INDIAN INSTITUTE OF TECHNOLOGY(BHU), VARANASI

Report submitted in fulfillment of requirement of Exploratory Project of

B.TECH SECOND YEAR

Pattern Recognition and Wrist Localization in Hand Silhouettes

---

Students
Khayati Mittal(15095042)
Samikshya Chand (15095062)
Sanya Singh(15095063)

Project Supervisor
Dr. K. Sarawadekar

Under Guidance of
Gourav Modanwal

Certificate

*This is to certify that the work contained in this report entitled "**Pattern Recognition and Wrist Localization in Hand Silhouettes**" being submitted by **Khayati Mittal (Roll No. 15095042), Samikshya Chand (Roll No. 15095062) and Sanya Singh (Roll No. 15095063)** carried out in Department of Electronics Engineering, Indian Institute of Technology (BHU) Varanasi, is a bonafide work of or supervision.*

Acknowledgement

*We would like to express our sincere gratitude to Dr. K. Sarawadekar and Gourav for giving us this opportunity to explore the enormous field of Machine Learning and Pattern Recognition by encouraging us to work on this project titled "**Pattern Recognition and Wrist Localization in Hand Silhouettes**", wherein we carried out the study of various algorithms for accurate detection of wrist points to ease out the process of hand gesture recognition.*

We would also like to thank Dr.. Amritanshu Pandey, Convenor of EC201:Exploratory Project for giving us the opportunity to work on this project and give us the opportunity to present our report.

Last but not the least, we would like to thank our parents for always encouraging us in our endeavors and being our support system in every adverse situation that we faced.

We have tried to work with full dedication on this project

Contents

1 Introduction

Hand gestures play an important role in the everyday lives. Not only do they support the verbal communication between humans, but they may constitute the primary communication channel for people with disabilities and simplify the human machine interaction. Recognizing hand gestures is not trivial and, due to its wide practical applicability, has been given considerable research attention over the years.

The methods in this project is validated using the database containing 899 images with the ground-truth data (http://sun.aei.polsl.pl/~mkawulok/gestures).

The algorithms analyzed in this project were implemented on MATLAB and run on an Intel Core i5 1.60 GHz (8 GB RAM) computer.

Furthermore, the failures of the presented algorithm are very often observed for hands with a long sleeve, where the wrist point is located on the contour.

Pattern recognition is a branch of machine learning that focuses on the recognition of patterns and regularities in data
The basic steps of pattern recognition is skin region detection and segmentation,hand detection and localization,hand feature points detection,hand modeling and pose estimation, tracking of the feature points in video sequences.

In this project, we have worked on the second and third step, namely - hand detection and localization and feature points detection.

We have detected the wrist points using our real-time wrist localization technique, Finding maximum inscribed circle in palm and the method based on the image moments calculation.

2 Algorithms

2.1 Longest Chord Method

The longest chord can be determined using a brute-force approach, in which every chord of the hand is verified. However, this results in an average processing time of 49 seconds per image. Thus, a randomized approach was applied which significantly reduces the search time (approx. 550 times). The algorithm for finding the longest chord is presented is initiated with a randomly chosen chord defined by two points P and Q from the hand contour. A new pair of points is chosen randomly from the contour and it is checked whether they form the longest and valid chord . If they do, then a new iteration is initiated. In every iteration, a maximal number of L random chords are found. If none of them updates the chord determined inthe previous iteration, then the local search is performed .It consists in checking additional eight chords, determined by the points that lie at a small distance along the contour from the current end points. If the chord is updated, then the next iteration starts. If the local search fails to improve the result, then the search is stopped.

The image is then rotated by the angle of the longest chord hence obtained. It may be observed that the contour of the palm is usually longer than the opposite part. Given this assumption, we verify the image orientation and start the profile analysis from the palm region.The longest chord PQ determines the direction, along which the widths of the silhouette are calculated. During the wrist search process, the local maximum only needs to be detected, rather than precisely located. Thus, its rough position is found by sampling the profile , until the extremum is determined. Thus, the wrist points are calculated from the extremum indices hence obtained.

