Offline Signature Verification System Using SVM Classifier with Image Pre-processing Steps and SURF Algorithm

Li Wen Goon¹, Swee Kheng Eng¹*
¹Faculty of Electrical Engineering Technology, Universiti Malaysia Perlis
Pauh Putra Campus, 02600 Arau, Perlis, Malaysia

*kheng@unimap.edu.my

Abstract. A signature is a mark or name that represents the identity of the people and the Signature Verification System (SVS) is used to validate the identity of people. The signature verification system is mostly used for bank cheques, vouchers, intelligence agencies and others. There are two types of SVS which are online and offline signature verification systems. The paper deals with an offline signature verification system. The proposed system consists of four main stages, (i) image acquisition, (ii) image pre-processing, (iii) feature extraction and (iv) classification. The image pre-processing steps involved binarization, noise removal using Gaussian filter and image resizing and thinning. In the feature extraction stage, Bag-of-Features with the Speeded Up Robust Features (SURF) extractor was utilized. In the third stage, the Support Vector Machine (SVM) classifier is used. Lastly, the confusion matrix and the verification rate were used to evaluate the performance of the classifier. In this paper, we implement and compare the performance of the signature verification system without entering the user ID and the signature verification system entering the user ID. For the ratio of 75% and 25% of the training and testing, respectively, the average accuracy for the signature verification system without entering the user ID is 71.36%, whereas the average accuracy for the signature verification system entering the user ID is 79.55%.

1. Introduction
Nowadays, the signature is very essential to represent someone's name or a mark to authenticate identity. The signature can be used as a security of data or document. Signature verification is a biometric technique used by a type of software to compare signatures and to authenticate the identity. Previously, manual signature verification has been used by some researchers [1], [2]. However, manual signature verification has some weaknesses/ disadvantages. If through the way of manual signature verification, humans cannot check that the signature is real or fake in a short time [3].

Handwritten signatures are widely used in authenticating the identity in daily life. Therefore, for technology today there are two major methods for signature verification which are online and offline signature verification. Online signature verification is dealing with dynamic features while offline signature verification uses the static feature from the signature image and it deals with the shape. Offline signature verification is more challenging than online signature verification. The online signature verification has complete dynamic information while the offline signature verification has issues with images such as noise, poor resolutions or other image processing issues [4]. Offline signature verification does not need special hardware when people are signing their signatures and it is ubiquitous.
Therefore, offline signature verification is important to verify the signatures found in the bank cheque or vouchers [5].

This paper proposes an offline signature verification system using a Support Vector Machine (SVM) classifier with image pre-processing step and SURF algorithm. In this work, after the stage of image acquisition, the pre-processing step is then implemented in order to enhance the quality of the signature in the images. The image pre-processing step involved binarization, noise removal using Gaussian filter and image resizing and thinning. After that, the Speeded Up Robust Features (SURF) extractor was utilized in the feature extraction stage. For the classification stage, there are a lot of classifiers such as Support Vector Machine (SVM), Convolutional Neural Network (CNN), k-Nearest Neighbors that can be used in signature verification. CNN is known to be outperformed with other classifiers [6] [7], but it has the disadvantages of having high complexity. Therefore, the SVM classifier which can perform high accuracy, but with lesser complexity in the implementation is employed during this stage.

2. Methodology

2.1. Proposed System

In this project, there are 4 main stages which are image acquisition, pre-processing steps, feature extraction and classification. The proposed method used for pre-processing steps is binarization, noise removal using Gaussian filter and image resizing and thinning. For the feature extraction, Bag-of-Feature (BoF) is used in the Speeded Up Robust Features (SURF) algorithm to extract the interesting points and the Support Vector Machines (SVM) is used in the classification stage. There are two types of the system proposed in the project, which are the offline signature verification system without entering the user ID and the offline signature verification system by entering the User ID. The flow chart of the system is shown in Figure 1.

2.1.1. Offline Signature Verification System Without Entering the User ID. For the offline signature verification system without entering the user ID, the datastore has split into training and testing sets. In this system, there are 11 users with 8 genuine and 8 forgery signature images that will implement for training and testing. A user will separate into two categories which are Forged_User and Real_User. This system has 3 different ratios of training and testing conditions as shown in Table 1. The 3 ratios for training and testing set, respectively, are ratio 1 (50%: 50%), ratio 2 (62.5%: 37.5%) and ratio 3 (75%: 25%).

