Comparison of Hand Feature Points Detection Methods

Tomasz Grzejszczak, Adam Gałuszka, Michał Niezabitowski, and Krystian Radlak

Silesian University of Technology, Faculty of Automatic Control,
Electronics and Computer Science, 16 Akademicka Street, 44-100 Gliwice, Poland
`{Tomasz.Grzejszczak,Adam.Galuszka,Michal.Niezabitowski,`
`Krystian.Radlak}@polsl.pl`

Abstract. This paper presents the research and comparison of four methods of hand characteristic points detection. Each method was implemented and modified in order to test their capabilities on database for hand gesture recognition. All methods are explained, tested and compared to others with other leading to final remarks. The main purpose of the research is to choose the best algorithm giving the most information about human hand that would lead to create a human – computer interaction program.

Keywords: Hand gestures, HCI, Characteristic points detection, Sign language.

1   Introduction

From the time when first machine was created, there was a need of controlling it. From the levers, through buttons to the present touchscreens, the way of controlling our devices is tending to be more and more intuitive. On the other hand the most intuitive way of communication between people were body language and gestures. Thus in order to create the future, there is a need of gesture detection algorithms. This lead to the question: is it possible to detect the human hand characteristic points accurate enough to classify the shown gestures? What methods of computer vision approach are the best for this problem?

This paper contains the first author's method of hand characteristic points detection [1]. Moreover, the authors of the paper compare the above mentioned method with 3 other algorithms [2 - 4]. All of presented methods were modified in order to fit them into the common testing environment that was performed on created by author gestures base [1, 5 - 7].

2   Relationship to Collective Awareness Systems

The presented solutions on hand feature points detection can be the core of human computer interaction program that can have many applications, such as sign language gestures classification or intelligent home control, and it means that the main contributions of our work coincide with the main theme of the conference "Collective Awareness Systems".

Presented research analyzes the hand detection algorithms which are the core of human computer interaction. Having the good algorithm of hand detection can lead to accurate computer control program, which can lead to various applications, from simple advertisement or information systems to control of a robotic arm manipulator in a sterile environment. All of those applications are the part of extended user communities, linking humans and smart objects.
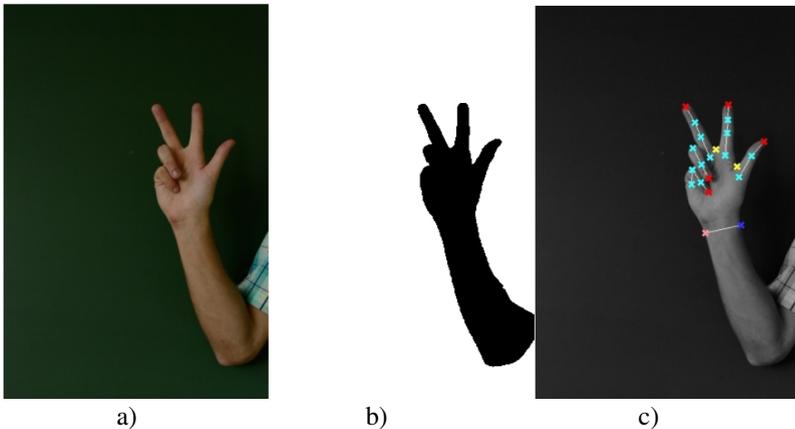
3   Data Base Tests

The presented solutions on hand feature detection were tested on database [7] for hand gesture recognition (HGR). The data are organized into three series acquired in different conditions containing over 1500 images with hands of over 30 individuals presenting about 30 gestures from Polish Sign Language, American Sign Language and some special gestures, such as pointing or grabbing.

Each single gesture that is referred as a one image from database is in fact a set of few things:

- original RGB images (jpg files);
- ground truth binary skin presence masks (bmp files);
- hand feature points location (xml files).

