Chapter 3

Spatial Domain based Off-line Signature Identification.

3.1 Introduction

The signature is a behavioural trait used to authenticate a person by hand writing styles of his/her name. In this chapter signature identification algorithms viz., (i) Efficient Offline Signature Identification based on Global Features (EOSIGF) and (ii) Cross-validation for Graph Matching based Offline Signature Verification (CGMOSV) for skilled and random forgeries are discussed. In EOSIGF, the normalization, noise reduction, thinning and skelitazition are the steps used in pre-processing. The global features used are aspect ratio, maximum horizontal histogram and maximum vertical histogram, horizontal centre and vertical centre of the signature, endpoints of the signature and signature area. The ED is used to compare test signature with signatures in database.

In CGMOSV algorithm, the database signatures are pre-processed, in which signature extraction method is used to obtain high resolution for smaller normalization box. The ED among the signature sample of each person in the database are computed. The Decision Factor (DF) of each sample is computed using mean of ED and SD. The Decision Factor Average (DFA) of each person is used for Cross-validation to set the reference signatures and optimum decision threshold value is determined for the reference signatures. The given test signature is pre-
processed and a test feature ie., Minimum ED of Test (M\_EDT) signature with reference signature is considered which is then compared with the Maximum ED of Reference (M\_EDR) signature value to authenticate the test signature.

3.2. Proposed model of EOSIGF

In this section the proposed EOSIGF model with block diagram as shown in Figure 3.1 is discussed in detail.

![Block diagram of EOSIGF model.](image)

**Fig 3.1: Block diagram of EOSIGF model.**

3.2.1 Signature Database:

The Grupo de Procesado Digital de Senales (GPDS) [127] database is considered and it has one hundred and sixty signers, with twenty four genuine signatures and twenty four forgery signatures per signer. The signature database for the proposed model is created by considering seventy signers with fifteen genuine samples per signer ie., one thousand and fifty genuine samples. The twentieth genuine signature sample of
each signer is considered as the test signature to compute FRR. The values of the FAR are computed by considering sixty signers with any one skilled forgery signature of each signer as the test signature. The genuine signatures and forgery signatures of one person is shown in Figure 3.2.

(a) Genuine Signature Samples of one person

(b) Forgery Signature Samples of one person

Fig. 3.2: Signature samples of GPDS database of one person.
3.2.2 Preprocessing:

The signatures are preprocessed to convert a random signature size into uniform size of signatures. The preprocessing stage has normalization, noise reduction and skeletonization.

(i) Normalization: The signature is rotated by angle $\theta$ about the centroid $(x, y)$ to maintain a uniform base line. The angle is computed by maximizing the deviation of data in one direction.

The mean $\mu_x$ of the $x$ series is calculated using Equation (3.1)

$$
\mu_x = \frac{\sum_{i=0}^{T} x_i}{T} \hspace{2cm} (3.1)
$$

Where $T$ is the total number of pixels

$x_i$ intensity of pixel

The deviation of $x$ series from $\mu_x$ is calculated using Equation (3.2)

$$
\sigma_x = \sqrt{\frac{\sum_{i=0}^{T} (x_i - \mu_x)^2}{T}} \hspace{2cm} (3.2)
$$

Equation (3.3) is maximized to obtain the maximum deviation.

$$
\sum_{i=0}^{T} \left( x_i' - \mu_x \right)^2 \hspace{2cm} (3.3)
$$

Where $x_i'$ indicates a rotated $x$ value. A rotation about a point is expressed as Equation (3.4)

$$
x_i' = (x_i - \mu_x)\cos(\theta)+(y_i - \mu_y)\sin(\theta)+\mu_x \hspace{2cm} (3.4)
$$
Equation (3.5) function $f(\theta)$ is obtained by substituting the Equation (3.4) in Equation (3.3)

\[
f(\theta) = \sum_{i=0}^{T} \left[ (x_i - \mu_x)\cos(\theta) + (y_i - \mu_y)\sin(\theta) + \mu_x - \mu_x \right]^2
\]

\[\text{.......................... (3.5)}\]

\[
= \sum_{i=0}^{T} a_i^2 \cos^2(\theta) + 2a_i b_i \cos(\theta)\sin(\theta) + b_i^2 \sin^2(\theta)
\]

\[
= \cos^2(\theta)\sum_{i=0}^{T} a_i^2 + 2\cos(\theta)\sin(\theta)\sum_{i=0}^{T} a_i b_i + \sin^2(\theta)\sum_{i=0}^{T} b_i^2
\]

\[
= \cos^2(\theta)p + 2\cos(\theta)\sin(\theta)Q + \sin^2(\theta)R
\]

\[\text{.....................(3.6)}\]

Where $a_i = x_i - \mu_x$ and $b_i = y_i - \mu_y$.

