Features Extraction and Verification of Signature Image using Clustering Technique

Samit Biswas[1], Tai-hoon Kim[2,*], Debnath Bhattacharyya[2]

[1]*Department of Computer Science and Engineering,*
Bengal Institute of Technology,
Kolkata- 700150, India.
samitbiet@yahoo.com

[2]*Department of Multimedia,*
Hannam University,
Daejeon, Korea.
taihoonn@empal.com, debnathb@gmail.com

Abstract **– Humans are comfortable with pen and papers for authentication and authorization in legal transactions. In this case it is very much essential that a person's Hand written signature to be identified uniquely. The development of efficient technique is to extract features from Handwritten Signature Image and verify the signature with higher accuracy. This paper presents a method for off line hand written signature verification with higher accuracy. In this paper we have introduced a procedure to extract features from Handwritten Signature Images. That computed feature is used for verification. Here we used a clustering technique for verification.**

Keywords **– Segmentation, Morphological operation, Thinning, Region of Interest (ROI), Clustering Techniques.**

1. Introduction

Handwriting is a skill that is highly personal to individuals and consists of graphical marks on the surface in relation to a particular language.

Many researchers have been done on this topic. Signatures of the same person can vary with time and state of mind [1]. A method proposed in [1] a signature verification system which extracts certain dynamic features derived from velocity and acceleration of the pen together with other global parameters like total time taken, number of pen-ups. The features are modeled by fitting probability density functions i.e., by estimating the mean and variance, which could probably take care of the variations of the features of the signatures of the same person with respect to time and state of mind.

Handwritten signature is a form of identification for a person A method is introduced by Md. Itrat Bin Shams [2] where a signature image is first segmented (vertical and horizontal) and then data is extracted from individual blocks. Here these data is then compared with the test signature.

Signatures are composed of special characters and flourishes and therefore most of the time they can be unreadable. Also intrapersonal variations and the differences make it necessary to analyze them as complete images and not as letters and words put together [3].

The research work proposed by Debnath Bhattacharyya, Samir Kumar Bandyopadhyay, Poulami Das, Debashis Ganguly and Swarnendu Mukherjee [3] was based on the collection of set of signatures from which an average signature was obtained based on the stated algorithm[3] and then taking decision of acceptance after analyzing the correlation in between the sample signature and the average signature.

*Corresponding author

Baseline is the imaginary or invisible line, which a signature is assumed to rest on. A baseline is the line on which the letter sits. In our daily life, a baseline must be imagined when signing or writing in an unlined sheet of paper. The straightness and direction of the signature can be changeable features in a signature [5]. A method was introduced by Azlinah Mohamed, Rohayu Yusof, Shuzlina Abdul Rahman, Sofianita Mutalib [5] to extract the baseline feature.

The system developed by Alan McCabe, Jarrod Trevathan [6] analyses both the static features of a signature (e.g., shape, slant, size), and its dynamic features (e.g., velocity, pen-tip pressure, timing) to form a judgment about the signer's identity.

Support Vector Machine (SVM) can also be used to verify and classify the signatures [7].

For verification of signature image after extraction of features from handwritten signature images many authors use different mathematical formulas. Such as correlation is used for verification between sample and test signature [3, 10].

In this paper for verification of signatures after extraction of features we used clustering techniques.

Clustering involves dividing a set of data points into non-overlapping groups, or clusters, of points, where points in a cluster are "more similar" to one another than to points in other clusters. The term "more similar," when applied to clustered points, usually means closer by some measure of proximity. Any particular division of all points in a dataset into clusters is called a partitioning [11].

Image segmentation is the decomposition of a gray level or color image into homogeneous regions [13]. In image segmentation, cluster analysis is used to detect borders of objects in an image.

A method is already there [14] for image clustering based on a shared nearest neighbours approach that could be processed on both content-based features and textual descriptions (tags).

