Performance Analysis of Codeword Histogram & Spatial Moments for Signature Recognition

H B Kekre[1] and V A Bharadi[2]

Abstract -- **In this paper we discuss the application of vector quantization to the problem of signature recognition. Vector quantization based methodology is used here to classify the signature template. Here we discuss a method for the codebook generation; this method is fast and simple. We use the codebook to generate a codeword histogram specific to the signature template. The spatial moments related to the codewords are also calculated. These parameters are used to classify the signature. Feasibility of this technique for signature recognition is discussed in his paper.**

Index Terms -- **Signature recognition, Vector quantization, Histogram, Spatial moments.**

I. INTRODUCTION

SIGNAT URE verification is an important research area in the field of person authentication [2][5]. We can generally distinguish between two different categories of verification systems: online, for which the signature signal is captured during the writing process, thus making the dynamic information available, and offline for which the signature is captured once the writing processing is over and, thus, only a static image is available. In this paper we deal with Offline signature Verification System. Here we try to develop a new set of parameters that and be used in any of the signature verification system foe classification of the signatures. This feature set is based on Vector Quantization.

Vector Quantization is a clustering technique mainly used for lossy image compression. Where we first generate a codebook having codewords that represents the image segments and then the list of codewords describes the data to be compressed. For encoding and decoding the codebook serves a very important role [1][3][4]. Here we use this approach to classify the signatures. We have built a codebook that is specific to the application and then we have used this codebook to generate a codeword histogram which is used to classify the signatures, we also used the spatial information related to the codeword. We have tested this parameter over a set of 1000 signatures, we present the result here.

The paper is organized as follows, section II is dealing with the signature template and codebook prerequisites. Section III will discuss the codebook generation process. In Section IV we discuss the process of formation of signature codeword Histograms and spatial information of the codewords. Section V is focussed on the classification process. We discuss the performance analysis & results in section VI and conclusion and future scope is discussed in section VII.

H B Kekre1is with Department of Computer Engineering at Thadomal Shahani Engineering College, Bandra (West), Mumbai-50.
hbkekre@yahoo.com 1

V A Bharadi2 is with Department of Electronics and Telecommunication at Thadomal Shahani Engineering College, Bandra (West), Mumbai-50.
vinu_bharadi@rediffmail.com 2

II. VECTOR QUANTIZATION

In vector quantization we segment the image to form a set of codewords and then we find the best match for the codeword from the codebook. The compressed data will consist of the codewords index. The better the codebook better is the compression and error (MSE) [2][3].This technique is as shown in the Fig. 1.

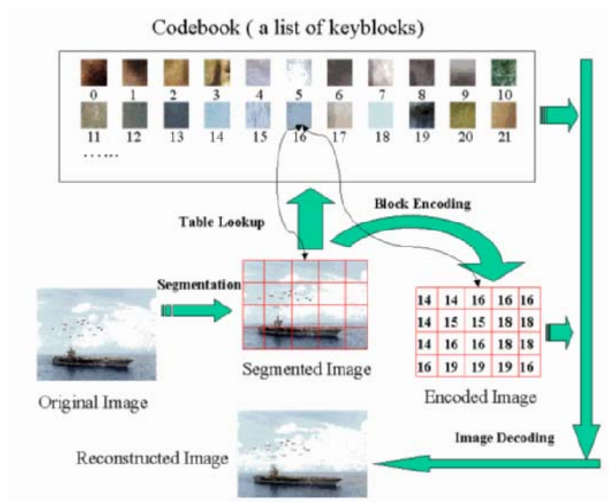


Figure 1. Encoding and Decoding of Image using Vector quantization [3].

Main Application of this technique is for data compression. Signature verification is a problem coming under the area of pattern recognition. The codebook is having the list of codewords that can be used to describe best possible match for the image segments. Signature template that is to be segmented is pre-processed [2]. We are having binary image that is obtained after pre-processing as shown in Fig. 2 and Fig3.

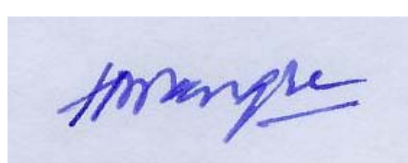


Figure 2 Original Scanned Signature.

