Offline Signature Verification and Forgery Detection Approach

Taraggy M Ghanim
Faculty of Computer Science,
Misr International University
Email: taraggy.ghanim@miuegypt.edu.eg

Ayman M Nabil
Faculty of Computer Science,
Misr International University
Email: ayman.nabil@miuegypt.edu.eg

***Abstract*—Signature verification and forgery detection is a challenging field with a lot of critical issues. Signatures forgery drives cooperates and business organizations to huge financial loss and also affects their security reputation. Highly accurate automatic systems are needed in order to prevent this kind of crimes. This paper introduce an automatic off-line system for signature verification and forgery detection. Different features were extracted and their effect on system recognition ability was reported. The computed features include run length distributions, slant distribution, entropy, Histogram of Gradients features (HoG) and Geometric features. Finally, different machine learning techniques were applied on the computed features: bagging tree, random forest and Support Vector Machine (SVM). it was noticed that SVM outperforms the other classifiers when applied on HoG features. The system was applied on Persian Offline Signature Data-set (UTSig) database and achieved satisfactory results in differentiating between genuine and forged signature.**

***Index Terms*—Signature Verification, Forgery Detection, Support Vector Machines, Histogram of Gradients.**

I. Introduction

Signature forgery in legal documents, bank checks, doctors prescriptions can lead to huge consequences .in this respect signature verification is an important application in the field of bio-metrics. Bio-metrics measure human behaviors and are used in constructing recognition systems. Such recognition systems are useful for authorization and achieving high degree of security.

Although signatures forgery are often manually detected by experts, but still high accuracy is not always achieved. there are many difficulties in the manual forgery detection due to variations in handwriting style and professionalism of forgers. Recognizing dissimilarity between genuine and forged signatures needs skilled professionals . Automatic recognition systems can play an effective role in verifying signatures with high accuracy and in differentiating between genuine and forged signatures. A new approach is introduced under the category of offline verification systems [1]. The system is tested on Persian Offline Signature Data-set (UTSig) [2]. The data-set is composed of 3105 genuine samples and 5175 forged signatures.

The introduced paper calculates the accuracy of different approaches based on extracting different types of features and applying different classifiers. Finally, a system is designed to be an effective and accurate tool for detecting forgery in signatures. The system was first trained by different classifiers; Support Vector Machines (SVM), Random Forest (RF),and bagging trees. Different types of features were computed to build a comparative study, to reach an optimal solution.

The paper starts by summarizing related work in section II. Section III introduces the new approach briefly, then experimental results will be shown in section IV. Finally, paper ends with conclusion.

II. Related Work

Several systems were proposed in the field of offline signature verification and recognition. Some approaches applied Support Vector Machine (SVM) for classification. Kruthi.C, Deepika.C.Shet [3] applied Support Vector Machine (SVM) using different kernel functions and Sequential Minimal Optimization (SMO) for Offline Signature Verification. Binarization, filtering, edge thinning and Canny Edge detector were applied on images for preprocessing and segmentation. Features were extracted based on aspect ratio, normalized area, horizontal and vertical profiles, vertical centroid, slant angle, edge histogram, edge direction histogram. The system was tested using 336 different signature samples and achieved classification error rate less than 7.16%. Another approach using SVM was introduced by [4] for offline signature verification. Features based on Gray Level Difference Matrix (GLDM) and Haar wavelets were extracted to differentiate between genuine and forgery signatures. The system achieved 7.533% error equal rate.

Soleimani [2] implemented a signature verification system using Linear SVM. They used data-set of 115 writer with 27 genuine sample for each writer. Signature envelope and interior stroke distribution in polar and Cartesian coordinates were extracted as features. In polar coordinates three different features were extracted including derivative of radius of signature envelope, its angle,and the number of black pixels that the radius crosses when rotated from one point to the next point. In Cartesian coordinates another set of features were extracted like height, width and the number of transitions from black to white or white to black pixels of signatures.