First, the images will go through pre-processing steps to enhance the image quality. Then, the interest point will be extracted and trained in the SVM classifier. After that, the testing set will be compared with the training set and used the Confusion Matrix and the verification rate to evaluate the accuracy of the classifier.

| Table 1. Ratios for training and testing sets |
|-----------------|-----------------|-----------------|
|                 | Ratio 1         | Ratio 2         | Ratio 3         |
| Training Sets   | 50.00% (88 images) | 62.50% (110 images) | 75.00% (132 images) |
| Testing Sets    | 50.00% (88 images) | 37.50% (66 images)  | 25.00% (44 images)  |
| Total           | 176             | 176             | 176             |

2.1.2. Offline Signature Verification System by Entering the User ID. For the offline signature verification system by entering the user ID, it can only verify one image at one time. First, the user ID is required to be keyed into the system and then the system will train that user’s genuine and forgery images. After training that user’s genuine and forgery images, the system can enter a testing image number to do the classification. The result will be displayed as genuine or forged. This system can only test 1 image every time and the system used the while loop until the key of “break” is pressed. This system is separated user by user in the training stage. Each time only a user will be trained, and a user
consists of 8 genuine and 8 forgery images. That means each time 16 images will be trained for a user. After that, pre-processing step, feature extraction and classification are performed.

![Diagram](image)

**Figure 1.** Offline signature verification system (a) Without entering the User ID (b) Entering the User ID

### 2.2. Image Acquisition

This project uses the public database of ICDAR 2009 Signature Verification Competition (SigComp2009) [8]. In the database, it contains genuine signatures from 100 writers and forgery signatures from 33 writers. However, only part of the images of the database will be used in our implementation.

### 2.3. Pre-processing Steps

Pre-processing plays an important role prior to feature extraction. It is implemented for enhancing the quality of the image. The details of the pre-processing steps are shown as follows.

#### 2.3.1. Binarization.

It is a process that transforms the colour image to the grayscale, and then converts the pixel image to a binary image. This process can lead to a clearer contour of the signature. In this step, Otsu’s method is used to minimize the within-class variance of the threshold black and white by choosing a threshold value [9]. The within-class variance is shown as Equation (1) where \(h_d(k)\) is the average intensity, \(h_c\) is the cumulative sum and \(k = 0, 1, \ldots, L-1\) [10]. The range of the gray level is from 0 until L-1.
\[ \sigma_w^2(k) = h_{c1}(k)[h_{a1}(k) - h_a(L-1)]^2 + h_{c2}(k)[h_{a2}(k) - h_a(L-1)]^2 \]  
(1)

2.3.2. **Noise Removal.** Gaussian filter is used to remove the noise and enhance the signature image structures. Gaussian smoothing is used also to blur the images and reduce the noise. Equation (2) is the impulse response of the 1-D Gaussian filter with a low pass where the \( \sigma \) is the standard deviation and \( n \) is the number of samples.

\[ h(n) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{n^2}{2\sigma^2}} \]  
(2)

2.3.3. **Resizing.** A process that resizes all the images to be the same size as 800 x 1500 pixels.

2.3.4. **Thinning.** A morphological operation is used to remove the selected foreground from the binary image and will lead to the skeleton of the object (signature contour) to be retrieved. This is very important since it enhances the feature extraction later by providing a significant reduction in the data.

2.4. **Features Extraction using Speeded Up Robust Features (SURF)**

Speeded Up Robust Feature (SURF) is very suitable for a signature verification system since it is fast and robust. For the signature verification system, the SURF algorithm is an approximation of the determinant of the Hessian as Equation (3). Equation (4) is the expression of local changes around the area of each point in direction x and y where the \( L_{xy}(x, \sigma) \) represents the convolution with Gaussian 2nd derivatives.

\[ \det(H_{approx}) = D_{xx}D_{yy} - (wD_{xy})^2 \]  
(3)

\[ H(x, \sigma) = \begin{bmatrix} L_{xx}(x, \sigma) & L_{xy}(x, \sigma) \\ L_{xy}(x, \sigma) & L_{yy}(x, \sigma) \end{bmatrix} \]  
(4)

2.4.1. **Bag-of-Features (BoF).** Bag-of-Feature (BoF) is also called bag of visual word model of the object which suitable used for managing the signature image collection and partition them to the training and testing sets. BoF can extract the features by using the extractor Speeded Up Robust Features (SURF) algorithms and it can train the image category classifier by using Support Vector Machine (SVM) [11].