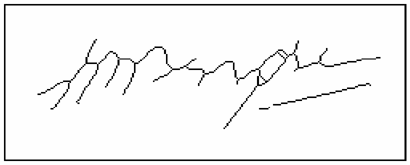


Figure 3 Normalized signature template.

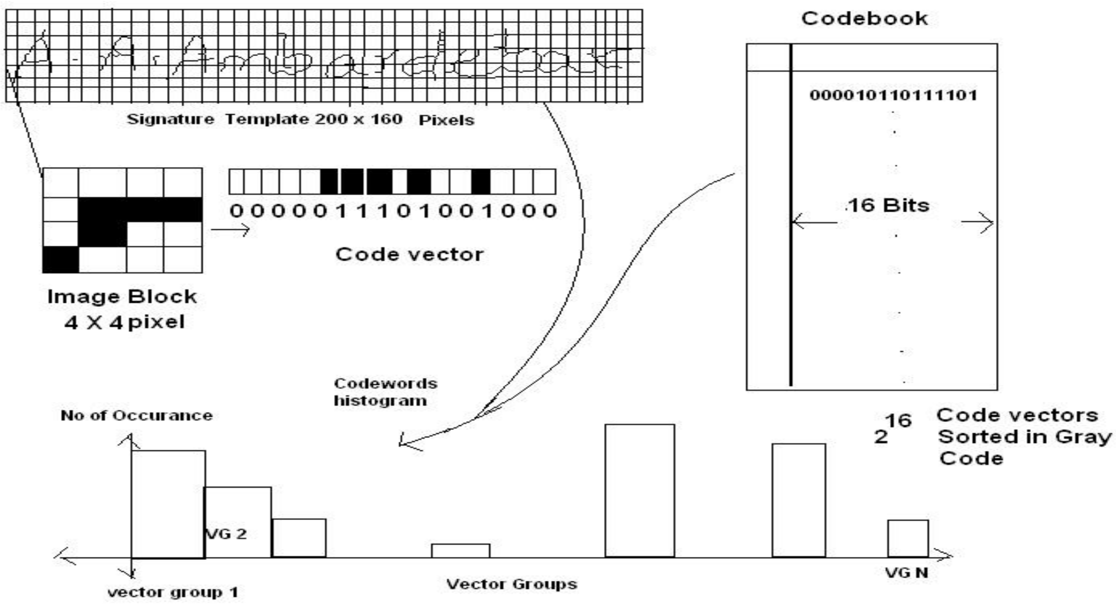


Figure 4. Vector Quantization of a signature template.

The normalized signature template is having following properties

1. It is having only black and white pixels
2. The signature is only one pixel thick, it having only single pixel thick runs.
3. The size of template is 200*160 Pixel.

For generation of codewords we segment the image into 4*4 pixel blocks, total 2000 such blocks are generated and then these blocks are used to generate the codeword histogram. This process is illustrated in the Fig. 4. The codeword histogram is used as a classifying feature; we match the codeword histograms to evaluate the similarity. This is vector quantization used for signature template. In the next section we discuss the codebook design process.

III. CODEBOOK DESIGN

The codebook design process is divided in to three parts codewords generation, codebook optimization & codewords grouping.

A. Codeword generation:

Here we have a binary image set as inputs, which are the normalized signature templates. We segment the signature template to for 4*4 pixel blocks. Total 2^{16} i.e. 65536 combinations are possible for this block size; since the signature is thinned it is a single pixel thick hence we must neglect the blocks which correspond to thickness of more then one pixel. Initially we generate a codebook having all the 65536 codewords. These codewords are all possible combinations of a 16 bit binary sequence, i.e. it starts for 0 to 65535. For generation of codeword Histogram we need to form groups of codewords which are similar or having minimum difference. To form the codeword groups, we go for codebook optimization and grouping process.

B. Codebook Optimization:

The codebook is having all the combinations of a 16 bit binary word. Two consecutive codewords may have variations in bit positions, hence to arrange the codebook in a manner that codewords are arranged with minimum bit difference in the consecutive positions, so that the consecutive blocks are similar, we rearrange the codebook. The rearrangement is done by sorting the codebook according to the GREY coding sequence. According to the property of GREY code two consecutive binary words are having difference in only one bit. Hence we have arrangement of codewords , where consecutive codewords have minimum hamming distance. Fig. 5 Shows a codeword group arranged as per the GREY coding and the Intra group distance, which is the Hamming distance of each codeword with the first codeword.