The original RGB image is a photo taken during the base creation. Basing on this image, an expert created a ground truth binary image, where value 0 is referred to background, and 1 is referred to skin region. Finally, expert marked up to 25 characteristic points, such as:

- wrist points;
- 5 finger tips;
- 14 knuckle points;
- 4 points representing space between fingers.

a)   b)   c)

Fig. 1. a) Original hand image, b) Binary mask, c) Hand feature points visualization

```
<IMAGE WIDTH="400" HEIGHT="604" COLOR="TRUE">
 <hand>
  <FeaturePoint x="312" y="333" category="Wrist" type="WristThumb"/>
  <FeaturePoint x="258" y="345" category="Wrist" type="WristPinky"/>
  <FeaturePoint x="309" y="260" category="Thumb" type="Knuckle1"/>
  <FeaturePoint x="328" y="228" category="Thumb" type="Knuckle2"/>
  <FeaturePoint x="347" y="206" category="Thumb" type="Fingertip"/>
       ......
```

File 1. Part of XML File with ground troth feature points localization

All tests of the presented methods were performed on the newest HGR1 database, containing 899 images of 24 gestures from Polish Sign Language.

It is also important to point out that all of the presented methods assume a binary mask as the initial image. In order to obtain a mask, one should perform a segmentation algorithm such as skin detection [6, 8- 15]. All of presented results are based on analyze of ground truth mask, because analyzing output of skin detection algorithm would carry the error of those algorithms.

4   Hand Feature Points

Human hand has 27 bones, where 14 of them are phalanges of fingers. In a simple approximation, hand position or gesture can be described by the orientation of hand, position of wrist and mutual relations between knuckles and fingertips. Human anatomy does not allow bone bending, the actual bending is performed in a joints. This fact can lead to some simplifications in hand modeling. The palm region is the biggest, inflexible object in hand, bounded by wrist, finger first knuckles and concave points between fingers. Fingers consist of 3 phalanges (2 in case of thumb), which can be descried as a constant length line segment bounded by two knuckle points or fingertip. In conclusion, those are the reasons why exactly 25 characteristic points have been chosen in database.

However, not all points can be detected. Some of them cannot be seen in particular hand orientation, other are limited by algorithm capabilities. Most of the presented algorithms are able to detect only the wrist points and finger tips.

4.1   Method 1. Analyze of Contour Points Distance from Wrist Point

The method was inspired by Nobuhiko Tanibata [3]. In this article author described the whole laboratory stand for Japanese Sign Language gesture recognition. The article contains the multiple solutions on whole body gesture detection. One of them is a method of counting raised fingers by analyzing the distance of contour points from wrist point. The method, however assumed that the user would wear a high sleeve, covering the hand, which gives easy palm segmentation. In order to have the method work in any conditions, a set of improvements needed to be added.

In summary, this whole method consists of two curtail steps. First step, developed by the first author and described briefly in paper [1] provides the hand orientation and

wrist detection. Second step, inspired by Nobuhiko Tanibata [3], was expanded in order to detect finger tips. The algorithm of the whole method is presented below:

1.   Obtain mask and its contour;
2.   Find the longest diameter of contour;
3.   Calculate the image profile. Each column of the profile image is a sum of pixels perpendicular to longest diameter in a given point;
4.   Find the local minimum of profile and obtain the wrist points in original image;
5.   Calculate the distance of each pixel of hand contour from wrist middle point;
6.   Analyze the extrema of distance function and obtain the fingertip points.

a)   b)   c)
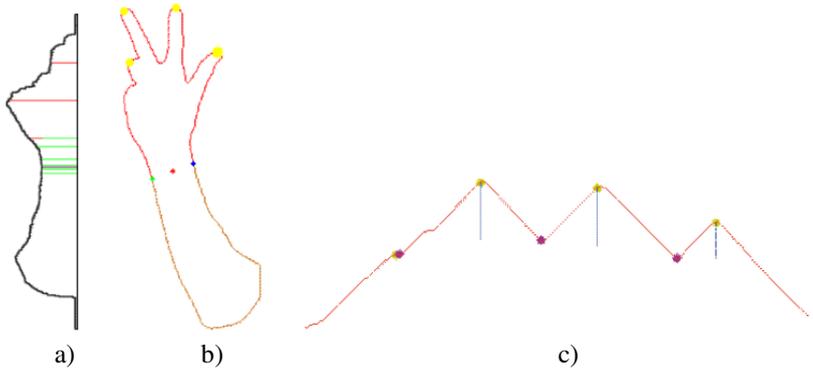
Fig. 2. a) Hand profile along the longest diameter, b) Hand contour with selected wrist region and detected fingers with use of extrema localization on c) function of distance of contour points from wrist point

