LETTER

# Hand Gesture Recognition Based on Perceptual Shape Decomposition with a Kinect Camera

**Chun WANG**[†a)], ***Nonmember***, **Zhongyuan LAI**[†b)], ***Student Member***, ***and*** **Hongyuan WANG**[†], ***Nonmember***

**SUMMARY** In this paper, we propose the Perceptual Shape Decomposition (PSD) to detect fingers for a Kinect-based hand gesture recognition system. The PSD is formulated as a discrete optimization problem by removing all negative minima with minimum cost. Experiments show that our PSD is perceptually relevant and robust against distortion and hand variations, and thus improves the recognition system performance.

***key words:*** *Kinect camerta, hand gesture recognition, perceptual decomposition, finger detection*

## 1. Introduction

Hand gesture recognition is important for human-computer interaction (HCI) in different areas, such as virtual reality, sign language recognition, and computer games [1]. Usually the shape feature is sufficient for successful recognition [2]. However, due to the nature of optical sensor, the quality of captured images is sensitive to the lighting conditions and cluttered backgrounds, which makes it very difficult to obtain hand shapes [3].

Thanks to the recent advent of the Kinect depth camera [4], new opportunities for hand gesture recognition emerge. Ren et al. were the first to develop a real-life hand gesture recognition system with a Kinect sensor [5], as shown in Fig. 1. With the assistant of depth cue, their system can segment hands robustly. However, due to the low-resolution of the Kinect depth map and the small size of hand image, the segmentation of hand is usually inaccurate and noisy, which significantly affects the recognition performance. In order to address this problem, they proposed the Finger-Earth Mover's Distance (FEMD), which only matches the fingers rather than the whole hand [5]. Thus, the finger detection has a significant impact upon robustness, accuracy and efficiency of their hand gesture recognition system.

Ren et al. presented two shape decomposition methods for finger detection. The first one is the Thresholding Decomposition (TD) [6]. This method decomposes outstretched fingers using a circle that is concentric with the maximal inscribed circle of the hand shape and of radius a specified threshold value. Despite the implementation simplicity and efficiency, this method is perceptually irrelevant, thus usually unable to detect fingers accurately. To address this problem, the second method called Minimum Near-Convex Decomposition (MNCD) was proposed [8], [9]. This method defines the mutex-pair as two contour points whose concavity is larger than a threshold, and resolves all the mutex-pairs by selecting cuts from the inner line segment set via Binary Integer Linear Programming (BILP). Despite the detection robustness and precision, this method needs to deal with a massive number of mutex-pairs, inner line segments and their complex relations, which significantly affects the efficiency of the overall system. Thus, a shape decomposition method that can better balance the robustness, precision and efficiency is required.


**Fig. 1** The framework of Ren's part-based hand gesture recognition system [5].

To meet this requirement, we propose a novel shape decomposition method called Perceptual Shape Decomposition (PSD). With the help of Discrete Contour Evolution (DCE) [10] and DCE-based skeleton pruning [11], we construct the negative minimum set and its symmetry set, which are of small size, high precision and robustness. Then we construct the candidate cut set, and decompose the shape by selecting cuts that can resolve all the negative minima with minimum cost. To improve the perceptual relevance, we impose three heuristic criteria, the negative minima rule [12], the short cut rule [13] and the part salience [14], [15], into the cost function. Experiments validate the advantages of our PSD embedded into part-based hand gesture recognition systems with a Kinect camera.

## 2. Perceptual Shape Decomposition

To define perceptual shape decomposition, first we give the following preliminary definitions.

**Definition 1**. For an object $O$, **shape partition** is to divide $O$ into $k$ subparts $\{P_i, i = 1, \cdots, k\}$,

$$SP(O) = \left\{ P_i | \ \cup_{i=1}^{k} P_i = O, P_i \cap P_j = \Phi \text{ if } i \neq j \right\}.$$

According to psychological studies in human vision [12], [14], the minima rule is the most widely used criterion in perceptual decompositions. Thus we have the following definition:

Manuscript received March 7, 2013.
Manuscript revised May 9, 2013.