Decimal value	Codevectors	Intra Group-Distance
8	0000000000001000	0
3	0000000000000011	3
6	0000000000000110	3
7	0000000000000111	4
12	0000000000001100	1
15	0000000000001111	3
14	0000000000001110	2
11	0000000000001011	2
25	0000000000011001	2
26	0000000000011010	2
30	0000000000011110	3
29	0000000000011101	3

Figure 5. Codeword Group showing arrangement of codewords according to GREY Coding and their Intra Group Distance (Snapshot of Final Codebook)

The codebook is having all the combinations that correspond to pixel runs of thickness of more than one pixel. Hence we must remove such codewords. The invalid codewords can be removed by thinning of the blocks.

We pass each block to the thinning function the output is the thinned block having only single pixel thick runs, as these are the possible blocks that can be a part of signature template, since we have normalized and thinned binary signature template. The Thinning operation is illustrated in Fig. 6 & Fig. 7. This shows a codeblock and output of thinning function.

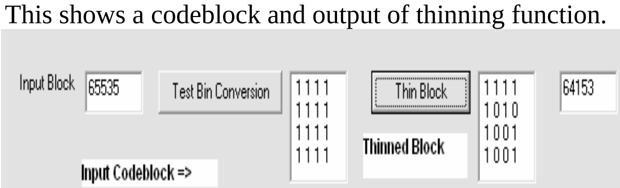


Figure 6. Thinning of a 4 X 4 pixel block

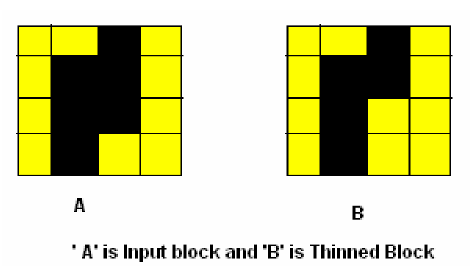


Figure 7. Graphical representation of thinning of pixel block

After thinning operation the invalid blocks are discarded and we get a set of valid codewords, we get total 11755 valid codewords. Next step is to form group the codewords to form vector groups.

C. Codewords Grouping:

In order to generate the codeword histogram we group similar codewords together to form a group, various combinations are tried in software code. Here we present grouping of 12 codewords to form total 980 groups. The grouping Algorithm is as follows

1. We start with first codeword in the codebook
2. We set initial distance to 1.
3. We search for the codeword in the codebook for match; if it is already grouped then we go for next vector.
4. If we find a codeword satisfying distance criteria then we, make the grouped flag for that vector to 1 and we find next codeword for matching.
5. This process is continued until one group is filled after that we start a new group.
6. If full codebook is covered and no vector satisfying the distance criteria is found then we increase the distance and search again.
7. In case of no vector found until we reach upper limit of distance we start a new group.
8. This process is continued until all codewords are covered.

The participants of group are codewords with minimum intergroup hamming distance and hence they represent a set of similar pixel blocks and hence similar template segments. We use this codeword groups to generate codeword Histogram. A typical Codeword Group entry is shown in Fig. 8, which lists a set of 12 codewords groups. We use two dimensional array of the size of (980, 12) to store codeword groups. Next step is to use this codebook for generation of codeword Histogram. This is discussed in detail in next section IV.

Serial	Group ID	Decimal value	Codevector	Distanc
Group	1			
1	0	8	0000000000001000	0
2	0	3	0000000000000011	3
3	0	6	0000000000000110	3
4	0	7	0000000000000111	4
5	0	12	0000000000001100	1
6	0	15	0000000000001111	3
7	0	14	0000000000001110	2
8	0	11	0000000000001011	2
9	0	25	0000000000011001	2
10	0	26	0000000000011010	2
11	0	30	0000000000011110	3
12	0	29	0000000000011101	3
Group	2			
13	1	22	0000000000010110	0
14	1	18	0000000000010010	1
15	1	17	0000000000010001	3
16	1	48	0000000000110000	3
17	1	52	0000000000110100	2
18	1	60	0000000000111100	3
19	1	56	0000000000111000	4
20	1	41	0000000000101001	6
21	1	42	0000000000101010	4
22	1	45	0000000000101101	5
23	1	44	0000000000101100	4
24	1	36	0000000000100100	3
Group	3			
25	2	37	0000000000100101	0
26	2	34	0000000000100010	3
27	2	33	0000000000100001	1
28	2	96	0000000001100000	3
29	2	97	0000000001100001	2
30	2	105	0000000001101001	3
31	2	104	0000000001101000	4
32	2	120	0000000001111000	5
33	2	112	0000000001110000	4