4.2   Method 2. Distance Transform

Method developed by Maciej Czupryna [2] is based on distance transform. New image is created with the same size as a source image, and each pixel from this new image has a value equal to distance to closest white pixel on source image.

a)   b)   c)
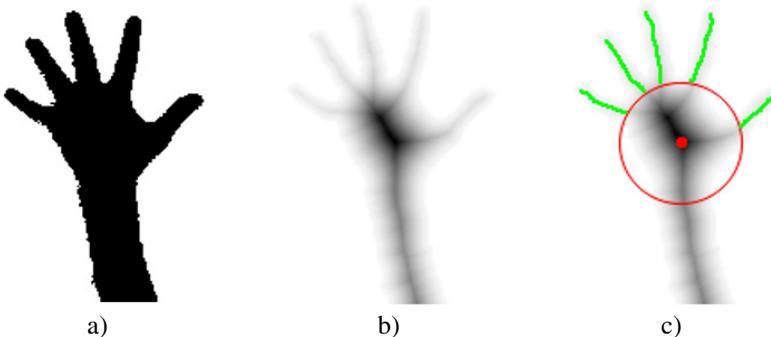
Fig. 3. a) Source image, b) distance transform of source image, c) analyzed image

As it can be observed on Fig. 3, a white pixel representing a background on source image is also represented as value 0 (white pixel). The pixels from palm region are located far from background. In conclusion, the farther from background, the higher value of output image pixel.

Basing on the distance transform image, one can easily obtain the palm region with center in the maximal value of distance transform, and radius equal to 1.3 times value of maximum. The fingers are detected as a constant line of non-rapidly changing local maxima of distance transform. The beginning of these lines are considered as a knuckle and the end as a fingertip.

4.3   Method 3. Template Matching with Use of Cross Correlation

In article [15] author describes a method of template matching. Observing that a fingertip is round ended, the algorithm is based on matching the template image presenting black circle into the source image with use of cross correlation.

a)   b)   c)
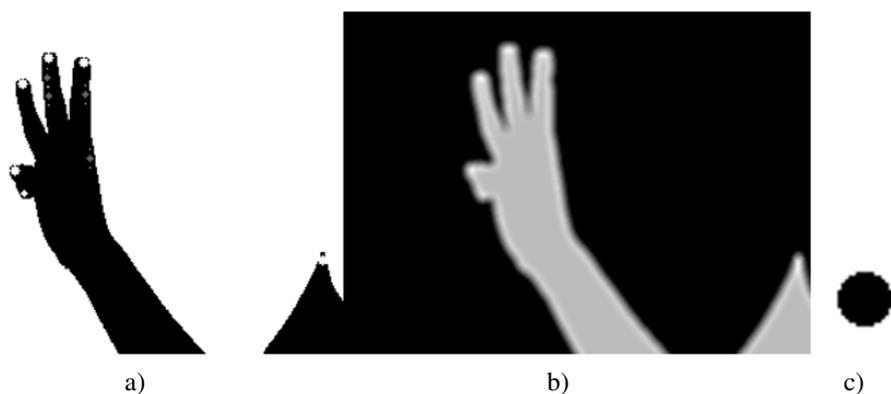
Fig. 4. a) Source image with annotated results (white – correctly detected points, grey – rejected points), b) Cross correlation of source image with c) template

The finger tips should be located in the local maxima of cross correlation output image. However it is important to filter the false positives. First method is based on closeness of detected points. Each detected point restricts an area in which no other points can be detected. Second method is based on the observation that a fingertip is surrounded by huge amount of background. While detecting a new local maximum, the test is performed if for a black pixel located on a border of template (about 20 pixels in radius from detected point) there exists a white pixel on an opposite side of template border. This rule can be observed on Figure5, where gray points are these which did not applied to the following rule.

