AN INTEGRATED APPROACH FOR OFFLINE SIGNATURE CLASSIFICATION USING ANN

Rahul Singh[1], Akhilesh Pandey[2]

[1]M.Tech Scholars, [2]Asst. Professor

Department of Computer Science and Engineering, Suresh Gyan Vihar University, Jaipur (India)

ABSTRACT

A great deal of work has been done in the area of off-line signature verification over the last two decades. Off-line systems are of interest in scenarios where only hard copies of signatures are available, especially where a large number of documents need to be authenticated. This paper is inspired by, amongst other things, the potential financial benefits that the automatic clearing of cheques will have for the banking industry. The purpose of this research is to develop a novel, accurate and efficient off-line signature verification system. In this proposed system we are collecting the sample from 480 people genuine and fake signature and for feature extraction use the DCT and IDCT technique. After getting the feature for training and testing we the neural network.

Keywords: DCT, IDCT, Neural Network, Feed Forward Network

I. INTRODUCTION

The main aim of research is to prepare a system which will automatically classify handwritten signature images whether it is authentic or fraudulent. For centuries, handwritten signatures have commercial transactions, contracts and agreements as an integral part. Handwritten signatures is unique, helps prove the identity of the signer, and acts to sign documents on behalf of the signer's acceptance of its terms, the content and the compilation of the documents is official, complete at the time it was signed. The image of the signature is representative of the personal style of people's handwriting, widely described by graphometry. Simultaneously, there should be sensible processing requirements that would make the adoption of such an automated system feasible.

This dissertation is inspired by the latent qualities that initiate the automatic cheque clearance and which is beneficial for the banking industry. Millions of bank cheques are processed daily in India. Generally, only those cheques are verified manually by an operator that exceeds some amount of threshold. Consequently, an automated system's performance is of great significance when compared to a human being. A system that is capable of screening casual forgeries has already been proved advantageous.

During one of the last century, the authentication on the help of signature has been greatly increased. One wants to know who signed a contract, or to check whether or not the real crime. This provides an easy to use application that can detect digital signature. Hand written signature authentication is one of the personal qualities of the most widely accepted. Since the consent and authorization of symbolism, especially in the field of credit cards and bank cheques, fraud has been the goal of a hand written signature. Therefore, the demand for personal identification and treatment is growing faster, more accurately, an automatic signature of a real system design challenge faced [1]. Signature verification is different than the character recognition, because signatures are often unable to read, and it is only an image of the person writing style with representatives of certain curves. Hand written signature is a special case, just a symbol. Therefore, it is wise and necessary signatures and pixel processing a complete image of the special distribution, on behalf of a particular writing style, rather than a collection of letters and words [2]. Suitable for off-line signature verification algorithm is a typical feature extraction method. In this method, it extracts a feature from the original signature and to make adjustments. The extracted feature is used as original and the forgery to be distinguished. In spite of large number of increasing electronic possibilities to paper cheques, fraud actions are being carried out at financial institutions in the India and it has become an international widespread undesirable phenomenon. In the year 2008 and 2014, Cheque Fraud Survey Reports, Indian Banker's Association pointed out that Paper based cheque fraud cost the financial institutions approximately $500 million annually. In cheque fraudulent, significant components were forged maker's signature and imitation of cheques. In 2020, nearly $2.2 billion worth of cheque fraud attempted against financial institutions. Several Indian banking institutions do not make a vigorous attempt to verify clients' signatures on cheques. Once when the cheque book of a client from the leading bank was stolen, the culprit went on a spending spree. At the early stage, forger went through a lot of difficulties to copy the clients' signature. As a result, very high quality forgeries were produced. These types of signatures are generally referred as skilled forgeries. In most cases, these are difficult to detect even by experts. Gradually, the forger signs the cheque for higher amounts and there is very less trouble in forging the client's signature. Possibly, some of the blank cheques have been sold by the forger. These forgeries do not resemble the client's signature or the client's name, these type are referred as random forgeries. Sometimes random forgeries can be easily detected but still all of these cheques are accepted by the banking institutions.

1.1 Recognition and Verification

There should be a clear distinction made between verification systems and recognition systems. A verification system is a system that just decides whether a specified entity belongs to a specific class or not. Whereas a recognition system is a system that has to decide in which a certain number of classes the entity belongs to.

Since this dissertation focuses on off-line signature verification, so here often classifiers and classes are mentioned. Although classifiers and classes are different, a verifier as a classifier may be interpreted with two classes. For an instance, a glass of genuine signatures i.e. positive class and a class of forgeries i.e. negative class. It can be seen that generally samples of forged signatures are inevitable. Signature recognition system differs a lot from signature verification systems.

