Chapter 5

Wrist Point Detection Algorithm using Geometric Features

In this chapter, we propose an automatic method for robust hand-forearm segmentation using geometric features. It is essential to segment and remove the forearm from the extracted image because there is no distinct information. Additionally, non-removal of the forearm region often leads to variation in extracted features affecting the correct recognition. Unfortunately, the majority of the works [146–149] have neglected hand-forearm segmentation by restricting users, i.e. wear band/full sleeve shirt. Imposing restriction on users is not appropriate as it hinders the naturalness of the interaction. Few researchers have [150–159] tried to solve this problem. But, these methods are not robust. Hence, a robust wrist point detection algorithm is proposed and detailed in this chapter.

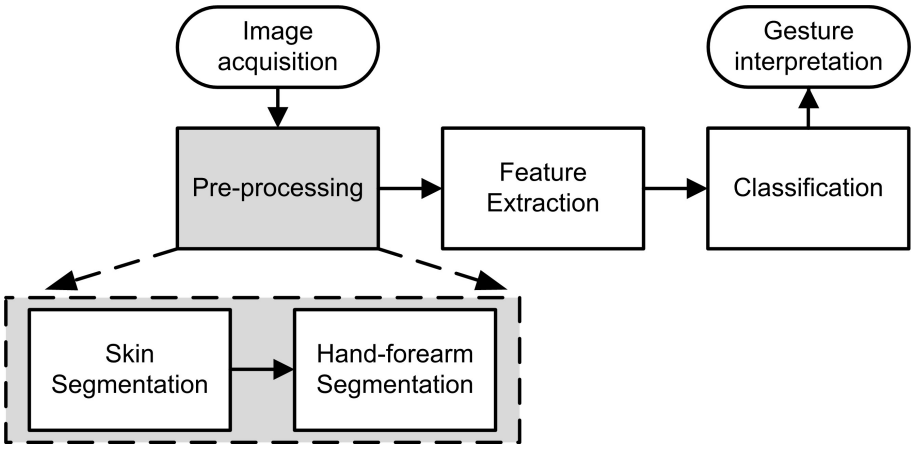

FIGURE 5.1: Abstract level diagram of a hand gesture recognition system.

5.1 Background

As discussed in Chapter 2, vision-based gesture recognition system has three essential steps: pre-processing, feature extraction and classification. Pre-processing step involves skin segmentation and binarization followed by hand-forearm segmentation (refer Figure 5.1). The basic idea of skin segmentation method is that the color of skin is unique and it forms a separate cluster in different color space. The boundary of these clusters is obtained empirically by analysing skin, non-skin pixels after color normalization. Some of these boundary based skin color models proposed in literature are RGB [55], HSV [56], YCbCr [57], etc. Most of these techniques work well in the constrained environment. However, in the case of the unconstrained environment, skin segmentation is quite challenging. Variation of illumination, interpersonal differences and a wide variety of races, ageing, etc. makes skin-color based segmentation a challenging task.

Skin segmentation is followed by a hand-forearm segmentation step. Most of the works have disregarded the hand-forearm segmentation step by assuming users wear

a band or a full sleeve shirt. Such restrictions simplify the hand-forearm segmentation. Usage of band allow researchers to separate hand from forearm assuming that hand is the largest object in the frame grabbed. This condition do not hold true in all cases as the captured image may be cluttered by other irrelevant objects which may be bigger than the hand. Few researchers have tried to solve this problem. Further details about hand-forearm segmentation methods are discussed in the next section of this Chapter.

5.2 Related Literature

Different parametric and non-parametric models have been proposed in the literature. Parametric models include skin modeling using Gaussian distribution [59]. Non-parametric models include skin modeling using k-means [60], Support Vector Machine [58], Neural Network [61], Random Forest [62], etc. In real-life scenario, the background object's color may be similar to skin, e.g. wooden furniture, dress, etc. causing false detection of background as skin region due to hard threshold. Hence, probabilistic approaches are suggested for real-life situations [63]. Jhon and Rehg [64] proposed a Bayesian classifier to obtain a Skin Probability Map (SPM) based on histogram of skin and non-skin pixels. Apart from this method, different skin segmentation [65–67] methods based on texture and spatial information have been proposed in the literature. These methods use additional features like texture and spatial information to improve the skin segmentation results. Work using such technique is presented in [65]. This work is further analysed the SPM obtained from [64]. Here, a distance transform is used for propagation of skinness in a combined domain of luminance, hue and skin probability. Similar work is presented in [66] which also use the texture features to improve the skin classification accuracy. Here,

textural features are obtained from the probability maps [64] rather than luminance channel. A method of propagating the skin/non-skin seed points is presented in [67]. It uses a semi-supervised learning method.