Taking the derivative of Equation (3.6) yields Equation (3.7)

\[
f'(\theta) = 2Q \cos^2(\theta) - 2P \cos(\theta)\sin(\theta)
\]

\[
+ 2R \cos(\theta)\sin(\theta) - 2Q \sin^2(\theta)
\]

\[\text{......................... (3.7)}\]

The roots obtained from Equation (3.7) are expressed as Equation (3.8) and Equation (3.9)

\[
\Theta = \pm \cos^{-1}\left\{ \pm \frac{1}{\sqrt{2}} \sqrt{1 + \frac{(P-R)}{\sqrt{P^2 + 4Q^2 - PR + R^2}}} \right\}
\]

\[\text{..........................(3.8)}\]

and

\[
\Theta = \pm \cos^{-1}\left\{ \pm \frac{1}{\sqrt{2}} \sqrt{1 + \frac{(R-P)}{\sqrt{P^2 + 4Q^2 - PR + R^2}}} \right\}
\]

\[\text{..........................(3.9)}\]

We adopt the smallest value for $\Theta$ from Equations (3.8) and Equation (3.9) which will result in a maximum value of $f(\Theta)$. 

55
(ii) **Skeletonization:**

Skeletonization makes the extracted features invariant to image characteristics like quality of pen and paper. A simplified version of a skeletonization technique is the image is segmented into $3 \times 3$ matrix which results with nine elements. The center element is considered as the reference element with eight neighbors. The matrix along the contour line of signature is considered with at least one white pixel in eight neighbors and two black pixels in eight neighbors are removed to make the image unitary matrix.

(iii) **Noise reduction:**

The dirt on cameras or scanner lens, imperfections in the scanner lighting etc., introduces noise in the scanned signature images. A Gaussian filtering function is used to remove noise in the signatures.

### 3.2.3 Feature Extraction:

(i). **Maximum Horizontal Histogram and Maximum Vertical Histogram:** The horizontal histograms are calculated for each row and highest value is considered as maximum horizontal histogram. The vertical histograms are calculated for each column and highest value is considered as maximum vertical histogram.

(ii). **Horizontal and Vertical centers** of the signature are calculated using Equation (3.10) and (3.11).
(iii). *Edge points of the signature:* The termination point if discontinuity in the signature.

(iv). *Signature Area:* The number of pixels in the skeletonized image area is considered as signature area.

(v). *Aspect Ratio:* It is the ratio of Signature height to Signature width. The signature height and width may vary, but the aspect ratio of the individual will remain approximately same.

### 3.2.4 Matching:

The extracted features of test image are compared with features of database image using ED given in Equation 3.12. The value of ED is compared with prefixed threshold value to validate signature.

\[
ED = \sqrt{\sum_{j=1}^{p} (x_j - y_j)^2} \\
\text{................................. (3.12)}
\]

Where,

\( p \) - Total number of features

\( x_i \) - Feature of test image and \( y_i \) - Feature of images in database
3.3 Algorithm:

The algorithm of EOSIGF system in which signature features are extracted based on global parameters is given in Table 3.1.

The objectives are

(i) To identify test signature by comparing with signature in the database.

(ii) To reduce percentage FRR and FAR.