Figure 8. Codeword Groups formed after grouping process

IV. GENERATING CODEWORD HISTOGRAM

The Histogram generation process is shown in Fig. 3. This shows the vector quantization process. To generate the codeword histogram we follow the following steps

1. Segment the signature template into 4*4 pixel Blocks. Here we have a 200*160 pixel size template hence we get a total 2000 signature segments
2. For each segment which is 2 dimensional, generate a 1 Dimensional codeword, just by rearranging the pixel values in a row, This forms a corresponding codeword for the segment
3. Now for each codeword find its participation group ID and store this information in an array

4. Now we have total 980 Groups, so start scanning the group ID of codeword and count the number of codeword (Frequency) for each Group ID, store the information in an Array
5. This will form the frequency distribution of the codeword in each groups
6. To add the spatial Information for Each codeword group we find the spatial moments, center of gravity and centre of inertia, Gx and Ix respectively. This will add the spatial information of codeword in each group.
7. We have the signature codeword Histogram, Spatial information of each codeword , That is specific to the signature and can be compared

Fig. 9 Shows how the codewords and their spatial information is related [1]. The Equation 3.1 and 3.2 are used to calculate the spatial moments.

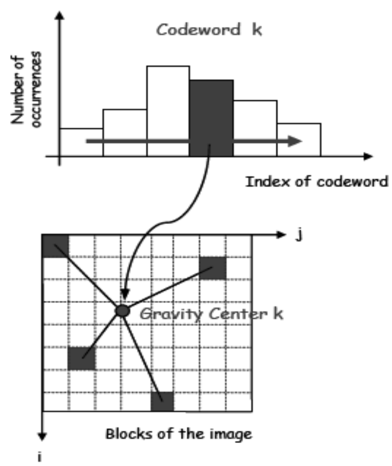


Figure 9. Codeword Histogram and Spatial information.

$$G_x = \frac{1}{M}\sum_{i=1}^{M} x_i \qquad G_y = \frac{1}{N}\sum_{i=1}^{N} y_i \qquad 4.1$$

$$I_x = \frac{1}{M}\sum_{i=1}^{n} x_i^2 \qquad I_y = \frac{1}{M}\sum_{i=1}^{n} y_i^2 \qquad 4.2$$

V. CLASSIFICATION PROCESS

We have codeword histograms and their associated spatial moments, to evaluate the distance between them we us an Euclidian distance based formula described in [32]. This is described below, given an encoded image having similar representation as a text document, image features can be extracted based on codewords frequency. The feature vector for signature template Ij and the feature vector for test signature q are given below,

For I1, It is given by I1= $\{W_{11}, W_{21}, \ldots W_{N1}\}$ 5.1

For I2, It is given by I2= $\{W_{12}, W_{22}, \ldots W_{N2}\}$ 5.2

In the histogram model, $W_{ij} = F_{ij}$, where F_{ij} is the frequency of group Ci appearing in Ij .Thus, the feature vectors I1 and I2 are the codeword histograms. The similarity measure is defined inIn the histogram model, Wij = Fij , where Fij is the frequency of Ci appearing in Ij . Similarly, Wi (q) equals the frequency of Ci appearing in q. Thus, the feature vectors I j and q are the codeword histograms. The similarity measure is defined in [3],[4] as

$$s(I2, I1) = \frac{1}{1 + dis(I2, I1)} \qquad 5.3$$

Where the distance function is

$$dis(I2, I1) = \sum_{i=1}^{N} \frac{|Wi1 - Wi2|}{1 + Wi1 + Wi2} \qquad 5.4$$

This formula is used to evaluate the similarity between two codeword Histograms, to evaluate the similarity between spatial information we use simple Euclidian distance. The testing results are discussed in next section.