4.4   Method 4. Analyze of Contour Points Distance from Wrist Point

The rather complicated method of characteristic points detection was presented in [4] where the authors applied the Self Organizing and Self Growing Neural Gas
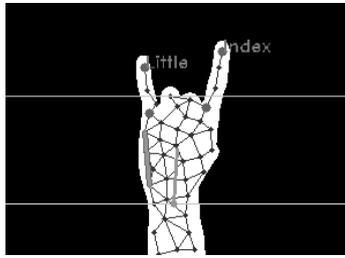
Fig. 5. Example of hand image with SGONG, hand slope detection, palm region detection and finger recognition

(SGONG) on the image of hand. The neural gas was spread on hand image and in stable state, it uniformly covered the hand region.

The neural gas was just a first step of characteristic points detection. The set of rules were applied in order to gather as much information as possible:

1.   Hand profile brings information about palm and wrist localization;
2.   Hand slope is calculated using 2 methods. Slope of leftmost neurons from palm region and a slope between palm and wrist;
3.   Neuron connections passing through background are removed;
4.   Neurons with only one neighbor are considered as fingertips;
5.   From finger tips, neurons with two neighbors are considered as mid-finger region, and with more neighbors are considered as finger base or knuckle;
6.   Fingers are identified with use of hand slope and 3 different measurements.

5   Results Comparison

All of the presented algorithms were implemented and adapted in order to work with our gesture data base. Each algorithm was tested on set of 899 images and the results are presented in Figures 6 and 7. On each image, few characteristic points were found, and these points positions were compared to those from database. The output data is presented in form of cumulative distribution function of detected points displacement. For example, method 3 with total of 3789 detected characteristic points has total of 3294 points with displacement error less than 14 pixels. This gives 89% of all points.
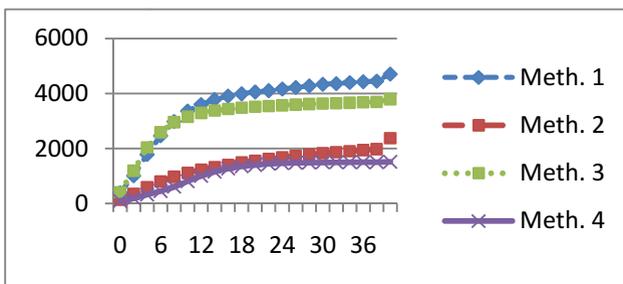
Fig. 6. Distribution of error of characteristic points detection for all presented methods
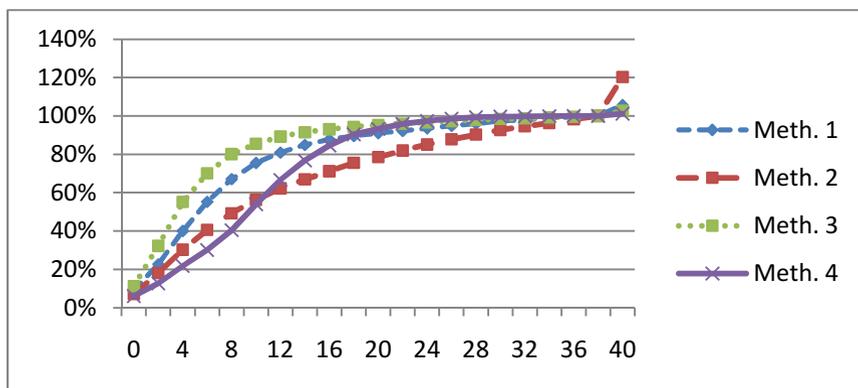
Fig. 7. Distribution of error of characteristic points detection for all presented methods. The values above 100% represents the rest of points with too high error.

6   Discussion of Results and Critical View

In conclusion to chapters 3 and 4, the following remarks have been done.