The failures of the presented algorithm are very often observed for hands with a long sleeve, where the wrist point is located beyond the wrist.

Algorithm 3 Wrist localization based on profile analysis

INPUT: $\boldsymbol{S}, \delta, \varepsilon, k, n_i$ ▷ a vector of the rotated mask profile S and the settings

$x \leftarrow 0; n \leftarrow n_i;$ ▷ x indicates the position in the profile
while $\boldsymbol{S}[x+\delta] > \boldsymbol{S}[x]$ **do**
    $x \leftarrow x+\delta;$ ▷ find the local maximum
end while
while $n > \varepsilon$ **do** ▷ find the local minimum
    if $\boldsymbol{S}[x+n] < \boldsymbol{S}[x]$ **then**
        $n \leftarrow n/k;$ ▷ contract the search step
    else
        $n \leftarrow -n/k;$ ▷ change the search direction
    end if
    $x \leftarrow x+n;$
end while

2.2 Moment Orientation Method

Instead of finding orientation and rotating the image with respect to longest chord, instead we can find the orientation of the hand with the help of the largest fitting ellipse. This is the method of moment.

2.3 Maximum Inscribed Circle

This is a mathematic model for handforearm segmentation. Then an accurate and effective approximation algorithm is proposed to obtain the solution of the model.
Distance transform is applied on the segmented images to find the center of the maximum inscribed circle. In most cases, the palm center is the point which is furthest away from the contour points of the hand.

Step 1: Make use of distance transform of the skin mask.The point where the distance transform is maximum is the palm center.

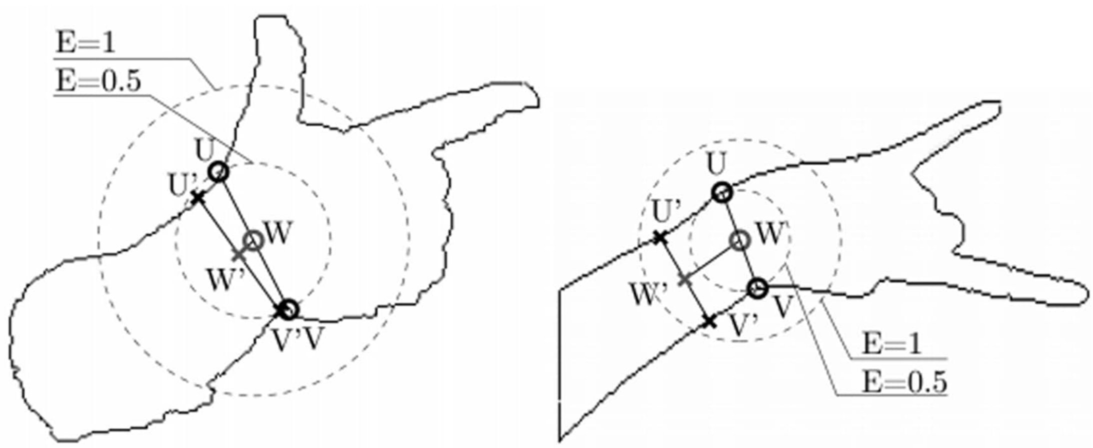
Step 2: Calculate the direction vector of hand using orientation function.

Step 3: Find the maximum radius of the circle inscribed with the palm center as the center of the circle.

Step 4: Go through all points in the contour of this circle and find the point where distance transform is maximum, this is the required wrist point.

3 CALCULATION OF ERROR

Silhouette with the ground-truth (U, V, W) and detected (U', V', W') points and possible wrist point areas.



