IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)
An International Open Access, Peer-reviewed, Refereed Journal

Handwriting Identifier with Help of Combined SVM-HMM Classifier Used With Curvelet Transformation

Poonam dubey Saraswat
M.Tech Scholar, Computer Science & Engineering, SIRTS, Bhopal (M.P)
pnm.saraswat25@gmail.com

Dr.Kalpana Rai
Professor, Department of CSE, SIRTS, Bhopal (M.P.)
kalpana.rai123@gmail.com

ABSTRACT: Handwriting Identifier is considered as important research field in the filled of forensic and biometric applications. It finds significance in fields like graphically which exploit the physiological performance of the human based on the handwriting. At this time too many technique are available for Handwriting Identifier. Although no one of the techniques is yet proved to be clarify for large number of object. That's also fact that all the pattern of writing will be differ of any human with the time. HMM is the best technique for the Identifier the writing for large number of object but its vector feature give differ patter Identifier like retina Identifier , used their training and test sample may vary. Therefore in this work we propose a technique for Handwriting Identifier with help of combined SVM and HMM. In this work curvelet transform are used predominantly for alphabet and numeric identify problem and hence are more suitable for this work. SVM is also given good efficiency but not in the large object. Hence we develop a new classifier and show that the method performs better than self-sufficient HMM and SVM classifier.

Keywords: Hidden Markov modal (HMM), Support vector machine (SVM).

I. INTRODUCTION

Handwriting is unique to each individual [8]. Some people handwriting may have similar for acquired copy write the matter. According to age of a person their handwriting will also changes. With a writing sample cannot determine the age of a suspected. Some give hints at some of these characteristics by the types of writing styles and detailed examination of an original document. The report of the work can be basically put as to verify the handwriting of the Human by using HMM. First Human handwriting is written on white paper which are scanned and given as input to the method. The value comprises of set of characters, say a line of word. System extracts the features and creates a database of element. When a sample scanned image of a person is given as input, it identifies the writer. [5]

We perform binarization and extract invariant moments from the image document followed by the curvelet transform. These features are used to build a HMM model which we use to classify the features of a test document. A users handwriting is the script and its placing on the page express the unique impulses of the being. Using a sample, an expert is able to identify the relevant features of the handwritten script, and the way of the features interact which provide the data for the analysis [6, 7].

It is show that the permutation of features, with the interaction each other so that enable a full and clear interpretation [6].

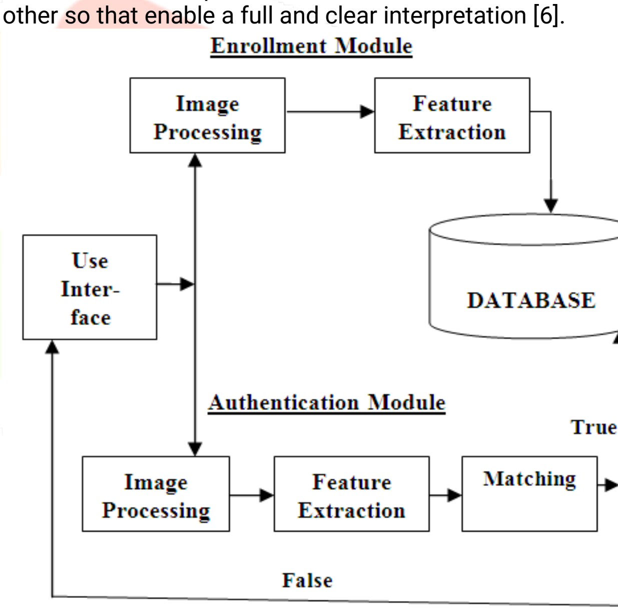


Fig 1: Process for handwriting verifier

HIDDEN MORKOV MODAL (HMM)

First we calculate the possibility matrix of emergence of 1 with respect to 0 and 1 and probability of occurrence of 0 with respect to 1 and 0. This pattern is unique for all the signatures. If the signatures are random then we consider a hidden markov model that can generate such a pattern. The resulting series of the velocity values for each probability serves as the sequence of observation symbols for training a discrete HMM. As outlined in Section 1, a HMM based signature verification system typically involves two phases – training and verification. The training phase consists of learning the features (R *V)* of a particular user from a set of signature samples of that user termed as the training sample set. The system then builds a reference HMM model for the user from this training set and stores it in the database. a HMM is quantitatively described by Equation 2[9].

