Chapter 4

Reduced Shape Signature

4.1 Background

Gesture recognition has provided an intuitive and natural interaction environment. With the advancement of image acquisition and recognition technology, hand gesture has gained popularity in human-machine interaction systems. In those system, accurate recognition of hand gesture is a key requirement for proper functioning. In order to recognise and classify hand gestures, we need to obtain the most relevant information from the input/processed image and represent that information in a lower dimensionality space, i.e. features. In this chapter, we have proposed a new feature extraction method—reduced shape signature—for effective hand gesture classification.

4.2 Related Literature

Once a hand is localized in the obtained image, features are extracted for effective shape classification. Numerous shape descriptors are found in the literature. These shape descriptors are generally classified into two major groups—region-based approach and boundary-based approach (also known as contour-based approach) [133].

In the region-based approach, shape descriptors are obtained using all the pixels of the corresponding shape. Region-based approach use moment descriptors to describe the shape. Some of these moment descriptors include geometrical moments [68], Zernike moments, Legendre moments [69], pseudo-Zernike moments [70] and Tchebichef moments [71]. It was found that Zernike moments are orthogonal, compact and rotation invariant. It outperforms the other moment based descriptors in terms of overall performance [134]. However, the region-based shape descriptors are extracted from spatial domain, and hence, they are sensitive to noise and shape variations. Apart from these approaches, few other region-based approach use histogram of oriented gradients [135], scale-invariant feature transform [136] and local binary pattern [137] based features obtained from texture and shape of the hand gesture.

In the boundary-based approach, shape descriptors are obtained using the boundary pixels of the shape. Some of the well-known boundary based shape descriptors includes chain code [77], finger geometric [138, 139], curvature scale space [76], shape context [140], localized contour sequence [141], time-series curve [82, 142], convex hull [143], shape signatures [72, 73]. Majority of these works were based on interpretation in space domain. Some work [73–75] with representation of descriptor on transform domain were also proposed.

Among these boundary based shape descriptors, shape signatures are of great importance in image retrieval and classification. Shape signatures are a 1D function used to represent a complex 2D shape in terms of compact and perceptual feature set. Shape signatures can be either real or imaginary. Different forms of shape signatures like radial distance, complex coordinates, angular function, triangular centroid area, angular radial coordinates, chord-length distance and farthest point distance are in existence [72, 73].

Existing shape signatures are sensitive to irregularities in contour, e.g. false cracks, twists, especially around the wrist region. For a hand gesture, it is observed that each shape differs only in finger attributes (i.e. number, distributions, etc.) while the palm region remains the same. Thus, instead of considering the entire boundary of a hand gesture, boundary corresponding to finger region only is considered in this work. This not only removes the aforementioned problem due to wrist region but also reduces the number of feature sets essential to describe a hand gesture.

4.3 Reduced Shape Signature

Reduced Shape Signature (RSS) is a $1D$ distance function denoted by $R(k)$. It is obtained from the reduced boundary $\beta_r(M_h)$ and a shifted centroid (Cx_i, Cy_i). It is denoted by $R(k)$ and is given by

$$R(k) = \sqrt{(x_r(k) - Cx_i)^2 + (y_r(k) - Cy_i)^2} \tag{4.1}$$

where $x_r(k)$ and $y_r(k)$ are the x and y coordinates of k^{th} number of boundary point in reduced boundary $\beta_r(M_h)$. Reduced boundary $\beta_r(M_h)$ is obtained by finding the appropriate number of intersection points (NIPs) between original boundary $\beta(M_h)$

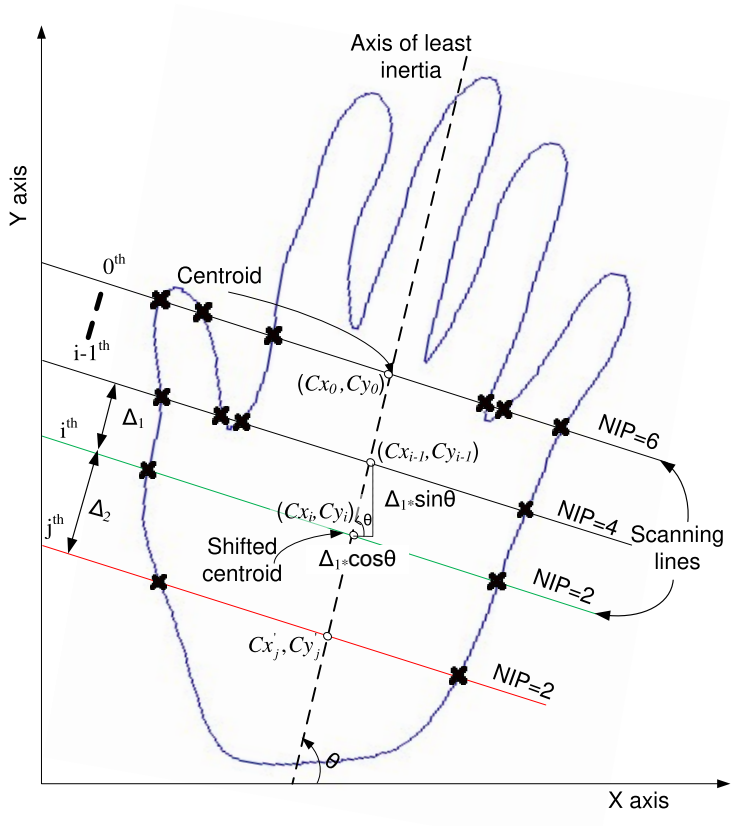

FIGURE 4.1: Illustration of finger extraction method.