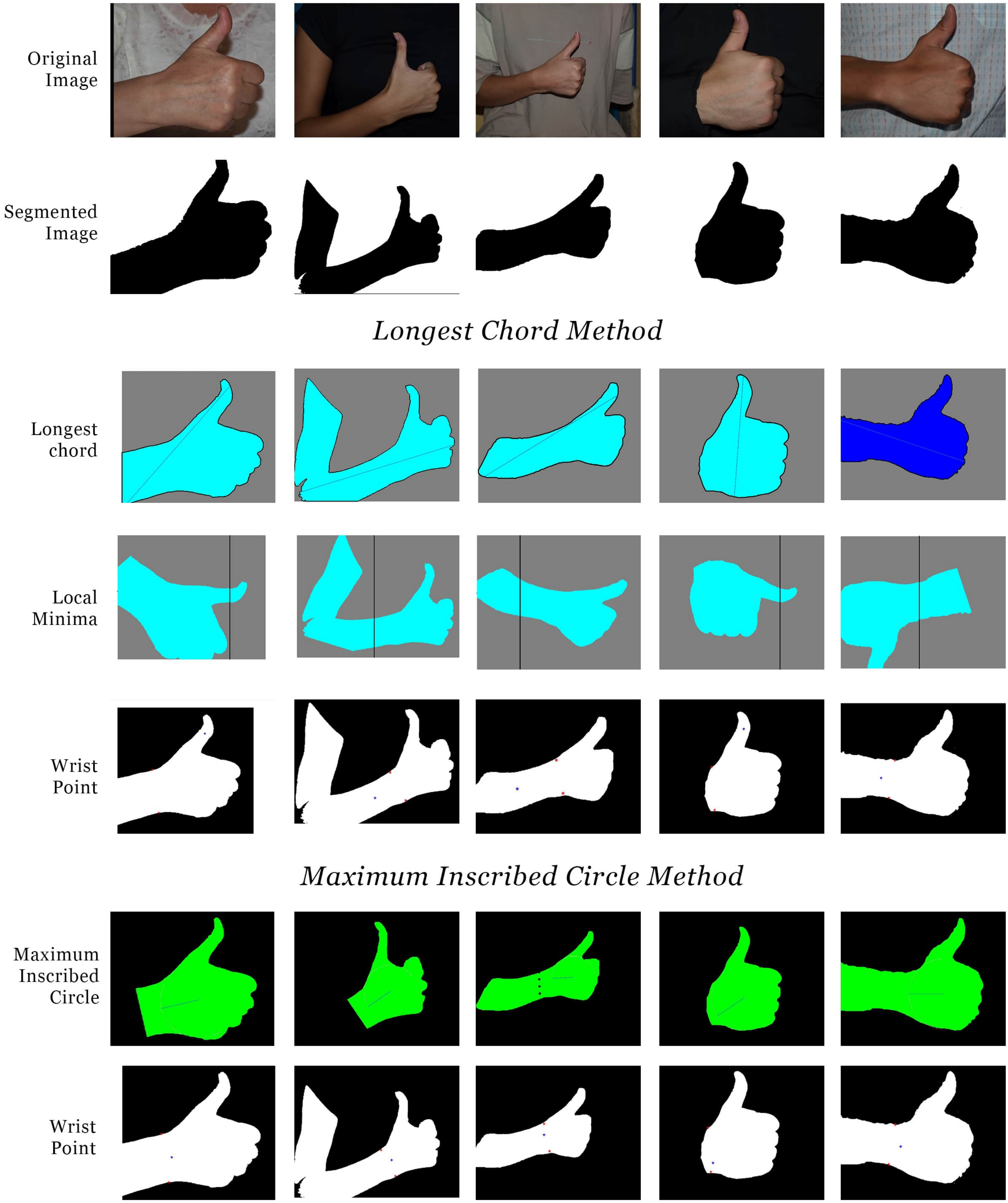
In order to verify the performance of our approach, we detected the points U', V' and W' for each image, where W is the wrist point, and compared them with the ground-truth points U,V andW. The detection error e is defined as e=|WW'|/|UV|.