2.5. **Classification**

Support Vector Machine (SVM) is a linear discriminant classifier that can use to verify the signature images [12]. It is used for classifying the genuine and forgery when enters a user ID. SVM is a great accuracy and simple classifier. There are two types of Support Vector Machine which are linear SVM and non-linear SVM. SVM is basically separates classes linearly by finding the best hyperplane. This project implemented multi-classes of linear SVM. In this project, the testing set of signature images is compared with the training set of signature images which have been trained by classifier SVM. Table 2 shows the 3 different ratios for the system without entering the user ID and Table 3 shows the number of images for training and testing for the system entering the user ID. The number of images is \((p \times 22)\) where the \( p \) is the number in the first \( p \) number that is used to train in a category and it multiplies with 22 categories.
Table 2. The offline signature verification system without entering user ID with 3 different ratios

| Image Sets     | Percentage | Number of images | Percentage | Number of images | Percentage | Number of images |
|----------------|------------|------------------|------------|------------------|------------|------------------|
| Training Set   | 50%        | 4x22=88          | 62.50%     | 5x22=110         | 75.00%     | 6x22=132         |
| Testing Set    | 50%        | 4x22=88          | 37.50%     | 3x22=66          | 25.00%     | 2x22=44          |
| Total          | 50%        | 176 images       | 75.00%     | 176 images       | 75.00%     | 176 images       |

Table 3. The number of images for training and testing for system entering user ID

|       | Training | Testing |
|-------|----------|---------|
|       | 11 users x (8 genuine + 8 forgery) | 11 users x (2 genuine + 2 forgery) |
|       | = 176 images | = 44 images |

2.6. Evaluation of Accuracy

Confusion Matrix and the Verification Rate have been adopted in this project to evaluate the performance of the project. In Confusion Matrix, it consists of 4 terms that will be used to evaluate the performance, which are true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN) and the accuracy is as Equation (5). Confusion Matrix and Verification Rate have been used in a signature verification system without entering the user ID. Confusion Matrix is a table in which column represents the actual label and row represents the predicted label. Verification Rate is used to calculate the total number that has been verified correctly per total number of signatures that has been verified as shown in Equation (6). The Confusion matrix and the verification rate will get the same accuracy since the confusion matrix is (TP+TN) divided by the total number of images tested. However, the system by entering the user ID only adopted the Verification Rate for evaluating the performance since it tests one image one time.

\[
\text{Confusion Matrix} = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \tag{5}
\]

\[
\text{Verification Rate} = \frac{\text{Total number that verify correctly}}{\text{Total number of signatures that verify}} \times 100\% \tag{6}
\]

3. Results and Discussion

Pre-processing steps for the signature images are an important role in this project. An example of the result of pre-processing steps is shown in Figure 2. In removing the noises and blur the images (in pre-processing step), the Gaussian filter performs better than the Median filter. This has been shown in the step of extracting the SURF features and evaluating the accuracy. For example, the Gaussian filter used in the system without entering user ID can extract up to 104388 SURF features from 176 images. However, only 82268 SURF features can only be extracted from 176 images by using a Median filter. The accuracy of using a Gaussian filter is higher than the accuracy of a Median filter. Hence, the Gaussian filter is more suitable to be used in this project. It is also mentioned that when more SURF features are extracted, it will probably lead to a better classification result.

An example of the strongest SURF features that are extracted in the Bag-of-Features (BoF) is as shown in Figure 3. The system without entering user ID at ratio (75%: 25%) has the highest average
accuracy (71.36%) as compared to the other two ratios as shown in Table 4. From the offline signature verification system without entering the user ID, it is known that when more images are used for training, then it will lead to higher accuracy in the classification. Besides that, the system entering the user ID obtained an accuracy of 79.55% and it is evaluated by using the verification rate as shown in Table 5. In Table 5, there are 35 out of 44 testing images are verified correctly. In addition, the remaining 9 testing images are verified wrongly, and it may be due to the forgery signature is almost the same as the genuine signature.

