中心軸距離値分析に基づく手の形状の認識手法

A Hand Shape Perception Method based on Medial-axis Distance Value Analysis

ジャン ポン* シオン フアイシン* リ タオ*
Peng ZHANG Huaixin XIONG Tao LI

要　旨

中心軸距離値分析に基づく手の形状の認識手法を開発した．中心軸とその対応する距離値がすべての形状情報を含んでいることから着想したものである．実際，手は3種類の重要な点，すなわち，指先，指の付け根，手のひらの中心，で表現できる．これらの点は中心軸上で急な距離値勾配変化を持つ特別な点となっている．手の形状のそのような点は指や手のロバストな検知に使えることが実験で示されており，さらに，手のジェスチャーの理解にも役立つと期待される．

ABSTRACT

In this paper, we develop a new hand shape perception method based on medial-axis distance value analysis. The main idea is that medial axis and its associate distance value contain all shape information. In fact, a hand can be described by three critical points, i.e. finger-tip, finger-root, palm center, and these critical points are all special points on the medial axis with abrupt distance slope changes. Such critical hand shape points will not only provide a robust finger/hand detection performance which is already shown by experiments, but also help us to understand the hand gesture.

* リコーソフトウェア研究所（北京）有限公司
Ricoh Software Research Center (Beijing) Co., Ltd.

1. Introduction

To control the computer directly by hand is very natural and convenient for our human beings. That is why the current multi-touch technology would be amazing. However, such kind of technology is always based on touch screen, which has many limitations, e.g. screen-size, portability and cost. On the other side, projector is widely used in office environment without the above limitations. So it would be exciting both technically and commercially, if the touch operation can be made on the projection screen. As a result, we can enjoy Human Computer Interaction (HCI) anytime and anywhere.

In fact, recent progress of computer vision makes it possible to simulate the touch function on projection screen. There have already been some efforts, e.g. Gesture-Tek, Microsoft Surface. Real-time and robust hand perception, or hand/finger detection, lies in the centre of this purpose.

However, most current hand or finger detection methods suffer from at least one of the two defects. First, they detect finger in a rather small range, e.g. a region around fingertip. Second, their detection feature lack of physical meaning, which means palm center, arm direction, and other important hand shape information are unknown in the detection stage. As a result, their detection performance is less reliable with many false alarms. In addition, since no enough hand shape information is provided in the detection stage, hand model fitting methods are needed to estimate the hand shape model parameter for further gesture recognition. Such model fitting methods are always very complex and time-consuming, which may not be necessary for some simple gesture recognition tasks.

To solve the above two problems, in this paper, we present a new hand perception method which detects finger in a finger-palm range with distinctive physical features derived from critical hand shape points, e.g. fingertip, metacarpophalangeal point (MCP point), i.e. joint between finger and palm, and palm center.

The HCI system is under an infra-red (IR) solution so that the projection content in visible light will be filtered out. As a result, we can focus on the input gray image of human body.

The most distinctive feature of our method is that it can provide hand shape information during detection process. Such information will not only lead to robust detection result, precise fingertip location, but also help us understand the hand, e.g. arm direction, palm region, and finger direction, which will benefit later high-level applications, such as gesture recognition.

2. Technology

2-1 Main Idea

The main idea behind this finger/hand method is as follows. On one aspect, medial axis and its associate distance value contain object shape information. The medial axis, or skeleton[1)], of an object is a set of points where the maximal inside-object circle centered on each point will include more than one boundary point. The radius of each medial axis point is its associated distance value. In fact, the object can be reconstructed by medial axis and its distance values. Fig.1 is an illustration for the above discussion. On the other aspect, a hand can be described by three kinds of critical hand shape point (CHS point), i.e. fingertip, MCP point, and palm center. These critical points are all special points on the medial axis with abrupt distance slope changes.

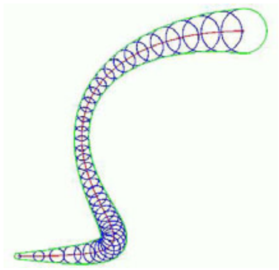

Fig.1 Medial axis and its distance value as the radius of maximal inside-object circle.

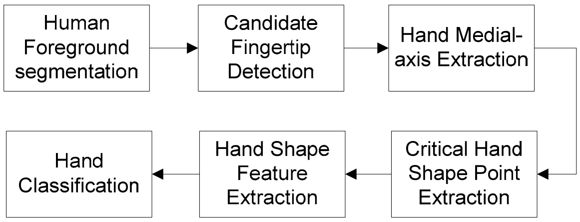

Fig.2 Flow chart of the hand perception method via medial axis distance value analysis.