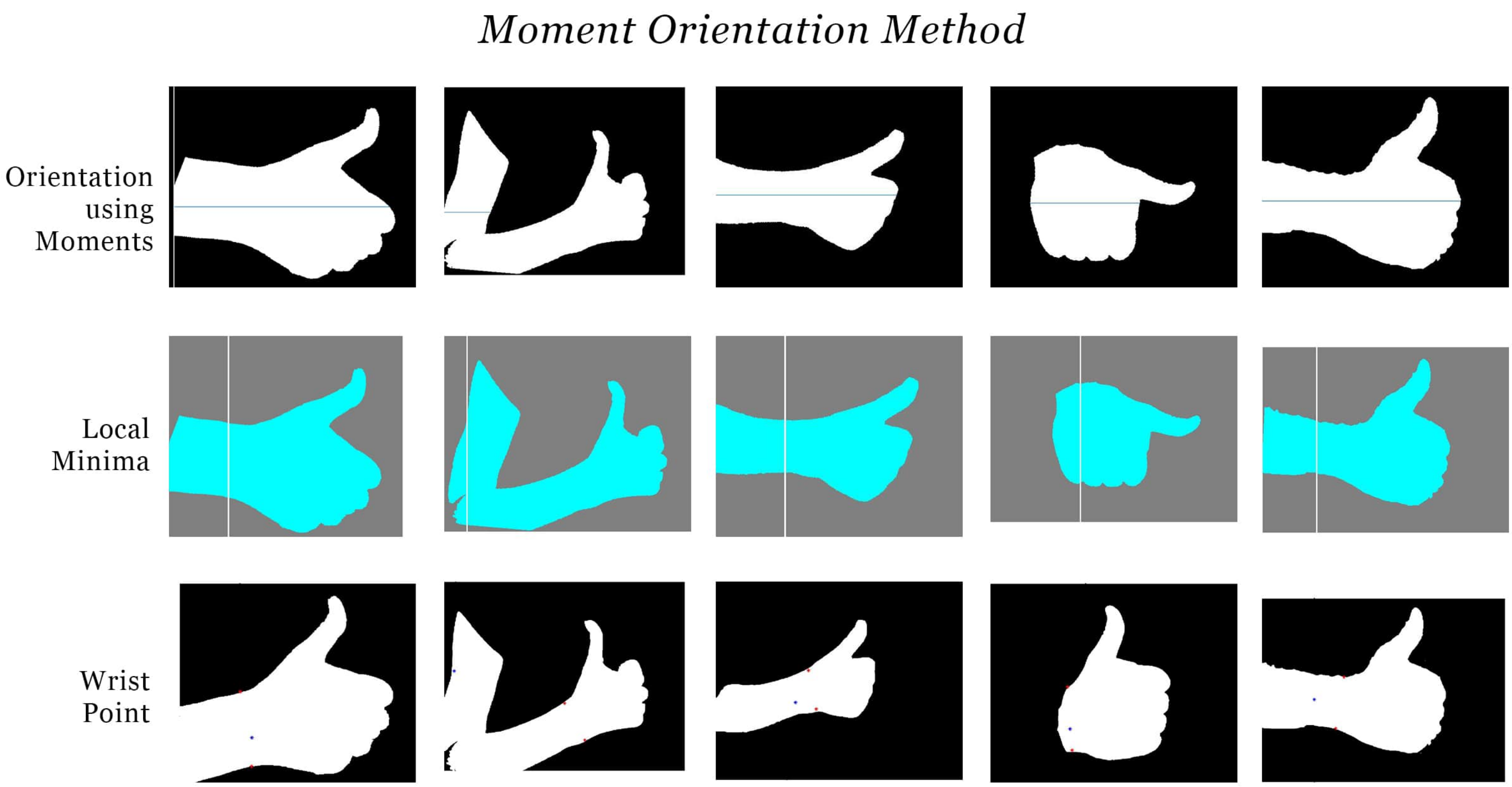
4 Analysis and Discussion

In this section we compare the accuracy of wrist detection for the ground-truth masks. We detected the wrist points using the longest chord method, moments calculation and largest enscribed circle method.

We can see that the maximum inscribed circle method outperformed the other two.

Wrist Detection	***e (value of error)***			
	0.5	***0.7***	***1.0***	***1.5***
Longest Chord Method	*0.39*	*0.44*	*0.56*	*0.65*
Moments	*0.40*	*0.46*	*0.55*	*0,69*
Maximum Circle Inscribed	*0.54*	*0.62*	*0.88*	0.99


Original Image
Segmented Image
Longest Chord Method
Longest chord
Local Minima
Wrist Point
Maximum Inscribed Circle Method
Maximum Inscribed Circle
Wrist Point


Moment Orientation Method
Orientation using Moments
Local Minima
Wrist Point

5 Ongoing Algorithms

5.1 Random Forest Algorithm

A set of fast-computable depth features for static hand posture classification from a single depth image can be computed. The proposed features, which are extracted from pixels on randomly positioned line segments, are specially designed as low-level cues to be used in random forest classifier which combines the cues to discover high-level unseen informative structure in an infinite dimensional feature space.