Some of the various approaches for Signature verification systems as reported by researchers are as follows:

For online signature verification following features considered in [1]

- The number of zeros in the velocity in X direction.
- The number of zeros in the velocity in **Y** direction.
  (or instead of the zeros it could be values that are a very small percentage of the peak values)
- The number of zero crossings in the acceleration in the **X** direction.
- The number of zero crossings in the acceleration in the **Y** direction. (These zero crossings indicate the change from increasing to decreasing velocity **or** from decreasing to increasing velocity).
- Total time taken in performing the signature.
- Total distance traveled by the pen. This is the sum of all the euclidean distances between all the points,
  $$D = \sum \{(X_i - X_{i+1})^2 + (Y_i - Y_{i+1})^2\}^{\frac{1}{2}}$$
  Where the summation in the previous expression runs
  from i = 0 to N-1.
- Total number of pen-ups (including or excluding the final pen-up).
- Total pen-up time, which is the total time for which the pen was up.
- The number of times the pressure **goes** above an upper threshold $\mathbf{T_u}$. and
- The number of times the pressure falls below a lower threshold $\mathbf{T_l}$ (Thresholds can be a certain percentage of the peak value).

Feature was extracted based on horizontal and vertical segmentation for signature verification [2] .

Vertical Segmentation: After pre-processing First the image is put inside a rectangle so that it is properly fitted inside this rectangle. Scanning is started from left most upper point. Here vertical scanning is performed and if any peak or crest in the image is found then a line is drawn through it. Similar process is performed by scanning from left most lower point and going
upwards while scanning. In this way total image is divided in vertical segments

Horizontal Segmentation: Horizontal segmentation is done on each vertical segment. Here scanning is done from upper left point and down to lower left point. In each case scanning is moved from left to right to find any peak or crest. If found a horizontal line is drawn through it. Similar operation is performed for right most upper point to lower point in each vertical segment.
After vertical and horizontal segmentation, signature have divided to small blocks, data have to be extracted from individual blocks.

Another signature verification method was introduced in [3]. An algorithmic approach was proposed in [3] for the verification of handwritten signatures by applying some statistical methods. The algorithm proposed in[3] for of handwritten signature verification in an offline systeml and security purpose.
The algorithmic approach has the flexibility of choosing the number of signatures, i.e., no_of_Sign for testing purpose to generate a signature as Avg_Sign containing the specialized mean features set from the test signatures set. After collecting the signatures for testing, the algorithm converts them into a set of 2D arrays of binary data values-0 and 1. From these binary arrays using statistical methods of calculating expected
mean an average data set is calculated using the formula given by

$$\mu_x = E(X) = \frac{\sum X_i}{N}$$

The algorithm for verification of handwritten signatures is required an additional input of Sample_Sign, which is to be tested for verification of acceptance or rejection. Here, another constant value is maintained, named as decisionValue, which is also calculated and set by professional statistician and security administrator deciding the security concern and policies of the organization after conducting surveys and testing over some sample experimental data set with already known results. The algorithm compares between two binary data sets obtained after analysis of Avg_Sign and Sample_Sign and calculates out the correlation coefficient, r_{xy} between them using the following statistical formula:

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x}) \sum_{i=1}^{n} y_i - \bar{y}}}$$

In turn, r_{xy} is compared with the decisionValue and accordingly return TRUE or FALSE as for acceptance or rejection respectively.

Another feature extraction method was proposed in[18] to perform offline signature verification based on intelligent techniques . Structural features are extracted from the signature's contour using the modified Direction Feature(MDF), and its extended version: the enhanced MDF.

Particular problems facing offline signature verification are the small number of genuine samples that may be used for training process and the system's ability to distinguish a genuine signature from different types of forgeries(random, simple, simulated, etc). Varying skill levels forgeries are listed as follows[18]:

- A forged signature can be another person's genuine signature.
- A forged signature is produced with the knowledge about the genuine writer's name only.
- A forged signature imitating a genuine signatures model reasonably well is categorized as a Simulated Forgery.
- Signatures produced by inexperienced forgers without the knowledge of their spelling after having observed the genuine specimens closely for some time are categorized as Unskilled Forgeries.
- Signatures produced by forgers after unrestricted by non professional forgers are categorized as Simple forgery/ Simulated simple forgery.
- Forgeries which are produced by a professional imposter or person who has experience in copying signatures are categorized as skilled Forgeries.

In this case , a random forgery refers to the 1st type and a targeted forgery refers to the 5th type.

*Corresponding author

This paper is organized as follows. Previous work is surveyed in section I. Proposed Algorithm is introduced in section II. Experimental result and comparative study is shown in section III. Finally, we conclude in Section IV.

