Offline Handwritten Signatures Classification Using Wavelet Packets and Level Similarity Based Scoring

Poornima G Patil^{#1}, Ravindra S Hegadi^{#2} ¹Department of Computer Science and Applications ²School Of Computational Sciences ¹Dayananda Sagar Institutions ²Solapur University ¹Shavige Malleswara Hills,Kumaraswamy Layout,Bangalore-78,India ²Solapur, Maharashtra-413255,India ¹poornima_g_patil@yahoo.com ²ravindrahegadi@rediffmail.com

Abstract—Offline Signature Classification has been extensively studied for many years. The challenge in this area is the correct classification of skilled forgeries which are the result of deliberate practice to imitate the signatures of any person. In this paper the preprocessed images of genuine handwritten signatures are subjected to analysis by Wavelet Packets. A regular wavelet like *db4* has been used to do the decomposition upto four levels. The resulting decomposed signal is further subjected to wavelet multiscale principal component analysis done for ten levels. The principal components are chosen according to the *kais* rule. The selected principal components consist of details at ten different levels and one approximation for each signature image. For a given test signature image the principal components are extracted in the same way and the principal components at each level are compared against the mean principal components of the genuine signatures at the corresponding level and the difference is within the permissible range, then a score is assigned. The collective score obtained due to all levels is used to classify the signature as genuine or forgery. The proposed system has a FAR of 12% and a FRR of 8%.

Keyword-Wavelet Packet, Principal Components, Details, Approximation, Score.

I. INTRODUCTION

Handwritten signatures have been used to authenticate a person since long. They are not only an accepted form of authentication in the society for every legal purpose but they are also a non invasive method. Handwritten Signature classification using computers is a really a challenging field because the signatures of the same person have variability. There can be an increase in the variability due to age, disease or emotional state of the person. Since the signature is a very small of information which adds to the complexity of the task.

Handwritten Signature classification systems are either offline or online. Offline system refers to the handwritten signatures usually scanned and stored as images in the computer system. Offline signature images do not contain any dynamic information like pressure, speed, velocity etc. Offline signature classification depends upon the static features of the signatures. Hence the classification accuracy is not high. The online system refers to the signature being captured on a tablet or a digitizing device which can record the dynamics of the signature during the act of signing. Hence can lead to higher classification accuracy. There are three types of forgeries.

A. Skilled Forgeries

The skilled forgeries are most difficult to handle because expert forgers practice it for some time and then they create them.

B. Casual Forgeries

The signer observes the signature for a while and then puts the signature in his/her own style without any knowledge of the spelling.

C. Random Forgeries

This is the crudest of all forgeries. The signer uses the name of the victim in his own style to create a simple forgery called as random forgery.

II. WAVELET TRANSFORM

A wavelet is a waveform which lasts for a limited duration and on an average its value is zero. A wavelet has a beginning and an end in contrast to the sinusoids which can extend form minus infinity to plus infinity. Wavelets facilitate the representation and analysis of signals at more than one resolution which is called as multiresolution ability. The advantage of multi resolution analysis is that the features which go undetected at one resolution may be easy to detect at another. Wavelets can analyse both stationary and non stationary signals. By stretching and shifting the wavelet, it can be made to correlate with any event which is of interest so that the frequency and time of the event can be exactly measured. When a signal is decomposed using the wavelet transform, both detail coefficients and approximation coefficients are obtained. When the wavelet is stretched, the longer portion of the signal is compared with it and they are low frequency components which are nothing but slowly varying parts of the signal. When a wavelet is shrunk, the smaller portion of the signal is being compared to it and they are high frequency components which are the rapidly changing parts of the signal. Both Continuous and Discrete Wavelet Transforms are possible.

A. Continuous Wavelet Transform

The scaled and shifted wavelet is multiplied with the signal and summed for the entire time of the signal. This transform is continuous in the sense that the signal is analyzed fully by the wavelet.

B. Discrete Wavelet Transform

Instead of analysing the signal at each scale and position here the analysis is done at dyadic scales and positions which are powers of two resulting in an accurate analysis.

C. Daubechies Wavelet Transform

Daubechies wavelets are having an order N and hence it is written as dbN where N stands for the order of the wavelet. It is a regular wavelet, orthogonal, has compact support. This wavelet supports both the continuous wavelet transform and discrete wavelet transforms.

