Offline Signature Verification and Identification using Dimensionality Reduction

Rashmi .B.N
Assistant professor
JSSATE, Bangaluru
Karnataka, India

ABSTRACT
In this paper we are proposed a novel approach to extracting the features from a hand-written off-line signature. The experiments are carried out on a user created data base. We are extracting the geometrical distance-metric features and pruned projection features. The extracted pruned projection features are huge in dimensions, it's difficult to process and analysis. To reduce the feature matrix dimensions without loss of information, existing stereographic reduction algorithm is used. The patterns are classified using the supervised Knn-classifier. FRR (False Rejection Rate) and FAR (False Acceptance Rate) for Identification by proposed approach is 6% and 7%. And that of Verification is 12.6% and 13 %.

General Terms
Dimensionality Reduction, K-nn classifier, Pattern recognition, Data base collection.

Keywords
Pruned projection, End-points, Distance between two end points, Angle made at each end point.

1. INTRODUCTION
Signatures are still widely used as the most vital structure of identification and verification of human beings. In Off-line approach stable dynamic variations are not used in the point of verification. In many parts of the world however it is considered more difficult than on-line verification due to the lack of dynamic information [1]. In Off-line signature system, signatures are treated as an image. The input image represents individual style of human handwritten signature. In a system the objective is to detect the forgeries, which are related to intra and inter-personal variability [3] [4].The first type is random forgery, is usually represented by a signature structure that belongs to a different writer of the signature model (see Fig.1a and Fig.1b).The second one, called simple forgery, is represented by a signature structure with the same shape of the genuine writer's form [5]. The last type is the skilled forgery, represented by a suitable imitation form of genuine signature model (see Fig.1c).

Fig. 1a	**Fig.1b**	**Fig.1c**

Signature identification refers to interpersonal variability; it means to differentiate the signatures from unlike classes, mainly dealing with random forgeries. Signature verification refers to the intra-personal variability; it means to distinguish the signatures within the same class in turn identifies the skilled forgeries of genuine samples. In this context we are exploring the novel approach for pruned projection feature extraction and a dimensionality reduction algorithm to reduce the dimensions of feature matrix.

2. DATA BASE CREATION
Off-line signature database is formed by collecting the signatures on a predefined paper script from the different users. Each individual has their own style of signing include English text and Kannada text and not readable form. 60 classes each class with 10 genuine, five skilled forgeries. The total count of samples in the data base is 900.

3. IMAGE PREPROCESSING STEPS
Step 1: Scan the pre-formatted document with 600dpi as a color image.

Step 2: Extract the signature from the document by cropping.

Figure 2a: Scanned image	**Figure 3a: Scanned image**

Step 3: Read the image and resize it to a standard size.

Step 4: Convert the color image (rgb) to black and white image (binary image).

Step 5: Applying thinning to reduce noise.

Step 6: Using auto-cropping to extract the region of interest (ROI) with a new size of $[m \times n]$.

Fig 2b: Preprocessed image	**Fig 3b: Preprocessed image**

4. PRUNED PROJECTION FEATURE EXTRACTION
Each and every classification, Verification and identification system invokes the feature extraction as a main module. Feature extraction means to a extracting the region of interest in the image which can further able to differentiate one from all other. Signatures are in readable form and much of them are not readable form because of this constraint, instead of treating signature as text we are treat it as whole image. In this approach we are extracting the distance to each on-pixel of the image which we encountered in the particular direction of traversal of the image, where each distance is a feature itself. The step wise procedure with respective of point of view [NEWS] is provided below.

Front View of Signature Image (East-to-West or Left-to-Right Traversal of image):

Step 1: Extract the region of interest from the image with five pixels off-set in each direction from the best fitting rectangular box.

Fig 4(a): L-R Projection.   **Fig 5(c): L-R Projection**

While we are finding the distance to a each first on-pixel from the front view of the image (4a), there are chances to reach the back-end view (tail-end) first as shown in image(5a), to avoid this erroneous situation we are taking the pruned image data, it means to 1/3 portions of the signature image from the respective direction.
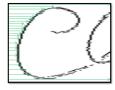

Fig 4(b): After pruning action. Fig 5(b): After pruning action.

$$F1=\{f11,f12,f13,f14,\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots f1n\}$$

F1 be the east to west (Left-to-right view), F2 be the west to east (Right-to-Left view), F3 be the TBV (Top-bottom view) pruned projection features set and F4 be the BTV (Bottom-top view) pruned projection features set.

$$F = \{F1, F2, F3, F4\}$$

$$F = \left\{ \begin{array}{l} (f_{11}, f_{12}, \ldots\ldots, f_{1n}), (f_{21}, f_{22}, \ldots f_{2n}), (f_{31}, f_{32} \ldots f_{3n}), \\ (f_{41}, f_{42}, \ldots f_{4n}) \end{array} \right\}$$

F is the pruned projection features matrix of single image which is large in dimensions.