2. Proposed Algorithm

First we have to collect sample signatures that we have to keep in database. Here For every person we have collected n signature samples for database. It is better if we can collect more signature samples for database. Then for verification collect test signatures against the sample signatures. These test signatures we have to verify if it is genuine or forgery. Each of the signatures (Samples and corresponding test) has to take within a same sized rectangular area on paper by pen and collect the image of that rectangular area. The signed paper is then scanned at 500 dpi resolution by a gray scale scanner. Therefore, the signature image will be processed in four stages in order to determine if it is genuine or forgery. The following flow chart in Figure. 1 shows these stages which will be presented in the following subsections. Left side of Figure. 1 shows the steps for extracting features from sample signature Image and right side of Figure.1 shows the steps for extracting features from test signature image. The last step (for both cases) is for clustering the extracted data.

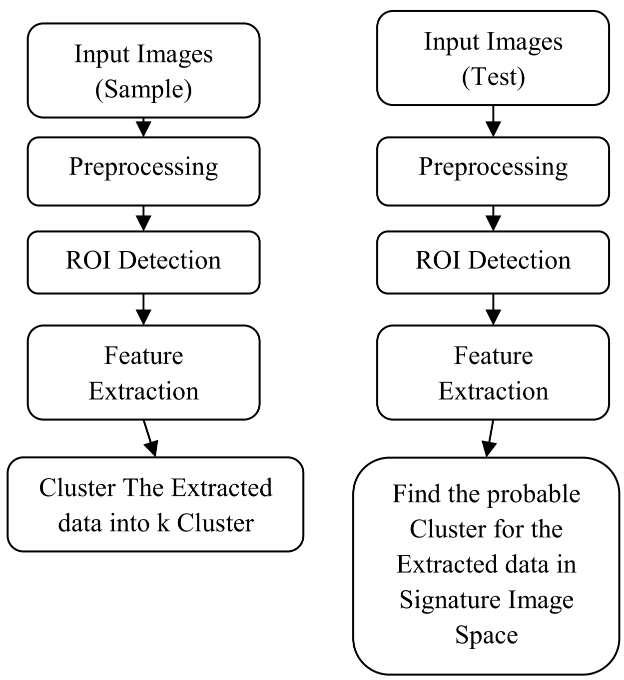

Figure.1. Flow Chart for Data Extraction from Signature image (Sample and Test) And Clustering Data.

A. Preprocessing Stage

The system deals with the static scanned image of the signatures. Unwanted images i.e. noises may include during the scanning process of the signature. The preprocessing of the signature images is related to the removal of noises, and thinning. The goal of thinning is to eliminate the thickness differences of pen by making the image one pixel thick. To remove noises and enhance, the images are preprocessed by filtering techniques [8]. For thinning Morphological operations [8] can be applied. Preprocessing have to be done for both sample and test signature images.

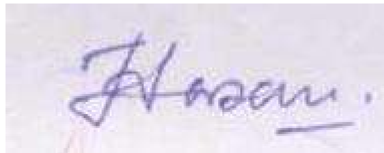

Figure. 2. Before Thinning the Signature Image.

Figure. 3. After thinning the Signature Image.

B. Region of Interest Detection and Scaling

In this step the signature area within the image is identified i.e. the region of interest (ROI) is identified. The Region of Interest detection of Figure. 3 is shown in Figure, 4. Region of Interest have to be identified from both Sample Signature and corresponding Test Signature. Then scaling is performed on both the sample and test signature. So stretching is performed to the input signature in case it is smaller than standard size or squeezing is done for being bigger. Normally all the signatures in the database are made to fit inside a rectangle of same height and width. Whole step is shown in Figureure.

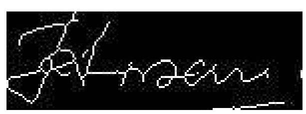

Figure. 4. Region of Interest Detection.

C. Feature Extraction Stage

Extracted features in this stage are used for clustering the signature images for verification stage. Features will have to be extract from both sample images and Test image. Three new Feature Extraction procedures for signature image verification is introduced here.

(a) *Signature Height Width Ratio:*

The ratio is obtained by dividing signature height to signature width. The height is the maximum length of the columns obtained from the cropped image. Similarly the width is also calculated considering the row of maximum length. Signature height and width can change. But height-to-width ratios of an individual's signatures are approximately constant.