The system succeeded with 70.67%. Another model [5] is introduced and applied on GPDS [6] and MCYT databases [7]. Histogram features were extracted to measure texture. The SVM classifier using the radial bases function (RBF) kernel was applied for classification.

An offline signature verification system [8] based on Artificial neural network (ANN) extracted six geometric features from binary cropped signature images. The extracted feature vector was composed of area, centroid, standard deviation, even pixels, kurtosis and skewness. Genuine and forged signatures were distinguished with accuracy 89.24% on the standard MCYT offline signature corpus data-set [7]. Suhail M. Odeh,Manal Khalil [9] also applied neural network to differentiate between original and fraud signatures. During pre-processing images were re-sized, signatures were then segmented and binarized. Fast Fourier transform (FFT) was applied then feature vector was computed based on eccentricity, skewness, kurtosis, orientation. A sigmoid function was used for activation of the multilayer perceptron (MLP) network.The system accuracy was 78.8% when tested on 300 signatures of GPDS database [6].

Another approach [10] based on feed-forward neural network was proposed in this field. The computed features include aspect ratio, directional and energy density features. A hyperbolic tangent sigmoid function was used for activation. The network was trained by small number of data samples and was inefficient due to rotation variance. A neural network based system [11] using Feed Forward Conic Section Function Neural Network (CSFNN) was applied as a classifier for signatures. Differential Evaluation Algorithm (DEA) was used for training CSFNN. The system was implemented on FPGA to be 105 times faster than similar systems.

Another verification approach [12] applied bagging trees classifier. The approach was applied on feature vector based on curves of the signatures. The calculated features were the total number of pixels in the signature, occupancy ratio and the minimum Eigen value of the signature curve. The used database consists of 500 signatures for 100 persons,each person having 5 signatures, represented in 500 patterns. 60% of the patterns were used for training and the rest 40% were used for testing. The achieved signature recognition rate was 79.8%.

Random forest classifier [13] was applied on ICDAR database [14] to construct an offline signature verification system. 800 genuine signatures were used for training and 200 genuine signature for testing. The system accuracy was 67%. An approach [15] based on pseudo dynamic features and principal component variance using eigen values was proposed for signature verification. The kinematic theory of rapid human movements [3] was studied to verify signatures. Analysis of finger movements styles was applied and features like number of strokes, their timing and similarity were computed. Similarity measure was achieved by string edit distance and dynamic time warping.

III. Approach

The introduced approach composed of the main phases of pattern recognition systems as shown in Fig. 1. Test images and database classes pass through preprocessing, feature extraction, classification and finally post-processing phase to determine whether input image is a genuine or fraud. The final optimized approach uses the HOG features and applies the SVM classifier. The experiment section compares between different types of implemented features where some geometric features proposed by [16], and the other type of feature is statistical features based on distributions, in addittion to the effect of Gabor wavelets and histogram of oriented gradients as features is included.

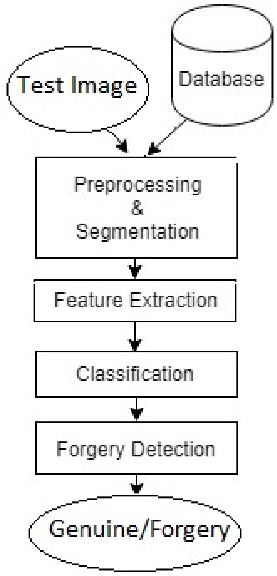
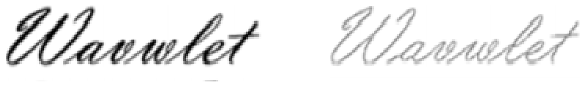


Fig. 1: System Overview

A. Preprocessing & Segmentation

Signatures are introduced to the system as scanned images of 200 dpi resolution. Images' size are normalized. Segmentation is achieved by binarization. Thinning algorithm is applied on signatures as shown in Fig. 2.