$$\lambda = \{A, B, \pi\} \quad \text{...... (2)}$$

Where, the state transition matrix is represented by A and it gives the possibility of transiting from one to another state. The observation probability matrix is represented by B and gives the possibility that a finicky inspection symbol occurs in a particular state. The initial probability distribution for the states is representing by pi. At the end of the training phase, the system evolves the most optimum values of A and B based on the input observation sequences (in this case, the *R V* values from each sample in the training set)

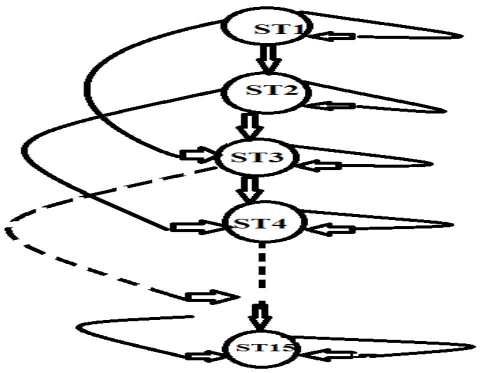


Fig 2: Initial probability distribution for the states

Since a Left-to-Right topology is used therefore, π is fixed at {1,0,0,0} and is not re-estimated in the training phase. As explained in ref [9], this Left-to-Right topology accounts for the time-dependent dynamic property of the online signature data. Furthermore, the presence of single skips in the LTR model gives more freedom in the number of transitions in the model and hence yields better performance in signature verification systems since it better incorporates the intra-user variance inherent in the training sample set of a particular user. The large number of states (at 15) is used because of large observation sequence lengths [9].

SUPPORT VECTOR MACHINES (SVM)

w1x1 + w2x2;:::; +wdxd + w0 = 0

SVM classifier based on the structural risk minimization standard (SRM). The SRM induction principle has two main objectives.

1. To control the empirical risk on the training data.
2. To control the capacity of the decision functions used to obtain this risk value. Second is used for classification (also, can be modified for regression and even for unsupervised learning applications) [10].

Multilayer Perceptions for achieving accuracy is given training data $D = \{(\bar{x}_i, y_i), i = 1 \dots N\}$ with $y_i \in \{-1, +1\}$ is separable by a hyper plane.[12]

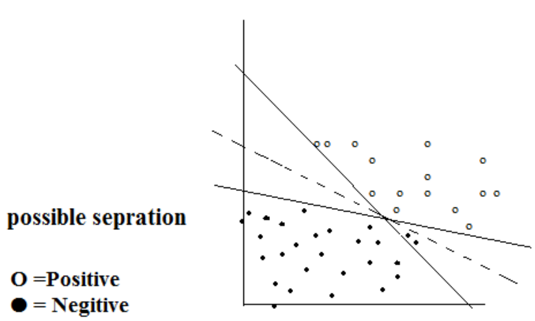


Fig 3: Linear Classifier

$$\hat{\mathbf{y}} = sign(\alpha(\mathbf{w}^T\mathbf{x} + \mathbf{b})) = sign(\mathbf{w}^T\mathbf{x} + \mathbf{b})$$

$f(\bar{x}) = \bar{w}^T\bar{x} + b \quad (= w_1x_1 + w_2x_2 + w_3x_3 + \dots + w_nx_n + b)$ While there can be an infinite number of hyper planes that achieve 100% accuracy on training data,

Given f(x), the classification is obtained as

$$\hat{\mathbf{y}} = sign(f(\mathbf{x})) = \begin{cases} +1 & f(\mathbf{x}) > 0 \\ -1 & f(\mathbf{x}) < 0 \end{cases}$$

w and b give result in the identical section. We can apply any scalar α such that:

$$\hat{\mathbf{y}} = sign(\alpha(\mathbf{w}^T\mathbf{x} + \mathbf{b})) = sign(\mathbf{w}^T\mathbf{x} + \mathbf{b})$$

Therefore there are many identical solutions [9].

