SYSTEM

In this section, we discuss the methods and processes implemented in our proposed verification system. Fig. 1 shows the overall system. There are two phases in our system: the enrollment phase and the authentication phase. In both phases, a binary seal imprint image is extracted, and the vertexes are detected. In the enrollment phase, the binary image and the vertexes are stored in the database, while in the authentication phase, we perform a geometric transformation to align the perspective of the two images. A range of predefined similarities are then calculated as the input of the SVM.

A. PREPROCESSING

1) Binarization

In the first step, we convert the seal imprint image to a binary image. The detection algorithm compares the form of the contours, as well as the features of the characters in the imprint. Since the color and hue of the image are basically composed of an ink smudge, the color information is not considered as a critical factor for identification. As a result, the color space for the input images is converted from the RGB color space to a bi-level (binary) image. Moreover, the binarization process will also remove some noise and stains on the paper. However, it is not easy to define a suitable threshold using RGB value. Consequently, we first convert the RGB to the HSV color space [11]. If the HSV values for each pixel are within a predefined range, the binary value for the pixel is set to 1. Otherwise, the value of the pixel is changed to 0. An HSV histogram analysis was performed in [4] to clarify following formula:

$$(0 < H < 15 \ \ or \ \ 330 < H < 360) \ \ and \ \ S > 0.5 \qquad (1)$$

The range defined in (1) is able to effectively separate the character region and the background region of the imprint. The original image and the binary image extracted from it are shown in Fig. 2(a) and (b), respectively.

(a) (b) (c) (d)

Fig. 2: The contours detected on the imprints. (a) The seal imprint image with the name “蘇侑晨” captured using a camera. (b) The binarization result of (a). (c) The image after the process of morphological operations.(d) The contour identified from (c).

(a) (b)

Fig. 3: Perspective transformation to align the perspective of the image. (a) The detected captured seal imprint image, which is subject to geometric distortion compared to the image shown in Fig. 2(a). (b) The perspective transformation of (a), which aligns the perspective to that of the image shown in Fig. 2(b).

2) Vertex detection

The borders of the image are detected to determine the vertexes of the seal imprint. Considering that the seal imprints may be incomplete, here we firstly perform the open operation [12] to maintain the closure of the seal imprint, such as that illustrated in Fig. 2(c). We then label the connected components and set all pixels to 0, except for the pixels within the region which has the most connected elements. After this process, we scan the image along the horizontal axis from top to bottom and the vertical axis from left to right. The contour of the borders can be then determined, as shown in Fig. 2(d). In order to improve the effectiveness and robustness of the process, we perform Hough transformation to identify the lines from the contour. We then categorize the lines into four classes, where each class corresponds to the borders. After categorizing the lines, we perform linear regression, which uses the least-squares method to fit the best line. Then, the vertexes are located by intersecting the lines.

3) Geometric transformation

To compare the slight difference between two imprints, we have to align the images. Consider the situation that camera is used to capture the imprint, as the position and orientation of the camera may be different, the perspective of the two images may not be exact as shown in Figs. 2(a) and 3(a). In other words, the detected imprint may be subject to geometric distortion when compared to the image stored in the database. Therefore, we apply a perspective transformation process to align the perspective.

After approximating the four borders through the process described above, we cross them to identify four points of intersection, and take these as the vertexes of the imprint. A perspective transformation from the four pairs of vertexes is then conducted to apply a deformation to the pixel grid, and this deformed grid is subsequently mapped to the destination image, as shown in Fig. 3(b).

B. SVM-based VERIFICATION

After performing the alignment of the seal, we calculate the predefined similarities as an input vector for the SVM. Utillizing the quantified edge difference [8] as the detection metric, we can measure the similarities between the original image and the detected image. To compare the similarities between the original image and the detected image, the distance between corresponding non-overlapped edges and their lengths are used as two parameters to quantify the geometric difference. To quantify the edge difference between the model (or reference) seal (MS) and the sample (or detected) seal (SS) and thereby calculate the two parameters, a left difference and a right difference image are defined, which are denoted as L_D and R_D respectively and defined as follows:

$$L_D(x,y) = \begin{cases} 1 & if\ E_{SS}(x,y)=0\ , E_{MS}(x,y)=1 \\ 0 & otherwise \end{cases} \tag{2}$$

$$R_D(x,y) = \begin{cases} 1 & if\ E_{SS}(x,y)=1\ , E_{MS}(x,y)=0 \\ 0 & otherwise \end{cases} \tag{3}$$