The above idea leads to a hand perception method summarized in Fig.2 which consists of the following 4 main steps. First, we get some candidate fingertip, which can greatly reduce the detection region for further verification. Second, we get the hand medial axis which starts from candidate fingertip point and is long enough to pass through palm center. Third, we find candidate CHS points along the medial axis, i.e. MCP point and palm center, via distance value analysis. In fact, critical points are all special points on the medial axis with abrupt distance value slope changes. To be specific, the distance value slope along medial axis is close to 0 in a neighborhood before the MCP point, and close to 1 in a neighborhood after MCP point. The slope is close to 1 before and close to 0 after palm center point (Fig.3). Finally, we can derive some distinctive feature of strong physical meaning from the CHS points, and feed it to a classifier for hand detection.

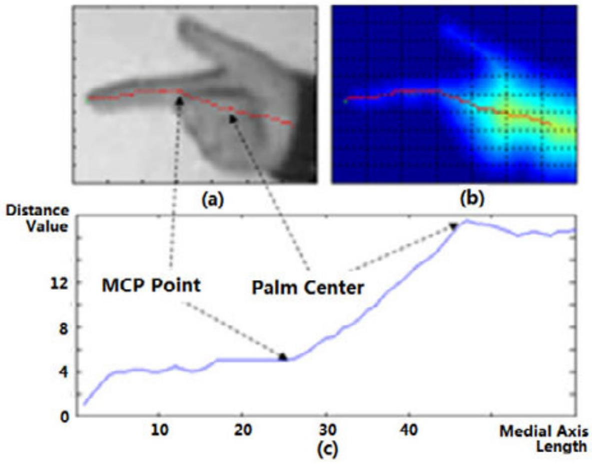

Fig.3 Illustration of critical hand shape (CHS) points and their distance values. (a) The hand medial axis from fingertip through palm center is overlaid on the original gray image. (b) The medial axis overlaid on distance transform image. (c) The distance value along the medial axis. MCP point and palm center are marked as CHS points.

With the CHS points and distance value, we can also get some hand gesture information, e.g. finger direction, finger number, arm direction, and etc., which will also benefit later gesture recognition.

Both the four detection steps and some hand geometric measurement will be introduced in the following subsections.

2-2 Candidate Fingertip Detection

This is the preprocessing stage of the whole method where we first convert the input IR gray image into binary image with human foreground. A binary image f(x) is an image consisting of only two kinds of pixels, i.e. uninformative background (f(x)=0), and informative foreground (f(x)=1).

Then, some fingertip candidates will be found for further verification. The binary human image can be got by background subtraction. The fingertip candidates may be detected by different methods. One possible idea is to define fingertip as a corner point, which can either be detected from gray image via Harris detector[2], or be detected from binary human image via Susan detector[2].

The main benefit of candidate fingertip detection is to save time, because the major computation cost of this method is hand medial axis extraction which is introduced later. So no matter what fingertip detection method is used, it should be radical enough to filter out as many edge points as possible for system efficiency, while it should also be reserved enough to keep all real fingertip points for system reliability.

2-3 Hand Medial Axis Extraction

This step will find a medial axis which starts from candidate fingertip and is long enough to pass the palm center.

However, direct medial axis transform or skeleton extraction method is very time-consuming, because it is performed in a sequential iterative style[1]. Luckily, medial axis is highly related to distance transform. The distance transform $Df(\mathbf{x})$ of a binary image $f(\mathbf{x})$ is the distance between current pixel $\mathbf{x}$ and its nearest background point. Distance transform can be achieved by forward-backward scan with an $O(N)$ complexity[1]. In fact, it is shown that the medial axis of $f(\mathbf{x})$ is just the ridge in $Df(\mathbf{x})$[3], which can be traced by maximal gradient ascending. Since the start candidate fingertip may not exactly lie in a ridge, the tracing method is different from 3), which is summarized in Fig.4.

2-4 Critical Hand Shape Points

A hand can be described by three CHS points, i.e. fingertip, MCP point, and palm center Fig.3(a). These CHS points are all special points with special slope change and large curvature value along medial axis distance value. To be specific, at MCP point, the slope will change from about 0 to 1; on a palm-center, the slope will change from about 1 to 0. One example on hand shape critical points and its physical meaning are illustrated in Fig.3. The algorithm following the above discussion on CHS point extraction is presented in Fig.5.

2-5 Hand Shape Feature and Classifier

We use the term **hand shape (HS) feature** for any feature derived from CHS points. HS feature has strong physical meaning and can be fed into a classifier for final hand perception.