and scanning line. Initially scanning line passes through the centroid (Cx, Cy) with an angle perpendicular to axis of least inertia as shown in Figure 4.1. Axis of least inertia is a line such that the sum of squared distances from all boundary points to that line is minimum. If a shape is represented by its contour, centroid [144] is given as

$$C_x = \frac{1}{6A}\sum_{i=0}^{N-1}(x(i)+x(i+1)) * (x(i)y(i+1) - x(i+1)y(i)) \tag{4.2}$$

$$C_y = \frac{1}{6A}\sum_{i=0}^{N-1}(y(i)+y(i+1)) * (x(i)y(i+1) - x(i+1)y(i)) \tag{4.3}$$

where N is the pixel count of the boundary $\beta(M_h)$, (x_i, y_i) are the x, y coordinates corresponding to i^{th} boundary point. Further, contour's area A is given by

$$A = \frac{1}{2}\left|\sum_{i=0}^{N-1}(x(i)y(i+1) - x(i+1)y(i))\right|. \tag{4.4}$$

The angle (α) of least inertia axis with respect to positive x–axis is calculated as discussed in [145].

$$\alpha = \frac{1}{2}\tan^{-1}\frac{b}{a-c}, \quad \frac{-\pi}{2} < \alpha < \frac{\pi}{2} \tag{4.5}$$

where

$$a = \sum_{i=0}^{N-1}(x(i) - C_x)^2 \tag{4.6}$$

$$b = 2\sum_{i=0}^{N-1}(x(i) - C_x)(y(i) - C_y) \tag{4.7}$$

$$c = \sum_{i=0}^{N-1}(y(i) - C_y)^2. \tag{4.8}$$

The equation of scan line passing through centroid point (Cx_i, Cy_i) and perpendicular to axis of least inertia is given as

$$y = m(x - Cx_i) + Cy_i \tag{4.9}$$

where m is the slope of line and it is equal to $1/tan\alpha$. The angle θ of axis of least inertia w.r.t. x–axis in anticlockwise direction is given as

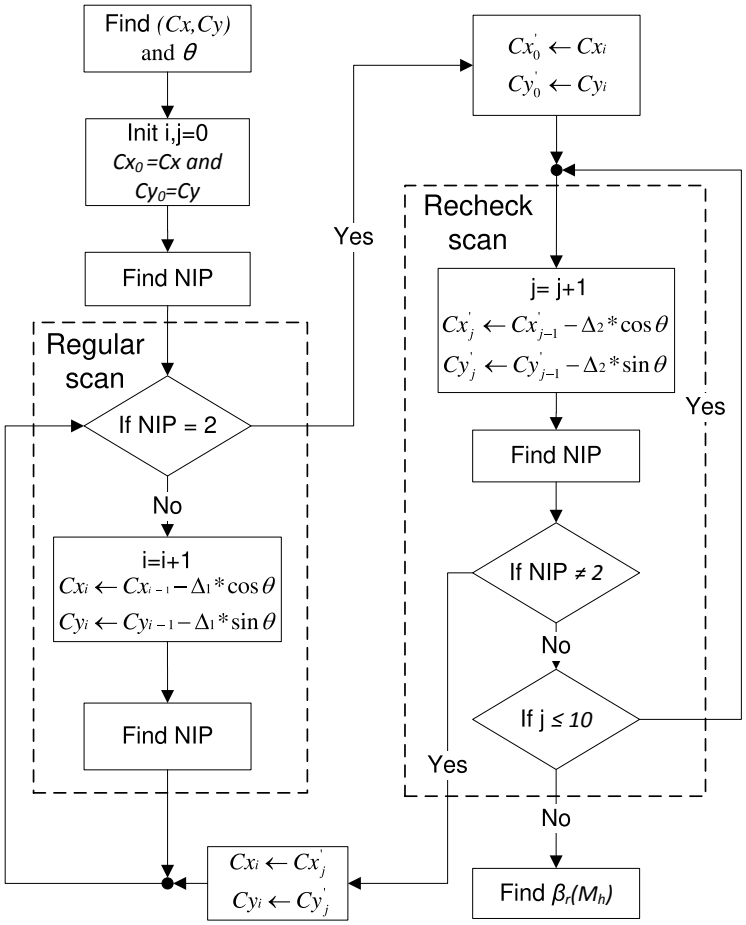

FIGURE 4.2: Flow diagram for steps performed in the finger extraction algorithm.

$$\theta = \begin{cases} 180 - |\alpha|\,, & \text{if } \alpha < 0 \\ |\alpha|\,, & \text{otherwise.} \end{cases} \tag{4.10}$$

If the appropriate condition of NIPs between $\beta(M_h)$ and scanning lines is not met, centroid/shifted-centroid is further shifted by a value Δ_k towards the wrist region. The corresponding shift in 'x' and 'y' coordinate is given $\Delta_k \cos\theta$ and $\Delta_k \sin\theta$, where θ is the angle of least inertia axis in an anti-clockwise direction as given in Eq (4.10) and index 'k' is 1 and 2 for regular and recheck scan, respectively. The value of