II. PROPOSED ALGORITHM

Algorithm for our new approach combined HMM-SVM classifier.

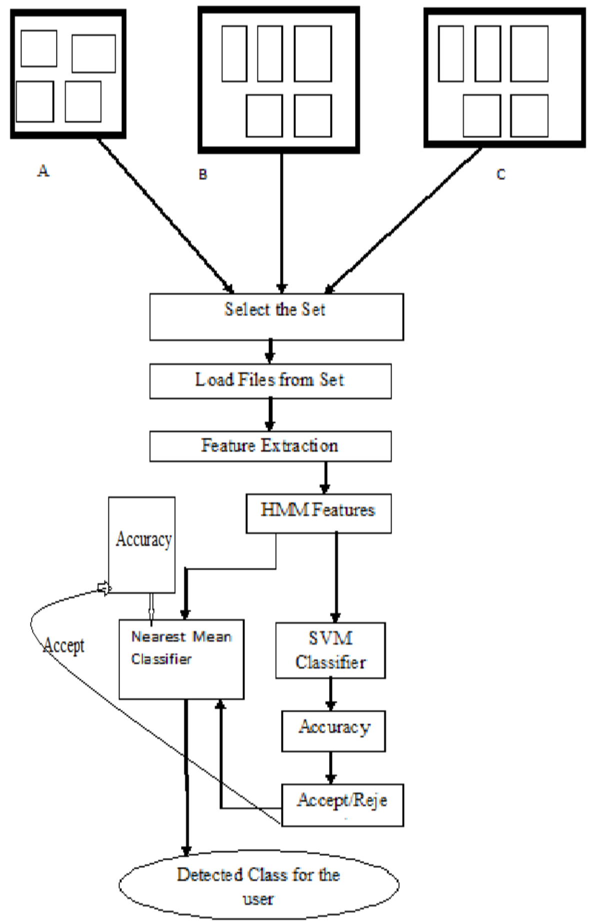


Fig 4: New Approach Of HMM-SVM

PROPOSED METHOLOGY

Preprocessing: Preprocessing is the first step to any document analysis. The steps are preceding bellow. First the sample text is scanned and saves as a picture. This is specifically a gray scale image. A binary image is generated out of this gray scale image and one bit binary noise is removed by using erosion. The subsequent steps are all for normalizing the document to a common axis [4].

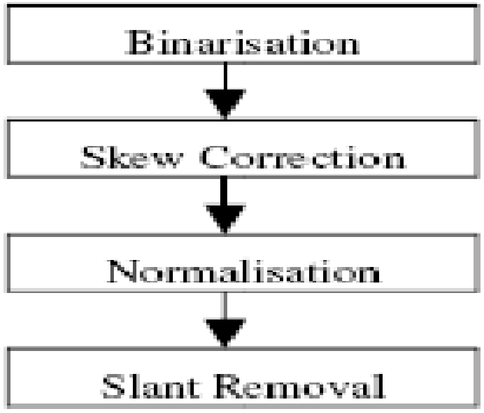


Fig 5: Component of Preprocessing

Due to the variation of stroke, normalization methods such as thinning or skeletonization are used. This normalizes the width of each stroke to 1 pixel.[14]

Segmentation*:* A simple ligature model is used; the features used to distinguish ligatures are very easy to detect and such features are the handwriting contour minima and the holes, which are white associated regions. Other possible parameter points has located exterior the holes belonging to the region between baseline and upper line (core region). In neat cursive handwriting system, holes are present in the following shown letters:

a b d e f g h l o p q

Dissection, or damage, is found by a growing algorithm on black pixels of the image. The process of the dissection is the cut the word image and to switch to white the black pixels below the parameter points. Each parameter point produces a cut. In most cases the cuts are led along vertical directions, but image cut is not a good way always. Image cut in the vertical direction may be long; hence making it is difficult to properly reconstruct the characters. If the two alphabet are alienated along the upright direction, so it will be hard to verify correctly the letter *a*. Slanted cut's direction is the correct one if the length of the vertical cut is over a length, parameterized by the width of the core constituency, other instructions are tested in the range ±45° around the perpendicular one. [13]

Binarization*:* An image is binarized using Natural Thresholding. The process of the Natural Thresholding is the mean gray level value of the images are extracted and values larger than the mean is map to white where as the values lesser than that are marked as black. Algorithm is used for handwritten text images with homogeneous or semi homogeneous backgrounds.