Here, $L_D(x,y)$ is the value of a pixel with position coordinates (x,y) in L_D. $R_D(x,y)$ is the value of a pixel with position coordinates (x,y) in R_D. $E_{SS}(x,y)$ is the corresponding pixel value at (x,y) in the detected image of SS, while $E_{MS}(x,y)$ is the corresponding pixel value at (x,y) in the registered image of the MS. Each connected component of the images for the left and right edge difference represents each element of a non-overlapped edge. Through connected component analysis, the area of i-th connected component in the images for the left or right edge difference is the length of each element of a non-overlapped edge, which is denoted as M_i, and the distance between each connected component to its corresponding seal edge is the distance between a pair of non-overlapped corresponding edges. For the imprints of the genuine seal, the difference in lengths of the non-overlapped edges caused by stamping conditions or by slight distortions are usually small. In contrast, forged seals, and even clone seals carved by a laser, usually contain continuous non-overlapped edges, since the combustion situation of the material is not completely the same. To reduce the effects of both noise and stamping conditions on the results of verification of the seal imprint, distance $d(x,y)$ from point (x,y) to its corresponding edge point is firstly calculated based on the distance transformation, as shown in Figs. **??**(a) and (b). Then, for every connected component, d_k denotes the distance d at the k-th pixel. The average of all the calculated distances d_k in the i-th connected component is regarded as D_i, i.e., the distance between that connected component and its corresponding seal edge. Formula (4) is the computational expression of the two parameters for the quantification of the edge difference.

$$\begin{cases} D_i(x,y) = \dfrac{1}{M_i}\displaystyle\sum_{k=1}^{M_i} d_k \\ L_i \qquad\quad = M_i \end{cases} \tag{4}$$

In [8], only the length of the longest non-overlapped edge L_{max} was adopted in order to verify whether or not the SS was genuine. The distance from the i-th element of the non-overlapped edge, D_i is only used to delete any erroneous edges which may be caused by the stamping conditions. We expect to improve the performance by exploiting D_i. Further, in order to gain statistical information, we identify the coordinates (x,y) that have the pixel values $1,2,3......n$, respectively in $d(x,y)$, and calculate the fraction of the pixels with a value of 1 in R_D at the corresponding coordinates. The fraction, denoted as R_p, is the rate of the number of pixels that differ from the MS at a certain distance d. Fig. 4 shows the estimated result, denoted as $R_{p,G}$ and $R_{p,F}$, respectively, reflecting the difference between the genuine seal imprint and the forgery to a certain extent. By treating R_p as a mapping table, the elements of the distance transformation for the reference image are mapped. An example is shown in Fig. 4. The generated matrix can then be used to weight the edge difference image from an element perspective. We define and calculate the similarities between the MS and the SS both globally and locally. That is, we apply the following formulas to both the whole seal imprints and the divided imprints which only include a character such as '蘇'. In the following discussion, the similarities between the divided seal imprints are notated as '$S_{n-\text{蘇}}$', ' $S_{n-\text{侑}}$', and '$S_{n-\text{晨}}$'.

$$S_1 = \frac{\sum_{x=1}^{N}\sum_{y=1}^{M} R_D(x,y) * D_i(x,y)}{\sum_{x=1}^{N}\sum_{y=1}^{M} R_D(x,y)} \tag{5}$$

$$S_2 = \sum_{i=1}^{K} D_i(x,y) * log(L_i) \tag{6}$$

$$S_3 = \max(L_i) \tag{7}$$

$$S_4 = \frac{\sum_{x=1}^{N}\sum_{y=1}^{M} R_D(x,y) * (1+d(x,y))^2}{\sum_{x=1}^{N}\sum_{y=1}^{M} R_D(x,y)} \tag{8}$$

$$S_5 = \frac{\sum_{x=1}^{N}\sum_{y=1}^{M} R_D(x,y) * R_{p,G}(x,y)}{\sum_{x=1}^{N}\sum_{y=1}^{M} R_D(x,y)} \tag{9}$$

$$S_6 = \frac{\sum_{x=1}^{N}\sum_{y=1}^{M} R_D(x,y) * log(\frac{R_{p,F}(x,y)}{R_{p,G}(x,y)})}{\sum_{x=1}^{N}\sum_{y=1}^{M} R_D(x,y)} \tag{10}$$