Table 3.1: EOSIGF Algorithm.

| • Input: Test Signature, Reference signature from database |
| • Output: Identified Signature. |
| i. Preprocessing the database signatures. |
| (a) Normalization |
| (b) Noise reduction |
| (c) Skelitization |
| ii. Global Features Extraction |
| (a) Max Horizontal Histogram and Max Vertical Histogram |
| (b) Horizontal and Vertical centre |
| (c) Edge point |
| (d) Signature Area |
| (e) Aspect Ratio |
| iii. Steps from i and ii are repeated for test signature |
| iv. Compare test signature features with database signature features using Euclidian Distance. |

3.4 Performance Analysis:

The signature database has seventy persons with fifteen genuine signatures per person. The FRR is computed by comparing genuine test signature with genuine signatures in the database. The value of FAR is
computed by comparing skilled forgery test signature with genuine signature in the database. The values of FRR and FAR for the proposed EOSIGF algorithm is compared with the existing algorithms such as Vu Nguyen et al.,[128] Jesus F Vargas et al.,[129] Javed Ahmed Mahar et al.,[130] Sepideh Afsardoost et al.,[131] and Tai-ping zhang et al.,[132] is given in Table 3.2. It is observed that the values of FRR and FAR are better in the case of proposed algorithm compared to existing algorithms.

Table 3.2: Comparison of Existing Algorithms with Proposed Algorithm

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Description of Features</th>
<th>%FRR</th>
<th>%FAR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vu Nguyen et al.,[128]</td>
<td>Texture features</td>
<td>17.25</td>
<td>16.90</td>
</tr>
<tr>
<td>Jesus F Vargas et al.,[129]</td>
<td>Histogram</td>
<td>5.05</td>
<td>7.35</td>
</tr>
<tr>
<td>S Afsardoost et al.,[131]</td>
<td>Geometric Features</td>
<td>14</td>
<td>15</td>
</tr>
<tr>
<td>Tai-ping zhang et al.,[132]</td>
<td>Signature Envelop</td>
<td>18.54</td>
<td>15.80</td>
</tr>
<tr>
<td><strong>Proposed algorithm</strong></td>
<td><strong>Global and Histogram features</strong></td>
<td><strong>5.4</strong></td>
<td><strong>4.6</strong></td>
</tr>
</tbody>
</table>

The performance of the proposed algorithm is better compared with existing algorithm since we have used both global features and histogram features.
3.5 Proposed Model of CGMOSV:

In this section the block diagram of Cross-validation for Graph Matching based Offline Signature Verification is shown in Figure 3.3 is discussed in detail.

Fig 3.3: Block diagram of CGMOSV system.
3.5.1 **Signature database:** The signature samples are considered from GPDS database. The signature database for the proposed model is created by considering seventy signers with fifteen genuine samples per signer i.e., one thousand and fifty genuine samples. The twentieth genuine signature sample of each signer is considered as the test signature to compute FRR. The values of the FAR are computed by considering sixty signers with any one skilled forgery signature of each signer as the test signature.

3.5.2 **Pre-processing:**

The principal objective of pre-processing is to obtain a transformed image, which is more suitable than the original image to enhance the quality of signature. The block diagram of pre-processing is shown in Figure 3.4 include (i) Noise removal (ii) Rotation (iii) Smoothing (iv) Thinning (v) Signature extraction and (vi) Normalization.

**(i) Noise removal:** The unwanted noise may be present in the signature due to transmission of image, dirt on the scanner etc., which results in salt and pepper noise, which can be removed using median filter. The signature samples with salt and pepper noise and after removal of noise are as shown in Figure 3.5.

**(ii) Rotation:** Rotation of a signature is necessary as time domain approaches are sensitive to angle variations compared to frequency domain approaches. It removes skewness, that is coincides the axis of mass of inertia of all the signatures to the same horizontal axis. The edge
of the signature is first detected using canny edge detector to which Radon transform [133] is applied and the angle of rotation is measured in anticlockwise direction. The signature is then rotated clockwise to remove skewness as shown in Figure 3.6.

![Block diagram of Preprocessing](image)

Fig 3.4: Block diagram of Preprocessing.

![Signature with salt and pepper noise](image) and ![Signature after noise removal](image)

Fig 3.5: (a) Signature with salt and pepper noise and (b) Signature after noise removal.

![Rotated Signature](image)

Fig 3.6: Rotated Signature.
(iii) **Smoothing:** The additive white noise due to presence of background after rotation of signature may be present and is removed using smoothing process. The adaptive filter is used for smoothing which preserves edges and high frequency components of the signature as shown in Figure 3.7.