(b) *Signature Occupancy Ratio:*

It is the ratio of number of pixels which belong to the signature to the total pixels in the signature image. This feature provides information about the signature density. The Signature Density, D_i for i^{th} sample signature image can be calculated as follows:

$$D_i = \frac{I_i}{X_i}$$

*Corresponding author

If n is the total number of sample image for a person then i=1,2,3, …..n. I_i is the number of pixels which belongs to the signature of i^{th} signature image sample. X_i is the total number of pixel in i^{th} sample signature. This feature provides information about the signature density.

(c) Distance Ratio calculation at boundary.

After cropping, the pixels in closest proximity to the boundaries (left, right, upper & bottom) are determined and their distance from the left & bottom boundaries are evaluated, i.e. for the upper leftmost pixel its distance from bottom boundary(L_1) & for the bottom left most pixel the distance from right boundary is calculated(L_2). These values are used later in verification process. Total procedure for calculation is shown in following Figureure.

Figure. 5. Distance Calculation.

The distance ratio, R is calculated as follows

$$R = \frac{L_1}{L_2}$$

The ratio should be approx same for all time for a person's signature.

(d) Compute the length and ratio of Adjacency Columns.

A new Feature Extraction procedure is introduced here. Features will have to be extract from both sample image and Test image. In this stage first we are computing the length of the adjacency columns from top and from bottom of the Sample signature image and store it in a 1D array L_T and L_B. Then compute the sum of all elements of L_T and sum of all elements L_B. Overall procedure for this feature extraction is shown in the following Figure 6 . Upper half of Figure.6 shows a portion of a signature Image. Lower half of Figure. 6 shows the length of the corresponding adjacency columns from top and from bottom.

Now Adjacency Ratio can be calculated as follows:

$$Adjacency\ Ratio = \frac{Sum(L_T)}{Sum(L_B)} \text{ x Signature Occupancy}$$

For every signature of a person Adjacency ratio will be more or less same. If small differences are there that will be negligible.

0	0	0	0	0	0	1	0	1	1	0	0	0	0
0	0	0	0	0	1	0	0	0	0	1	0	0	0
0	0	0	0	1	0	0	0	0	0	0	1	1	1
0	0	0	1	0	0	0	0	0	0	0	0	0	1
0	0	1	0	0	0	0	0	0	0	0	0	0	1
0	0	1	0	0	0	0	0	0	0	0	0	0	1
0	1	0	0	0	0	0	0	0	0	0	1	1	1
0	1	0	0	0	0	0	0	0	0	1	0	1	0
1	0	0	0	0	0	0	0	1	1	0	0	0	1
1	0	0	0	0	0	0	1	0	0	0	0	0	1

Length From Top (L_T):													
8	6	4	3	2	1	0	9	0	0	1	2	2	2
Length from Bottom (L_B):													
0	2	4	6	7	8	9	0	1	1	2	3	2	0

Figure. 6. Calculation of length of adjacency columns from top and bottom.

(e) Compute the number of spatial symbols within the signature Image.

Every person in their signature uses some spatial symbols, such as they uses some 'x' marks (cross marks), star marks or other symbols. The total number of spatial symbols of a person's signature is unique. For calculating the total number of spatial symbols in a signature image we have to preprocess the image upto thinning. Then If we find that one pixel having more than two neighbors each of which get the values 1 then those pixels will form a Spatial symbol. Such types of pixels are shown in Figure. 7

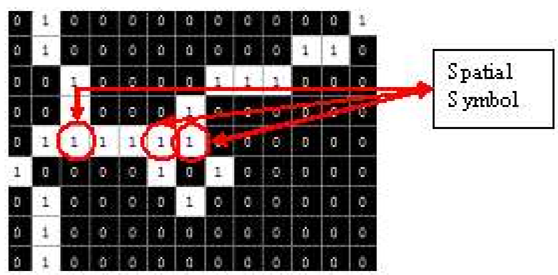

Figure. 7. Spatial Symbols in a Signature Image.