The proposed features, while being simple, can effectively capture both hand geometry shape and depth texture information.

5.1.1. Hand Geometry Values and Orienataion

To extract the said features from a depth image I(x) of a human hand, we first need to know the following geometry measurement values of the hand in I(x):
a) the approximate location Xc of the centre of the hand,
b) the approximate distance h between the wrist and the furthest part of the hand from the wrist, which varies from posture to posture,
c) the approximate in-plane orientation θ of the hand, which is defined as the angle between the 'axis' of hand and vertical direction of the image. The hand 'axis' is defined as a line that passes through the centre of wrist and separates the hand into two parts with equal area.
d) the foreground region, i.e. the set of pixels that belong to the hand.

A coordinate system is set with the origin at Xc, the x-axis being parallel to the hand axis and 1 unit equalling to h pixels in length. A line segment is parametrized by four parameters based on the coordinate system. The centre of the line segment is placed at (α, β), where α and β are random values and belong to $[-0.5, 0.5]$. The angle between the line segment and x-axis is defined as random value γ and the length of the line segment is defined as random value δ, where γ is $(-90°, 90°)$ and δ is $(0, 2.8)$. The upper bound of δ is set to be 2.8 so that when (α, β) is placed the furthest from the origin, the line segment will still be able to thoroughly pass the hand region.

5.1.2 Feature Extraction

For a given image I(x) and a line segment L (α, β, γ, δ) on it, the following four features are calculated based only on depth values of pixels intersected by the line segment. Let A(L) denote the set of all pixels that are intersected by the line segment, and S(L) the set of foreground pixels in A(L). The first feature is defined as the percentage of foreground pixels on the line segment,

$$f_1(\mathrm{I}; L) = \frac{|\mathcal{S}(L)|}{|\mathcal{A}(L)|}$$

The second feature is defined as the average depth value of foreground pixels on the line segment,

:

$$f_2(\mathrm{I}; L) = \frac{1}{|\mathcal{S}(L)|} \sum_{\mathbf{x} \in \mathcal{S}} \mathrm{D}(\mathbf{x})$$

where D(x) is a shifted and scaled function of the original image I(x). The range of[0, 1] in D(x) corresponds to the range of[I − 100, I + 100] in I(x), where I denotes the average depth value of foreground pixels.

$$\mathrm{D}(\mathbf{x}) = \frac{\mathrm{I}(\mathbf{x}) - \bar{\mathrm{I}} + 100}{200}$$

then the third feature is defined as the position of x(l) on the line segment L, i.e.

$$f_3(\mathrm{I}; L) = \frac{|\mathcal{R}(L)|}{|\mathcal{A}(L)|}$$

where 9(L) is the set of pixels on the line segment that have smaller xcoordinate values than xl, i.e.

$$\mathcal{R}(L) = \{\mathbf{x} \in \mathcal{A}(L) | \mathbf{x}_1 < \hat{\mathbf{x}}_1\}$$

where x1 denotes the x-coordinate of x and xl1 denotes the x-coordinate of xl. The last feature measures the degree of fluctuation along the line segment, and is defined as the sum of the absolute depth value changes along the line segment,

$$f_4(\mathrm{I}; L) = \sum_{(\mathbf{x}_1, \mathbf{x}_2) \in \mathcal{N}} |\mathrm{D}(\mathbf{x}_1) - \mathrm{D}(\mathbf{x}_2)|$$