III. WAVELET PACKET METHOD

The wavelet packet method differs from the wavelet transform in the fact that in a wavelet transform the decomposition at each level produces approximation and detail coefficients and only the approximation coefficients are decomposed at each subsequent level whereas in the wavelet packet method the details are also decomposed at each subsequent level resulting in richer analysis. The wavelet packets can be used for expanding a given signal in many ways and select the best decomposition based on an entropy measurement.



Fig. 1. Wavelet Packet Analysis of a Signal S

IV. RELATED WORK

There are lot of applications of Biometrics and one most important one is establishing the identity of an individual [1]. Handwritten Signatures have been used to identify an individual since long.Different methods have been used in handwritten signatures verification. Template matching is suitable for rigid matching to detect genuine signatures. However these methods are not very efficient in detecting skilled forgeries [7]. Neural networks have been used in signature verification and the recognition rates are high [5]. Neural networks are most commonly used classifiers for pattern recognition problems. They offer very promising results with extremely low FAR and FRR[7]. The geometry-based approach uses many features like calibration, proportion, guideline and base behaviours. In addition, other features have been applied in this approach, like pixel density, pixel distributions. However, static features do not describe adequately the handwriting motion. Therefore, it is not enough to detect skilled forgery [8]. The Hidden Markov Models used in signature verification have shown that the error rates for simple and random forgery are low and close to each other, but the type II error rate in skilled forgery signatures are high. However, though Structural techniques are suitable for detecting genuine signatures and targeted forged signatures they are exhaustive due to demand for large training sets and computational efforts [8]. Statistics based methods have been used which can give good classification results[6]. Methods based on the statistical approach are generally used to identify random and simple forgeries. The reason for this is that these methods have proven to be more suitable for describing characteristics related to the signature shape[2].

The survey of various methods used suggest that the highest accuracies can be achieved using graph matching [3] and the Discrete Wavelet Transform (DWT) [4]. Wavelet transforms are being extensively used in the domain of image processing. Wavelet transforms are linear in nature and are orthogonal. Orthogonality refers to the fact that each term is independent of the other which eliminates redundancy. Wavelets can perform multiresolution analysis. One of the most appropriate global features extraction techniques is wavelet transform, since it extracts time-frequency wavelet coefficients from the signature image [9]. Wavelet Transform is especially suitable for processing an off-line signature image where most details are hardly represented by functions, but could be matched by the various versions of the mother wavelet with various translations and dilations [10]. An offline signature verification system using texture features has reported good results [11].

V. PROPOSED SYSTEM

The proposed system is an offline signature classification system where the genuine signatures of 640 subjects have been decomposed using wavelet packets. The resulting coefficients are subjected to principal component analysis using multiscale wavelet principal component analysis for ten levels. The principal components selected at each level are compared against the principal components of the test signature image at the corresponding level. The difference out of comparison is checked against the minimum and maximum deviations in principal components existing for genuine signatures. If the difference is within the allowable range then a score is assigned for that level. The comparison is repeated for all the levels and a final score is obtained for each test signature image. The final score value will be on the higher side if the differences obtained for details at the higher level are within the allowable range. Thus the final score is used to classify the signature as genuine or forgery.



Fig. 2. Signature Images before Preprocessing

A. Preprocessing

The handwritten signature images are from a standard database called GPDS. There are signatures belonging to 640 persons. There are 24 genuine signatures and 30 forgeries for each person. All the signatures are put the bounding rectangles.

They are binarized and then are thinned. The size normalization is done for the genuine signatures only. For each person, the aspect ratio of each signature is calculated. Among the genuine signatures the one with the maximum width is identified and all the signatures are resized to the maximum width and the corresponding height maintaining the aspect ratio. The resizing of the images has been done using bicubic interpolation method.



Fig. 3. Signature Images after preprocessing

B. Feature Extraction

The preprocessed images are subjected to wavelet packet analysis using *db4* upto four levels. The coefficients generated are subjected to multiscale principal component analysis upto ten levels using *db4*. In order to reduce the dimensionality, the principal components are chosen according to *Kaiser's* rule. *Kaiser's* rule selects only the components associated with eigen values greater the mean of all eigen values. There are total eleven principal components of details for each of the ten levels and one approximation at tenth level. The principal components at first three levels are ignored by making their values zero because the lower levels usually contain noise. So denoising is achieved by this step. Excluding the first three levels the remaining eight levels are used to compute some statistical measures using the principal components.