As per hand anatomy, width values increase significantly at the wrist points while traversing from the forearm to the palm [160]. This is illustrated in Figure 5.2. Most of the previous literatures [150–155] use this fact to determine the wrist point. Work presented in [150] use the narrowest location of ridges as a wrist point in the distance transformed hand mask. However, it is hard to find an appropriate threshold to obtain good ridges. Additionally, this method performs the worst when the hand is rotated at wrist point. A wrist locating technique using width information along the orientation of hand is explained in [151]. Here, orientation is obtained from image moments. Work in [152] assumes that the region below the maximum width of the hand is forearm. This maximum width information is also used in [153] to locate the forearm. Here, the forearm cut line is considered as a fixed offset to the maxima point obtained in the vertical projection along the orientation of the hand. However, both methods [152, 153] do not provide the exact distinction between palm and forearm. Work in [154] presents a wrist localization technique based on hand profile which is obtained along the direction of the longest cord fitted inside the

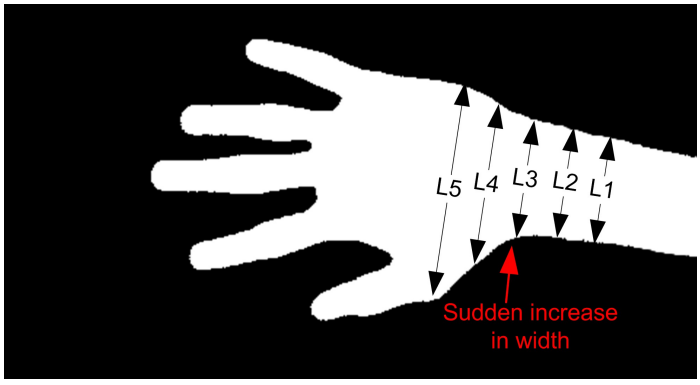

FIGURE 5.2: Illustration of sudden increase in width at the wrist [160].

hand silhouette. Radon transform is used in [155] to find palm region by obtaining local minima in the projection along the detected wrist line.

Few works [156–159] have considered geometric properties of hand mask to obtain the wrist point. Use of geometrical properties is proposed in [156] to detect the wrist line. This method uses convex hull to obtain depth points. The wrist line is located as the line passing through the closest depth point from the centroid of the arm which is perpendicular to the arm contour. Another solution based on convex hull is described in [157]. This method searches candidate lines and locates one end of the wrist line as the point on the contour with the maximum distance from the candidate lines. The other end of the wrist line is obtained using the geometric characteristics of an inscribed rectangle. A wrist detection method based on Harris corner features is presented in [158]. It assumes that wrist points are the nearest corner to the centre of gravity towards the side which contains the highest number of corners. This assumption works fine on hand with a forearm. However, if forearm region is missing, the centre of gravity lies inside the palm causing the wrong detection of wrist point. This issue is addressed in [159]. It proposes Skin Region Verification (SRV) ahead of wrist segmentation to identify the presence of forearm. In the presence of forearm approach similar to [158] is used otherwise the case is dealt separately.

In this chapter, we propose a systematic approach for wrist point detection based on the geometric features of the binary hand mask. The proposed method is rotation, translation and scale invariant. The wrist point detection accuracy is improved to 84% for an acceptable error ($e < 0.5$). Additionally, we implemented some of the skin segmentation algorithms and investigated its effect on the wrist point detection accuracy. The remainder of the chapter is organised as follows. Section 5.3 describes the proposed algorithm for robust wrist point detection and Section 5.4 presents the

parameters used to evaluate its performance and Results are discussed in Section 5.5. Section 5.6 concludes the paper.

5.3 Proposed Algorithms

As per the anatomy of the human hand, it is observed that the hand palm can be approximated using the largest circle inscribing the hand mask. The wrist point will lie on the boundary of the hand palm. Therefore, the wrist point must lie on the largest circle inscribing the hand mask. Additionally, it is a fixed point. Ulnar and radial deviation occur around an axis that passes through the capitate bone which is the central position in the wrist. Therefore, the wrist point is a point which is located at the centre of forearm-palm joint regardless of ulnar/radial deviation. Hence, a point which satisfies both the conditions is known as wrist point. It is denoted by w_r and is equal to that value of w_i for which the following expression produce the maximum value.

$$d = Max(|D(M_h, w_i)|) \tag{5.1}$$

where w_i belongs to the set of points sampled on the largest circle inscribing the hand mask (M_h). Thus, in order to select the wrist point, we first need to find the largest circle [161] inscribing the hand mask. This is obtained by using the distance transform (D) of the hand mask. Further, we seek for a point with the largest value in the distance transform (D), refer Figure 5.3(b). This obtained point is referred as palm centre (P_c) of the hand and the largest value is radius (R) of the largest inscribed circle. Palm centre is independent of rotation, translation, and scaling and it is fixed relative to the shape.

