Offline Signature Verification based on Euclidean distance using Support Vector Machine

Isha[1], Pooja[2], Varsha[3]

[1]Research Scholar, Department of CSE/IT, CTIEMT Campus, Shahpur, Punjab, India
[2,3]Assistant Professor, Department of CSE/IT, CTIEMT Campus, Shahpur, Punjab, India

***Abstract*— *In this project, a support vector machine is developed for identity verification of offline signature based on the matrices derived through Euclidean distance. A set of signature samples are collected from 35 different people. Each person gives his 15 different copies of signature and then these signature samples are scanned to have softcopy of them to train SVM. These scanned signature images are then subjected to a number of image enhancement operations like binarization, complementation, filtering, thinning, edge detection and rotation. On the basis of 15 original signature copies from each individual, Euclidean distance is calculated. And every tested image is compared with the range of Euclidean distance. The values from the ED are fed to the support vector machine which draws a hyper plane and classifies the signature into original or forged based on a particular feature value.*

Keywords*— *Offline signature verification, Support Vectors, SVM, Euclidean Distance, SMO, EDH, Kernel Perceptron, Large-Margin-Hyper plane.

I. INTRODUCTION

BIOMETRICS plays important role in personal identification and authentication systems. Several traits developed in this area are based on fingerprints, face, iris, palm, voice, the handwritten signature, hand, etc. Handwritten signatures have a very special place in this wide set of biometric technologies. The main reason is tradition: handwritten signatures have long been established as the most common means of personal verification. Signatures are generally accepted everywhere, by governments and financial institutions as a legal means of verifying identity. Moreover, verification by signature analysis requires no extra measurements and people are used to this thing in their day to day activities. Automatic signature recognition has many real time applications like credit card validation, security systems, cheques, contracts, property papers etc. There are two types of systems: signature identification systems and signature verification systems. A signature verification system only decides whether a given signature belongs to an authorized person or not. A signature identification system, on the contrary, has to decide a given signature belongs to which one of a certain number of persons. There can be two modes of signature acquisition: online and offline. Online signature records the motion of the pen while the signature is written, and it includes velocity, location, acceleration, and pen pressure, as functions of time. Online systems use this information captured during acquisition. These dynamic characteristics are specific to every individual and sufficiently stable and repetitive. Off-line data is 2-D image of the signature therefore processing Off-line signature is complex due to the absence of stable dynamic characteristics. It is hard to segment signature strokes due to highly unconventional and stylish writing variations. In signature verification systems, generally two common classes of forgeries are considered: skilled and casual. A casual forgery is produced by only knowing the name of the person, and without experience of the genuine signature. When forger uses his own signature or try genuine signature of another person as a casual forgery, it is called a substitution forgery. So, style differences are common in casual forgeries. In skilled forgeries, the forger has experience of genuine signature. Since skilled forgeries are very similar to genuine signatures, some appropriate features for detection of casual forgeries are ineffective in detection of skilled forgeries.

The system proposed can be divided into two major parts: training phase and testing phase. The block diagram of the proposed system is as shown in the figure 1.

The proposed system can be measured by applying following metrices:

1. PSNR (Peak signal- to- noise ratio) - PSNR is the peak signal-to-noise ratio measured in decibels (dB). The PSNR is only useful for data encoded in terms of bits per pixel, or bits per sample. For example, an image with 8 bits per pixel contains integer value from 0 to 255.
2. MSE: The mean square error (MSE) is the squared norm of the difference between the data and the approximation divided by the number of elements.
3. MAXERR: Maximum squared error- MAXERR is the maximum absolute squared deviation of the data, from the approximation.
4. L2RAT: Ratio of squared norms- L2RAT is the ratio of the squared norm of the image or signal

approximation, to the input signal or image.