At each level k, the minimum principal component and maximum principal component are obtained. The mean absolute deviation is calculated for each level k. The standard deviation and mean principal component values are calculated

For each level the coefficient of variation is calculated using the mean and standard deviation. Then the consistency at each

level k is calculated as

consistency(k)=1-coefficient of variation(k) - Eq(1).

Thus consistency is calculated for each level k, k=4 to 10 for details and also for approximation at tenth level.

Then the minimum deviation of principal component value with respect to the mean principal component value is calculated for each level k. Similarly the maximum deviation value with respect to the mean principal value is computed for each level k. These are computed to understand the allowable range for the principal component values at each level k against which a test signature image can be compared.

The novel idea used in this paper is the comparison of principal component values of the test signature image with the principal component values of the representative training image for each level. Since the details at higher levels have more discriminating power compared to the details at lower levels, different weights are assigned for the similarity between test signature and trained signature at different levels. The weights assigned for the levels from 4 to 11 have been normalized within the range 0 to 1. The weights for the details at each level are higher than the weights at the previous level and the weight given to details at each level is higher than weight assigned to the approximation. The weights for each detail and approximation were determined with a trial and error approach by deliberately increasing gradually the weight from lower detail level to higher detail level and assigning the lowest weight to approximation so that the sum of all the weights put together is equal to 1. The weights determined for each level are shown in the table I. Intuitively, even the consistency calculated at each level *k* justifies these weights.

	Weights for		
Level	Detail	Approximation	
4	0.04		
5	0.06		
6	0.08		
7	0.1		
8	0.14		
9	0.22		
10	0.34	0.02	

TABLE I Weights Assigned for Details and Approximation

C. Classification

Level based scoring forms the basis for classification. The score is computed by summing the weights at each level if the test and the representative training signature image are similar.

Similarity is tested with the following steps.

- 1. A variable score is initialized to zero.
- 2. At level k, the principal component of the test signature is checked as to whether it lies between the minimum and maximum principal component value of the representative training signature at that level.
- 3. If the value lies in the limits as said in the step 1, then it is compared against the mean principal component value and the deviation from the mean principal value is calculated. If the deviation lies between the minimum and maximum deviation of principal component values or in other words, if it lies in the allowable range then weight for the level k as shown in the table I is added to the score variable.
- 4. Steps 2-5 are repeated for each level and a final score is obtained.
- 5. The final score thus obtained indicates that if the score is high, then similarity is high.
- 6. The final score is compared against the coefficient of variation calculated previously. If the difference is large, it suggests two things.
 - The test signature closely resembles the genuine signature because the score is high
 - Small value of coefficient of variation further supports classification of test signature as genuine.
- 7. On the other hand, if the difference calculated as done in step 7 is small, it supports the classification of test signature as forgery but this needs to be further ascertained.
- 8. But to classify the signature is as genuine or forgery, we have to determine what is called as threshold. For this we depended upon the sum of weights in last four details i.e from levels 7 10. If similarity is

found in these levels then definitely the final score will be greater or equal to 0.8 which can be verified by looking at the values in Table I.

9. So, if the final score is greater than or equal to 0.8 then the signature is classified as genuine, otherwise forgery.

The threshold selected depends upon the higher level details since they have higher discriminating powers and also the wavelet packet analysis which has been used to decompose the signature images provides richer analysis compared to wavelet transforms. The features which go undetected in lower levels generally are detected in the higher levels which makes the multiscale analysis of a signal highly desirable.

VI. RESULTS

The proposed system has been developed in Matlab software. The results are tabulated in the Table II. There are signatures of 640 persons in the database consisting of both genuine signatures and forgeries. There are total 24 genuine signatures and 30 forgeries for each person. Therefore the total number of signatures in the database is 34560. All the genuine signatures have been preprocessed. They are thinned to hide the effect of using different pens and size is normalized using bicubic interpolation method as said in the preprocessing stage. But only five genuine signatures of each person have been used for feature extraction. The training signature images have been decomposed using wavelet packet method upto four levels. Then principal component values are obtained for the features at ten levels which make the final feature set. Then the various statistical measures like mean, maximum, minimum values of principal components are calculated as explained in the feature extraction section. Similarity found at higher detail levels have been given higher weightage and the score for the similarity is calculated. The threshold for classification is determined based on the level based score. Only genuine signatures have been used for training. The forgery signatures have not been used for training. The nineteen genuine signatures excluding five which are used for feature extraction become the test signatures. All the thirty forgery signatures are used for testing. The results show a FAR of 0.12 and a FRR of 0.08. Results have been tabulated in Table II. The results are quite encouraging. The proposed system has been trained and tested on a large database consisting signatures of 640 persons. The wavelet packet method used for feature extraction holds lot of promise for related research interests.