The definitions of the similarities are as follows: S_1 is the mean of the distance between the non-overlapped edges and their corresponding seal edges, for $M \times N$ images; S_2 is the product of the distance and the length of the non-overlapped edges, where K denotes the number of connected components; S_3 is the longest non-overlapped edge; S_4 to S_6 are modified forms of the distance weighted correlation [1], where S_4 only uses the distance to decide the weight of the pixels. When attempting to determine a more proper weight for the pixels, the estimated rate R_p, $R_{p,G}$ and $R_{p,F}$ are used to calculate S_5 and S_6. Compared to S_5, the similarities S_4

and S_6 are more sensitive for detecting those edges which have a larger distance, while S_5 is more conducive to detecting the edges which are close to their corresponding seal edges.

0	0	0	0	0
0	1	1	1	0
0	1	1	1	0
0	0	0	0	0
0	0	0	0	0

(a)

1	1	1	1	1
1	0	0	0	1
0	1	1	1	0
1	1	1	1	1
2	2	2	2	2

(b)

.31	.31	.31	.31	.31
.31	0	0	0	.31
0	.31	.31	.31	0
.31	.31	.31	.31	.31
.29	.29	.29	.29	.29

(c)

Fig. 4: Estimated density R_p at certain distance for genuine and forged imprints

The vector representation of the calculated similarities would be [$S_{i-\text{蘇}}$, $S_{i-\text{侑}}$,$S_{i-\text{晨}}$, $S_{i-total}$], where i is from 1 to 6. After inputting the information into the SVM, the SVM classification score would be outputted, then the SS would be verified as either genuine or a forgery based on a threshold of the score. A Gaussian kernel is used in the SVM where $\sigma = 4.5$ and the penalty factor $C = 1$.

III. EXPERIMENT RESULTS

We collected and produced the test data for the seal imprints from physical seals. Seal imprints that had a rectangular shape were used for the experiment. The forged seals used in our experiment were Type S forgery [8] which means that the seal was forged by scanning the genuine seal imprint into a computer and generating a forged seal from the scanned genuine seal template. A Type S forged seal is also called a clone seal. Moreover, two pairs of genuine and forged seals were carved at the same seal carving store using a laser carving machine based on fixed parameters. Consequently, these seal are more similar to each other than the other pairs which were carved at a different seal carving store. We constructed the test data using 850 rectangular images based on the genuine seal and 850 rectangular imprint images based on the forged seal. Of these, 50 rectangular genuine imprint and 50 rectangular forged imprint images were constructed by stamping using intaglio seals (yin carved seals), while the remaining imprints were obtained with yang carved seals.

The test data was imprinted from a specific physical seal in various situations. For example, when stamping with ink on paper, the seal was pressed at different strengths, and rotated different angles. When capturing the image, we positioned the camera lens from different perspectives and distances. Therefore, we were able to obtain data from many different compositions of translation, scaling and rotation. From the perspective of the data from a forged imprint, if an imprint was judged to be from a forged seal, it means that the similarity between the original imprint and the detected imprint was low. This could be because the two imprints have different components of in the name characters, or even have the same name but with different structures in the detail between the characters. In our experiment, the SS is aligned to the MS and the similarities between them is calculated as the input vector for the SVM. 60% of the images were used as a training set, and the remaining images were used as the testing set. To observe the performance, we plotted the ROC (receiver operating characteristic) curve using different thresholds. From Fig. 5, it can be seen that the recognition rate using the proposed method is up to 92.5% correct, while the false acceptance rate (FAR) and true positive rate (TPR) were 10% and 95%, respectively. Consequently, it can be deduced that the proposed method performs significantly better than other methods.