D. Signature Images Clustering

A number of methods have been proposed by many authors for clustering data. Hierarchical clustering, self-organizing maps, K-means, and fuzzy c-means have all been successful in particular applications. A person may have many sample signature images. We create separate clusters for set of sample signatures for each person. Here we use K-Nearest Neighbors' (KNN) clustering Technique for verifying a test signature belongs which cluster. The detailed algorithm for verification is as follows.

Steps on how to compute k-Nearest Neighbors k-NN algorithm:

Input:

A clustered data set X (extracted feature from sample Signature image) of n points in d dimensional space in to K Clusters. Where K is the total number of persons, n is the total number of sample signature for each person and d is the total number of extracted feature.

$X : \{x_1, x_2, \ldots\ldots\ldots\ldots, x_n\}$

where $x_i = [x_{i1}, x_{i2}, \ldots, x_{id}]$ for i=1,2,…,n.

A test data set T(extracted feature from Test Signature image)

$T: \{t_1, t_2, \ldots\ldots\ldots\ldots, t_n\}$

where $t_i = [t_{i1}, t_{i2}, \ldots, t_{id}]$ for i=1,2,…,n.

Output:

Find the cluster for the dataset T.

*Corresponding author

Algorithm:

Step 1. Determine parameter K = number of nearest neighbors.
Step 2. Calculate the distance between the query-instance and all the training samples.
Step 3. Sort the distance and determine nearest neighbors based on the K-th minimum distance
Step 4. Gather the category of the nearest neighbors.
Step 5. Use simple majority of the category of nearest neighbors as the prediction value of the query instance

If k = 1, then the object is simply assigned to a new class i.e create a new cluster for this data point .

Here, another constant value is maintained, named as decision Value (distance between the query-instance and all the training sample), which is also calculated and set by professional statistician and security administrator deciding the security concern and policies of the organization after conducting surveys and testing over some sample experimental data set with already known results.

Signature Acceptance Percentage (%) can be calculated as follows:

$$\text{Acceptance Percentage} = \frac{M_s}{T_c} \times 100\ \%$$

Where M_s is the number of nearest neighbors in the majority cluster as the prediction value of the query instance and T_c is the total number of signature or data point in the majority cluster.
s

3. Experimental Result

In this section we show a few experimental results to illustrate the performance of the proposed method. The algorithm proposed in this paper consists of two distinct divisions. First features from the sample signature image and Test signature image have to be computed. Then a decision will be made based on the clustering results between the computed feature of Sample signature image and Test Signature Image.

The detailed extracted data is shown in Table-I. Here we have considered five peoples signature. We have taken ten sample signatures from each of five peoples. Extracted data for genuine test signature and verification result is shown in Table-II and for forgery signatures is shown in Table-III.

4. Discussion and Conclusion

A new method to extract features from handwritten signature and recognition of handwritten signature is presented here. Achieved results are encouraging and suggest the adequacy of the selected features. This proposed algorithm will help community in the field of signature verification, signature analysis and signature recognition. This work studies an image clustering process based on a k nearest neighbours approach enabling to handle clusters of different sizes and shapes. This types of Image clustering techniques can also be used in the field of Face recognition and Thumb impression recognition.

The proposed system promises a very simple yet reliable solution to the problem of signature verification. Experimental results clearly show that this method can indeed differentiate forgery with actual ones.

Table I. Extrtacted Features from Sample Signature Image.

Person Sl	F_1	F_2	F_3	F_4	F_5	Cluster Index
1	0.23585	0.067937	26	0.06727	4	1
	0.18487	0.077536	23	0.10724	7	
	0.21138	0.072879	27	0.16114	9	
	0.22	0.087817	23	0.20664	8	
	0.17424	0.074561	24	0.16602	8	
	0.25472	0.073765	28	0.14931	8	
	0.18033	0.078473	23	0.14865	3	
	0.176	0.074189	23	0.11347	9	
	0.1811	0.070638	24	0.12843	7	
2	0.33803	0.067778	1.0476	0.093693	2	2
	0.3	0.071111	1.2222	0.088208	8	
	0.34483	0.055352	1.1739	0.065907	1	
	0.26966	0.064	0.55556	0.085758	2	
	0.27174	0.052936	1.05	0.075477	1	
	0.30769	0.068354	0.5	0.086575	1	
	0.27059	0.064922	0.55556	0.054611	2	
	0.25263	0.06	0.95238	0.059183	1	
	0.31646	0.067788	1.2222	0.077703	6	
	0.27619	0.051887	0.52	0.065378	1	
3	0.41975	0.070732	0.52632	0.21338	6	3
	0.5	0.068052	0.37838	0.11818	6	
	0.3956	0.06698	0.51282	0.16905	2	
	0.34409	0.069632	0.34884	0.095108	4	
	0.34021	0.081032	0.3913	0.12189	7	
	0.41379	0.070946	0.32432	0.15692	7	
	0.35484	0.066959	0.18421	0.11855	4	