Results	Obtained for GP	DS Signature D	atabase		
Number of persons:640					
Number of Genuine signatures per person:24 Number of Forgeries per person:30					
Number of genuine signatures used for training:5					
Number of forgeries used for training :nil					
Number of	Number of	Number	Number		
genuine	forgeries	of genuine	of		
signatures	used for	signatures	forgeries		
used for	testing	classified	classified		
testing		correctly	correctly		
12,160	19,200	11,187	16896		
(19*640)	(30*640)				
FAR	0.12%				
FRR	0.08%				

TABLE II Results Obtained for GPDS Signature Database

VII.CONCLUSION

The proposed system is based on a novel idea of level by level comparison of the features of training signature and the test signature and scoring based on similarity. Similarity at each different level is assigned a different weight based on the discriminating power of each level. Consistency of each genuine signature has been measured at each level and an average consistency measure has been calculated for each genuine signature. The weights to be assigned can be further investigated per user based on the individual's features instead of using a common set of weights. Genetic algorithms can help in finding the optimal weights. The proposed system has achieved good classification results using the simple statistical measures.

ACKNOWLEDGEMENT

I sincerely thank Dr.Ravindra S Hegadi, Associate Professor, School of Computational Sciences, Solapur University, Solapur, Maharashtra, India for his constant support and guidance for my all my research endeavours.

REFERENCES

- Jain, A. Ross, and S. Prabhakar, "An Introduction to Biometric recognition," Circuits and Systems for Video Technology, IEEE Transactions on, vol. 14, no. 1, pp. 4–20, 2004.
- [2] Jain, F. Griess, and S. Connell, "On-line signature verification," Pattern Recognition, vol. 35, no. 12, pp. 2963–2972, 2002.
- [3] Chen and S. Srihari, "A New Off-line Signature Verification Method based on Graph Matching," in Proc. 18th International Conference on Pattern Recognition (ICPR'06), 2006, Volume 02, pp. 869–872.
- [4] W. Tian, Y. Qiao, and Z. Ma, "A New Scheme for Off-line Signature Verification Using DWT and Fuzzy Net," Eighth ACIS International Conference, 2007, vol. 3, no. 2, pp. 30–35.
- [5] S. Srihari, A. Xu, and M. Kalera, "Learning strategies and classification methods for offline signature verification," in Proc. 7th Int. Workshop on Frontiers in handwriting recognition (IWHR),2004,pp.161–166,.
- [6] Dimauro, G., Impedovo, S., Modugno, R., Pirlo, G., Sarcinella, L, "Analysis of stability in hand-written signatures," in.Proc. Internat. Workshop on Frontiers in Handwriting Recognit. (IWFHR), 2002, 259–263.
- [7] Meenakshi S Arya, Vandana S Inamdar, "A Preliminary Study on Various Off-line Hand Written Signature Verification Approaches," International Journal of Computer Applications (0975 – 8887) Volume 1 – No. 9, 2010.
- [8] Neeraj Shukla, Dr. Madhu Shandilya, "Invariant Features Comparison in Hidden Markov Model and SIFT for Offline Handwritten Signature Database, "International Journal of Computer Applications (0975 – 8887) Volume 2 – No.7, June 2010.
- [9] V. Nalwa, "Automatic on-line signature verification," Lecture Notes In Computer Science, in.Proc, Third Asian Conference on Computer Vision, 1998, p.p 10 - 15.
- [10] Sing-Tze Bow,"Pattern recognition and image preprocessing", Marcel Dekker, Inc, chapter 15, 2002.
- [11] J.F.Vargas, M.A.Ferrer, C.M.Travieso, J.B.Alonso,"Off-line signature verification based on grey level information using text features," Pattern Recognition, ISSN:0031- 3203,vol 44, no.2,pp.375--385, 2011.