First method of finger detection with use of distance between contour points and wrist point brings the highest amount of points detection. However, there are images, where the algorithm identify the hand orientation in opposite way, leads to detection in region of elbow. These points are located in last column, with error values above 40.

Second method based on distance transform did not well in images with joined fingers. Bringing the small amount of characteristic points with the lowest accuracy, this method is considered as fast bus inaccurate.

Third method with template matching used to match a finger in sharp mask edges, as it can be shown on example on Fig. 4 where a template is matched on shoulder. This caused the error to be similar to method 1, however with much less characteristic points detected.

Fourth method, that used SGONG, brings the best characteristic points matching, with nearly flat line, while expanding of distribution function. However this method has many limitations with proper hand orientation, causing for example horizontally oriented hand not to be considered in tests. This brings the smallest detection rate. Moreover this algorithm is much slower in comparison to other.

In further work the first algorithm would be used and improved as it brings the most information and detects the highest amount of characteristic points. The improvement should deal with proper hand orientation leading to less false points detection. Moreover an reasoning algorithm would be developed, able to classify the viewed gesture into the proper gesture type.

References

1. Grzejszczak, T., Nalepa, J., Kawulok, M.: Real-time wrist localization inhand silhouettes. In: Proceedings of the 8th International Conference on Computer Recognition Systems CORES 2013, pp. 439–449. Springer, Heidelberg (2013)
2. Czupryna, M., Kawulok, M.: Real-time vision pointer interface. In: Proceedings of International Symposium ELMAR 2012, pp. 49–52 (2012)
3. Tanibata, N., Shimada, N., Shirai, Y.: Extraction of hand features for recognition of sign language words. In: Proceedings of the International Conference on Vision Interface, pp. 391–398 (2002)
4. Stergiopoulou, E., Papamarkos, N.: Hand gesture recognition using a neural network shape fitting technique. Engineering Applications of Artificial Intelligence 22, 1141–1158 (2009)
5. Kawulok, M., Kawulok, J., Nalepa, J.: Spatial-based skin detection using discriminative skin-presence features. Pattern Recognition Letters (2013) (article in press)
6. Kawulok, M.: Fast propagation-based skin regions segmentation in color images. In: Proceedings of 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition, pp. 1–7 (2013)
7. Kawulok, M., Grzejszczak, T., Nalepa, J., Knyc, M.: Database for hand gesture recognition (2012), http://sun.aei.polsl.pl/~mkawulok/gestures/
8. Kawulok, M., Kawulok, J., Nalepa, J., Papiez, M.: Skin detection using spatial analysis with adaptive seed. In: Proceedings ofIEEE International Conference on Image Processing (ICIP 2013), pp. 3720–3724 (2013)
9. Kovac, J., Peer, P., Solina, F.: Human skin color clustering for face detection. IEEE REGION 8 EUROCON 2003, Computer as a Tool 2, 144–148 (2003)
10. Jones, M., Rehg, J.: Statistical color models with application to skin detection. International Journal of Computer Vision 46, 81–96 (2002)
11. Kakumanu, P., Makrogiannis, S., Bourbakis, N.G.: A survey of skin-color modeling and detection methods. Pattern Recogn. 40, 1106–1122 (2007)
12. Phung, S.L., Chai, D., Bouzerdoum, A.: Adaptive skin segmentation in color images. In: IEEE International Conference on Acoustics, Speech, and Signal Processing, vol. 3, pp. 353–356 (2003)
13. Soriano, M., Martinkauppi, B., Huovinen, S., Laaksonen, M.: Skin detection in video under changing illumination conditions. In: 15th International Conference on Pattern Recognition (ICPR 2000), vol. 1, pp. 839–842 (2000)
14. Solar, J.R., Verschae, R.: Skin detection using neighborhood information. In: 6th IEEE International Conference on Automatic Face and Gesture Recognition, pp. 463–468 (2004)
15. Sato, Y., Kabayashi, Y., Koike, H.: Fast Tracking of Hands and Fingertips in Infrared Images for Augmented Desk Interface. In: Proceedings of 4th IEEE International Conference on Automatic Face and Gesture Recognition, pp. 462–467 (2000)