![Smoothed Signature](image)

Fig 3.7: Smoothed signature.

(iv) **Thinning:** The morphological operations are used in thinning for the reduction of data, which in turn reduces computational time. A fast parallel Zhang-Suen algorithm [134] is used for thinning as it preserves pixel connectivity and end points. It aimed at deleting the boundary points and the corner points by comparing foreground and background pixels. It reduces the signature to a skeleton of unitary thickness as shown in Figure 3.8.

![Thinned Signature](image)

Fig 3.8: Thinned Signature.

(v) **Signature Extraction:** The smallest box which contains only signature is extracted by removing an additional background created due to rotation. It is determined by the height and width of the signature and is then cropped to the measured dimensions. The allowance for little background is given in all directions so that the signatures do not touch
the boundary of the box as shown in Figure 3.9. It increases the occurrence of foreground signature compared to background space i.e., high resolution of signature for smaller normalization box and hence reduces error rate.

![Extracted Signature]

Fig 3.9: Extracted Signature.

(vi) **Normalization:** is required to resize the exact signatures size to uniform size for all signatures in the database. The normalized image size of 32*64 for the proposed method is considered and is as shown in Figure 3.10.

![Signature resized to 32 * 64.]

Fig 3.10: Signature resized to 32 * 64.

### 3.5.3 Signature Features:

The signature varies with time and mood which results in slight variations in shape of genuine signatures available. To measure the variations the following features are computed:

1. Signature height
2. Signature width
3. Aspect ratio
4. Signature area
5. Maximum horizontal and vertical histograms
6. Horizontal and vertical centre

64
3.5.4 Selection of Reference Signatures:

The reference signatures are selected from fifteen genuine signature samples of each signer in the database template using mean and Standard Deviation (SD) of ED.

The following steps are used to select reference signatures.

(i) Let the genuine signature samples per signer are $S_i$, where $i = 1$ to $n$.

The features of $S_1$ are compared with features of every signature sample of signer using $ED$ to compute $ED_{1,1}$, $ED_{1,2}$, \ldots $ED_{1,n}$ as shown in Figure 3.11.

\begin{figure}
\centering
\includegraphics[width=0.5\textwidth]{ed_comparisons.png}
\caption{Computations of ED between samples.}
\end{figure}

(ii) The discrepancies among signature sample $S_i$ of a signer are observed by computing Mean, SD and Decision Factor (DF) of ED’s using Equation 3.13, Equation 3.14 and Equation 3.15.

\[
ED_M (1) = \frac{ED_{1,1} + ED_{1,2} + ED_{1,3} + \ldots + ED_{1,n}}{n}
\]
\[
ED_M (2) = \frac{ED_{2,1} + ED_{2,2} + ED_{2,3} + \ldots + ED_{2,n}}{n}
\]

\[
ED_M (n) = \frac{ED_{n,1} + ED_{n,2} + ED_{n,3} + \ldots + ED_{n,n}}{n}
\]

In general

\[
ED_M (i) = \frac{\sum_{j=1}^{n} ED(i, j)}{n}
\]

........................................ (3.13)

Where \( n \) is number of signature samples of single signer

\[
SD_i = \sqrt{\frac{\sum_{j=1}^{n} (ED_{i,j} - ED_M (i))^2}{n-1}}
\]

........................................... (3.14)

For SD of signature \( i = 1 \) with other samples of signature ie., \( j = 1 \) to \( n \)

\[
SD_1 = \sqrt{\frac{(ED_{1,1} - ED_M (1))^2 + (ED_{1,2} - ED_M (2))^2 + (ED_{1,3} - ED_M (3))^2 + \ldots + (ED_{1,n} - ED_M (n))^2}{n-1}}
\]

The decision factor is the ratio of mean and SD of each signature sample.

\[
DF_1 = \frac{ED_M (1)}{SD(1)}
\]

\[
DF_2 = \frac{ED_M (2)}{SD(2)}
\]

\[
DF_n = \frac{ED_M (n)}{SD(n)}
\]

In general

\[
DF_i = \frac{ED_M (i)}{SD(i)}
\]

........................................... (3.15)
(iii) Compute Average DF i.e., Decision Factor Average (DFA) using Equation 3.16.

\[
DFA = \frac{DF_1 + DF_2 + \ldots + DF_n}{n}
\]

(iv) Compare DF’s of each signature with DFA. The signature with value of DF less than or equal to DFA are considered as reference signatures, else signature samples are discarded. The value of FRR and FAR decreases by selection of reference signatures say \(RS_i, RS_2, RS_3, \ldots, RS_m\). The total number of reference signatures say \(m\) of single signer are less than or equal to total number of original signature samples of single signer.