[†]The authors are with the Department of Electronics and Information Engineering, Huazhong University of Science and Technology, Wuhan 430074, China.
a) E-mail: wangchun1022@gmail.com
b) E-mail: laizhy@gmail.com

DOI: 10.1587/transinf.E96.D.2147

**Definition 2**. For an object $O$, **negative minimum set** is the set of points on object contour $\Omega(O)$, whose curvature has negative sign and local minimum value in its contour neighborhood:

$$\begin{aligned} V(O) &= \{v_i | v_i \in \Omega(O), Curvature(v_i) < 0 \\ &\quad and\, Curvature(v_i) \\ &= \min(Curvature(Neighborhood(v_i)))\} \end{aligned}$$

The sizes of neighborhoods measure the significances of negative minima, which are used to filter out noise and preserve natural part boundaries. In perceptual decomposition, all negative minima must be decomposed.

According to [13], part cuts across symmetrical axis with short cut length are preferred. Then we define the symmetry function as follows:

**Definition 3**. The **Symmetry function** $Symmetry(\cdot)$ defines the symmetric relation to medial axis $MA(O)$ for a pair of contour points. The image and the preimage of symmetry function are the contour points closest to the same axial point:

$$\begin{aligned} &\forall v_i, v_j \in \Omega(O), v_i = Symmetry(v_j) \Leftrightarrow \exists p \in MA(O), \\ &\text{s.t.} \|v_i - p\| = \|v_j - p\| = \min_{v \in \Omega(O)} \|v - p\|. \end{aligned}$$

The part cuts are extracted based on definition 2 and 3.

**Definition 4**. For an object $O$, the **candidate cut set** is the set of inner line segments, whose endpoints are selected from either negative minima $V(O)$ or symmetry relation $Symmetry(.)$:

$$\begin{aligned} C(O) &= \{c_j | c_j \subset O, Endpoints(c_j) \in (V(O) \times V(O)) \cup \\ &\quad (V(O) \times Symmetry(V(O)))\}. \end{aligned}$$

Definition 4 provides two complementary ways to construct $C$, connecting a pair of negative minima or connecting a negative minimum and its symmetry. Our final cut set is a subset of $C$. Next, we present two general definitions to deal with part cuts and negative minima. Their implementation details are described in Sect. 3.

**Definition 5**. The **perceptual cost** of part cut $c$, denoted by $Cost(c)$, is defined as the degree of inconsistency between $c$ and human decomposition behavior.

A small value of $Cost(c)$ signifies that part cut $c$ is friendly to human vision and being selected with high likelihood. We can use the following criteria to define $Cost(.)$ heuristically: (1) the minima rule [12], (2) the short cut rule [13], and (3) the part salience [14], [15].

**Definition 6**. The negative minimum $v_i$ is **removed** by part cut $c_j$, if and only if $v_i$ is no longer the negative minimum after decomposition by $c_j$:

$$v_i \in Remove(c_j) \Leftrightarrow v_i \in V(O), \forall P_j \in SP(O),\ v_i \notin V(P_j),.$$

If $v_i$ is removed by part cut set $\{c_{j_1}, \cdots, c_{j_N}\}$, we denote $v_i \in Remove(\{c_{j_1}, \cdots, c_{j_N}\})$.

Now we present our definition of perceptual shape decomposition.

**Definition 7**. For an object $O$, the **perceptual shape decomposition** $PSD(O)$ is to select a set of part cuts $C^*$ from $C(O)$ for a shape partition $SP(O)$, with all negative minima *removed* at minimum *perceptual cost*.

Suppose $CS(O)$ is a set of subsets of $C(O)$ that removes all negative minima in $V(O)$:

$$CS(O) = \{cs_i \,|\, cs_i \subseteq C(O), \forall v \in V(O),\ v \in Remove(cs_i)\},$$

then $PSD(O)$ is the element of $CS(O)$ with minimum cost:

$$\begin{aligned} &PSD(O) = \{P_i \,|\, \cup_{i=1}^{k} P_i = O, P_i \cap P_j = \Phi \text{ if } i \neq j, \\ &C^* = \text{argmin}_{cs_i \in CS(O)} \sum\nolimits_{c \in cs_i} Cost(c)\,\}. \end{aligned}$$

## 3. Implementation

In this section, we describe in detail the implementation of our definitions in Sect. 2.

### 3.1 Curvature

As stated in [16], Discrete Contour Evolution (DCE) and its relevance measure reflects polygonal definition of global curvature, thus concave DCE vertices are located near negative minima. We regard the concave DCE vertices set as negative minimum set $V(O)$ in Definition 2 (red points in Fig. 2 (a)).