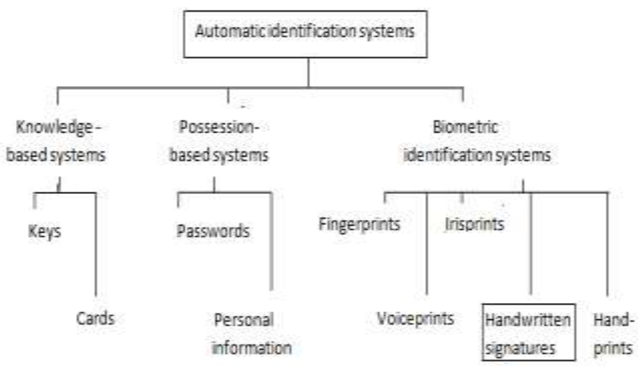


Figure 1: Automatic Identification System

1.2 Statistical Pattern Recognition

In this type of approach, each entity is usually represented by a sequence of P, d-dimensional feature vectors, which is an observation sequence. Each entity can also be denoted by a single feature vector, i.e. P=1. A measurement is represented by a component of a feature vector. A pattern can therefore be viewed as a point in a d x P dimensional feature space. Aim is to select those features that allow a sequence of observation. It belongs to different classes and it covers compact and disjoint regions in the feature space. The potency of the representation space is to decide by observation sequences from which various classes can be varied.

Statistical pattern recognition systems consists minimum distance classifiers, neural networks and Bayesian classifiers. When considering the case of minimum distance classifiers, each class is represented by a Gaussian probability density function (PDF). The height of the PDF for which it is greatest, the observation sequence is assigned to that or vice-versa. Two types of minimum classifiers are appropriate for this work. They are simple distance classifiers (SDCs), for which P=1 and HMMs for which P>1.

Simple distance classifiers: In this case, each pattern is represented by a single observation; it means a single d-dimensional feature vector. Each pattern class is identified by a Gaussian PDF in a d-dimensional feature space. In this feature space, each PDF is special and is defined by the mean vector and covariance matrix which belongs to a particular class.

Data scarcity: Although when there is only one sample available, the mean vector for each class can be calculated. Still, a full covariance matrix can only be reliably estimated when the amount of samples available is much greater than the dimension of the feature vectors. When this criterion is not fulfilled, then one can assume that the off-diagonal elements of each covariance matrix are equal to 0, in such type of cases only diagonal entries are estimated. So here the assumptions made by the features cannot be correlated. It means that the data distribution is modelled less accurately. It means there are no requirements made for the possible data rotation. In the extreme case, when the quantity of available samples is much less than the dimension of the feature vectors, the covariance matrix may not be estimated at all. It means that each covariance matrix has to be kept equal to the identity matrix. This is done because each class is modelled by only one mean vector for that class. Therefore distribution is data thus here is ignored.

Distance measures: Classification is based on the two type of distance. One is Mahalanobis distance and other is Euclidean distance. Mahalanobis distance is measured when the full covariance matrix is estimated for each class. And, when only the mean vector is estimated for each class, division is based on the Euclidean distance.

Dynamic time warping: This includes a wide area of dynamic programming algorithms tht are used to non-linearly align feature vectors. The aligned feature factors then are compared to non-linearly aligned features.

Hidden Markov models: HMMs are used to design a series of observations and their relationship to each other. These are extremely used in speech recognition systems and also in on-line signature verification systems. HMMs are best suited for the above applications; this is because they are time independent. But an off-line signature contains no such information, and in this type extraction of time – dependent observations is of less importance.

HMM consists of X states, and each state has two principal elements: PDF and a histogram. A PDF describes the nature of a group of observations which is related with the state and a histogram describes the probability of creating a transition to any of the X states- thus it creates a probability of remaining in the same state. In the applications where classifications are required an HMM is designed for each pattern class. First each and every HMMs are initialised and then it is being trained with a set of training patterns. When the classification of observation sequence is to be test patterned, assumptions are made that the observation sequence is produced by one of these HMMs. In order to decide which one of the HMM, each of the observation sequence is computed.

In systems where automatic identification is to be done requires verification, for doing this an HMM is trained for each client. The pattern is accepted when the likelihood is greater than a certain threshold otherwise it is rejected.

II. PROBLEM STATEMENT

We described DTW and HMM based system is the development of this article extracted from the original signature image feature vector sequences. The feature extraction technique, these two systems are very similar, and differ only in each observation sequence alignment of the DTW system extracted with a sense and a reference sequence, to ensure rotation invariance. This alignment is not required for HMM based systems. For a system using the DTW, the reference sequence, as each template corresponds author's signature.