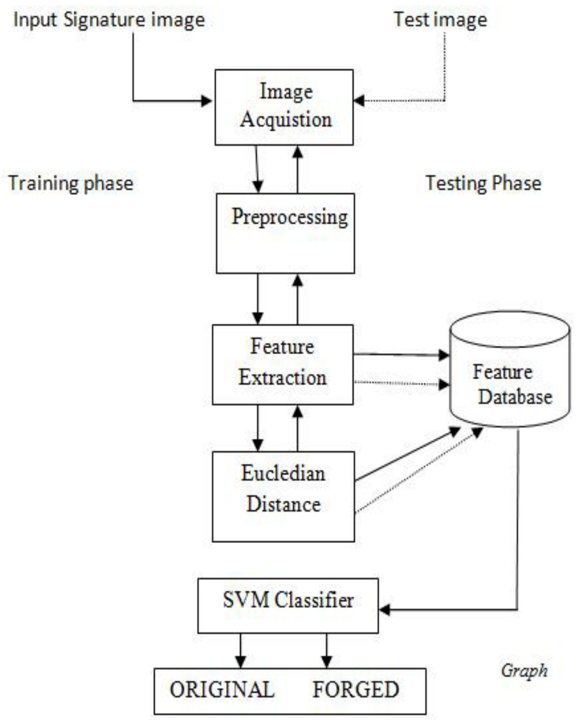

Fig.1: Block diagram of the proposed system.

II. SIGNATURE ACQUISITION

Signatures of an individual are obtained by using ball, gel or sketch pens on plane sheet. These sheets are then scanned using a scanner and the image is saved in jpg format. Each signature was then tight cropped manually using Microsoft Office Picture Manager.

III. PREPROCESSING

The scanned images might contain certain kind of noise, or might be subjected to damage because it"s an old signature or the signatures under consideration may vary in thickness and size. Hence some kinds of image enhancement operations are performed to remove noise and to make signature fit for further operations to be applied.

A. Preprocessing stages of a signature sample

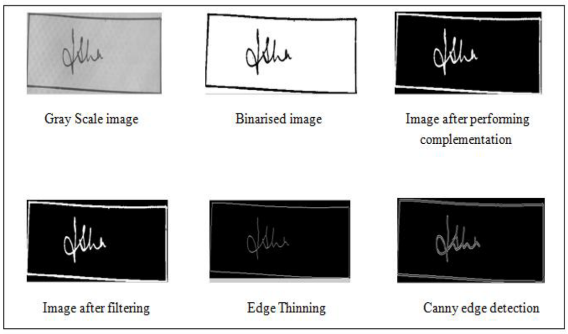

Fig.2: Preprocessing Steps

B. Grayscale Conversion

As the scanned images are stored in database as a color image, a three dimensional image (MXNX3) is not suitable for further processing, so it must be converted into a grayscale image to represent each pixel in the range from 0 to 255.

C. Binarization

It is done by extracting lightness (brightness, density), each pixel in an image is converted into one bit either '1' or '0' depending upon the mean value of all the pixel. If greater than mean value then its '1' otherwise its '0'.

D. Complementation

Complementation is done as the natural tendency to have data in form of 1s. In the complement of a binary image, 0 is changed to 1 and 1 ischanged to 0; white and black are reversed. In the output image, dark areas become light and light areas become dark.

E. Filtering

Filtering is done in order to remove the noise that might be introduced during the scanning process. The median filter is nonlinear and is normally used to remove noise. It considers each pixel in the image and then looks at its nearby neighbors to decide whether or not it is representative of its surroundings and then replaces it with the median of neighbor pixels.

F. Edge thinning

Edge thinning is used to remove the outcast extra points on the edge of the signature image. The edge operator has been applied (like sobel, canny) to detect the edges and the edges are smoothed using an appropriate threshold.

G. Canny Edge detection

Edge is a boundary between two homogenous surfaces. Applying an edge detection algorithm (either using laplacian or gradient method) to an image which shows acute variation in brightness or, which has layoffs, significantly reduces the amount of data that is to be processed and thus filters out the non structural, less relevant information of the image and preserves only the important structural properties.