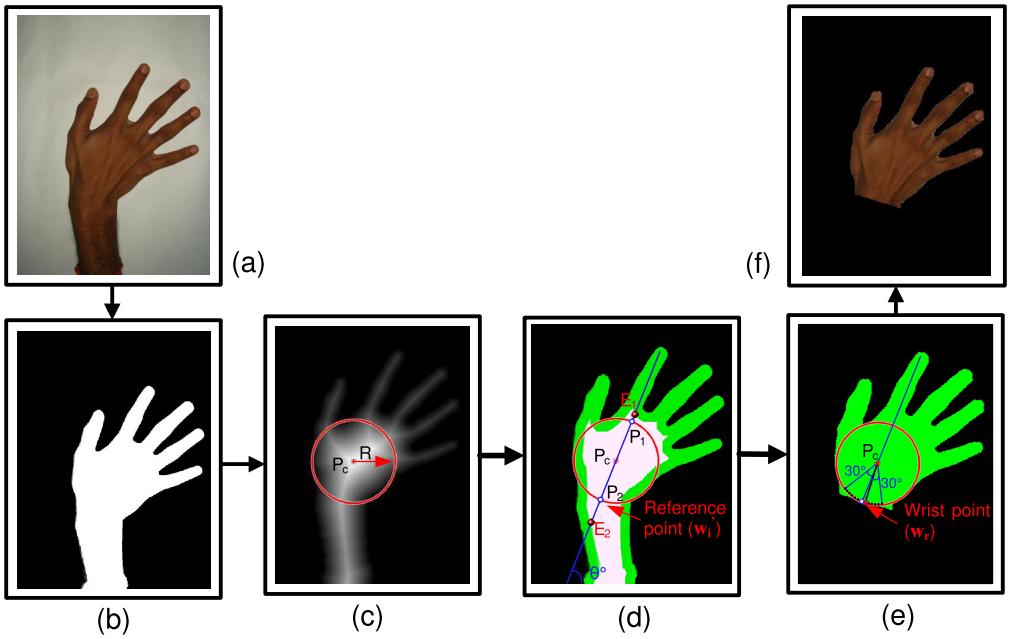

FIGURE 5.3: Illustration of steps (a) Input RGB image (b) Binary hand mask (M_h) (c) Distance transform of M_h (d) Eroded image (e) Forearm segmented mask (f) Forearm segmented RGB image.

A hand can be flexed either towards the radial or ulnar bone of the arm. This flexion occurs around an axis that passes through the capitate bone of the wrist. The range [162] of the radial (Δ_r) and ulnar (Δ_u) deviation is 15° and 30°, respectively as shown in Figure 5.4. So, the wrist point must lie in this range and hence, instead of searching for all w_i points, one needs to search for points within the range Δ_θ. In this work, the value of Δ_θ is selected as the maximum of the radial and ulnar deviation, i.e. $\Delta_\theta = max(\Delta_r, \Delta_u)$. Thus, the wrist point should lie within $\Delta_\theta = \pm 30°$ around the orientation axis of the hand mask. The orientation (θ) of hand mask is obtained as the second order central moment given by the following expression

$$\theta = \frac{1}{2} * arctan(\frac{2\mu_{11}^{'}}{\mu_{20}^{'} - \mu_{02}^{'}}) \tag{5.2}$$

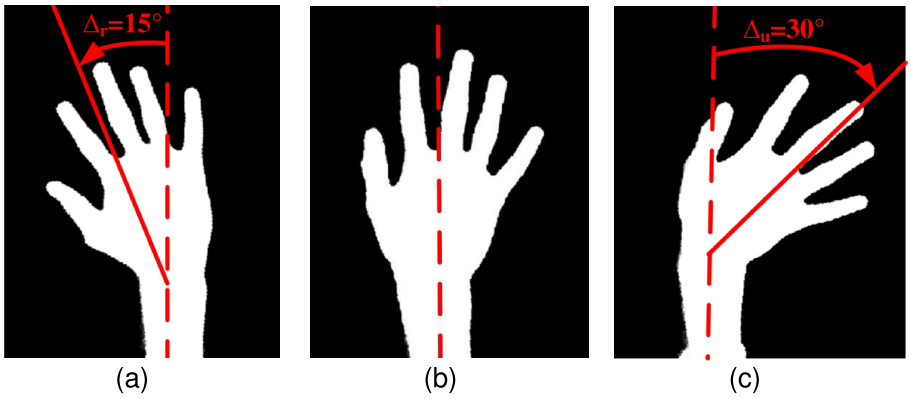

FIGURE 5.4: Illustration of movement of the wrist joint (a) Radial deviation (b) Neutral position (c) Ulnar deviation.

where,

$$\mu_{ij}^{'} = \frac{\sum_x \sum_y (x-\bar{x})^i (y-\bar{y})^j M_h(x,y)}{\sum_x \sum_y M_h(x,y)}. \tag{5.3}$$