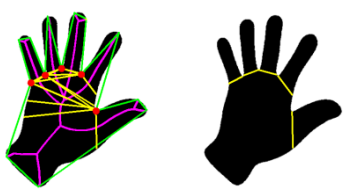
The DCE is originally proposed for 2D digital shape simplification. For two consecutive line segments $s_1$, $s_2$ and their common point $A$, their relevance measure $K$ is given by:

$$K(A) = \frac{\beta(s_1, s_2) l(s_1) l(s_2)}{l(s_1) + l(s_2)} \tag{1}$$

where $\beta(\cdot, \cdot)$ is the turning angle and $l(\cdot)$ is the segment length, respectively [16]. In every evolutional step, a pair of consecutive line segments $s_1$, $s_2$ with smallest relevance measure is replaced by a single line segment joining the endpoints of $s_1 \cup s_2$. We evolve the object contour to a proper stage where all significant vertices are preserved (vertices of green polygon in Fig. 2 (a)). Each vertex is assigned to a relevance measure.

### 3.2 Symmetry

We adopt Bai's skeleton [11] to realize our symmetry func-



**(a)** **(b)**

**Fig. 2** Hand decomposition. (a) DCE vertices (vertices of green polygon), negative minima (red points), Bai's skeleton [11] (pink segments) and candidate part cuts (yellow part lines). (b) Selected part cuts.

tion in Definition 3, due to its nice property and DCE basis. As shown in Fig. 2 (a), $v_i = Symmetry(v_j)$ means that $v_i$ and $v_j$ are tangent points on the same local maximum inscribed circle centered at a pruned skeleton point [11] (pink segments). Then the candidate cut set in Definition 4 is constructed by negative minimum set $V(O)$ and its symmetry set $Symmetry(V(O))$ (yellow part lines).

### 3.3 Perceptual Cost

According to perceptual decomposition behavior [14], we define the **perceptual relevance measure** of part cut $c$ as follows:

$$\varphi(c) = \frac{\max\left(K(A), K(B)\right)}{|K(A) - K(B)| + \min\left(|K(T) - K(A)|,\ |K(T) - K(B)|\right)} \tag{2}$$

where $A, B \in Endpoints(c)$, $T = \text{argmax}_{t \in \widehat{AB}} K(t)$. The numerator and the first term of denominator reflect the minima rule, favoring significant negative minima on both endpoints. The second term of denominator reflects the matching degree between part salience and boundary strength [14]. Combining Eq. (2) with the short cut rule [13], we define the perceptual cost of part cut $c$ as:

$$Cost(c) = \frac{dist(A, B)}{1 + \alpha \cdot \varphi(c)} \tag{3}$$

where $dist(\cdot, \cdot)$ is the cut length, and $\alpha$ is the parameter balancing perceptual relevance measure and cut length. Thus Eq. (3) meets the heuristic criteria in Definition 5.

### 3.4 Removable Relation

After decomposition, we require that $\forall v_i \in V(O)$, and $\forall P_j \in PSD(O)$, $v_i$ has either positive curvature or insignificant negative curvature that can also be quantified by DCE relevance measure in Eq. (1). As shown in Fig. 3 (a), to remove $B_i$ in left part, either $\alpha_1 < 180°$ or $K\left(\overline{A_1B_i}, \overline{B_{i-1}B_i}\right) \le \tau$ must hold, where $\tau$ is set to $\min_i K\left(\overline{B_{i-1}B_i}, \overline{B_iB_{i+1}}\right)$. The case of two part cuts removing one negative minimum is shown in Fig. 2 (b). Similar constraint is required. This procedure implements the function $Remove(.)$ in Definition 6.

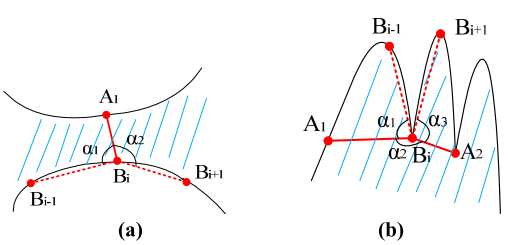



**Fig. 3** Removable relations between negative minima and part cuts, where $\{B_j,\ j = 1, 2\cdots\}$ is the set of DCE vertices. (a) $B_i \in Remove\left(\overline{A_1B_i}\right)$. (b) $B_i \in Remove\left(\left\{\overline{A_1B_i}, \overline{A_2B_i}\right\}\right)$.