3.5.5 Maximum Variations among reference signatures:

The ED between pairs of single signer in the database are computed. The total number of ED’s for \(m\) reference signatures is given in Equation 3.17

Total number of ED’s = \(^nC_2 = \frac{n!}{(n-2)! \times 2!}\) \hspace{1cm} \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots (3.17)

For example if reference signatures of single signer say three, then total number of ED’s is \(^3C_2 = 3\) i.e., \(ED_1, ED_2, ED_3\).

The maximum value of ED’s of reference signatures are considered as maximum variations among reference signatures, which acts as final
feature of single signer ie., Maximum ED of References (M_aEDR).
Similarly compute M_aEDR for all signers in the database.

### 3.5.6 Test Signature Feature:

The test signature is preprocessed and features, such as signature height, signature width, aspect ratio, maximum horizontal histogram, maximum vertical histogram and horizontal and vertical centers are computed. The test signature features are compared with features of reference signatures of each signer using ED. For example, if three reference signatures of single signer are considered, then ED’s between test signature and reference signatures are d_1, d_2 and d_3. The Minimum ED of Test (M_iEDT) is noted, which gives the feature of test signature. Similarly the values of M_iEDT’s for all signers are computed.

### 3.5.7 Matching:

The value of M_iEDT is compared with the values of M_aEDR. The test signature is matched with signer in the database, if M_iEDT less than M_aEDR the test signature is a matched, else miss match.

### 3.6 Algorithm of CGMOSV:

The genuine signature identification using proposed algorithm is given in Table 3.3

Given test signature and large Signature database.

The objectives are:

- To recognize a person effectively using signature
- To decrease the value of EER.
Table 3.3: CGMOSV algorithm.

- **Input:** Test signature, genuine and skilled signature database.
- **Output:** Matched signature.

1. The signatures in database are preprocessed to obtain exact signature area.
2. The features such as signature height, signature width, aspect ratio, maximum horizontal histogram, maximum vertical histogram and horizontal and vertical centers are computed.
3. The ED among signature samples of each person are computed.
4. The values of mean, SD and DF are computed for each sample of each signer.
5. Compute DF average DFA of all samples of single user.
6. The reference signatures from original signature set of single signer are considered by comparing DF’s with DFA.
7. The ED between reference signature pairs are computed. The $M_{aEDR}$ of reference signature is considered as final features.
8. The steps 1 and 2 are performed for test signature.
9. The test signature features are compared with features of reference signature in the database using ED say $d_1, d_2, d_3, ..., d_m$.
10. The $M_{aEDT}$ is noted, which is the feature of test signature.
11. The test signature is matched with signature in the database if $M_{iEDT} \leq M_{aEDR}$, else not match.
3.7 Performance Analysis:

The proposed algorithm is tested with performance parameters such as FRR, FAR_S, FAR_R and EER based on genuine, skilled forgery and random forgery signatures tests. The performance parameters are also computed with different sizes of normalization boxes i.e., 8*16, 16*32, 24*48, 32*64, 40*80 and 48*96 based on preprocessed signature.

(i) Genuine Signature Test: The genuine signatures database is created by considering seventy signers with fifteen samples per person i.e., one thousand and fifty signatures in the database. The twentyeth signature of each person is considered in the test section to compute FRR.

(ii) Skilled forgery signature test: The skilled forgery samples of seventy persons with one sample per person are considered in test section to compute FAR_S.

(iii) Random forgery signature test: The random forgery signatures of seventy persons with one sample per person are considered in the test section to compute FAR_R.