(a) Image before thinning (b) Image after thinning

Fig. 2: A sample image before and after thinning

B. Feature Extraction

Different types of features were extracted in this phase. The effect of each type of feature on classification is studied and shown in the experiment section later. This phase is an essential part of computation that precedes classification phase. It aims to extract valuable information from input images that concerns region of interest.

1) Geometric Features: This type of features extracted in the approach [16] represents different lines types. Line type is determined from the relation between each pixel and its neighbours. Lines are classified in images to left and right diagonal, horizontal and vertical. Direction vector is extracted from images by tracing pixels and representing line type using the direction matrix shown in Fig. 3, where C is each edge pixel in the segmented region of interest.

$$\left[\begin{array}{c|c|c} 4 & 5 & 6 \\ \hline 3 & C & 7 \\ \hline 2 & 1 & 8 \end{array}\right]$$

Fig. 3: Kernel to determine Line type

The shown 3x3 direction matrix is used to determine line type with respect to the center pixel according to the following rules:

- Maximum occurrence of 2 or 6 is an indication to right diagonal line type.
- Maximum occurrence of 4 or 8 is an indication to left diagonal line type.
- Maximum occurrence of 1 or 5 is an indication to vertical line type.
- Maximum occurrence of 3 or 7 is an indication to horizontal line type.

Each Input image is divided in 3 equally horizontal zones and 3 vertical equally zones. From each zone, a feature vector is extracted of length equal 8. These eight features are number of horizontal, vertical, right diagonal, reft diagonal lines and their respectively normalized lengths. Normalized number per line type is defined by

$$value = 1 - ((numberoflines/10)x2) \tag{1}$$

Normalized length per line type is defined by

$$length = (numberOfPixelsPerLineType)/(ZoneArea) \tag{2}$$

2) Statistical Features: Statistical features is popular type of features in the field of signature verification and recognition [1]. In the introduced system, statistical features were extracted including edge-hinge distribution, slant distribution, run-length and entropy[17].

a) Slant distribution: Slant distribution is computed by extracting edge information using Sobel edge detector. Thresholding is then applied to remove undesired edge values. Each valuable edge pixel is now centered in a square neighborhood kernel. Edge orientations are then recorded as shown in Fig. 4.

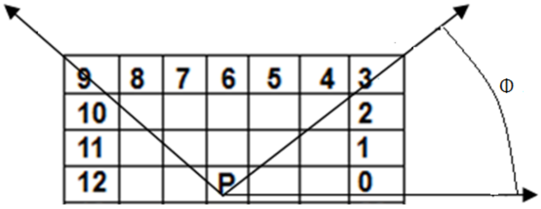


Fig. 4: Extraction OF Edge-Slant distribution

A histogram counts all the verified orientation instances. This histogram is normalized to form a probability distribution that provides probability information about the computed orientations.

b) Edge-hinge distribution: Two edge fragments are considered between neighbor pixels and each central pixel. Joint probability distribution of the two fragments orientations [18] are computed as shown in Fig.5.

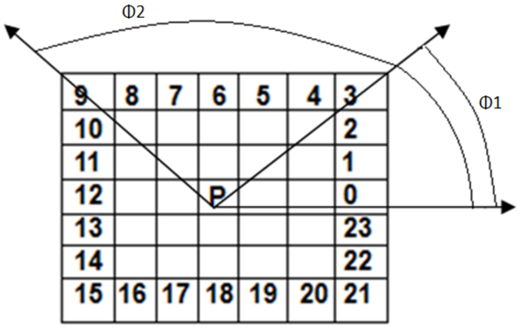


Fig. 5: Extraction OF EDGE-Hinge distribution

The Edge-hinge distribution characterizes the gradients of a writing stroke in handwritten signatures. It is extracted using a sliding window over an edge-detected binary image. The directions of the two edge fragments are measured and stored as pairs. The joint probability distribution is obtained from such pairs [18].

c) Run Length Features: Run lengths matrix (Petrou et al (2006)) are applied on the computed binary image to represent number of runs with pixels of value i and length j. Signatures are scanned horizontally along the rows and vertically along the columns. The run length feature vector is normalized and is interpreted as a probability distribution [18] [17]. Features like short runs and long runs emphasis, gray Level run length matrix and non-uniformity run percentage are computed to represent binary textures in the input signature images.

d) Entropy: Entropy as a feature evaluates the uncertainty of data distribution as shown in equation 3. It is a quantitative description to the internal extracted gradient features of signature images.

$$Entropy(I)) = \sum_{k} P(k) \log P(k) \tag{3}$$

where I is the input image while k is the possible gradients values.