Analysis with a large number of variables generally requires a large amount of memory and computation power or a classification algorithm which over fits the training samples and generalizes poorly to new samples. To reduce the dimensions of the feature matrix we are applying exiting one of the dimensionality reduction algorithms named Stereographic projection algorithm [6].

Major steps of existing Stereographic projection reduction algorithm:

Step 1: Obtain the doublets from the features set by applying combinations to each set of features without repetitions.

$$F = \{(f11, f12, \ldots\ldots, f1n), (f21, f22, \ldots f2n), (f31, f32 \ldots f3n)\}$$

Doublets for each set of features are:

$$F = \left\{ \begin{array}{l} [(f_{11}, f_{12}), \ldots (f_{1n-1}, f_{1n})], [(f_{21}, f_{22}), \ldots (f_{2n-1}, f_{2n})], \\ [(f_{31}, f_{32}), \ldots (f_{3n-1}, f_{3n})], [(f_{41}, f_{42}), \ldots (f_{4n-1}, f_{4n})] \end{array} \right\}$$

Step 2: covert the each doublet into a triplet using the formulas provided by the algorithm [7].

$$X = \frac{a_p^2}{a_p^2 + a_q^2 + 1}$$

$$Y = \frac{a_q^2}{a_p^2 + a_q^2 + 1}$$

$$Z = \frac{a_p^2 + a_q^2}{a_p^2 + a_q^2 + 1}$$

Where a_p and a_q are $F_1 = \{f1 \ldots fn\}$ chosen and its next arrival of the one feature set.

Step 3: using the triplet combinations finds the hue, saturation and intensity value by the assist of formula.

$$Hue = \sqrt{X^2 + Y^2}$$

$$Saturation = \tan^{-1}\left(\frac{X}{Y}\right)$$

$$Intensity = 1 - Z$$

Step 4: Using raine's equation find the pseudo RGB components of the image [6].

$$X' = Saturation \times \sin(Hue).$$

$$B = Saturation \times \cos(Hue).$$

$$Blue = \frac{2}{3} \times B + \frac{1}{3}V$$

$$Green = \frac{1}{(\sqrt{3} \times X')} + \frac{1}{3}V - \frac{1}{3}B$$

$$\operatorname{Re}d = \frac{-1}{(\sqrt{3} \times X')} + \frac{1}{3}V - \frac{1}{3}B$$

$\sum R \quad \sum B \quad \sum G$ Be the pseudo-RGB components of the particular point of view of the image.

Step 6: Now we have only 12 features from an image with respect to all four points of view $[3 \times 4]$.

With respect to our data set we have feature matrix of size $[900 \times 12]$, where 900 is the number of samples and 12 is the features.

5. CONCLUSION

In this project context we dealt with Off-line signatures. The experiments carried out on Identifying and Verifying the Off-line signature data base which is created by the user. From the analysis of the system results which desire the best case performance among the different experiments conducted with differing thresholds, k-value, and varying training and testing samples ratio on Identification and Verification System with K-nn classifier, its clears that proposed method shows FAR 6% and FRR 7% for Identification and that of Verification is FAR 12.2% and FRR 13.8%

7. REFERENCES

[1] Meenakshi S Arya, Vandana S Inamdar,."A Preliminary Study on Various Off-line Hand-Written Signature Verification Approaches", ©2010 International Journal of Computer Applications (0975-8887) Volume 1 – No. 9

[2] Edson J. R. Justino, A. El Yacoubi, F. Bortolozzi and R. Sabourin "An Off-Line Signature Verification System Using HMM and Graphometric Features", DAS 2000, 4th IAPR International Workshop on Document Analysis Systems, Riode Janeiro, Brazil, (2000), pp 211--222.

[3] Drouhard, J.P., R. Sabourin, and M. Godbout, "A neural network approach to off-line signature verification using directional PDF", Pattern Recognition, vol. 29, no. 3,(1996), 415--424.

[4] "Off-line Signature Verification Using HMM for Random, Simple and SkilledForgeries", Edson J. R. Justino 1 , Flávio Bortolozzi 1 , Robert Sabourin 1, 21 PUCPR - Pontifícia Universidade Católica do Paraná ,R. Imaculada Conceição, 1155 CEP:80215-901 - Curitiba - PR – Brazil - {justino, fborto }@ppgia.pucpr.br 2ETS - Ecole de Technologie Supérieure, 1100, rue Notre-Dame Ouest - Montréal (Québec) H3C 1K3 – Canada.

[5] Dubey RB, Sachdeva S."An Offline Signature Verification Technique". WebmedCentral MISCELLANEOUS 2011:2(5):WMC001919 doi:10.9754/journal.wmc.2011.001919.

[6] Written by Paul Bourke EGG data courtesy of Dr. Per Line.