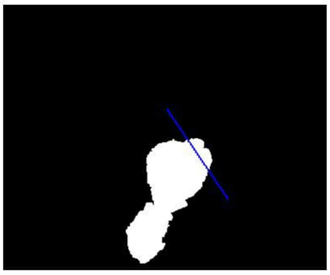
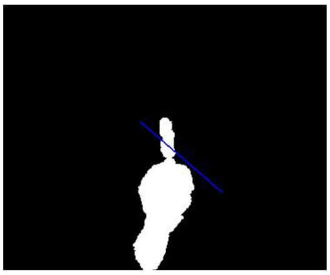
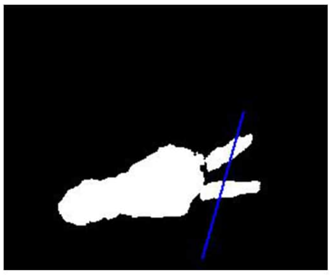
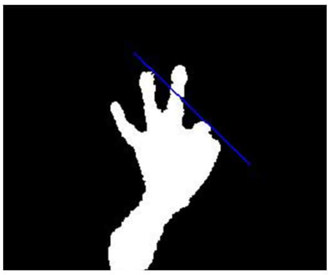
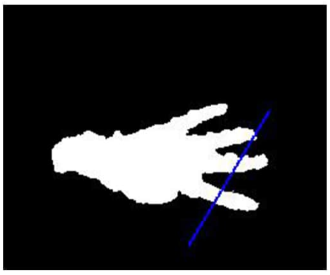
5.1.3 Classification using Random Forest

In our algorithm, a random tree is a binary decision tree in which each test image travels from the root node to one of the leaf nodes. Each node, except for leaf nodes, is a decider on whether the image will be sent to its left child node or right child node, according to whether the decision function stored on the node outputs a value greater than zero or not when taking the image as input. The leaf node that the image reaches will make predictions on the probabilities of the image belonging to each class.

A random forest is an assemble of random trees and each tree is trained separately. Predictions from all trees are averaged to generate the final prediction by the random forest.

5.1.4 Training

We have trained over 1000 depth images for 10 decision trees out of which 7 decision trees gave high accuracy. However, our overall accuracy tends to be only 63% and hence we will work on better features which can give higher accuracy

Original Image	After Depth Thresholding	f(1)	f(2)	f(3)	f(4)
		0.3455	0.1203	0.3455	0.0215
		0.1720	0.0606	0.2581	0.0050
		0.2652	0.0669	0.1894	0.0090
		0.1415	0.0789	0.3208	0.0075
		0.2523	0.0769	0.1892	0.0115

5.2 Elbow Removal

The method used of circle inscribed to find out the wrist points one of the most effective method of all. However, the circle can appear in the elbow when the user is wearing a short-sleeved shirt. To avoid this, two different circles are found, the maximum inscribed circles with center in each half of the blob.

At this point, the skin segments are computed. At first, intersections between the hand contour and a circumference concentric to the palm circle are extracted. Skin segments are the segments between two consecutive intersections that run over a skin region. Due to the arm anatomy, in the hand region one to six skin segments can appear whereas in the elbow region only one or two. In the vast majority of cases two and only two segments appear in the elbow zone if a circle proportional to the one of the palm has been detected. Therefore, further comparison is needed only when there are two segments in each region. In that case, the sum of the segments' longitude must be greater in the elbow region.

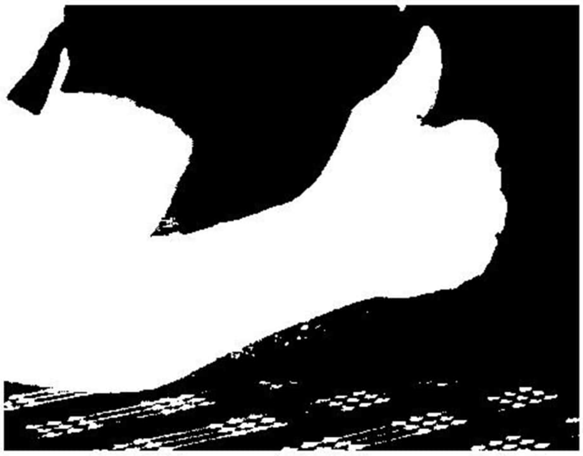
Number of segments		
Circle A	**Circle B**	**Selected Circle**
$\neq 2$	2	A
2	$\neq 2$	B
2	2	Lower $\sum$ longitude