In this paper, the development of feature extraction techniques for system DRT based calculation is done. Image processing is also necessary to remove speckle noise and normalize each signature that is to ensure that the observed order is the corresponding translation, rotation and scale invariant representation signature image. It requires a lot of floating-point calculations from DRT operating in feature extraction process.

2.1 Feature Extraction

In this development based on DTW and HMM based system using the same feature extraction algorithm to obtain the initial observation sequence from the original signature images. For HMM based system, each writer's signature is represented by the HMM model. These Hidden Markov models are in such a manner that each initial observation sequences have rotation invariance constituting said corresponding signature. No further treatment is required.

Daejeon- based system, while on the other hand, representatives of each writer's signature with single observation sequence, which serves as a template. Each initial observation sequence with the author of the reference sequence, in order to ensure the alignment rotation order is the corresponding signature image. Each order constitutes a final observation sequence.

We captured signed as Stellenbosch data set a standard flatbed scanner. Each of these signatures produced within the specified bounding box of paper blanks pages. However, the signature Dolfing data set was originally captured on the line. Then each signature image is transformed into a dynamic data, using only space writing coordinates. Therefore, one believes that this section of the binary image has been acquired for each signature.

On an average, a signature has a width of 400 to 600 pixels and a height of 200 to 400 pixels. The dimensions of the images are not normalised and median filtering is applied to remove speckle noise. Each column of the DRT represents a projection or shadow of the signatures at a certain angle. Once these projections are processed and normalised, they represent an initial set of feature vectors for the signature.

2.2 DTW-Based System: Observation Sequence Alignment

In order to achieve rotational invariance when each initial observation sequence is compared with a reference sequence aligned correctly writer's rights, they are pre-compared. This alignment is not necessary for HMM-based systems. Both methods can be used to achieve this alignment. Linear approach is straight aligned by moving all of the appropriate number of initial observations of the observation sequence to the left or right times. Alternatively, the non-linear approach can be used in the sequence alignment as observed nonlinear methods.

Orientation is achieved in two steps. In first step, a reference observation sequence obtained for each writer. In the second step, the initial observation order for each input signature could be a training or test signature. It can be better understood below in following steps:

Step 1: Get a Reference Sequence

For every writer, a reference sequence is chose from a writer's training set preliminary observation sequence. This sequence can be randomly selected or most representatively training sequence may be used. One of the way to obtain the most representative of each training sequence is relatively all other sequences of the training sequence. This differs from the minimum order of other training sequence is selected.

Two training sequences can be one of the following comparisons. When one assumes the train signature is normalised with respect to the rotation of more or less change-this can be a realistic assumption, because usually all the training signatures for the respective views- an admissions obtained during conference signature is compared DTW alignment feature vector. Thus, observation of the first non-aligned is a simple distance average of the corresponding distances between the two signatures observation. When one assumes that the training did not return a signature rotation with respect to changes in the respective ratio of the first and the observed sequence, then by comparing the matching of the alignment of the classes. This type of alignment is also used when a test sequence matches the reference sequence.

Step 2: Observation Sequence Alignment

Linear method: the linear sequence alignment preliminary observation of entering the signature and the reference sequence. One way to achieve this is to first compare DTW characters using the average value of these observations is the distance between them. Each input observer is then shifted one position to the right. The first replacing the second observation, then second observation by the third replacement and likewise, from the cyclical nature of the DRT when calculated through 360^o, so that we can replace the first observation and the last observation one. Now comparing mobile observation sequence with the most representative sequence and the average distance is DTW. The process is repeated until the initial observation sequence is shifted a total of N times, where N is the length of each observation sequence.

When the shift amount is the best, it leads to the alignment for which the distance between the observed sequences is minimal. The distance between these best sequences from the characteristic ratio expressed as discussed further. When way to align the two observations sequence linearly, $4N_{\Phi}^{2}$DTW distance must be calculated.

III. IMPLEMENTATION

When a system aims to detect only random forgeries, subsets of other writers' training sets can be used to model "typical" forgeries. Since the systems developed in this dissertation aim to detect only skilled and casual forgeries, and since models for these forgeries are generally unobtainable, we are not able to utilize any of the above-mentioned impostor validation techniques. We also do not use any subset of genuine signatures for validation purposes. When a claim is made that a test pattern was produced by a specific writer, the systems first match this test pattern with a trained model of the writer's signature. Statistics for the claimed writer's signature are used to obtain a threshold value. The dissimilarity between the test pattern and the model is then calculated. When this dissimilarity value is greater than the threshold, the test pattern is rejected as fraudulent; otherwise it is accepted as authentic.