The values of $\bar{x}$ & $\bar{y}$ in Equation (5.3) are calculated as follows

$$\bar{x} = \frac{\sum_x \sum_y x M_h(x,y)}{\sum_x \sum_y M_h(x,y)} \tag{5.4}$$

$$\bar{y} = \frac{\sum_x \sum_y y M_h(x,y)}{\sum_x \sum_y M_h(x,y)}. \tag{5.5}$$

Once the orientation of the hand mask is known, we further seek for the reference point $w_i{}'$ on the orientation axis around which the wrist point needs to be searched. It is crucial to find reference point because the circle does not give information related to the position of the forearm relative to the hand. This information is obtained by finding the intersection of the circumscribed circle with the line passing through the palm centre at an angle θ, i.e. orientation axis. This line intersects the circle at points P_1 and P_2 as shown in Figure 5.3(c). The selection of reference point between points P_1 and P_2 is made using an eroded version of M_h. The point P_i for which $dist(P_c, E_i) - R$ is greater is selected as a reference point; where E_1 and E_2 are

the points obtained by the intersection of the line with the boundary of the eroded image. Wrist point is searched within $\Delta_\theta = \pm 30^\circ$ around this reference point on the boundary of the inscribed circle (refer Figure 5.3(d)). The point satisfying Equation (5.1) is selected as wrist point w_r. Once wrist point w_r is obtained, wrist line is located as line perpendicular to another line joining palm centre P_c and the wrist point (w_r). The palm region is selected using connected components labelling with palm centre inside it and neglecting the other regions. The final forearm segmented RGB image is shown in Figure 5.3(f).

As per the assumption, the hand palm can be approximated using the largest circle inscribing the hand mask. However, sometimes due to the presence of arm the largest inscribing circle is found near the elbow, as shown in Figure 5.5. To avoid this false selection, two largest non-overlapping inscribed circles with radius R_1 and R_2 are drawn ($R_1 \geq R_2$). The selection of palm circle between these two circles is done by growing both the inscribed circles to twice of their radius and calculating the width of skin segments as shown in Figure 5.5. One to six skin segments are obtained with the palm circle, whereas only one or two skin segment/s is/are found with elbow circle. It means a circle with more than two skin segments is definitely a palm circle. Since the width of fingers is small as compared to the forearm, it

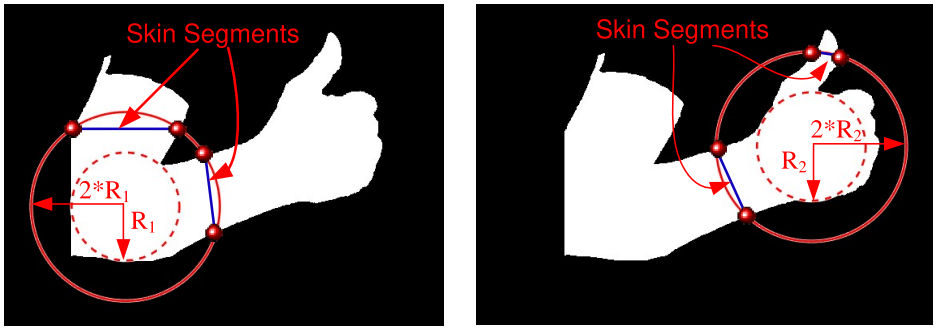

FIGURE 5.5: Illustration of selection of palm circle (a) Largest inscribing circle (b) Second largest inscribing circle.

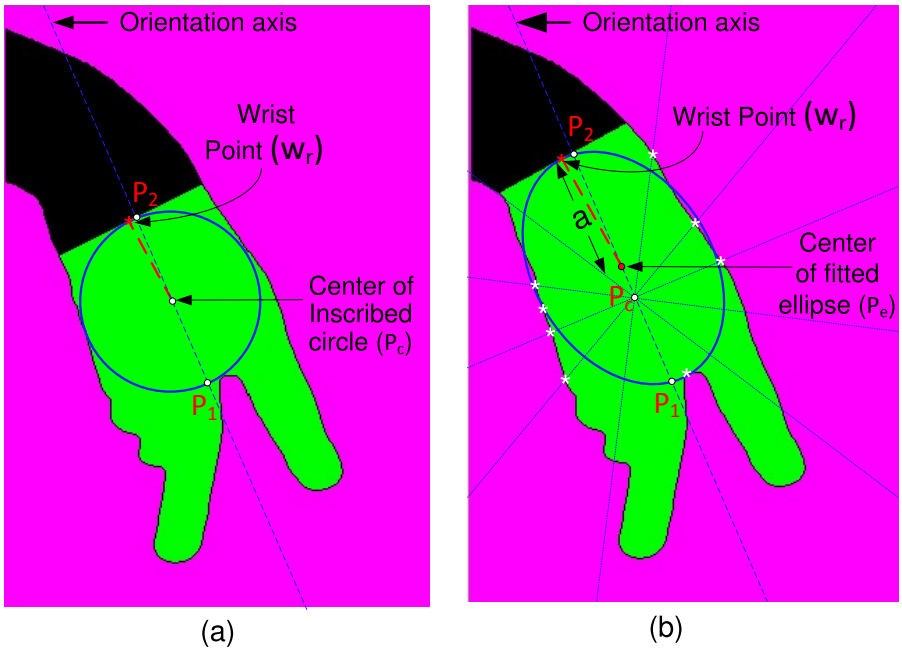

FIGURE 5.6: Example cases (a) Circle fit method (b) Ellipse fit method.

is possible to identify the palm circle when there are only two skin segments. The circle for which the value of sum of skin segments width is less is considered as palm circle.