The variations of FRR, FAR_S and FAR_R with threshold are plotted in Figure 3.12 for normalization box sizes of 8*16, 16*32, 24*48, 32*64, 40*80 and 48*96. As threshold increases FRR decreases and both the FAR’s increases. We noticed that, the values of FAR_R are lower than the values of FAR_S, since the impostors providing skilled forgeries have knowledge of original signature while the random forgers do not have it. The values of EER_S for skilled forgeries and EER_R for random forgeries
are noted from Figure 3.12 for different normalization box sizes. The
variation of EER_S and EER_R for different normalization box sizes is
shown in Figure 3.13. The values of EER_S and EER_R are normally
decreases with increase in the size of normalization box. The values of
$EER_S$ and $EER_R$ for existing algorithm Offline Signature Verification
using Graph Matching (OSVGM) and proposed algorithm are tabulated in
Table 3.4 and Table 3.5 respectively for different normalization box sizes.
The EER values for both skilled and random forgeries are less in the case
of proposed algorithm compared to existing algorithm for all sizes of
normalization boxes since graph matching and cross validation principle
are used. The values of EER for skilled forgery of proposed algorithm are
compared with existing algorithms presented by I A Ibrahim [135] and
Jon Almazan et al., [136] are tabulated in Table 3.6. It is observed that
the proposed algorithm has lower EER value compared to existing
algorithms.

Table 3.4: Comparison of EER of skilled forgeries of existing and
proposed method.

<table>
<thead>
<tr>
<th>Normalization box</th>
<th>OSVGM[135]</th>
<th>Proposed method</th>
</tr>
</thead>
<tbody>
<tr>
<td>8 × 16</td>
<td>37.5</td>
<td>32.67</td>
</tr>
<tr>
<td>16 × 32</td>
<td>35.2</td>
<td>25.33</td>
</tr>
<tr>
<td>24 × 48</td>
<td>35.0</td>
<td>24.0</td>
</tr>
<tr>
<td>32 × 64</td>
<td>29.0</td>
<td>21.33</td>
</tr>
</tbody>
</table>
Fig 3.13: Variations of EER with size of normalization box.

Table 3.5: Comparison of \( \text{EER} \) of Random forgeries of existing and proposed method.

<table>
<thead>
<tr>
<th>Normalization box</th>
<th>OSVGM[135]</th>
<th>Proposed method</th>
</tr>
</thead>
<tbody>
<tr>
<td>8 x 16</td>
<td>28.0</td>
<td>25.0</td>
</tr>
<tr>
<td>16 x 32</td>
<td>18.5</td>
<td>15.88</td>
</tr>
<tr>
<td>24 x 48</td>
<td>20.0</td>
<td>13.63</td>
</tr>
<tr>
<td>32 x 64</td>
<td>15.3</td>
<td>11.75</td>
</tr>
</tbody>
</table>

Table 3.6: Comparison of existing algorithms with proposed algorithm

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Description of Features</th>
<th>EER</th>
</tr>
</thead>
<tbody>
<tr>
<td>I A Ibrahim [135]</td>
<td>Graph Matching</td>
<td>29.0</td>
</tr>
<tr>
<td>Jon Almazan et al.,[136]</td>
<td>Non Rigid Feature</td>
<td>24.16</td>
</tr>
<tr>
<td><strong>Proposed algorithm</strong></td>
<td><strong>Cross Validation and Graph Matching</strong></td>
<td><strong>21.33</strong></td>
</tr>
</tbody>
</table>

The advantage of proposed algorithm

(i) The number of samples per person in the database are reduced since most similar signature samples retained by rejecting dissimilarity signature samples using reference signature set.
(ii) The memory required is less to store database template.

The minor limitations may be little increase in false rejection rate of genuine signatures.

3.8 Summary:

In this chapter two signature identification algorithms based on global features with histogram values and graph matching with cross validation are discussed. In the first signature identification algorithm global features such as height, width, aspect ratio, edge point, signature area and signature centre along with horizontal and vertical histogram values are used in identifying a signature. The test signature features are compared with features of signature in database using ED. In the second identifying algorithm the signatures are compared using graph matching and ED to measure similarity between them. The cross validation principle is used to eliminate few signature samples of single person having more values of dissimilarity. The performance parameters such as FRR, FAR and EER are computed for both signature algorithms.