The figure 2 shows the preprocessing stages of a signature sample

IV. FEATURE EXTRACTION

The feature extraction method is most important step in any recognition system, because the recognition accuracy totally depends on the features extracted. The main objective of a feature extraction technique is to accurately retrieve the features. The extracted features such as angle of rotation, stop watch timer and difference are calculated through Euclidean distance. Stop watch timer is used to calculate the training time of signature, testing time taken by svm and computation time by the project. If the tested image is tilted then it is rotated at some angle to match the training image and numbers of angles are calculated.

A. Classification using Euclidean Distance

The Euclidean distance or Euclidean metric is the "ordinary" distance between two points that can be measured with a ruler, and it is given by the Pythagorean formula. By using this formula as distance, Euclidean space (or even inner product space) becomes a metric space. The associated norm is called the Euclidean norm. The Euclidean distance between point's p and q is the length of the line segment connecting them. In Cartesian coordinates, if p = (*p*1, *p*2...*pn*) and q = (*q*1, *q*2... *qn*) are two points in Euclidean *n*-space, thenthe distance from p to q, or from q to p is given by equation 1. Using Euclidean distance metric, similarity score between any two feature set can be obtained in terms of the extracted features.

$$\mathrm{d}(\mathbf{p},\mathbf{q}) = \mathrm{d}(\mathbf{q},\mathbf{p}) = \sqrt{(q_1-p_1)^2 + (q_2-p_2)^2 + \cdots + (q_n-p_n)^2} = \sqrt{\sum_{i=1}^{n}(q_i-p_i)^2}. \quad \text{-- (1)}$$

The distance is used as the matching criterion, i.e. a signature is matched if this distance lies in a range of subjective threshold. However, using the Euclidean distance we generate matching value by matching a test signature with all the trained database of signatures. Classifying the images using Euclidean distance is quite easier and convenient task for a beginner.

B. Developing SVM

A support vector machine (SVM) is a machine learning task to deduce a function called classifier, from supervised training data. They are a specific class of algorithms which are classified by usage of kernels and optimize it with an algorithm that is very fast in thelinear case, acting on the margin on number of support vectors. Asupport vector provides several computational benefits to classify to present the solution by providing simple hypothesis using extracted test points. SVM contains some main features like maximum margin classifier: a decision strategy which separates the training data with the maximal margin and a nonlinear function that controls the input parameters to find a linear separating hyper plane which do not depend upon high dimensional feature space. This type of classification approach depends on certain activation values. Areview was done and SVM is supposed to be better classifier to verify a signature. It gives more accurate results than other classifying techniques.

C. Kernel functions

Kernel function svm train uses to map the training data into kernel space. The default kernel function is the dot product. The kernel function uses following function handle:

'Linear' — Linear kernel, meaning dot product.

Method used: „SMO" (Sequential minimal optimization) --- because memory consumption is controlled by the kernelcachelimit option. The SMO algorithm stores only a sub matrix of the kernel matrix, limited by the size specified by the kernelcachelimit option. However, if the number of data points exceeds the size specified by the kernelcachelimit option, the SMO algorithm slows down because it has to recalculate the kernel matrix elements. SMO is a simple algorithm that quickly solves the SVM QP (Quadratic Programming) problem without invoking an iterative numerical method for each sub-problem and without any extra matrix storage. SMO decomposes the overall QP problem into QP sub-problems. SMO chooses to solve the smallest possible optimization problem at every step.