3) Gabor Wavelet Features: The Gabor transformation extracts textural information from images. It is defined by:

$$g(x, y; \lambda, \theta, \phi, \gamma) = \exp\left(-\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2}\right) \cos\left(2\pi\frac{x'}{\lambda} + \phi\right), \tag{4}$$

where

$$x' = x\cos\theta + y\sin\theta, y' = -x\sin\theta + y\cos\theta. \tag{5}$$

where λ is the wavelength, θ is orientation and represents angle of the normal to the sinusoidal, ϕ is the phase which is the offset of the sinusoidal. Finally, $gamma < 1$ is the aspect ratio.[19].

4) Histogram of Oriented Gradients (HOG): Histogram of Oriented Gradients (HOG) is one of the most effective features [20] to represent information of local gradient directions [21]. This is achieved by counting occurrences of gradient orientations in localized zones of an image. It is applied on a dense grid of uniformly spaced blocks. The Locally normalized HOG descriptors [20] outperform other features including wavelets [22] [23].

C. Classification

In this phase, three different classifiers were applied on the different types of computed features. Their effect is recorded in the experiment sections.

1) Support Vector Machine : SVM classifier is a supervised learning technique that can applied using different kernel types. In the proposed system, linear kernel achieves satisfactory recognition accuracy when applied on the data-set [2]. The proposed system applied multi-class SVM to perform classification by constructing an n-dimensional hyper-planes. These constructed hyper-planes differentiate between feature vectors of writers' classes. Applying multi-class SVM can be implemented by two different models, either one-versus-one model or one-versus-many model. In the introduced system, the one-versus-many approach is implemented.

2) Bagging Trees: Bagging trees classifier [24] works on random distribution of training classes. It is used for classification and regression. The classification trees are constructed during training stage. At each node a binary decision either true or false is computed. The results of many binary decision trees are combines to decrease over-fitting. The approach selects the branches that satisfies best optimization criterion that subjects to the minimum leaf constraint.

3) Random Forest: Random forest classifier [13] constructs a forest of decision trees. Each tree is constructed based on the feature vectors of the training samples. Set of feature variables is chosen at each node to provide the best tree binary split. Sets of feature variables differ from one node to the other in the same tree.

D. Forgery detection

The SVM classifier measures divergence between each input image and all the data-set training signatures. Degree of divergence is a value that represents the probability the signature is fraud or genuine. Fig. 6 is a histogram of all the calculated divergence measurements during testing. It was found that a continuous range of values discriminate the true positive and true negative samples from the fraud signature samples. Accordingly, a threshold is determined to distinguish between genuine and fraud signatures.

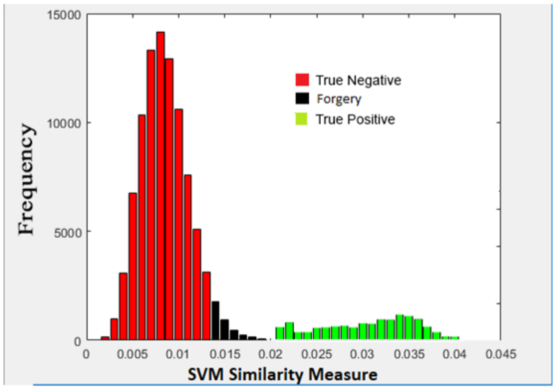


Fig. 6: Histogram of probabilities for SVM detecting forgery.