It is sometimes observed that the circle does not approximate the palm of hand mask. Such a case arises when posed hand is not perpendicular to the camera axis. The circularity of hand palm gets affected by perspective projection. Apart from this, the circularity of palm also gets affected in thin gestures (usually without thumb and index), as shown in Figure 5.6. A more generalized model is required to overcome this issue. An ellipse is fitted to obtain a better approximation of the hand palm as suggested in [163].

The hand mask is divided into 12 equal sectors (i.e. 30 °) with respect to orientation axis. The points on the hand mask (shown by white asterisk) are obtained by

considering two adjacent sectors at a time, i.e. sweeping through two sectors. Next, the point on the boundary of the hand mask within these two sectors which is the nearest to the centre of the inscribed circle (P_c) is searched and marked. This process is repeated for each 30° shift in angle. Once all these points are obtained, a least-square fitting method [164] is used to obtain the ellipse which represents an approximation of hand palm. Overlapping sectors are used to find these boundary points. It is quite probable that same point is obtained as the nearest boundary point in two consecutive sweeps.

The reference point w_i' is obtained by finding the intersection of the fitted ellipse with the line passing through its centre with an angle θ, i.e. orientation axis. The point towards the wrist is selected similar to the circle fit method. The point for which $dist(P_e, E_i) - a$ is greater is selected as a reference point and wrist point is searched within $\Delta_\theta = \pm 30\,^\circ$ around it on the boundary of fitted ellipse. The point which satisfies Equation (5.1) is selected as wrist point w_r. Once wrist point w_r is obtained, wrist line is located as a line perpendicular to another line joining ellipse centre and wrist point w_r. The palm region is obtained using connected components labelling with ellipse centre inside it and neglecting the other regions.

The proposed method is translation, rotation and scale invariant. Invariance to translation is achieved by shifting the polynomial basis of moment into the palm centroid. The wrist point is located around the orientation axis—line with an angle θ passing through P_c. The parameter θ changes with the rotation of the hand, keeping the orientation axis fixed for the hand, making the proposed method invariant to rotation changes. If a hand mask is re–sized by any scale factor S, the fitted circle/ellipse also scales by the same factor. Thus, the proposed method is invariant to scale changes too! Moreover, with the proposed method users are free to move their hand at any location and orientation within the camera's field of view. Additionally,