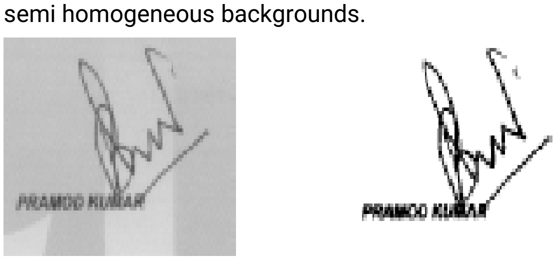


Fig 6: Process of binarization of a sample of a signature

Noise Removal: After the binarization of the image the second noise is removed with different type of transformation. The process of the noise removal is the binarized image a simple 5X5 matrix form then applied morphological filter to remove the spatial noise. This clean binarized image is cropped from both the directions. We have not implemented rotation invariance. We have same pattern image then some of the works the signatures are rotated to a single plane. [3]

Feature Extraction: it has following states

1. Read the matrices from image after preprocessing
2. Extract HMM features
3. Extract statistical data from HMM states such as mean and standard deviation

Verifier: Each state now represents a matrix which can define a probability matrix as calculated during the HMM. The mathematical mean and standard divergence of each value is taken as the training vectors for a SVM. [5]

Curvelet Transform: The curvelet transform is mostly used in Computational Harmonic Analysis [1, 2]. It is a Computational Harmonic Analysis tool used for transformation. In this transform technique curvelet frames were at the start designed to endow with optimally meager representations for objects that are smooth except along smooth curve-like discontinuities (e.g. image with edges). In these techniques Curvelets are obtained by partitioning the 2D Fourier plane into dyadic coronae and sub-partitioning

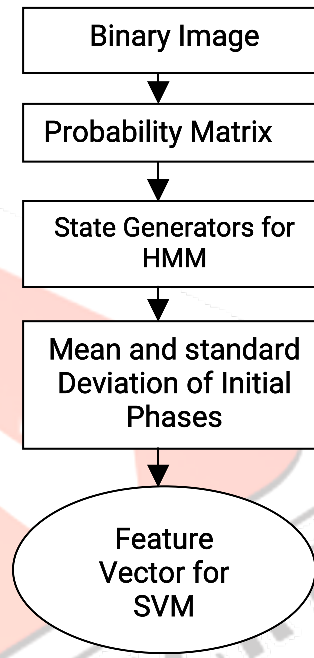


Fig 7: Overall System

those into angular wedges. It followed a parabolic scaling law – at scale 2^{-j}, each element has an effective support of length $2^{-j/2}$ and width 2^{-j}. The output frames are highly anisotropic, multi-directional, and localized in frequency and space. On the practical side, exists high-speed *O(*M*og*N*)* algorithms that tolerate for a decomposition of *n*-by-*m* picture *f* with *x*y=N*. These algorithms are mathematically tough and have an explicit erection for the ad- joint that equals the ersatz-converse

$$C^{*} \cdot = C^{+} \cdot \text{ and } C^{*} C = I \tag{1}$$

where C = curvelet transform. The algorithm is given below:

1. Initialize threshold Γ_0, first guess x_0, set number of CG iterations among thresholdings J and number of iterations $L_{\max}$
2. For $i = 1 : L_{\max}$
   - $\bar{x}_i = CG\,(\tilde{x}_{i-1}, J)$,
   - Soft-threshold $\bar{x}_i$ with the threshold Γ_i and obtain $\tilde{x}_i$.