Algorithm 1 Finger extraction algorithm

Input: Hand gesture boundary $\beta(M_h)$

Output: Reduced boundary $\beta_r(M_h)$

procedure
  Find m, θ and (Cx, Cy)
  Initialize i = 0, j = 0 and $Cx_0 \leftarrow Cx$
  Find *NIP* between $y = m(x - Cx_0) + Cy_0$ and $\beta(M_h)$
Regular:
  while $NIP \neq 2$ **do**
    i=i+1
    $Cx_i \leftarrow Cx_{i-1} - \Delta_1 \cos\theta$
    $Cy_i \leftarrow Cy_{i-1} - \Delta_1 \sin\theta$
    Find *NIP* between $y = m(x - Cx_i) + Cy_i$ and $\beta(M_h)$
Recheck:
  repeat
    $Cx'_{j+1} \leftarrow Cx'_j - \Delta_2 \cos\theta$
    $Cy'_{j+1} \leftarrow Cy'_j - \Delta_2 \sin\theta$
    Find *NIP* between $y = m(x - Cx'_{j+1}) + Cy'_{j+1}$ and $\beta(M_h)$
    if $NIP \neq 2$ **then**
      $Cx_i \leftarrow Cx'_{j+1}$
      $Cy_i \leftarrow Cy'_{j+1}$ and **goto** *Regular*.
    j=j+1
  until $j \leq 10$
  Obtain $\beta_r(M_h)$

Δ_1 and Δ_2 is 1% and 2.5% of the height of the hand shape, respectively. This facilitates fine scanning of the hand shape. Steps of the finger extraction algorithm are presented in Algorithm 1.

RSS is translation invariant because these distances are calculated with respect to the shifted centroid. Scale invariance is obtained by normalizing $R(k)$. RSS is also invariant to rotation because scanning lines are always perpendicular to the axis of

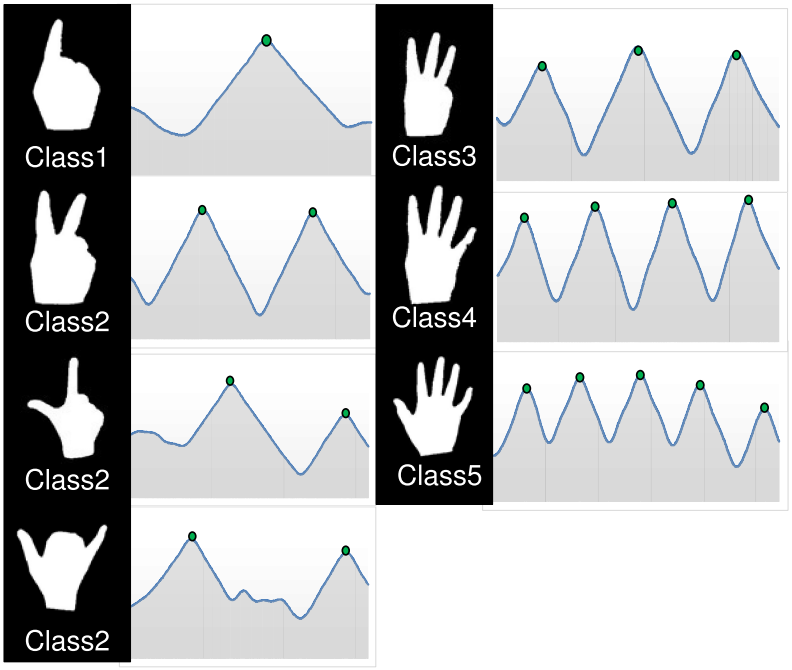

FIGURE 4.3: Various gesture classes and their reduced shape signature.

least inertia. Experimental results show that the number of peaks (NOPs) obtained in RSS are nothing but the fingertips. RSS and its corresponding peaks for different gesture classes are plotted in Figure 4.3. It should be noted that a class of gestures is defined on the basis of number of finger/s presented in the gesture. Gestures with different finger count are categorised as inter-class gestures, whereas gestures with the same number of fingers are considered as intra-class gestures.

The advantage of using RSS is the reduction of computational overheads, as well as the increase of interpretation in temporal domain. Even the effect of the band position does not influence the reduced shape signature, whereas, in the case of simple radial distance signature, unnecessary peaks are obtained due to the wrist portion. Peaks encircled in Figure 4.4 represent the unnecessary peaks and apparent change in radial distance signature.