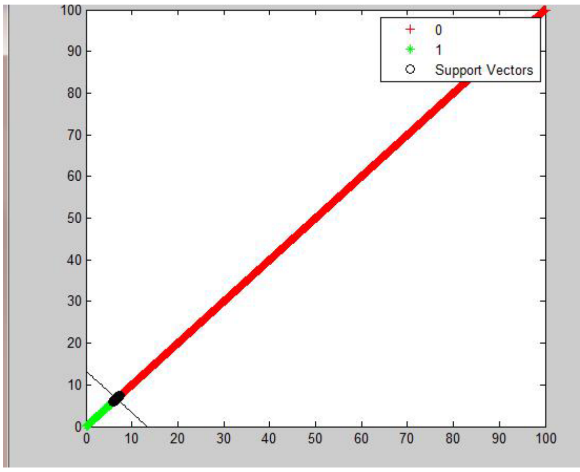

Fig.3: Graph Showing Training Time Taken by SVM using SMO Algorithm

V. RESULT ANALYSIS

The support vector machine was successfully developed with SMO algorithm. The values obtained for metrices (PSNR, MSE, MAXERR, L2RAT) in the process are tabulated as shown below.

TABLE.1: Metrices values using SMO algorithm.

Images	PSNR	MSE	MAXERR	L2RAT
1	20	580	190	0.9
2	23	270	175	1.1
3	22	380	160	0.9
4	21	460	178	0.94

TABLE 2: Time Taken by Process using SMO algorithm.

Images	Training Time	Testing Time	Computation Time
1	26.7240	0.5202	0.5
2	30.3185	0.5466	0.5
3	36.9084	0.9288	0.92
4	36.1370	0.4881	0.48

In the above table training time, testing time and computation time by SVM using SMO algorithm is tabulated.

VI. CONCLUSION AND FUTURE WORK

In this project we are trying to verify whether a signature sample is forged or not using support vector machine. We have acquired the signature samples of different persons, pre-processed them using techniques like gray scale conversion, binarization, complementation, thinning, filtering and edge detection using canny edge detector. Further from these pre-processed signatures features such as Euclidean distance, angle of rotation and time of operation are extracted. These feature set are separately passed through the support vector machine developed using SMO algorithm which are tested against both linear and polynomial kernel. This project actually differentiate that signature is forged or original.

In future work it would be extended to verify a signature on the basis of false acceptance rate (FAR) and false rejection rate (FRR). A new feature extraction algorithm will be imposed which tell at what difference the signature is said to be forged. This is done so, as there may be variation in two original signatures in this case we are to calculate the percentage to accept the signature or reject it.

REFERENCES

[1] Kruthi.C, Deepika.C.Shet, "Offline Signature Verification Using Support Vector Machine" Fifth International Conference on Signals and Image Processing, IEEE (2014) DOI10.1109/ICSIP.2014.5, pp. 1-7

[2] S. A. Angadi, Smita Gour, Gayatri Bajantri "Offline Signature Recognition System Using Radon Transform", Fifth International Conference on Signals and Image Processing, IEEE (2014) DOI 10.1109/ICSIP.2014.13, pp. 56-61

[3] S. A. Angadi, Smita Gour "Euclidean Distance Based Offline Signature Recognition System Using Global and Local WaveletFeatures" Fifth International Conference on Signals and Image Processing, IEEE(2014), DOI 10.1109/ICSIP.2014.19 pp. 87-91

[4] Ibrahim S. I. ABUHAIBA "Offline Signature Verification Using Graph Matching", Turk J Elec Engin, VOL.15, NO.1 2007, c TU¨BI˙TAK, pp. 89-104

[5] Prof. Sachin Patel (et.al) "Comparative Analysis of Different Approaches for Offline Signature Verification "International Journal Of Advanced Electronics & Communication Systems Approved by CSIR NISCAIR ISSN NO: 2277-7318, ISSUE 1 VOL 3, JAN-FEB, 2014 PAPER ID SM-1134.

[6] Vahid Kiani, Reza Pourreza & Hamid Reza Pourreza "OfflineSignature Verification Using Local Radon Transform and Support Vector Machines", International Journal of Image Processing (IJIP) Volume(3), Issue(5) pp. 184-194