The inverse transform of reflectivity estimate is given by $\tilde{m} = C^{*}\tilde{x}$. Here the coefficient x is considered as the feature vector for the classifier

III. SIMULATION RESULT

In the HMM – SVM combined classifier experimental results is to find the healthier accurateness. For this intention several different techniques such as HMM are tested features. Experimental results shown better for each feature, is combined of HMM and SVM classifier with optimum efficiency., in the study of HMM classifier when we applied with 300 sample of thirty user accuracy is found 97.31 and In this study, the HMM –SVM combined classifier with different degree was tested, and the result is when we applied on 300 sample of thirty users then accuracy also increase. In my experiments with 300 samples of thirty users accuracy is found 98.44 which is optimum in all previously exist handwriting verifier algorithm those are applied with 300 samples.

Tabulation result of 300 samples of 30 user

The values of consequent presentation parameter for handwriting verifier are tabularized as shown below Table 6.1

No. of Sampl es	HMM Accuracy (%)	HMM-SVM Accuracy (%)	SVM Accurac y (%)
10	90	90	In SVM all 300 samples are executin g in one time
20	94.30	94.73	
30	96.29	96.73	
40	97.22	97.20	
50	97.86	97.87	
60	96.42	98.24	
70	95.45	98.30	
80	97.33	98.60	
90	97.66	98.83	
100	96.80	98.94	
110	97..08	99	
120	97.32	99.11	
130	96.69	98.36	
140	96.18	98.48	
150	96.42	98.58	
160	96.66	98.66	
170	96.83	98.73	98.44
180	96.98	98.87	
190	97.11	98.87	
200	97.31	98.92	
210	97.35	98.98	
220	97.32	99.12	
230	97.36	98.21	
240	97.37	98.91	
250	97.31	98.93	
260	97.67	99.01	
270	97.44	98.12	
280	97.56	98.23	
290	97.39	98.81	
300	97.31	98.93	

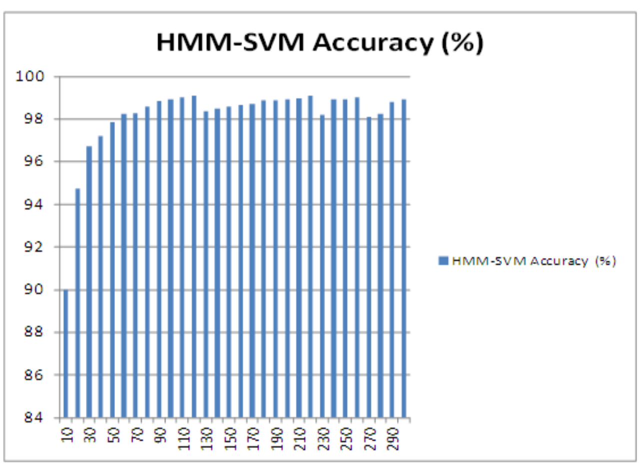


Fig 8: Graphical represntation between samples and accuracy

Tabulation result of 20 samples of 2 user

In this result we represent outcomes 0and 1 are represented by false and true respectively.

Actual class (user)	Dected class	True(1)/false(o)
1_1.jpg	1	1
1_2.jpg	1	1
1-3.jpg	1	1
1_4.jpg	1	1
1_5.jpg	1	1
1_6.jpg	1	1
1_7.jpg	1	1
1_8.jpg	1	1
1_9.jpg	1	1
1_10.jpg	14	0
2_1.jpg	2	1
2_2.jpg	2	1
2_3.jpg	2	1
2_4.jpg	2	1
2_5.jpg	2	1
2_6.jpg	2	1
2_7.jpg	2	1
2_8.jpg	2	1
2_9.jpg	2	1
2_10.jpg	no match	missed data

Fig 9: Graphical represntation of feature extraction of a sample.