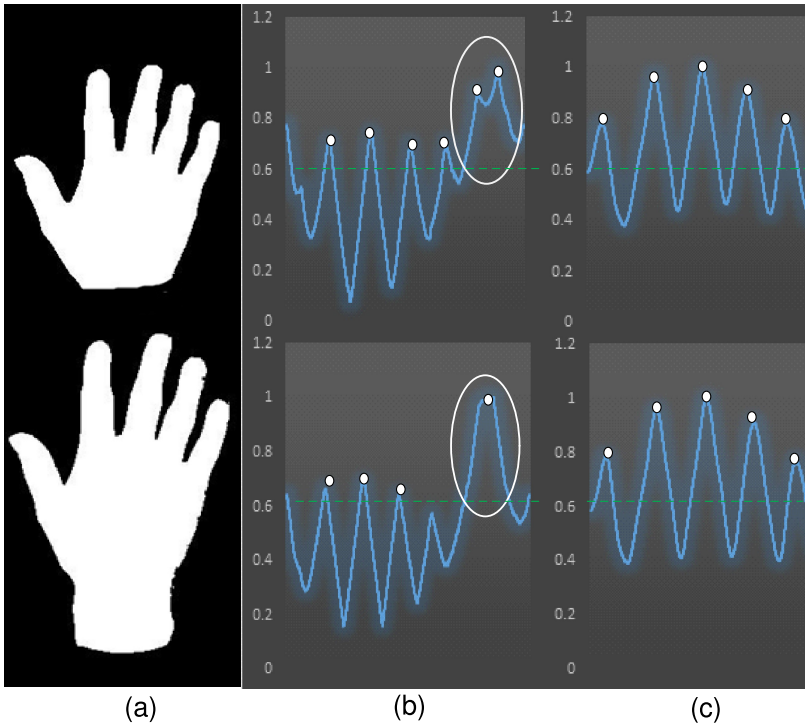

FIGURE 4.4: Effect of band location (a) Extracted hand shape (b) Radial distance signature (c) Reduced shape signature.

Another advantage of RSS over other shape signatures is that it is compact. It reduces the number of feature-sets essential to describe a hand gesture. The number of feature sets in radial distance signature and RSS for all the 14 gestures are calculated for a randomly selected user and their plot is presented in Figure 4.5. Percentage reduction in feature set for each gesture is calculated for 25 different users and on an average, nearly 35% reduction in feature sets is observed.

However, it is sometimes observed that there might be few peaks in the RSS, which don't represent the tip of the fingers. These peaks are nothing but the insignificant peaks obtained due to wrong or casual gesture posing. To differentiate between such significant and insignificant peaks, experiments were performed to obtain an optimal value of peak threshold (T_p) (refer to Figure 4.11(a)). It is observed that

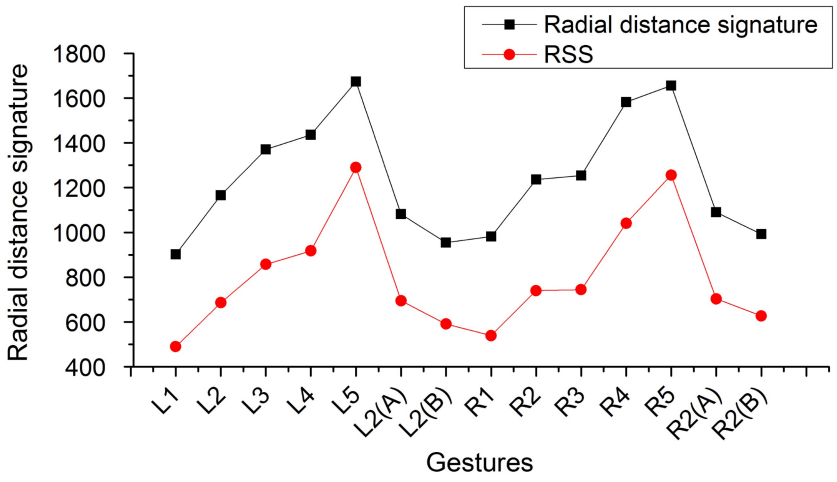

FIGURE 4.5: Comparision of number of feature sets.

with $T_p = 0.62$, maximum recognition rate can be achieved (refer Section 4.6.4). It is also found that recognition rate decreases in either direction because any increase in its value would cause the removal of the significant peak while any decrease would cause the addition of insignificant peaks. Hence, if the peak of RSS is greater than the peak threshold value, it corresponds to the significant peak otherwise, it is an insignificant peak. Finger count decision is taken by excluding such insignificant peaks to avoid misinterpretation.

In some cases, even peak threshold may not be able to differentiate some of the peaks as shown in Figure 4.6. Under such circumstances, local property around the peak of unnormalized RSS is investigated. For this purpose, either 100 points around each peak or all points between peak and valley are considered. Next, set with a minimum number of points between them is selected. Two lines are fitted around the peaks with the help of least square fitting, as shown in Figure 4.6(b) and intersection angle (ϕ) between them is obtained. It is observed that by selecting a threshold angle (T_a) equal to 130, such insignificant peaks can be omitted (refer