VI. RESULTS

We consider signature from same as well as different users and then perform the codeword histogram generation and distance calculation. Fig. 10 shows standard test signature 1, test 1, test 2 and Fig. 11 shows their codeword histogram.

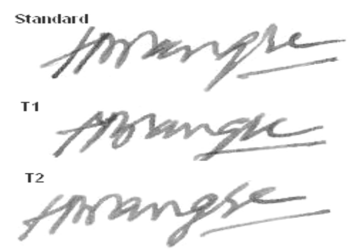


Figure 10. Standard signature and test Signature 1 & 2

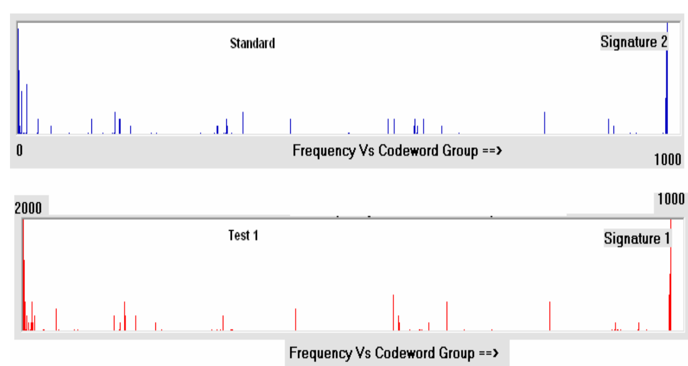


Figure 11 Frequency Vs Codeword Group Histogram

We use signatures from other persons also these signatures are as shown in Fig. 12. Now for Each Signature in test signature group we match the histogram with standard signature shown in Fig 10. We calculate the S-Score by Equation 5.3 & 5.4 and the Euclidian Distance d – for frequency. We present Euclidian distance for frequency of codewords and the moments of inertia Ix, Iy. The result is shown in Table I

TABLE I
SIMILARITY SCORE, EUCLIDIAN DISTANCE FOR FREQURNCY, AND VERTICAL AND HORIZONTAL SPATIAL MOMENTS OF INERTIA

Sign	S-Score	Euclidian Distance	Ix	Iy
T1	3.2316	15.4596	6418.207	2668.473
T 2	3.3409	18.8149	5477.747	2687.285
D1	2.4871	47.7700	7770.988	3255.156
D2	2.5723	19.1572	5806.319	3283.935
D3	2.6023	21.8860	8233.13	3504.646
D4	2.1200	33.3766	7820.745	3753.719
D5	2.6731	22.2261	7541.738	3199.428
D6	2.8159	19.3132	8500.541	2938.301
D7	2.7948	26.7207	7902.172	3157.305

One thing should be noted is that the similarity Score is Higher is better and the Euclidian distance is lower is better. The results clearly indicate that the similarity score for the codeword histograms is higher (S > 3) for signature from same user (T1 and T2) and it is low (S < 2.8) indicating dissimilarity for signatures from different user (D1 to D7).

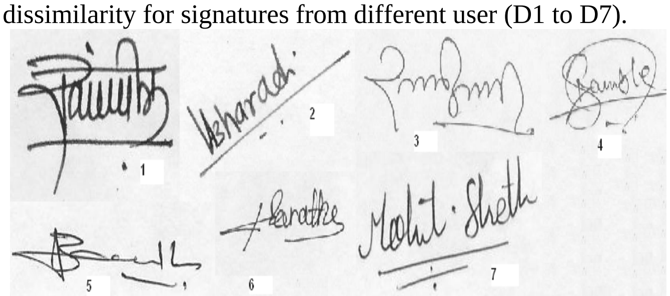


Figure 11. Signatures used for testing.