### 3.5 Perceptual Shape Decomposition

From Definition 7, based on the negative minimum set $V(O)$, the candidate cut set $C(O)$, the perceptual cost and the removable relation, we formulate PSD as the selection of subset from $C(O)$ that can remove $V(O)$ with minimum cost as follows:

$$\mathbf{min}_{\boldsymbol{x}} \boldsymbol{P}^{\mathrm{T}}\boldsymbol{x},\ \text{s.t.}\ \boldsymbol{A}\boldsymbol{x} \ge \mathbf{1},\ \boldsymbol{x}^{\mathrm{T}}\boldsymbol{E}\boldsymbol{x} = 0,\ \boldsymbol{x} \in \{0, 1\}^n, \tag{4}$$

where $\boldsymbol{x}_{n\times 1} = \{x_i\}$ is the binary vector to indicate whether $c_i$ is selected, $\boldsymbol{P}_{n\times 1} = \{p_j\}$ is the vector to record part cuts' perceptual cost, i.e., $p_j = Cost(c_j)$. $\boldsymbol{A}_{m\times n} = \{a_{ij}\}$ represents the removable relation, where $\sum_{k=1}^{N} a_{ij_k} = 1$ if $v_i \in Remove(\{c_{j_1}, \cdots, c_{j_N}\})$. $E_{n\times n} = \{e_{ij}\}$ is the compatible relation between part cuts. $e_{ij} = 0$ means $c_i$ and $c_j$ can be selected simultaneously, otherwise $e_{ij} = 1$. The objective of Eq. (4) is to minimize the perceptual cost of selected part cuts. The inequality constraint means all negative minima in $V(O)$ must be removed, and the equality constraint indicates that any pair of selected part cuts must be compatible. We implement PSD by first constructing matrices in Eq. (4) and then solving Eq. (4).

Let us denote binary $\boldsymbol{R}_{m\times n} = \{r_{ij}\}$ where $r_{ij} = 1$ if $v_i \in Remove(c_j)$ and binary matrix $\boldsymbol{M}_{m\times n} = \{m_{ij}\}$ where $m_{ij} = 1$ if $v_i \in Endpoint(c_j)$. For negative minimum $v_i$, we only consider two cases: (1) $v_i$ is removed by a single part cut, as shown in Fig. 3 (a), which leads to $\sum_{j=1}^{n} r_{ij} \ge 1$, and (2) $v_i$ is removed by two part cuts, as shown in Fig. 2 (b), which leads to $\sum_{j=1}^{n} m_{ij} \ge 2$. Summarizing these two cases, we have $(\boldsymbol{R} + \boldsymbol{M}/2)\boldsymbol{x} \ge 1$. Therefore, $\boldsymbol{A} = \boldsymbol{R} + \boldsymbol{M}/2$. For the case that two adjacent part cuts cannot remove their common negative minimum, we consider these cuts inefficient and do not select them simultaneously. This case is recorded in $\boldsymbol{E}_{n\times n} = \{e_{ij}\}$. Another source of incompatibility between part cuts is intersection.

Equation (4) is a binary programming problem with a quadratic constraint. The solution can be found efficiently by using standard discrete optimization techniques. To facilitate comparison, we follow Ren's work [9] to use CPLEX. An example of selected part cuts is shown in Fig. 2 (b).

### 3.6 Computational Complexity

There are two main procedures in PSD: (1) computing negative minima, candidate part cuts and their relations, and (2) solving the problem formulated in Eq. (4). In the first procedure, the computational complexity of DCE and skeleton pruning are $O(N \log N)$ and $O(N)$ [11], thus the total complexity of the first procedure is $O(N \log N)$, where $N$ is the number of contour points. In the second procedure, the complexity is $O(2^{n^2})$, where $n$ is the number of negative minima and $n \ll N$. Compared with MNCD [9] that requires $O(N^2)$ time for calculating mutex pair and $O(2^{N^2})$ for solving BILP problem, our method improves the computational efficiency in both procedures.

**Fig. 4** Examples of ten hand gestures.

**Fig. 5** The decomposition of hand shapes from [9]. The three rows show the results from TD [6], [7], MNCD [8], [9] and our method, respectively. The parts in ellipses are either redundant or undetected.