As shown in Fig. 6, there are three different areas. The Red bars represent the true negative samples, the green bars represent the true positive, while the black bars represent the forgery signatures. The horizontal axis is the measured divergence values, while the vertical axis represents their histograms.

IV. Experiments

The System was tested over 115 different writers. Each writer provides 20 training images (total 3200 images for all writers) and 7 testing images (total 805 for all writers).

All the classifiers were applied independently in the system. The effect of Bagging Trees, Random Forest and SVM classifiers were tested. Support vector machine (SVM) classifier achieves the best accuracy, which is **94%** in detecting forgery in signatures. Table I shows the experimental results of applying bagging trees on different types of features. Different numbers of trees were tested. The optimum performance was achieved when the bagging trees were applied on feature vector composed of slant distribution, entropy and run length features. In this experiment 100 bagging trees achieved the highest recognition rates relative to other experiments with different numbers of bagging trees.

TABLE I: Performance of Bagging Trees

Feature Vector	No. of trees	accuracy
Slant distribution, entropy, run length	100	**79.7 %**
Slant distribution, entropy, run length	200	77.7%
Slant distribution, entropy, run length	80	76.2%
HoG features	100	78.9%
HoG features	200	71.2%
HoG features	80	77.1%
Geometric features	100	47.9%
Geometric features	200	45.1%
Geometric features	80	54.8%

The performance of Random Forest was better than bagging trees due to the feature selection phase added to this type of classifier. The HoG features with 200 trees achieved best results as shown in table II.

TABLE II: Performance of Random Forest

Features	No. trees iterations	Accuracy
HOG Features	200	**86%**
HOG Features	100	81.8%
HOG Features	80	80.5%
Slant distribution, entropy, run length	200	80.2%
Slant distribution, entropy, run length	100	76.8%
Slant distribution, entropy, run length	80	72.2%
Geometric features	200	70.2%
Geometric features	100	66.9%
Geometric features	80	68.4%

Experiments shown in table III summarize the results of applying SVM classifier in the system. SVM was trained using different types of features. It achieves superior results with the Histogram of gradients features. The Linear kernel is applied as it fits the handwriting recognition systems [1].

TABLE III: Performance of Support Vector Machine

Features	Accuracy
HOG Features	**94%**
Geometric features and Slant distribution, entropy, run length	77%
Slant distribution, entropy, run length	64%

V. Conclusion

The paper introduces a recognition system which is based on HOG features and SVM classifier for offline signature verification and forgery detection. The performance of SVM, bagging trees and Random forest classifiers was measured using different types of features. Finally, SVM classifier achieved superior recognition rates compared to the other classifiers, when applied on HoG features.

Acknowledgment

We would like to thank Diana Salama, Hatem Herzawy, Sylvia El Maraghy and Caroline Kerlos for their greatly assistance in this research

References

[1] T. M. Ghanim, M. I. Khalil, H. M. Abbas, Phog features and kullback-leibler divergence based ranking method for handwriting recognition, in: 8th IAPR TC3 Workshop on Artificial Neural Networks in Pattern Recognition, IEEE, 2018.

[2] A. Soleimani, K. Fouladi, B. N. Araabi, Utsig: A persian offline signature dataset, IET Biometrics 6 (1) (2016) 1–8.

[3] C. Kruthi, D. C. Shet, Offline signature verification using support vector machine, in: Signal and Image Processing (ICSIP), 2014 Fifth International Conference on, IEEE, 2014, pp. 3–8.

[4] N. RamyaRani, S. Veerana, D. Prabhakaran, Texture based offline signature verification system.

[5] M. A. Ferrer, J. Vargas, A. Morales, A. Ordóñez, Robustness of offline signature verification based on gray level features, IEEE Transactions on Information Forensics and Security 7 (3) (2012) 966–977.

[6] F. Vargas, M. Ferrer, C. Travieso, J. Alonso, Off-line handwritten signature gpds-960 corpus, in: Document Analysis and Recognition, 2007. ICDAR 2007. Ninth International Conference on, Vol. 2, IEEE, 2007, pp. 764–768.