**Figure 2.** Pre-processing steps (a) Binarization, noise removal using Gaussian filter and resizing (b) Thinning
Figure 3. SURF features detected in signature image

Table 4. The accuracy for signature verification system without entering user ID with 3 different ratios

| Training Set | Testing Set | Accuracy 1 | Accuracy 2 | Accuracy 3 | Accuracy 4 | Accuracy 5 | Average |
|--------------|-------------|------------|------------|------------|------------|------------|---------|
| 50.00%       | 50.00%      | 0.5114     | 0.5114     | 0.5568     | 0.5114     | 0.5455     | 0.5273  |
| 62.50%       | 37.50%      | 0.6061     | 0.6818     | 0.6970     | 0.6667     | 0.6364     | 0.6576  |
| 75.00%       | 25.00%      | 0.7500     | 0.7273     | 0.7045     | 0.6818     | 0.7045     | 0.7136  |

Table 5. The accuracy for signature verification system by entering user ID

| Verification | Accuracy | \( \frac{35}{44} \times 100\% = 79.55\% \) |
|--------------|----------|------------------------------------------|
| Correct      | 35       |                                          |
| Wrong        | 9        |                                          |
| Total        | 44 images|                                          |

4. Conclusion and Recommendation
This project has two different ways for verification of the signature which the offline signature verification system without entering the user ID and the offline signature verification system by entering a user ID each time. For the Offline signature verification system without entering the User ID, the highest average accuracy is 71.36% (at a ratio 75%; 25%). For the offline signature verification system by entering a User ID, it achieved an accuracy of 79.55%. The accuracy can be improved in future work. A lot of parameters that affect the performance of the system. In future work, the more different public databases can be adopted to increase reliability. In addition, the different features extraction algorithms and the classifiers can be also adopted to improve the accuracy of the classification.

Acknowledgements
The authors would like to show their appreciation to the Universiti Malaysia Perlis for the opportunity and assistant given.
References

[1] J Coetzer, B M Herbst and J A du Preez 2006 Tenth Int. Work on Off-Line Signature Verification: A Comparison between Human and Machine Performance (South Africa: University of Stellenbosch) pp 1–6

[2] Aythami M, Derlin M, Julian F and Ruben V-R 2017 IET Biometrics on Signature Authentication based on Human Intervention: Performance and Complementarity with Automatic Systems vol 6 (Spain: Universidad Autonoma de Madrid) pp 307–315

[3] S Priya, A K R N Supreeth, K Somesh and A Hruday Kumar 2019 Signature Verification System using Different Algorithms (Blue Eyes Intelligence Engineering and Sciences) pp 107–109

[4] Mustafa S K, Mamoun J M and Hayden A 2021 Indones. J. Electr. Eng. Comput. Sci. An Accurate Signature Verification System based on Proposed HSC Approach and ANN Architecture vol 21, pp 215–223

[5] Mustafa B Y 2015 Offline Signature Verification with User-Based and Global Classifiers of Local Features (Sabanci University)

[6] Hsin H K and Che Y W 2020 Appl. Sci. on An Offline Signature Verification and Forgery Detection Method based on A Single Known Sample and An Explainable Deep Learning Approach vol 10 (Switzerland: MDPI, Basel)

[7] Eman A, Belal A M A, Bassem S A-N, Ahmed J K, Musleh M M, Alaa M B and Samy S A-N 2019 Int. J. Acad. Multidiscip. Res. on Handwritten Signature Verification using Deep Learning vol 3, pp 39–44

[8] Dr. Phil. nat. Marcus Liwicki ICDAR 2009 Signature Verification Competition (SigComp2009) -TC11 Available: http://iaprtc11.org/mediawiki/index.php?title=ICDAR_2009_Signature_Verification_Competition_(SigComp2009)

[9] Nobuyuki O 1979 IEEE Trans. Syst. Man Cybern. on OTSU Paper vol 20, pp 62–66

[10] D Sundararaan 2017 Digital image processing vol 19 (Singapore: Springer Nature Singapore Pte Ltd)

[11] Mathworks 2016 Image Category Classification Using Bag of Features pp 1–10, Available: https://la.mathworks.com/content/dam/mathworks/tag-team/Objects /i/88400_93009v00 _Image_Class_Bag_Features_2016.pdf

[12] J R Parker 2013 Algorithms For Image Processing And Computer Vision vol 53 (Indiana: Wiley Publishing,Inc., Indianapolis)