**Table 1** The mean decomposition time on the NTU-Microsoft Kinect Hand Gesture Dataset [17].

| | Original MNCD [8] | Improved MNCD [9] | PSD |
|---|---|---|---|
| Decomposition time | 45.6132s | 3.9705s | 1.875s |

## 4. Experiments

In this section, we substantiate that compared with existing decomposition methods our PSD better balances the robustness, precision and efficiency and thus improves the performance of part-based hand gesture recognition system. We choose the challenging real-life NTU-Microsoft Kinect Hand Gesture dataset [17]. This dataset contains 10 gestures for number 1 to 10, each of which has 100 cases with variations in orientation, scale, articulation, etc. We pre-segment the hand shapes using the method proposed in [6], [7], where some examples of segmented hands are shown in Fig. 4. We make sure that the implementation environments and parameter settings of [6], [7] and ours are the same as far as possible.

Firstly, we compare PSD with existing decomposition methods in terms of robustness and perceptual naturalness. In Fig. 5, three rows show the decomposition results from TD [6], [7], MNCD [8], [9] and PSD, respectively, and the parts being redundant or undetected are marked in ellipses. As we can see, PSD only produces a redundant part in the first gesture, while both TD and MNCD produce incorrect decomposition in multiple hand shapes, demonstrating that PSD is more robust against orientation, scale and articulation changes. Moreover, the end points of part cuts in TD are perceptually irrelevant, which have a negative effect on the subsequent processing, whereas PSD always locates end points in negative minima or their symmetries.

Secondly, we compare PSD with existing decomposition methods in terms of efficiency, as listed in Table 1. As reported in [9], the decomposition time for original version of MNCD [8] is 45.6132 s, which is extremely time-consuming, and the improved version [9] applies CPLEX to solve BILP problem and reduce the decomposition time to 3.9705 s. By using CPLEX, we can further reduce the implementation time to 1.875 s. This comparison verifies the theoretical analysis on complexity in Sect. 3.6.

**Table 2** The confusion matrix of our PSD-based hand gesture recognition system.

| | | Predicted Class | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| Actual Class | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | 2 | .01 | .95 | .03 | .01 | 0 | 0 | 0 | 0 | 0 | 0 |
| | 3 | .02 | 0 | .87 | .04 | .07 | 0 | 0 | 0 | 0 | 0 |
| | 4 | 0 | 0 | 0 | .94 | .05 | 0 | 0 | 0 | .01 | 0 |
| | 5 | .01 | .01 | .01 | .04 | .91 | .02 | 0 | 0 | 0 | 0 |
| | 6 | 0 | 0 | 0 | .01 | .01 | .96 | 0 | .02 | 0 | 0 |
| | 7 | 0 | 0 | 0 | .01 | 0 | 0 | .93 | 0 | .01 | .05 |
| | 8 | 0 | 0 | .08 | .01 | 0 | 0 | 0 | .91 | 0 | 0 |
| | 9 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| | 10 | .01 | 0 | .01 | .02 | .02 | 0 | 0 | 0 | 0 | .94 |

**Table 3** The mean accuracy and the mean recognition time for various hand gesture recognition systems. This table excluding the last row is extracted from [7].

| | Mean Accuracy | Mean Recognition Time |
|---|---|---|
| Shape Context without bending cost [18] | 83.2% | 12.346s |
| Shape Context with bending cost [18] | 79.1% | 26.777s |
| Skeleton Matching [19] | 78.6% | 2.4449s |
| TD+FEMD [7] | 93.2% | 0.075s |
| MNCD+FEMD [7] | 93.9% | 4.0012s |
| PSD+FEMD | 94.1% | 1.967s |

Thirdly, we confirm that the PSD embedded part-based hand gesture recognition system achieves performance improvements. The confusion matrix of our system is shown in Table 2. In theory, we should achieve a higher performance than TD-based FEMD (93.2%) [7], because our decomposition is constructed on the basis of human perceptual behavior. We should also have a lower recognition time than MNCD-based FEMD, due to that we have much smaller number of candidate part cuts and size of searching space of optimization. The mean accuracy and mean recognition time of various shape recognition systems is shown in Table 3. As expected, the PSD embedded recognition system achieves the highest mean accuracy (94.1%) at the second lowest mean recognition time cost (1.967 s), demonstrating a better tradeoff between accuracy and efficiency.