*Corresponding author

Person Sl	F_1	F_2	F_3	F_4	F_5	Cluster Index
	0.29032	0.074088	0.39535	0.14497	4	
	0.29787	0.076588	0.33333	0.12362	3	
	0.4125	0.076616	0.7	0.14488	9	
4	0.42188	0.072527	2.2857	0.1207	4	4
	0.34375	0.07291	2.3333	0.099313	2	
	0.3375	0.053351	1.6667	0.085706	2	
	0.31034	0.064935	2.5714	0.078716	4	
	0.2907	0.067639	1.2727	0.091198	1	
	0.30263	0.079545	1.25	0.10304	5	
	0.2809	0.065385	2.125	0.093432	4	
	0.2375	0.083333	2	0.12449	3	
	0.28571	0.069176	1.375	0.13324	1	
	0.3913	0.076531	2.125	0.11829	3	
5	0.55385	0.067158	4.6	0.10148	3	5
	0.48052	0.070513	2.625	0.078737	7	
	0.5	0.074737	3.6	0.10549	3	
	0.47887	0.068651	1	0.073083	1	
	0.47826	0.069328	2	0.053906	5	
	0.58333	0.065626	1.75	0.064644	7	
	0.43678	0.066725	3.6667	0.12478	4	
	0.55224	0.071594	2	0.065406	3	
	0.44578	0.069549	2	0.069995	3	
	0.47945	0.061186	2.375	0.062193	5	

Table II: Extracted features from Genuine Test Signature Images and Acceptance Result.
(Query Distance Considered:2)

Person Sl	**F_1**	**F_2**	**F_3**	**F_4**	**F_5**	**Cluster & Acceptance Percentage**
1	0.19008	0.076161	24	0.14058	15	Cluster:1 % : 40
2	0.36905	0.055515	0.7368	0.055457	2	Cluster:2 %: 80
3	0.41667	0.071569	0.3437	0.12637	5	Cluster: 3 % : 50
4	0.47541	0.076882	2.4286	0.090737	3	Cluster: 4 %: 70
5	0.56	0.074737	3.6	0.10549	4	Cluster:5 %: 70

Table III: Extracted features from Forgery Test Signature Images and Acceptance Result.
(Query Distance Considered:2)

Person Sl	**F_1**	**F_2**	**F_3**	**F_4**	**F_5**	**Cluster & Acceptance Percentage**
1	0.235	0.067937	26	0.067	5	Cluster: 1 % : 10
2	0.34524	0.063922	1.091	0.080165	6	Cluster: 3 %: 40
3	0.42857	0.068362	5.25	0.059312	1	Cluster: 5 % : 0
4	0.37209	0.055033	2.572	0.11197	5	Cluster: 4 %: 40

*Corresponding author

Person Sl	F_1	F_2	F_3	F_4	F_5	Cluster & Acceptance Percentage
5	0.51807	0.058983	3.25	0.060097	5	Cluster: 5 %: 30

Here meaning of F_1,F_2,F_3,F_4,F_5 will be as follows

F_1 = Signature Height Width Ratio

F_2 = Signature Occupancy Ratio

F_3 = Distance Ratio at boundary

F_4 = ratio of Adjacency Columns

F_5= number of spatial symbols in the signature image.

References

[1] G. K Kiran, R. Srinivasa Rao Kunte, Sudhaker Samuel, "On-line Signature Verification System Using Probabilistic Feature Modelling", International Symposium on Signal Processing and its Applications (ISSPA), Kuala Lumpur, Malaysia, 13 - I6 August,2001.