Input: *FingerTip* // finger-tip candidate
Output: *medAxis* // output medial axis from finger-tip to palm centre
— *medAxis*.push_back(*iniPoint*); //Initialize *medAxis* by pushing *FingerTip* into it.
— **While** *medAxis*.size() <= L_{trace}
　— *curPoint* = *medAxis*.back()// the last point of *medAxis*;
　— *chkPointList* . clear (). // *chkPointList* is candidate of next *medAxis* point
　— **For** *chkPoint* ∈ *N*8(*curPoint*) // *N*8(**x**): 8-neighbourhood of **x**
　—*chkPointList*.push_back(*chkPoint*), if *chkPoint* is an unchecked foreground point, and the linkage between *curPoint* and *chkPoint will not lead to direction change ≥90 for medAxis.*
　—**End**
　—**If** Df(*curPoint*) ≥ max{*Df*(*chkPointList*)} // Df(x): distance transform of binary image f(**x**);
　　— sort *chkPointList* by their distance to *curPoint* in a decreasing order
　　— push the **first ridge point** of the ordered *chkPointList* into *medAxis;*
　　— if there is no ridge point in *chkPointList*, stop tracing process;
　— **Else**
　　— sort *chkPointList* by their directional derivative to *curPoint* on Df(**x**).
　　— push the **first point** of ordered *chkPointList* into *medAxis*
　— **End**
— **End**

Fig.4 An algorithm of Medial axis extraction via maximal ridge tracing on distance map.

```
Input: medAxis // hand medial-axis start from candidate fingertip
Output: MCPList, palmList // MCPList¥palmList is a list of candidate MCP-point¥palm-centre in medAxis
— Let distList be Df(medAxis); // distList is the distance value along medAxis
— For 1 ≤ curIdx ≤ length(medAxis) // curIdx is the current check index.
  — Let forSlope be the line fitted from distList in a neighborhood after curIdx;
  — Let bakSlope be the line fitted from distList in a neighborhood before curIdx;
  — Let curvature be (forSlope – bakSlope);
  — push medAxis[curIdx] point into MCPList, if forSlope~1, bakSlope~0, and curvature>Tc
  — push medAxis[curIdx] point into palmList, if bakSlope~1, and curvature>Tc
        // slope~y means abs(slope – y) <Ts
— End
— for adjacent points in MCPList or palmList, only retain the one with highest curvature
```

Fig.5 An Algorithm of Critical Hand Shape (CHS) point extraction.

It is clear from the CHS detection algorithm in Fig.5 that there can be more than one candidate palm-center and MCP points in a hand medial axis starting from fingertip. As a result, we can explore all possible combinations, and find the true point by classifier. To be specific, we generate a set of CHS-point triples (*fingertip*, *MCP-point*, *palm-center*) based on the combination of each candidate MCP and palm center. Each triple will generate a HS feature for hand classification. We first introduce three parameters which is very important for HS feature derivation. First, finger-width (FW) is the distance value on MCP-point, i.e. Df(MCP-point). Second, the palm-width (PW) is the distance value on palm center, i.e. Df(palm-center). Third, finger-length (FL) is the medial axis length from fingertip to MCP-point. Although there can be a lot of possible choices on HS feature, we think the following triple feature is very distinctive:

$$[FW, FL/FW, PW/FW] \qquad (1)$$

The above HS feature has strong physical meaning, i.e. finger-width, normalized finger-length, and normalized palm width. It should be noted that other variables can be introduced into HS feature, e.g. the line-fitting slope and error of the medial axis distance values from MCP-point to palm-center. However, we think the above 3 variables in (1) are essential for a robust classification result.

Hand classification step aims to make a decision whether the hand medial axis really is a hand based on its HS feature. There are a lot of classifiers for this purpose, e.g. SVM, neural network, and decision tree[4). However, considering the physical meaning of our hand shape feature, decision tree may be the most appropriate one, which also presents the best experiment result. Of course, decision tree ensemble classifiers, e.g. Ada-boost, random forest[4), may lead to better result. Also it should be note that in some cases, there may be more than one positive CHS-point triple in a hand medial axis after classification. This situation can be solved by considering the consistency with the detection result in last frame, i.e. tracking, in either a deterministic or probabilistic framework.

2-6 Hand Gesture Information

The CHS points, e.g. fingertip, MCP point, and palm-center, and their distance values contain so many hand shape knowledge that we can get some hand gesture information for later gesture recognition. Different gesture information may be needed for different application task. We will give some important example in the below.

Number of visible fingers. The number of visible fingers is very important for gesture recognition. With the knowledge of palm center and palm width, we can associate each hand medial axis to its original hand.

Therefore, we find the identity of all hands and their related fingers.

Finger direction. The finger direction is useful for gesture recognition, and it can be estimated from the medial-axis segment from MCP point to fingertip candidate via line-fitting (Fig.6).