## 5. Conclusion

We present a novel shape decomposition method for a part-based hand gesture recognition system. We formulate the shape decomposition problem as an optimization problem, which efficiently remove all the negative minima with minimum cost. Theoretical analysis and experimental results show that our PSD is robust to shape variations, friendly to perception and more efficient than MNCD [8], [9]. Therefore, the PSD embedded hand gesture recognition system with a Kinect camera [5] outperforms the state-of-the-art in accuracy at moderate computational cost without complicated device, resource-consuming procedure or training.

## References

[1] G.R.S. Murthy and R.S. Jadon, "A review of vision based hand gesture recognition," Int. J. Information Technology and Knowledge Management, vol.2, no.2, pp.405–410, July–Dec. 2009.

[2] J.P. Wachs, M. Kolsch, H. Stern, and Y. Edan, "Vision-based hand-gesture applications," Commun. ACM, vol.54, no.2, 60–71, Feb. 2011.

[3] A. Erol, G. Bebis, M. Nicolescu, R.D. Boyle, and X. Twombly, "Vision-based hand pose estimation: A review. Computer vision and image understanding," vol.108, no.1-2, pp.52–73, Oct.–Nov. 2007.

[4] Microsoft Corporation, "Kinect for XBOX 360," Redmond WA, 2010.

[5] Z. Ren, J. Meng, J. Yuan, and Z. Zhang, "Robust hand gesture recognition with Kinect sensor," Proc. ACM International Conference on Multimedia, pp.759–760, Scottsdale, Arizona, USA, Nov.–Dec. 2011.

[6] Z. Ren, J. Yuan, and Z. Zhang, "Robust hand gesture recognition based on finger-earth mover's distance with a commodity depth camera," Proc. ACM International Conference on Multimedia, pp.1093–1096, Scottsdale, Arizona, USA, Nov.–Dec. 2011.

[7] Z. Ren, J. Yuan, J. Meng, and Z. Zhang, "Robust part-based hand gesture recognition based on finger-earth mover's distance," IEEE Trans. Multimed., vol.15, no.5, pp.1110–1120, 2013.

[8] Z. Ren, J. Yuan, C. Li, and W. Liu, "Minimum near-convex decomposition for robust shape representation," Proc. IEEE International Conference on Computer Vision, pp.303–310, Barcelona, Italy, Nov. 2011.

[9] Z. Ren, J. Yuan, and W. Liu, "Minimum near-convex shape decomposition," IEEE Trans. Pattern Anal. Mach. Intell, 2013.

[10] L.J. Latecki and R. Lakämper, "Polygon evolution by vertex deletion," in Scale-Space Theories in Computer Vision, Proc. International Conference on Scale-Space, ed. M. Nielsen, P. Johansen, O.F. Olsen, and J. Weickert, vol.LNCS 1682, pp.398–409, Corfu, Greece, Sept. 1999.

[11] X. Bai, L.J. Latecki, and W. Liu, "Skeleton pruning by contour partitioning with discrete curve evolution," IEEE Trans. Pattern Anal. Mach. Intell., vol.29, no.3, pp.449–462, March 2007.

[12] D.D. Hoffman and W.A. Richards, "Parts of recognition," Cognition, vol.18, no.1-3, pp.65–96, Dec. 1984.

[13] M. Singh, G. Seyranian, and D.D. Hoffman, "Parsing silhouettes: The short-cut rule," Perception & Psychophysics, vol.61, no.4, pp.636–660, Jan. 1999.

[14] D.D. Hoffman and M. Singh, "Salience of visual parts," Cognition, vol.63, no.1, pp.29–78, April 1997.

[15] Z. Lai, W. Liu, F. Zhang, and G. Cheng, "Perceptual distortion measure for polygon-based shape coding," IEICE Trans. Inf. & Syst., vol.E96-D, no.3, pp.750–753, March 2013.

[16] L.J. Latecki and R. Lakämper, "Convexity rule for shape decomposition based on discrete contour evolution," Comput. Vis. Image Understand., vol.73, no.3, 441–454, March 1999.

[17] http://www.ntu.edu.sg/home/renzhou/HandGesture.htm

[18] S. Belongie, J. Malik, and J. Puzicha, "Shape matching and object recognition using shape contexts," IEEE Trans. Pattern Anal. Mach. Intell., vol.24, no.4, pp.509–522, April 2002.

[19] X. Bai and L.J. Latecki, "Path similarity skeleton graph matching," IEEE Trans. Pattern Anal. Mach. Intell., vol.30, no.7, pp.1–11, July 2008.