[2] Md. Itrat Bin Shams, "Signature Recognition by Segmentation and Regular Line Detection" TENCON 2007 - 2007 IEEE Region 10 Conference Volume , Issue , Page(s):1 – 4, Oct. 30, 2007- Nov. 2, 2007.

[3] Debnath Bhattacharyya, Samir Kumar Bandyopadhyay, Poulami Das, Debashis Ganguly and Swarnendu Mukherjee , "Statistical Approach for Offline Handwritten Signature Verification" , Journal of Computer Science 4 (3): 181-185, 2008 ISSN 1549-3636, 2008

[4] Samir Kumar Bandyopadhyay**,** Poulami Das**,** Debnath Bhattacharyya , **"**Statistical Analysis towards Image Recognition" International Journal of Multimedia and Ubiquitous Engineering,Vol. 3, No. 3, July, 2008

[5] Azlinah Mohamed, Rohayu Yusof, Shuzlina Abdul Rahman, Sofianita Mutalib, "Baseline Extraction Algorithm for Online Signature Recognition", WSEAS TRANSACTIONS on SYSTEMS, Issue 4, Volume 8, ISSN: 1109-2777, April 2009

[6] Alan McCabe, Jarrod Trevathan,"Handwritten Signature Verification Using Complementary Statistical Models", JOURNAL OF COMPUTERS, VOL. 4, NO. 7, pp: 670-680,JULY 2009

[7] Emre Özgündüz,Tülin Şentürk and M. Elif Karslıgil,"Off-line signature verification and recognition by support vector machine", Eusipco-2005, 4-8 September, 2005, Antalya, Turkey, pp. 113-116. http://www.eurasip.org/Proceedings/Eusipco/Eusipco2005/deferent/papers/cr2010.pdf

[8] Rafael C. Gonzalez, Richard E. Woods "Digital image Processing ", ISBN 81-7808-629-8

[9] B.Chanda, D. Dutta Majumder, "Digital Image Processing and Analysis", ISBN: 81-203-1618-5

[10] Samit Biswas, Sudipta Bhattacharya, Sandipan Sahu, "A method to extract features from handwritten signature image for signature verification", in Proceedings of National Conference on Emerging Trends and Application in Computer Science (NCETACS-2010), St. Anthony's College, Shillong, 9-10 April, 2010, ISBN:978-81-910147-0-9, pp. 275-277

[11] Vance Faber , "Clustering And the Continuous K-means Algorithm", *Los Alamos Science,* Number 22 , 1994

[12] P.S.Bradley, Usama M.Fayyad, "Refining Initial Points for K-Means Clustering", *International Conference on Machine Learning (ICML98)*, J. Shavlik (ed.), pp. 91-99, 1998

[13] Dorin Comaniciu, Peter Meer, " Mean Shift: A Robust Approach Toward Feature Space Analysis ", IEEE transactions on Pattern Analysis and Machine Intelligence, Vol. 24, No. 5, May, 2002

[14] Pierre-Alain Moëllic, Jean-Emmanuel Haugeard, Guillaume Pittel, "Image Clustering Based on a Shared Nearest NeighborsApproach for Tagged Collections",Proceedings of the 2008 international conference on Content-based image and video retrieval, pp:269-278 ISBN:978-1-60558-070-8 , 2008

[15] Ujjwal Maulik, Sanghamitra Bandyopadhyay, "Performance Evaluation of Some Clustering Algorithms and Validity Indices", IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 24, NO. 12, DECEMBER 2002

[16] H. Liu1, J. Li and M. A. Chapman, "Automated Road Extraction from Satellite Imagery Using Hybrid Genetic Algorithms and Cluster Analysis", Journal of Environmental Informatics 1 (2) 40-47, 2003
[17] M.Liwicki and H.Bunke, "Combining Online and off-line Systems for Handwriting Recognition", 9th International conference on Document Analysis and Recognition (ICDAR 2007) , Curitiba, Brazil, Vol-I, pp:372-376
[18] Vu Nguyen, Michael Blumenstein, Vallipuram Muthukkumarasamy,Graham Leedham, "Offline Signature Verification Using Enhanced Modified Direction Features in Conjunction with Neural Classifiers and Support Vector Machines", 9th International conference on Document Analysis and Recognition (ICDAR 2007) , Curitiba, Brazil, Vol-II, pp:734-738

*Corresponding author