[7] J. Ortega-Garcia, J. Fierrez-Aguilar, D. Simon, J. Gonzalez, M. Faundez-Zanuy, V. Espinosa, A. Satue, I. Hernaez, J.-J. Igarza, C. Vivaracho, et al., Mcyt baseline corpus: a bimodal biometric database, IEE Proceedings-Vision, Image and Signal Processing 150 (6) (2003) 395–401.

[8] A. S. Shah, M. Khan, F. Subhan, M. Fayaz, A. Shah, An offline signature verification technique using pixels intensity levels, International Journal of Signal Processing, Image Processing and Pattern Recognition 9 (8) (2016) 205–222.

[9] S. Odeh, M. Khalil, Apply multi-layer perceptrons neural network for off-line signature verification and recognition, International Journal of Computer Science Issues (IJCSI) 8 (6) (2011) 261.

[10] M. Tomar, P. Singh, A directional feature with energy based offline signature verification network, International Journal on Soft Computing 1 (6) (2011) 48–57.

[11] A. R. YILMAZ, B. Erkmen, O. YAVUZ, Accelerating handwritten signature recognition using intelligent algorithm based embedded system., Sigma: Journal of Engineering & Natural Sciences/Mühendislik ve Fen Bilimleri Dergisi 34 (3).

[12] D. Darwish, Assessment of offline digital signature recognition classification techniques, International Journal of Computer Networks & Communications Security 1 (4).

[13] M. Thenuwara, H. R. Nagahamulla, Offline handwritten signature verification system using random forest classifier, in: Advances in ICT for Emerging Regions (ICTer), 2017 Seventeenth International Conference on, IEEE, 2017, pp. 1–6.

[14] V. Märgner, H. El Abed, Icdar 2009 arabic handwriting recognition competition, in: Document Analysis and Recognition, 2009. ICDAR'09. 10th International Conference on, IEEE, 2009, pp. 1383–1387.

[15] J. Arunalatha, C. Prashanth, V. Tejaswi, K. Shaila, K. Raja, D. Anvekar, K. Venugopal, S. S. Iyengar, L. M. Patnaik, Pcvos: Principal component variances based off-line signature verification, in: Recent Trends in Information Systems (ReTIS), 2015 IEEE 2nd International Conference on, IEEE, 2015, pp. 195–199.

[16] D. D. Gaurav, R. Ramesh, A feature extraction technique based on character geometry for character recognition, arXiv preprint arXiv:1202.3884.

[17] M. Bulacu, L. Schomaker, L. Vuurpijl, Writer identification using edge-based directional features, in: null, IEEE, 2003, p. 937.

[18] M. Konstantakis, E. Yannakoudakis, A writer identification system of greek historical documents using matlab, International Journal of Emerging Technology and Advanced Engineering 4 (10) (2014) 609–617.

[19] J. Wen, B. Fang, Y.-Y. Tang, T.-P. Zhang, H.-X. Chen, Offline signature verification based on the gabor transform, in: Wavelet Analysis and Pattern Recognition, 2007. ICWAPR'07. International Conference on, Vol. 3, IEEE, 2007, pp. 1173–1176.

[20] H. Bristow, S. Lucey, Why do linear svms trained on hog features perform so well?, arXiv preprint arXiv:1406.2419.

[21] J. Hu, Y. Chen, Offline signature verification using real adaboost classifier combination of pseudo-dynamic features, in: Document Analysis and Recognition (ICDAR), 2013 12th International Conference on, IEEE, 2013, pp. 1345–1349.

[22] A. Mohan, C. Papageorgiou, T. Poggio, Example-based object detection in images by components, IEEE Transactions on Pattern Analysis & Machine Intelligence (4) (2001) 349–361.

[23] P. Viola, M. J. Jones, D. Snow, Detecting pedestrians using patterns of motion and appearance, in: null, IEEE, 2003, p. 734.

[24] J. S. Rao, W. J. Potts, Visualizing bagged decision trees., in: KDD, 1997, pp. 243–246.