We use the codeword histogram of the signature as a classifying feature of the signature. To evaluate similarity we use the S-Score as discussed above well as the Euclidian Distance between the histograms. We evaluate FAR and FRR for both the methods simultaneously. Besides this we find the performance of the spatial moments of the codewords. For all these features we have performed total 140 tests each. We present the summary for the test for

1. VQ Codeword Histogram S-Score - (VQS) – Threshold used -3.0
2. VQ Histogram Euclidian Distance- (VQED) – Threshold used - 22
3. Spatial Moment of Gravity- (SPMG) – Threshold -6400
4. Spatial Moment of Inertia-(SPMI) - Threshold - 3090

The test results summarized in Table XXII

TABLE II
SIGNATURE VERIFICATION RESULTS FOR VQ BASED FEATURES

Case	VQS	VQED	SPMG	SPMI
Cases that Should be Accepted	70	70	70	70
Cases that actually Accepted	51	61	43	45
Cases that Should be Rejected	70	70	70	70
Cases that actually Rejected	45	47	41	44

TABLE III
PERFORMANCE METRICS FOR VQ BASED FEATURES

Sr.	Parameter	VQS	VQED	SPMG	SPMI
1	FAR	20.00	32.85	41.43	37.15
2	FRR	24.14	12.86	38.57	35.72
3	TAR	72.85	87.14	61.43	64.28
4	TRR	64.28	67.15	58.57	62.85
5	CCR	68.57	77.14	60.00	63.57
6	FCR	31.43	22.86	40.00	36.42

Performance Metrics for VQ Codeword features

We have evaluated the performance of VQ-Codeword Histogram based module with metrics such as FAR, FRR [1] [3] [4] for S-Score, Euclidian distance, the moments Gravity & Inertia. The results are presented below. Fig. 12 shows FAR, FRR plots. Equal Error Rate (EER) of 22% is reported for FAR FRR. Fig.13 shows the above mentioned plots for Euclidian Distance of VQ codeword histogram. Here we can see that the EER for FAR & FRR is 21%.

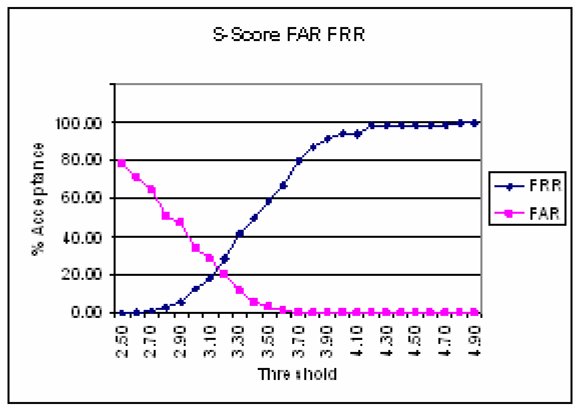


Figure 12 FAR, FRR Plot for VQ-Euclidian Distance.

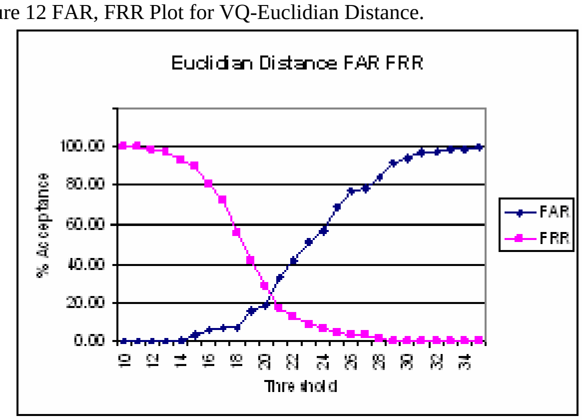


Figure 13 FAR, FRR Plot for VQ-Euclidian Distance.

The VQ codeword moments are also used as classifying feature in the signature recognition system. We evaluate the performance of these features separately the results are as shown in Fig.14 & 15.

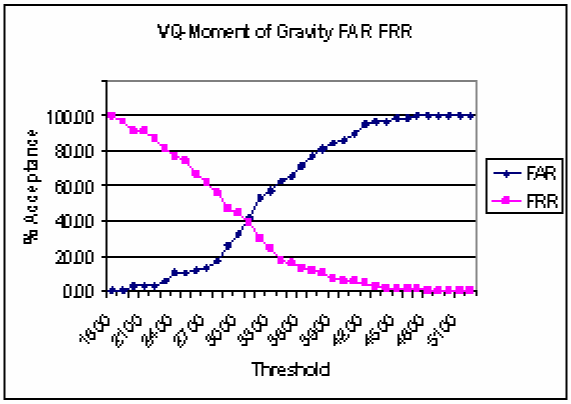


Figure 14. FAR, FRR Plot for VQ-Moment of Gravity.