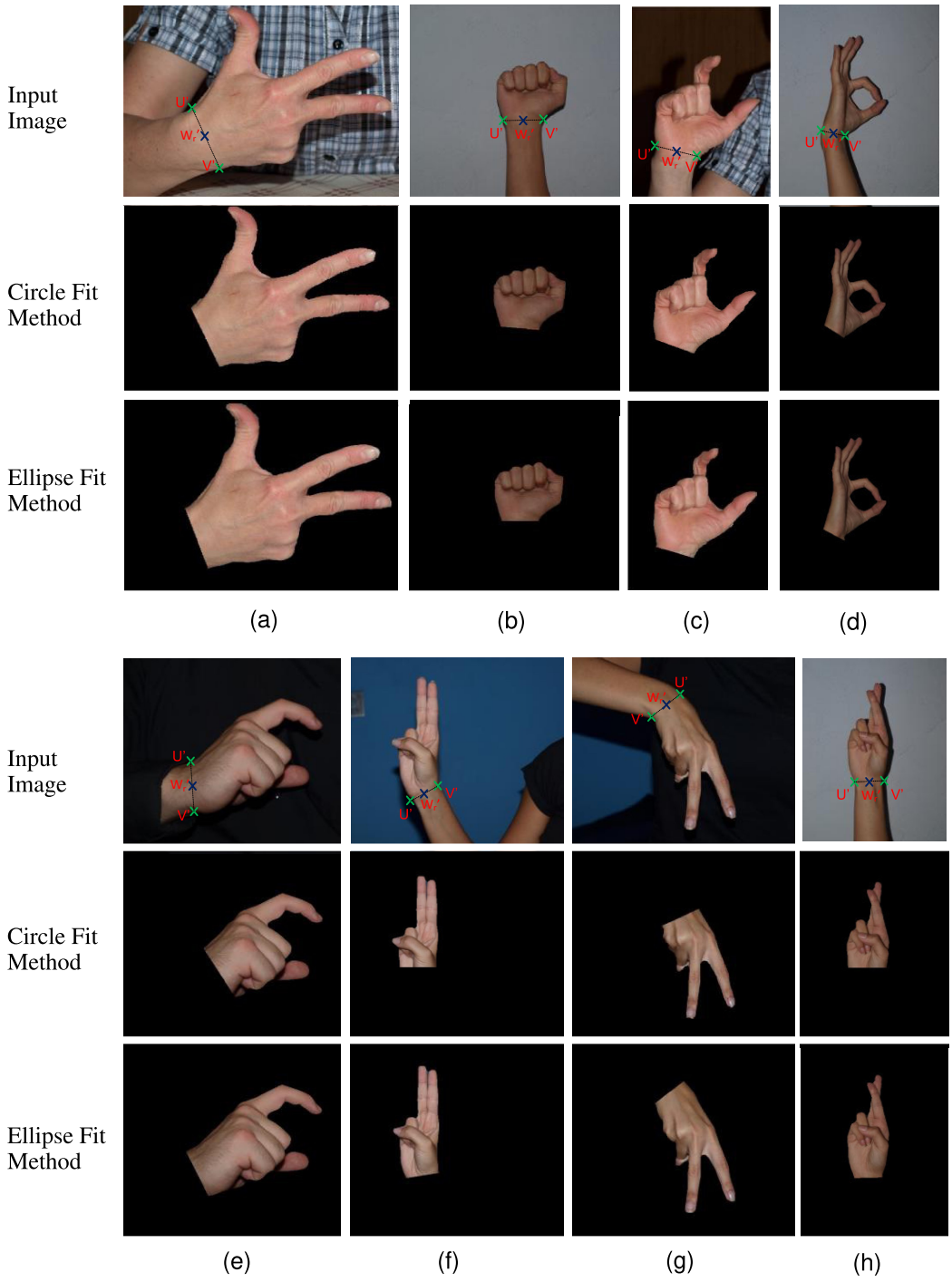

FIGURE 5.7: Wrist point localization results on different gestures in HGR1 database. Wrist point ($w_r{}'$) by expert is shown in blue cross while the green cross represent end points (U' and V') of the line separating hand and forearm in the ground truth data.

there is no need to wear any band/ a full sleeve shirt. The proposed method does not require any initial training or calibration. Some results of hand-forearm segmentation on sample images of the HGR1 database are shown in Figure 5.7. It can be observed that for the cases where the hand is exactly perpendicular to the camera axis (refer Figure 5.7(a-c)), both the method separate hand and forearm effectively. However, for the cases where gestures are not perpendicular to camera axis (refer Figure 5.7(d-e)) and cases with thin gesture (refer Figure 5.7(f-h)), ellipse method performs better than the circle fit method.

5.4 Performance Evaluation Parameters

In order to test the robustness of the proposed method, we have used the HGR1 database provided by [154]. It consists of 899 images acquired in uncontrolled lighting conditions with labelled ground-truth skin mask along with landmark points annotated by the expert. These ground truth skin mask and landmark points are used to assess the performance of the skin segmentation as well as the wrist point detection algorithm, respectively.

Performance metrics—accuracy, precision, recall and F-measure—are calculated to compare the performance of the different skin-segmentation algorithm. These measures [67] are given as below.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{5.6}$$

$$Precision = \frac{TP}{TP + FP} \tag{5.7}$$

$$Recall = \frac{TP}{TP + FN} \tag{5.8}$$

where,

True Positive (TP) = Number of skin pixel detected correctly as skin,

True Negative TN = Number of non-skin detected correctly as non-skin,

False Positive (FP) = Number of non-skin detected as skin,

False Negative (FN) = Number of skin detected as non-skin.

Precision gives us information about algorithm performance with respect to false positives while recall gives us information about its performance with respect to false negatives. A low precision indicates a large number of false positives whereas a low value of recall indicates many false negatives. F-Measure conveys the balance between the precision and the recall as it considers both false positives and false negatives. It is obtained using the following expression.

$$F\text{-}Measure = 2 * \frac{Precision * Recall}{Precision + Recall}. \tag{5.9}$$

To evaluate the performance of wrist detection methods an error (e) is calculated as shown below

$$e = \frac{dist(w_r, w_r')}{dist(U', V')} \tag{5.10}$$

where U', V' are the endpoints of the line separating hand and forearm in the ground truth data, w_r is the detected wrist point, w_r' is the wrist point by the expert which is exactly midway between U' and V' (refer Figure 5.8). To compensate the discrepancy

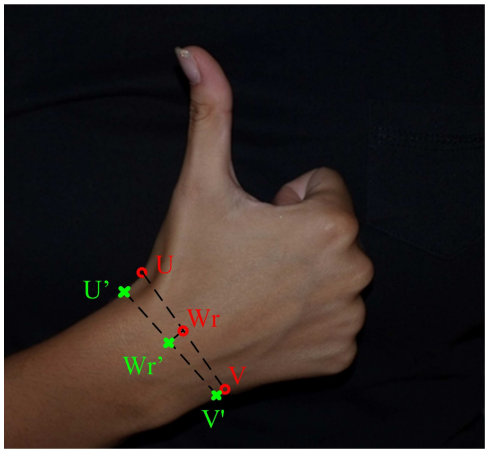

FIGURE 5.8: Illustration of error (e).

due to human factors and ground truth data annotator software, the wrist detection error (e) less than 0.5 is considered to be correct and acceptable [154].

5.5 Result and Discussion

Usually, skin segmentation is performed before wrist detection and in such multi-stage algorithm, the effectiveness of the preceding stage affects the performance of the succeeding stage. Therefore, the efficiency of wrist detection method depends on the outcome of the skin segmentation algorithm. In this Section, we first compare the performance of all the wrist detection methods on the ground truth skin mask of the HGR1 database (entire 899 images). Next, the performance of these methods is also tested in a real-life scenario.