5.2.1 Accuracy

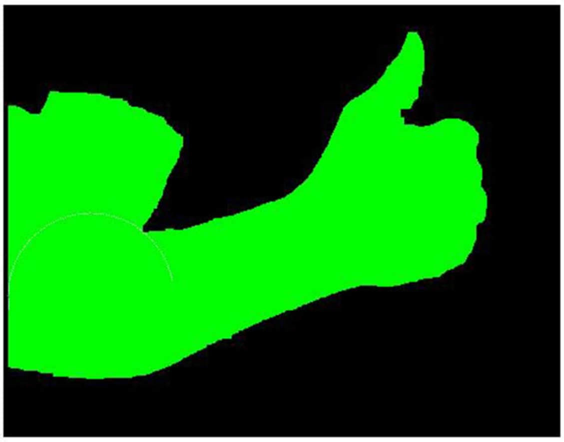
However, it has been observed that the orientation in many of the images is not correct. Also there might be situation where the above conditions for elbow is also satisfied for palm inscribed circle and hence the test fails for some of the images.

The value for the maximum width of the inscribed segment needs to be calibrated so as to obtain the minimum error.

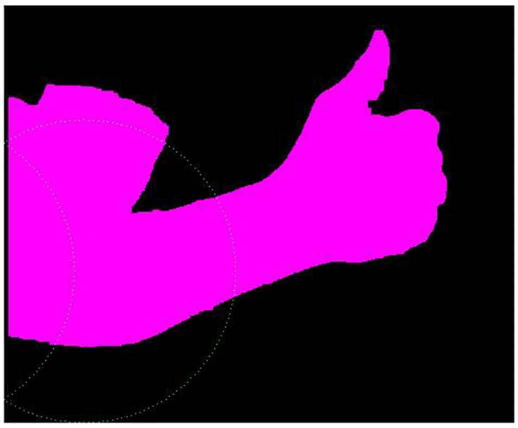
Original Image



Segmented Image

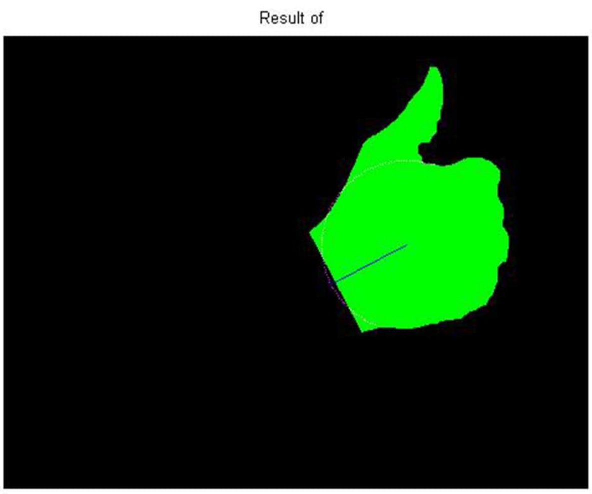


Elbow Incribed Circle



Bigger circle segments

Processed Image



Wrist Point Detection

References

[1] Bosheng Wang, Jiaqi Xu.*Accurate and Fast Hand-Fore-arm Segmentation Algorithm Based on Silhouette.* Multimedia Communication and Pattern Recognition Labs, Beijing University of Posts and Telecommunications, Beijing 100876, China

[2] Tomasz Grzejszczak, Jakub Nalepa, and Michal Kawu-lok.*Real-Time Wrist Localization in Hand Silhouettes*

[3] Jakub Nalepa, Tomasz Grzejszczak, and Michal Kawu-lok.*Wrist Localization in Color Images for Hand Gesture Recognition*

[4] Weizhi Naia, Yue Liua, David Rempelb, Yongtian Wan-ga.*Fast hand posture classification using depth features extracted from random line segments*. Beijing Institute of Technology, 5 South Zhongguancun Street, Beijing 100081, PR China.UC Berkeley, 1301 South 46th Street, Building 163, Richmond, CA 94804, USA