Table 1: Comparison of time complexity

	[1]	[7]	[8]	This work
Distance Transformation	0.064	0	0.064	0.032
Skeletonization	1.506	0	0	0
Connected-component labelling	0	0	0.324	0
Remaining	0.025	0.021	0.032	0.103
Total Time(sec)	1.757	0.021	0.420	0.135

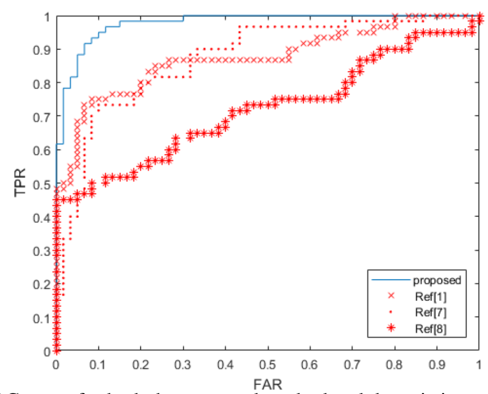

Fig. 5: ROC curve for both the proposed method and the existing methods

The comparison of computational complexity between the feature extracting methods is shown in the Table 1. The average cost time in each main process is calculated and recorded in the table. Note that, in [1] and [7], distance transform is performed twice for both E_{MS} and E_{SS} while in our method it is only performed on $E_M S$. The processes noted as Remaining containing the detail calculation like summation, array multiplication, logarithm, division and weighting are basically linear time complexity for each feature.

IV. CONCLUSION

To automate and improve the efficiency of the seal identification process, we have proposed a method based on morphological operations, Hough transformation, and regression analysis to approximate the imprint borders, and use geometric transformation to adjust the perspective of the detected imprint image. After the alignment of two imprints at the four corners, the similarities are calculated as the input vector for the SVM. The imprint can then be classified. This method is simple, but has the ability to enable accurate verification of the seal imprints, even if they were obtained from the seals carved using the same laser carving machine based on the same parameters.

V. References

[1] T. Fan and W. Tsai, "Automatic chinese seal identification," *Computer Vision, Graphics, and Image Processing*, vol. 25, pp. 311–330, 1984.

[2] W. Pan, J. Hu, M. Liu, X. Qi, and H. Lang, "Seal imprint segmentation based on color feature classifier," in *2012 International Conference on Audio, Language and Image Processing*, July 2012, pp. 837–840.

[3] K. Ueda, "Extraction of signature and seal imprint from bankchecks by using color information," in *Proceedings of 3rd International Conference on Document Analysis and Recognition*, Aug 1995, vol. 2, pp. 665–668 vol.2.

[4] X. Qi H. Lang, C. Xie and H. Ling, "Seal forgery detection by geometric consistency," *Information: An International Interdisciplinary of Journal*, 2012.

[5] Wei-Ho Chung, Mu-En Wu, Yeong-Luh Ueng, and Yi-Hsuan Su, "Forged seal imprint identification based on regression analysis on imprint borders and metrics comparisons," in *2018 IEEE Conference on Dependable and Secure Computing (DSC)*, 2018, pp. 1–5.

[6] X. Wang and Y. Chen, "Seal image registration based on shape and layout characteristics," in *2009 2nd International Congress on Image and Signal Processing*, Oct 2009, pp. 1–5.

[7] T. Horiuchi, "Automatic seal verification by evaluating positive cost," in *Document Analysis and Recognition, 2001. Proceedings. Sixth International Conference on*. IEEE, 2001, pp. 572–576.

[8] Z. Hao and J. He, "Automatic seal imprint verification by quantifying edge difference," in *Optoelectronic Imaging and Multimedia Technology*. International Society for Optics and Photonics, 2010, vol. 7850, p. 78500D.

[9] Richard O Duda and Peter E Hart, "Use of the hough transformation to detect lines and curves in pictures," *Communications of the ACM*, vol. 15, no. 1, pp. 11–15, 1972.

[10] Corinna Cortes and Vladimir Vapnik, "Support-vector networks," *Machine learning*, vol. 20, no. 3, pp. 273–297, 1995.

[11] Rafael C Gonzalez, Richard E Woods, et al., "Digital image processing," 2002.

[12] Jean Serra, "Introduction to mathematical morphology," *Computer vision, graphics, and image processing*, vol. 35, no. 3, pp. 283–305, 1986.