Ground-truth skin mask and landmark points provided in the HGR1 database are used to compare the performance of the wrist detection methods. The results of the proposed wrist detection methods along with the state of the art methods are furnished in Table 5.1. It can be observed that the methods proposed in this work

TABLE 5.1: Wrist point detection accuracy on ground truth skin mask of HGR1 database

Wrist detection method	Wrist point detection accuracy			
	e<0.5	e<0.7	e<1.0	e<1.5
Moments [151]	0.59	0.69	0.75	0.83
Longest chord † [154]	0.64	0.76	0.84	0.89
Corner method ‡ [158]	0.37	0.45	0.58	0.77
Corner method + SRV [159]	0.32	-	0.67	-
Circle fit (proposed)	0.68	0.80	0.87	0.91
Ellipse fit (proposed)	**0.84**	**0.86**	**0.88**	**0.91**

† Result for the longest chord is averaged over 11 iteration.

‡ Result presented only for selected 437 images.

outperform existing methods. The best detection accuracy for an acceptable error ($e < 0.5$) is obtained for ellipse fit method followed by circle fit method. Since wrist point is assumed to lie on the boundary of the hand palm, accurate estimation of the hand palm becomes crucial. The ellipse fit method provides a better approximation of palm even in the case where circularity of hand palm gets affected by perspective projection. Thus, leading to better results. It should be noted that all the results presented in Table 5.1 are obtained with 899 images of HGR1 database except corner method [158]. Corner method presented their result on selected 437 images only.

In a real-life scenario, ground truth skin mask is unavailable. This entails that the performance of these methods should also be tested on skin masks obtained using skin segmentation algorithm. Therefore, skin segmentation algorithms [64–67] mentioned in Section 2.4.3.1 are also implemented. Next, we have applied wrist detection method on the outcome of these algorithms. The performance of the skin segmentation algorithm along with the wrist detection error is furnished in

Table 5.2. It can be observed that the MPMG [67] method outperforms other methods. Wrist point detection methods, when applied on skin mask obtained from the MPMG method results in the highest detection accuracy (almost comparable to ground-truth), whereas the least accuracy is obtained with the Bayesian method, [64]. Moment [151] and longest chord method [154] have lower accuracy as compared to the other methods. However, for few cases of high error $e < 1.0$ and $e < 1.5$ (shown by an asterisk), longest chord method shows slightly better result than the ellipse fit.

The histogram of error (e) is furnished in Figure 5.9. It can be conjectured that the proposed ellipse fit method provides more precise and accurate results. Almost 84% (753 out of 899) of the wrist points are detected accurately with an acceptable error ($e < 0.5$) for ground truth skin mask (Refer Table 5.1). Among these correctly detected wrist points, the majority (480 out of 753) belongs to error bin $e < 0.2$ (refer Figure 5.9(a)). A similar observation can be made for other skin masks. The ellipse fitting method has the largest number of occurrences with $e < 0.2$ followed by circle fit method. It is also observed that the ellipse fit method has less number of occurrences with $e > 1.0$ except for DSPF [66] and FPSD [65]. For DSPF [66] and FPSD [65], longest chord method provides slightly less number of occurrences with $e > 1.0$ as compared to ellipse fit method. From Figure 5.9(a-e), it can be seen that the longest chord method has the highest number of occurrences for all error bins except for $e < 0.2$ and $e > 1.0$. Almost 69.96% (629 out of 899) and 70.10% (639 out of 899) of the error (e) belongs to 0.2-1.0 bin for DSPF [66] and FPSD [65], respectively. This shows that the longest chord method is ineffective for cases where very low error ($e < 0.2$) is required. However, it performs satisfactorily for moderate error.

TABLE 5.2: Performance comparison of the skin segmentation & wrist point detection accuracy in real-life scenario

Skin Segmentation	Accuracy	Precision	Recall	F-Measure	Wrist Detection	Wrist point detection accuracy			
						$e < 0.5$	$e < 0.7$	$e < 1.0$	$e < 1.50$
MPMG [67]	0.97	0.96	0.95	0.96	Longest chord [154] †	0.61	0.74	0.83	0.89
					Moments [151]	0.56	0.67	0.73	0.82
					Circle fit	0.63	0.74	0.83	0.89
					Ellipse fit	**0.79**	**0.83**	**0.86**	**0.89**
DSPF [66]	0.97	0.95	0.94	0.95	Longest chord [154] †	0.59	0.73	**0.83***	**0.89***
					Moments [151]	0.54	0.65	0.72	0.81
					Circle fit	0.63	0.73	0.80	0.87
					Ellipse fit	**0.71**	**0.76**	0.80	0.85
FPSD [65]	0.96	0.93	0.95	0.94	Longest chord [154] †	0.59	0.72	**0.81***	**0.88***
					Moments [151]	0.53	0.64	0.72	0.80
					Circle fit	0.61	0.72	0.79	0.85
					Ellipse fit	**0.70**	**0.75**	0.77	0.82
Bayesian [64]	0.96	0.94	0.92	0.93	Longest chord [154] †	0.53	0.66	0.77	0.85
					Moments [151]	0.45	0.55	0.65	0.77
					Circle fit	0.61	0.71	**0.79**	**0.85**
					Ellipse fit	**0.71**	**0.76**	0.78	0.83

† Result presented for the longest chord is averaged over 11 iteration.