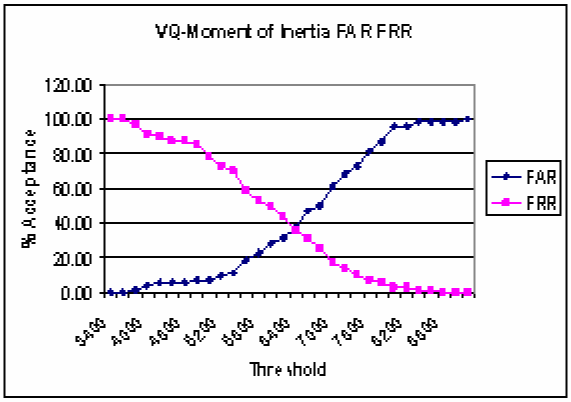


Figure 15. FAR, FRR Plot for VQ-Moment of Inertia.

These features have high EER. The codeword moment of gravity has EER of 60% for FAR, FRR, the moment of inertia has EER of 36% for FAR, FRR.

VII. CONCLUSION

In this paper we discussed application of vector quantization technique for the problem of pattern recognition. Here we consider the problem of signature recognition. The normalized signature template is segmented to form a codeword histogram. We use a new codebook generation and optimization approach to design the codebook. We form codewords groups for the purpose of generation of histograms. We have added the spatial information of the codewords, i.e. the spatial moments to the histograms. We have used existing distance measurements approach based on Euclidian distance to evaluate the similarity between the histograms.

The results clearly indicate that the S-Score for the signature from same user is higher and the Euclidian distance is lesser as compared to the signatures from different user. Hence the patterns are classified. We can use this technique to classify the signatures.

We have achieved the Correct Classification Ratio (CCR) in the range of 60 to 77%. Here we have not used any training mechanism we are just analyzing the feature set over a database collected from 100 persons. This accuracy can be improved by using training mechanism based kohenen's Self organized feature maps (SOFM) [1] and neural networks based approach these parameters can be used for signature recognition.

VIII. REFERENCES

[1] Ordonez, G. Cazuguel, J. Puentes B. Solaiman, C. Roux, "Spatial-textural Medical Image Indexing based on Vector Quantization" Proceedings of the 25th Annual International Conference of the IEEE EMBS, 2003,0, 802- 805,A,A,GB.

[2] A. Kholmatov , "Biometric Identity Verification Using On-Line & Off-Line Signature Verification", Master of Science thesis, Sabanci University,S 2003

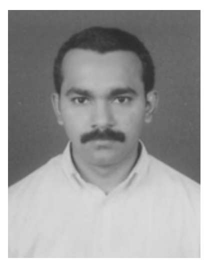
[3] L. Zhu, A. Rao & A. Zhang, "Theory of Keyblock-based Image retrieval", ACM Journal, Volume V, No. N, March 2002, PP 1-32

[4] L. Zhu, A. Rao & A. Zhang, "Using Thesaurus To Model Keyblock Based Image Retrieval", IEEE International Conference on Multimedia and Expo. 2001, ISBN 0-7695-1198-8/01

[5] A. K. Jain, A. Ross, S. Prabhakar, "An Introduction to Biometric Recognition", IEEE Transactions on Circuits and Systems for Video Technology, Vol. 14, No. 1, January 2004

IX. BIOGRAPHIES

Hemant Kekre - was born in India on April 4 1935. He graduated from the Jabalpur university, India, and completed his M. Tech from IIT Mumbai in 1960. He has completed PhD from IIT Mumbai in the field of system identification. His employment experience included the IIT Mumbai, Thadomal Shahani Engineering College. His special fields of interest included signal processing, Image processing, statistical theory. He has published over 150 papers in various national as well as international level conferences on varios topics such as System Identification, Simulation and Modelling, Digital Signal Processing, Image Processing, Computer Networking and Communications.



Vinayak Bharadi - was born in India on March 25 1981. He graduated from the Mumbai university, India, in Aug 2002 and pursuing his M. E. Electronics & telecomm from Mumbai university . His employment experience included the Government Polytechnic Mumbai, Thadomal Shahani Engineering College. His special fields of interest included signal processing, Image processing, computer programming. He has published 4 papers in various national level conferences and 5 papers in international level conferences on Signature recognition, Cluster based features, Image & Signal processing.