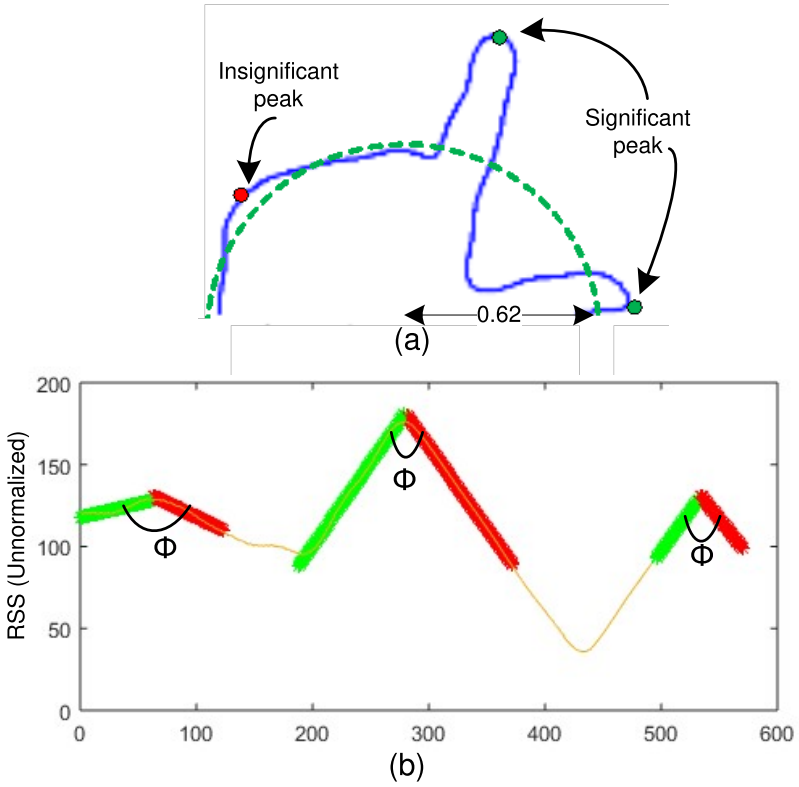

FIGURE 4.6: Illustration of insignificant peaks (a) Finger detection with $T_p = 0.62$ (b) Formation of threshold angle.

Section 4.6.4). Finger count decision is taken by excluding such insignificant peaks to avoid misinterpretation.

RSS can distinguish interclass gestures easily as it solely depends on finger count attribute. However, for the case of intraclass gestures, as shown in Figure 4.3, it will generate equal NOPs and it will get confused because the finger count is the same. But, their distribution is different. To distinguish symbols among intraclass gestures, two unique features—difference angle and polygonal area—are introduced in this work.

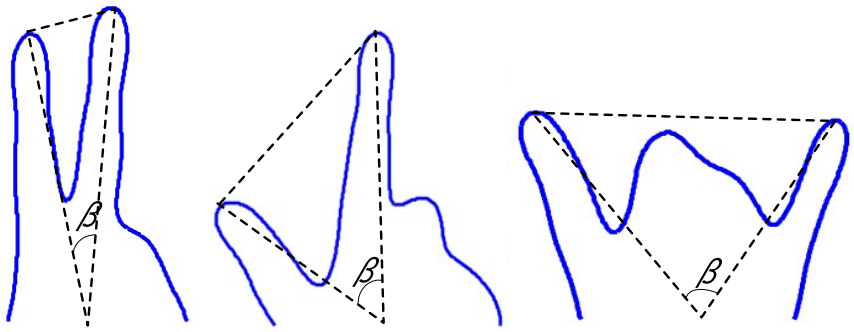

FIGURE 4.7: Illustration of polygonal area and difference angle.

4.4 Difference Angle and Polygonal Area

RSS provides significant peaks which correspond to number of fingers posed. Let us take a case of intraclass gestures of class 2 as shown in Figure 4.3. Here, RSS will generate two peaks for all gestures irrespective of finger position and distribution. To differentiate between these gestures difference angle and area of the polygon is calculated. Difference angle (β) is the angle formed between tips of fingers at the shifted-centroid, as shown in Figure 4.7. Similarly, polygonal area is the area formed by the polygon whose vertices are at fingertip coordinates and new shifted-centroid. The polygonal area is capable of discriminating intraclass gestures of the same size. To make this polygonal area robust to scaling, it is normalized by area of the largest circle fitted inside the boundary of hand gesture.

4.5 Classification

Interclass gestures shown in Figure 4.3 can be easily classified using RSS based on NOPs. However, to discriminate intraclass gestures of class 2, polygonal area and difference angle are required along with RSS.

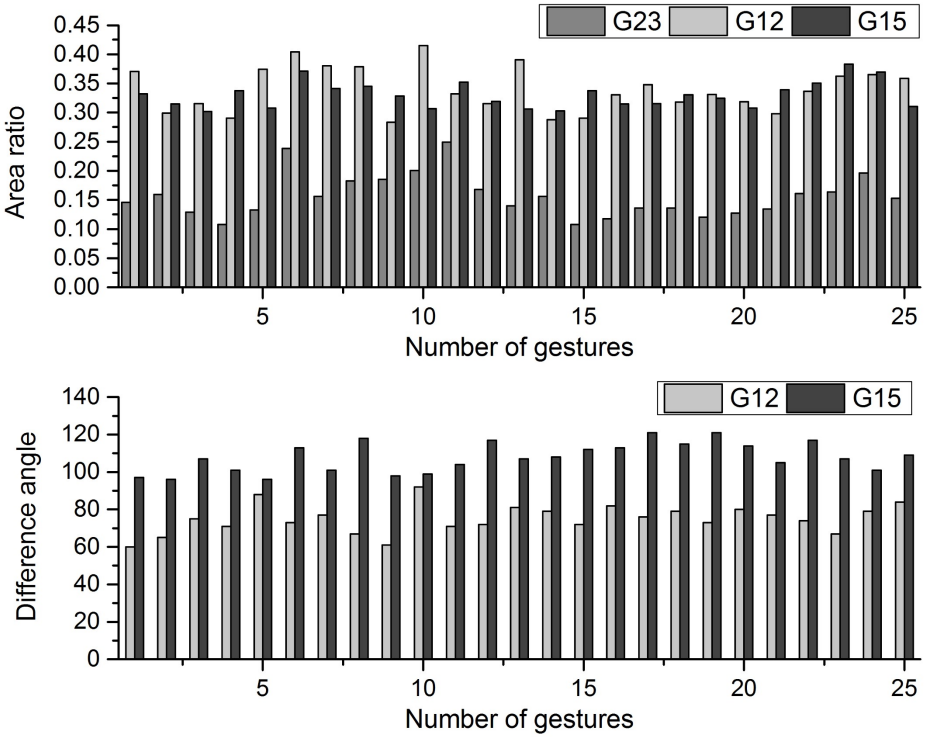

FIGURE 4.8: Bar-graph showing (a) Area-ratio (b) Difference angle.