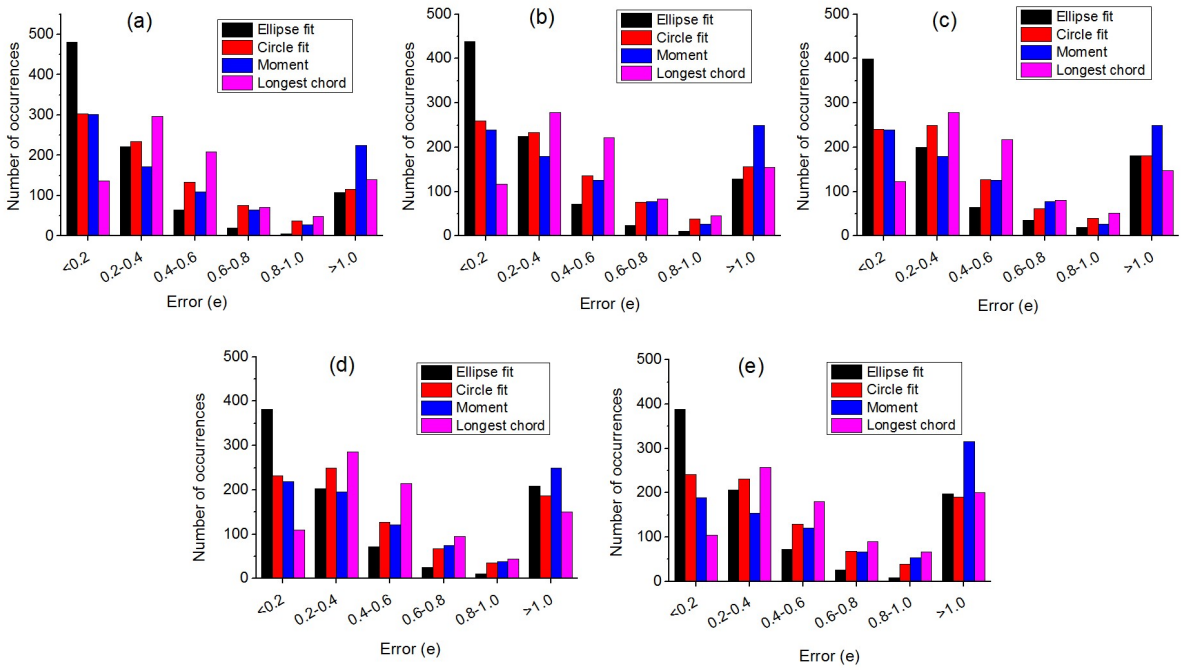

FIGURE 5.9: Error histogram of the wrist detection methods on the skin masks obtained from (a) Ground Truth (b) MPMG [67] (c) DSPF [66] (d) FPSD [65] (e) Bayesian [64]

The possible reasons for the mediocre performance of the longest chord method are that there exist many gestures with long forearm region. Such gestures contain a large number of contour points around the forearm region as compared to the palm region. The assumption made in [154] fails for the aforementioned gestures and this makes the longest chord method less effective which is also reported by the authors themselves. Additionally, the roughness of the hand silhouette due to improper segmentation causes detection of minima slightly away from the wrist region. This gives rise to a considerable number of occurrences with moderate error (e) between 0.2 to 1.0 and only a few numbers of occurrences with error $e < 0.2$ and $e > 1.0$.

5.6 Concluding Remarks

In this Chapter, we proposed a robust wrist point detection algorithms. Two methods—circle-fit and ellipse fit—are presented to approximate the hand palm. Next, a wrist point detection algorithm is proposed using geometric features of the binary hand mask. This method segments the hand region without imposing any restrictions on the users, i.e. users do not require to wear any band/full sleeve shirt. Additionally, they can also pose their hand in any orientation. The proposed algorithm is rotation, translation and scale invariant. Based on the experimental results, it can be concluded that the proposed ellipse fit method is accurate and effective as compared to the state of the art methods. Almost 84% (753 out of 899) of the wrist points are detected accurately with an acceptable error ($e < 0.5$) for ground truth skin mask of HGR1 database. Among these correctly detected wrist points, the majority (480 out of 753) belongs to error bin $e < 0.2$.

Furthermore, the performance of the proposed methods is evaluated in a real-life scenario. For this purpose, different skin segmentation algorithm—Multi-seed Propagation in Multi-layer Graph (MPMG), Discriminative Skin Presence Features (DSPF), Fast Propagation-based Skin Detection (FPSD), Bayesian modelling —are used and skin masks are obtained. The outcome results are compared with the ground truth skin mask of HGR1 database. We observe that the results obtained with multi-seed propagation in multi-layer graph (MPMG) method are almost comparable to ground truth skin mask.