For this work we have collected the data from the different writer. They make the two type of signature:

1. Original Signature
2. Fake Signature

Figure 2: Data Collection from Different Writer

3.1 Proposed Algorithm

Step 1. Read Image

Step 2. Apply DWT transform and get all 4 energy components

Step 3. Apply IDWT using Vertical, Horizontal and Diagonal component to get the image

Step 4. Apply Radon transform from 0 to 179 degree on the image.

Step 5. Take max value at each angle.

Step 6. Hence a feature vector of length 180 is obtained.

Step 7. Use Classification technique

a) For classification KNN, Feed Forward Neural Network, and Fitting Neural Network and LDA is used.

All paragraphs must be indented as well as justified, i.e. both left-justified and right-justified.

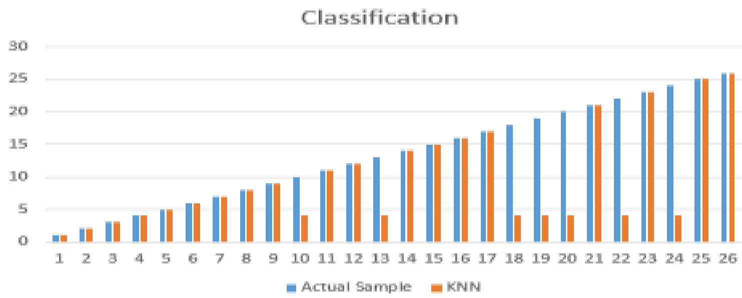
Figure 3: Cropped Image of Genuine Signature of different Writer

Figure 4: Cropped Image of Fake Signature of different Writer

IV. RESULT

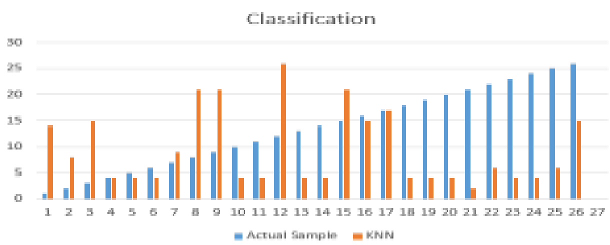
4.1 Real Signatures

Initially results of KNN are considered, In case of confusion it is matched with FITNET. 8 samples were misclassified by KNN which were correctly classified by FITNET. Hence, decreasing the False Negative ratio to 5/26. Hence we obtain an overall accuracy of 85% given the limited number of samples in test.



4.2 Fake Signatures

The total False Positive Ratio is 2/26 giving a total of 92% recognition rate.



In the figure above fake sample number 4 and 17 are identified as similar to their real sample. Since all the samples are fake sample they should have been classified as a different sample, rather than their original ones. Hence a false positive ratio is observed.

V. CONCLUSION

During testing same algorithm was used to get results of KNN, FITNET and feedforwardnet. I found that since training data and test data is too less neural networks do not work well (since it needs really good training data around 1000+ for each 100 test data). feedforward network has lot of false negative calls, so instead of Majority voting, i used a new algorithm. Which is takes in to account the dependability of classifiers.

REFERENCES

[1] Ismail, M.Hajjar and M.Quafafou, “Off-line signature verification uses single and parallel processing”, in Proc. 2009 Intl Conf: Sciences of Electronic, Technologies of Information and Telecommunications, 2009, pp. 1-6.

[2] S.Biswas, T-H Kim and D. Bhattacharyya, “Features extraction and verification of signature image using clustering technique”, International Journal of Smart Home, vol.4, no.3, pp. 43-55, 2010.

[3] Fang, B., Leung, C.H., Tang, Y.Y., Kwok, P.C.K., Tse, K.W., and Wong, Y.K. (2002). Off-Line Signature Verification with Generated Training Samples. IEE Proceedings -Vision, Image and Signal Processing, vol. 149, no. 2, pp. 85 - 90.

[4] Fang, B., Leung, C.H., Tang, Y.Y., Tse, K.W., Kwok, P.C.K., and Wong, Y.K. (2003). Off-line signature verification by tracking of feature and stroke positions. Pattern Recognition, vol. 36, pp. 91-101.

[5] Huang, K. and Yan, H. (2002). Off-Line signature verification using structural feature correspondence. Pattern Recognition, vol. 35, pp. 2467-2477.

[6] Xiao, X. and Leedham, G. (2002). Signature verification using a modified Bayesian network. Pattern Recognition, vol. 35, no. 5, pp. 983-995.