A rule-based classifier is used for class 2 intraclass gestures classification. The polygonal area is used to differentiate between gestures with thumb (i.e. special gesture) and without thumb (i.e. basic gesture). Any gesture with the thumb pointing away from the palm would have a larger area as compared to that of without thumb. To investigate this fact, an experiment is conducted with 25 users having different hand size. The ratio of polygonal area to the area of the largest circle for these gestures is

TABLE 4.1: Illustration of classification rules.

Classification rules
Area ratio<0.25 → G23
Area ratio>0.25 & Difference angle < 95→ G12
Area ratio>0.25 & Difference angle > 95→ G15

calculated, and it is presented in Figure 4.8(a). It is observed that with a threshold value equal to 0.25, it is possible to discriminate basic gestures from the remaining two special gestures. Similarly, special gestures with thumb are differentiated based on difference angle as shown in Figure 4.8(b). Classification rules for class 2 gestures are furnished in Table 4.1.

4.6 Evaluations

4.6.1 Dataset

A separate database of the proposed dactylology is created which comprises of gestures from 25 users (15 male and 10 female) belonging to a different race, age, and sex. Two sets of all gestures are captured per user. Hand gestures in the database have been acquired using a camera with a resolution of 2048*1536 ($\approx$ 3.2 M-pixels) under an indoor environment and a uniform background. Figure 4.9 presents sample gestures in the database.

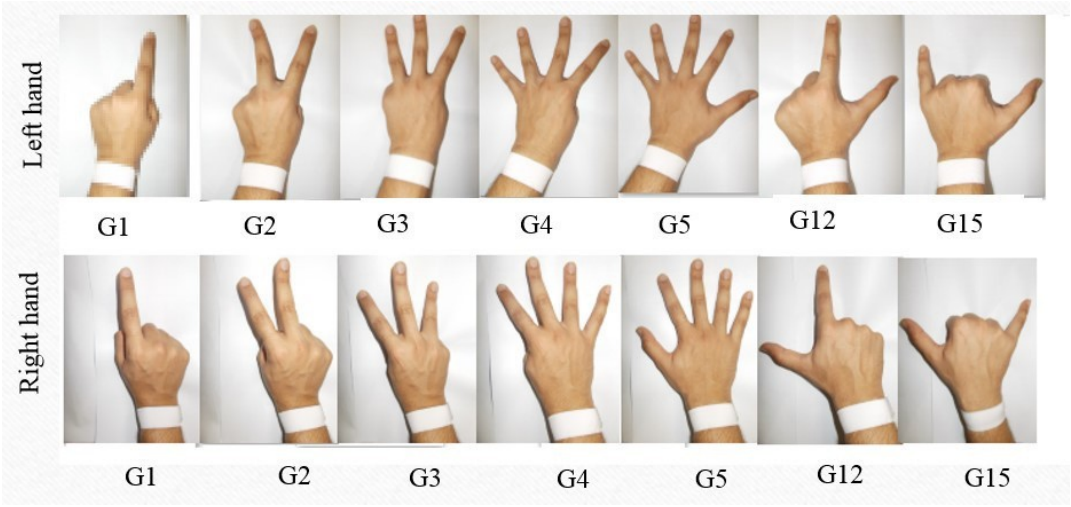

FIGURE 4.9: Samples of gestures within the database.

4.6.2 Verification of User Independence

User independence is the robustness of the algorithm to a gesture or its posing variations by the user. This parameter is verified by testing the algorithm with 25 different users belonging to a different race, age, and sex. Variation in hand geometry, the flexibility of fingers and posing angles are tested by considering two sets of gestures from each user.

4.6.3 Robustness to Rotation, Scale, Translation

In a real-time environment, a hand gesture can have different orientation, scale, as well as its location. But, it should be within the field of view of the camera. The proposed system is not only invariant to scale, translation, rotation but also insensitive to boundary distortion. Illustration of robustness of RSS to orientation and scale changes are demonstrated in Figure 4.10. It is clear that the gestures in Figure 4.10(a) and Figure 4.10(c) are of different size. Because RSS is normalized, both gestures are identified as similar one. Further, gestures in Figure 4.10(a), Figure 4.10(b) and Figure 4.10(d) are having a different orientation. However, the

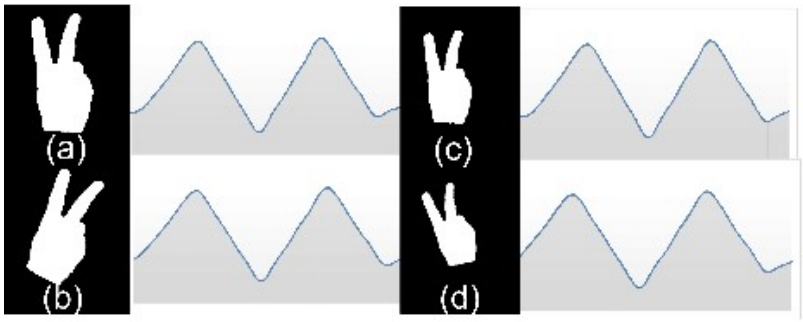

FIGURE 4.10: Illustration of robustness to orientation and scale changes (a) Original gesture (b) Rotated gesture (c) Scaled gesture (d) Rotated and scaled gesture.

plot of RSS is almost same because scanning lines are chosen to be perpendicular to least inertia axis.