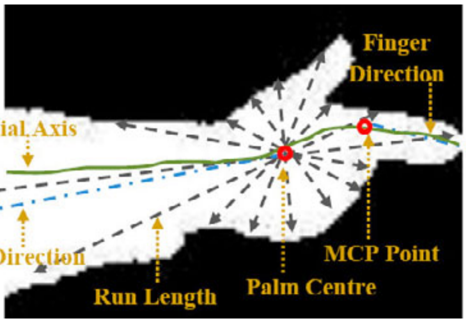

Fig.6 Hand gesture information based on CHS point. Finger direction estimation via line-fitting from medial axis between MCP point and Fingertip; arm direction estimation via maximal run-length.

Fingertip position. The candidate fingertip may not be very spatially precise due to different detection methods. Therefore, a re-location step may be necessary in the application of fingertip positioning. In fact, fingertip can be estimated as the furthest foreground point along finger direction.

Arm direction. Arm direction can be derived from the ray starting from palm center and passing through arms. Its estimation is easy, since the palm center is already known. Start from the palm center, we test a series of rays form 0-degree to 360-degree. For each ray of -degree, let its run-length in the foreground region be RunLen(θ), which is the distance between palm center and the first background point on this ray. The arm axis should be the ray starting from palm center on the following direction θ*:

$$\theta^* = \operatorname{argmax}_\theta \ \text{RunLen}(\theta). \tag{2}$$

Fig.6 is an example on run length and arm direction estimation. Arm direction is very important to identify fingers. To be specific, fingers may be differed from each other according to the intersect angle between arm direction and the segment from palm-center to its MCP point. Such angle for finger identification is called **MCP-Arm angle**.

3. Results

The major benefit of our hand perception method is that its CHS points and HS feature has clear physical meaning. Therefore, we can specify the finger-root, palm center, and arm direction very clearly in the detection result in Fig.7, where the index finger is distinguished from thumb via MCP-Arm angle.

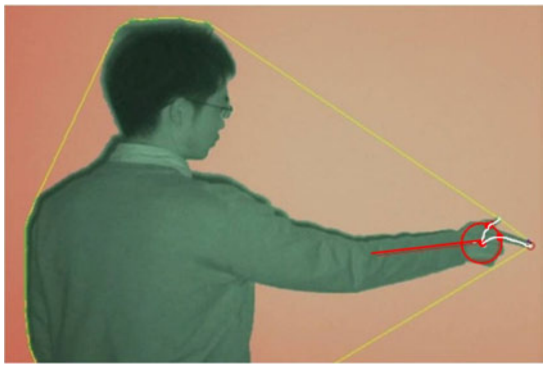

Fig.7 Hand detection result with hand medial axis, CHS points, arm direction, and palm circle.

We use our hand perception method for finger detection and test it on a database consist of 8 videos. We compared this method with our last finger detection method, i.e. Finger-Mouse II[6], which is based on convex-hull point and some local features. The comparative in Table 1 shows that the HS feature derived from the whole hand region is very helpful for a reliable detection performance, especially a low false alarm rate.

Table 1 Finger detection result.

		Positive	Negative	Total
	Sample	854	116	970
Finger-Mouse II	False	23	70	93
	Accuracy	97.31%	39.65%	90.41%
Our Method	False	15	2	17
	Accuracy	98.24%	98.27%	98.25%

The operation speed of our hand perception method is real-time. The whole system achieves a speed of about 32ms per frame in a platform of Intel Core 2 Quad CPU Q4900 2.66-GHZ.

4. Future work

Our hand perception method (Fig.2) can be further improved on the following aspects.

First, better human foreground segmentation method is necessary, since our hand perception method is based on the output binary image. Current background subtraction with automatic global threshold will fail in the case of non-uniform luminance or overall gain variation. Background updating and edge fusion may be helpful.

Second, better medial axis tracing method is helpful. Current tracing method is just a greedy method based on local heuristic information in an 8-neibour region. It may fail in some extremely rare cases. Global heuristic information may be helpful.

Third, a tracking-by-detection[5)] step will also help to improve the hand perception performance, especially a lower false alarm rate, by associating the detection result between adjacent frames.

References

1) M. Sonka, V. Hlavac, R. Boyle: Image Processing, Analysis, and Machine Vision, 3rd ed., Thomson Engineering, (2007).

2) T. Tuytelaars, K. Mikolajczyk: Local invariant feature detectors: a survey, foundations and trends, Computer Graphics and Vision, Vol. 3, No 3, (2008), pp. 177-280.

3) F. Shiha, C. Pu: A skeletonization algorithm by maxima tracking on Euclidean distance transform, Pattern Recognition,Vol. 28, Issue 3, (1995), pp. 331-341

4) T. Hastie, R. Tibshirani, J. Friedman: The Elements of Statistical Learning, 2nd ed., Springer-Verlag,(2008).

5) M. Breitenstein et al.: Robust tracking-by-detection using a detector confidence particle filter, IEEE Intl. Conf. on Computer Vision, (2009).

6) H. Xiong: A fast fingertip detection method, Chinese patent, 201010279511.3, (2010).