Snapshot:

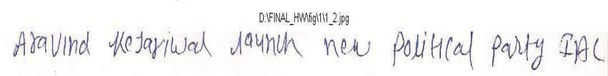
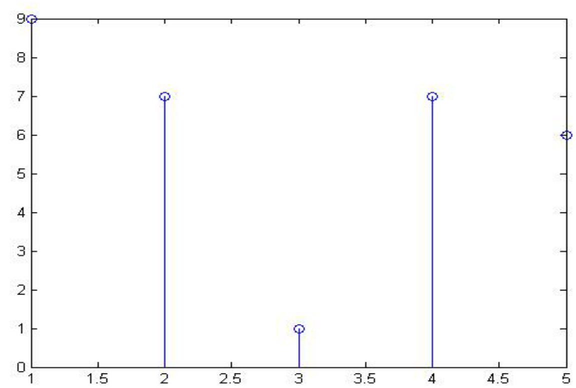


Fig 10:

Sample picture



Fig

11: Curvelet view

Fig 12:

HMM+SVM View

IV. CONCLUSION

Handwriting identifier and analysis are part of larger domain of work which finds application in graphology and forensic science. Handwriting identifier is a challenging aspect as same user's handwriting tends to differ depending upon type of Pen being used, the writing surface and so on. Beside, Handwriting is not considered as unique biometric property. It is rather a pattern associated with different users. Therefore handwriting verifier differs from other similar verifier like signature biometric. In this work we have presented a Novel technique of Handwriting verifier by first extracting statistical moments and curve let features from the user's handwriting pattern and then forming a statistical state machine with HMM. Further the technique was improved by the HMM-SVM combined classifier. HMM-SVM classifier improve the accuracy also resolve the problem of multiple detection of class.

The system can be improved by incorporating more features that describes shapes and slants of the characters of handwritten text like Zernike moments, shape contexts and so on.

V. REFERENCES

[1] Casey R.G. and Lecolinet E., A survey of methods and strategies in character segmentation. *IEEE Trans. PAMI* **18** (7) 1996 pp. 690-706.

[2] Cesar M. and Shingal R., Algorithm for segmenting handwritten postal codes. *Int'l J. Man Machine Studies* **33** (1) 1990 pp. 63-80.

[3] Baird H.S., Kahan S. and Pavlidis T., Component of an Omnifont Page Reader. *Proc 8th ICPR* Paris 1986 pp. 344-348.

[4] Yanikoglu B. and Sandon P.A., Segmentation of off-line cursive handwriting using linear programming. *Patt. Recog.* **31** (12) 1998 pp. 1825-1833.

[5] Bozinovic R.M. and Shrihari S.N., Off-line cursive script recognition. *IEEE Trans. PAMI* **11** (1) 1989 pp. 68-83.

[6] Kimura F. et al., Improvements of a Lexicon Directed Algorithm for Recognition of Unconstrained Handwritten Words. *Proc 2nd ICDAR* Tsukuba October 20-22 1993 pp. 18-22.

[7] Romeo-Pakker K. et al. A New Approach for Latin Arabic Character Segmentation *3rd ICDAR*, Montreal, August 14-16, 1995, pp 874-877.

[8] David A. Katz, Handwriting Analysis.

[9] Javed Ahmed Mahar, Mohammad Khalid Khan, Prof. Dr. Mumtaz Hussain Mahar," Off-Line Signature Verification of Bank Cheque Having Different Background Colors"

[10] Bikash Shaw, Swapan Kumar Parui, Malayappan ShridharOffline Handwritten DevanagariWord Recognition: A holistic approach based on directional chain code feature and HMM.

[11] Edson J. R. Justin0 Flfivio Bortolozzi ', Robert Sabourin ', Off-line Signature Verification Using HMM for Random, Simple and Skilled Forgeries

[12] Yousri Kessentini, Thierry Paquet1, and Abdelmajid Benhamadou' Off-Line Handwritten Word Recognition Using Multi-Stream Hidden Markov Models published in "Pattern Recognition Letters 31, 1 (2010) 60 - 70" DOI: 10.1016/j.patrec.2009.08.009

[13] Rahul Kala, Harsh Vazirani, Anupam Shukla, Ritu Tiwari 'Handwriting reorganization by using genetic algorithm' International Journal of Computer Science Issues, Vol. 7, Issue 2, No 1, March 2010

[14] Roman Bertolami _ Matthias Zimmermann 1 Horst Bunke' Rejection strategies for offline handwritten text line recognition' ACM Portal, Vol. 27, Issue. 16, December 2006

[15]www.mathwork.com