4.6.4 Parameter Sensitivity

In this sub-section, two important parameters—the peak threshold (T_p) and angle threshold (T_a) are evaluated. Experiments are performed to select optimal value of these parameters. For this purpose, we have varied T_p and obtained gesture recognition rate. The plot of the same is presented in Figure 4.11. It can be seen that the maximum recognition rate is achieved for $T_p = 0.62$. It is also found that recognition rate decreases in either direction because any increase in threshold value would cause the removal of the significant peak while any decrease would cause the addition of insignificant peaks. Hence, peak threshold $T_a = 0.62$ is selected. Similarly, the maximum recognition rate is obtained for threshold angle (T_a) equals to 130°.

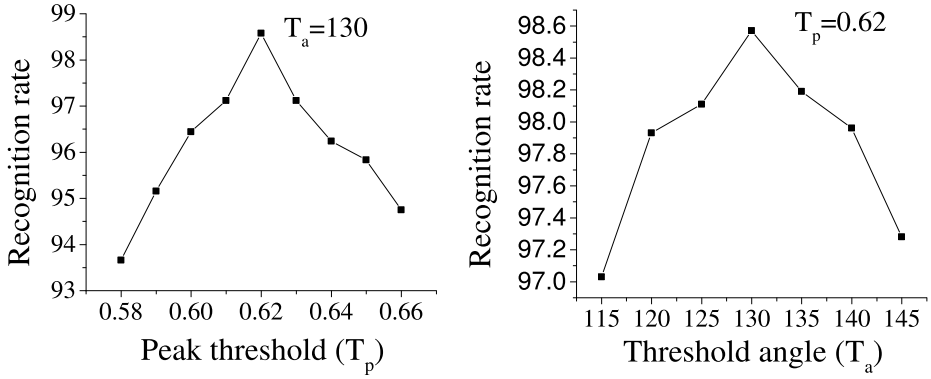

FIGURE 4.11: Illustration of parameter sensitivity.

4.6.5 Recognition Accuracy

The proposed approach is applied on the aforementioned database. A total of 59 symbols, as shown in Figure 3.6, are considered in this experiment. Symbol recognition accuracy is calculated by testing symbols obtained from 25 users. Each user has contributed two sets of data. So, each symbol is tested 50 times. While testing 2950 (59×50) symbols, 2908 are recognized correctly. From this, it can be conjectured that symbol recognition rate of the proposed algorithm is 98.57%. Further, it is observed that symbol recognition rate for basic hand gestures with single- and double- hand is 100%. This rate is lower in the case of special hand gestures because sometimes intraclass gestures may be confused. Symbol recognition accuracy for special gestures with single–and double–hand is 96.50%.

The proposed approach performs well because it neglects the palm region and considers finger through an iterative scheme. As a result, it prevents the misclassification arising due to wrist region. Apart from this, the presented method is also capable of removing insignificant peaks thus leading to improved performance.

4.6.6 Comparison with Other Methods

For comparison purpose, exactly similar work is not observed in the literature. Hence, some papers are selected for comparison purpose on the basis of a number of the symbol used, features or techniques used, and accuracy of the system. Since the intention of the proposed work is to produce more symbols, the emphasis is given on number of symbols coded using a particular technique. From Table 4.2, it is clear that so far maximum 32 gestures are used [81] in an experiment, and the symbol recognition rate is 75.15%. Similarly, the maximum accuracy achieved is

99.85% [86] with only 10 gestures. But, is possible to recognize 59 symbols with 98.57% accuracy using the proposed technique.

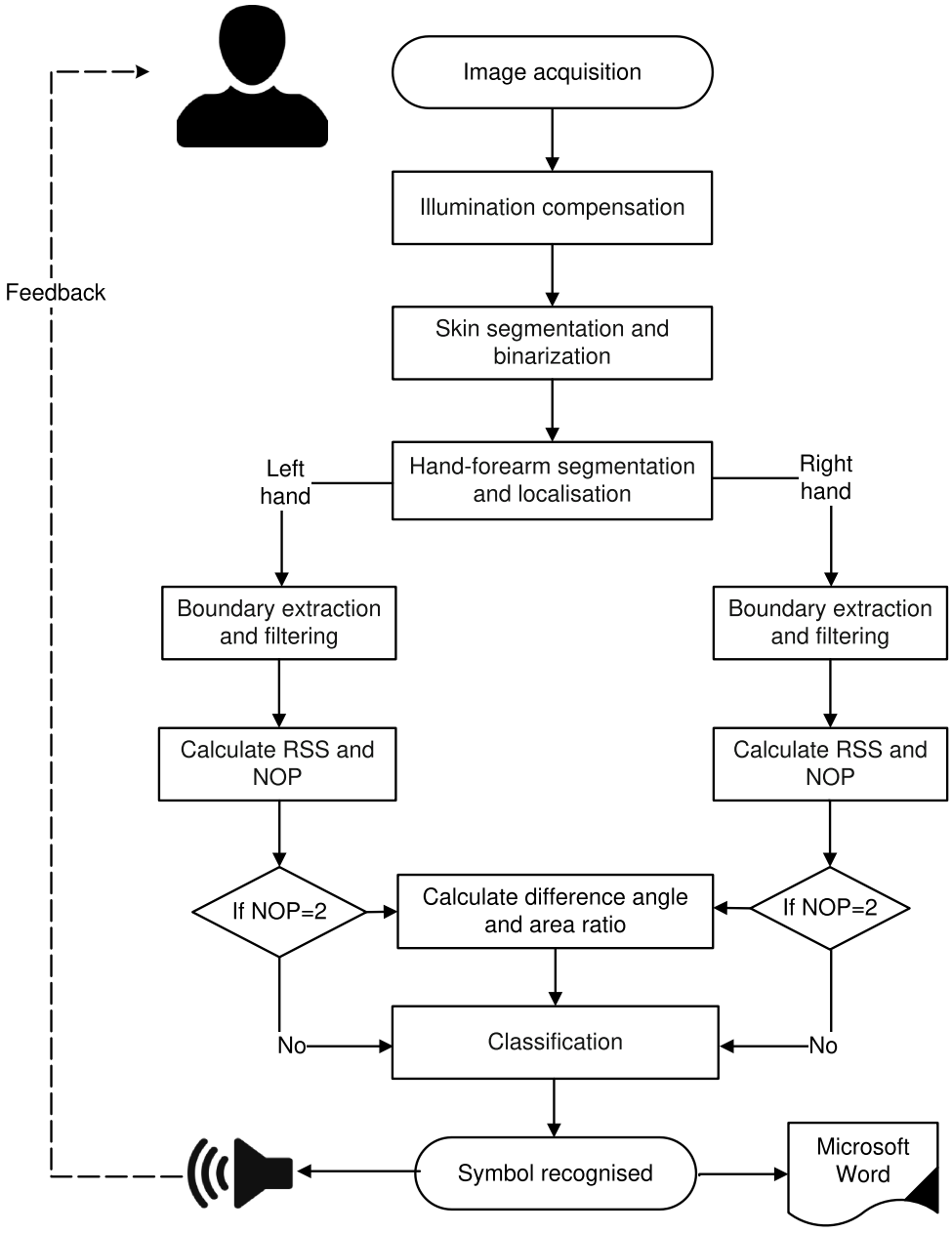

FIGURE 4.12: Workflow diagram of the proposed real-time writing support system.

TABLE 4.2: Performance comparison with other hand gesture recognition systems.

References	Number of symbols/gestures	Features/techniques used	Classification techniques	Accuracy (in %)
[80]	10	Geometry based normalizations and Krawtchouk moments	Minimum distance	96.94
[85]	12	Multiple feature descriptors such as distance, elevation, curvature and palm area	Support vector machine	93.80
[88]	10	Random forests	Binary decision tree	97.80
[86]	10	Histogram of Oriented Gradient features	Hierarchical elastic graph matching	99.85
[82]	10	Thresholding Decomposition and Finger Earth Mover's Distance	NA	93.20
	10	Near-Convex Decomposition and Finger Earth Mover's Distance	NA	93.90
[83]	10	Elastic bunch graph and adaptive boosting	Weighted elastic graph matching	97.08
[78]	10	Multi-feature fusion and template matching	Euclidean distance	91.00
[79]	10	Multiple shape-based feature (Area, perimeter, radial profile etc.)	Rule-based	94.40
[81]	32	Locality Preserving Projections	Support vector machine	75.15
[84]	24	Scale-Invariant Feature Transform, Bag-of-Visual-Words	Support vector machine	91.26
Proposed work	59	Reduced shape signature, Polygonal area and Difference angle	Rule-based	98.57

4.7 Workflow

A real-time writing support system was developed based on the proposed hand gesture recognition system. A complete workflow diagram of the proposed system is depicted in Figure 4.12. Image acquisition step captures a symbol that seems to be static for five consecutive frames. This prevents false images capturing in between posing of two symbols. The captured symbol is processed using the proposed algorithm as explained in Section 4.3. Once the symbol is recognized, a command is sent to Microsoft Word. The proposed system also provides an audio feedback. This will enable visually impaired user to listen the actual symbol entered and correct errors, if any. This ensures accurate and reliable data entry.

4.8 Concluding Remarks

In this Chapter, a new feature extraction method called reduced shape signature (RSS) is introduced. It uses simple feature extraction and non-complex classification technique. This method processes only finger boundary instead of entire hand gesture boundary. Consequently, it reduces the number of feature sets essential to describe a hand gesture. It is observed that on an average 35% reduction in feature sets is achieved. Additionally, this method is rotation, translation and scale invariant as well as simple, compact, computationally efficient and robust to irregularities around the wrist region. Differential angle and polygonal area are computed along with RSS to discriminate intra-class gestures, and it is observed that 2908 out of 2950 symbols are recognized correctly. It means symbol recognition rate of the proposed algorithm is 98.57%. The proposed method is independent of user characteristics, and it does not require any kind of algorithm training.

In this work, we have constrained the user by asking them to wear a band while posing symbols. Imposing restrictions on users is not a good idea as it hinders the naturalness of the interaction. Hence, to remove this constraint, we propose an automatic hand-forearm segmentation method. The method is further detailed in Chapter 5.