



# 중국식 숫자 손 제스처 인식을 위해 개선된 다단 형판 정합

임 종 관\*

## Chinese Hand Number Gesture Recognition by Enhanced Multi-stage Template Matching

Jong Gwan Lim\*

요 약

한 손으로 10가지 제스처를 표현하는 중국식 한 손 숫자 제스처는 중국에서 자주 사용되는 의사소통 수단이다. 한 손만 사용하기 때문에 5, 6가지 제스처 인식을 위해 제안된 기존 방법들은 제약이 있어 문헌상 보고된 인식률도 91.15%로 저조하다. 이를 위해 제시된 다단 형판 정합 방법에서 제기된 문제점을 개선하기 위해, 손 영역 검출과 손바닥 영역 검출을 수정하고 Major/Minor 그룹 분류 기준에 손바닥 넓이를 추가한다. 끝으로 Minor 그룹 내에서 클래스 7과 10을 분류하기 위해 HoG 정합을 대안으로 제시한다. 크기, 방향, 조명이 다양한 2000개 손 제스처 영상에 대해 성능 검증을 실시한 결과 제안하는 방법은 최종적으로 93.65%의 인식률을 기록한다.

Abstract

Chinese number gestures are a common means of communication using one hand in China. Conventional methods proposed for recognizing 5 or 6 gestures are limited due to the representation of 10 gestures with one hand, and the recognition rate reported in the literature is also low at 91.15%. In this paper, we improve multi-stage template matching by modifying hand and palm segmentation, proposing hand area as another Major/Minor group categorization criterion and using the HoG(Histogram of Oriented Gradients) for the classification in the Minor group. As a result of performance verification with 2000 hand gesture images with various scales, orientations, and lights, the proposed method finally records 93.65% recognition rate varying cell size and orientations.

Keywords

HoG, chinese number gesture, gesture recognition, image processing

\* 목원대학교 지능로봇공학과 교수  
- ORCID: <https://orcid.org/0000-0003-1223-0279>

• Received: Jul. 23, 2019, Revised: Aug. 21, 2019, Accepted: Aug. 24, 2019  
• Corresponding Author: Jong Gwan Lee  
Dept. of Intelligent Robot Engineering, Mokwon University,  
88, Doanbuk-ro, Seo-gu, Daejeon, Korea,  
Tel.: +82-42-829-7520, Email: [jonggwanlim@gmail.com](mailto:jonggwanlim@gmail.com)

### 1. 서 론

손 제스처의 유용한 실례로서 중국식 손가락 숫자는 독특하다. 4와 10의 유사한 발음과 지역별 방언 때문에 중국인은 의사 소통시 한 손 숫자 표현을 사용한다. 그림 1에서 보듯 1-10까지 한 손으로 표현한다.

기존 손가락 제스처 인식 연구는 1-5까지 혹은 10을 포함한 6개 동작의 인식을 시도하므로 주로 펼친 손가락의 개수를 센다. 손의 내접원과 외접원을 계산 후, 두 원의 반지름 사이 값을 반지름으로 삼는 원 둘레와 펼친 손가락 간 교차점을 확인, 손가락 개수를 센다[1][2]. 또는 손끝을 검출하여 손가락 개수를 추정하는데 손의 윤곽에 convex hull 알고리즘을 적용하거나[3], 손끝 부분을 형판(Template)으로 저장해 형판 정합을 시도하거나[4], 손 윤곽의 세 연속점으로 각도를 계산한다[5]. 혹은 손바닥 중심에서 손 윤곽 간 거리를 2차 좌표공간으로 변환시키는 거리 그래프를 이용해 펼쳐진 손가락 개수를 확인한다[6].

펼친 손가락의 개수 검출을 통한 중국식 숫자 제스처 인식은 한계가 명확하다. 우선, 대부분의 연구는 중국식 숫자를 대상으로 성능을 제시하지 않는다. 둘째, 펼친 손가락의 개수 파악으로는 동일한 손가락 개수를 갖는 중국식 숫자 1과 9, 2, 6, 8을 판별해 낼 수 없다. 또한 손가락을 모두 접은 7과 10의 구분도 불가능하다. 문헌상 확인되는 중국식 숫자 인식 연구 역시 2-3개 클래스에 대한 부분적인 성과이거나[7][8], 10개 클래스에 대해 낮은 인식률을 보인다[3]. 중국식 숫자 제스처 인식 연구 중 측면 촬영 영상과 RGBD 카메라의 깊이 정보를 활용하는 연구는 논외로 한다.

### II. 연구 동기

우리는 중국식 손 제스처에 대한 다단 형판 정합(Multi-stage Template Matching)을 제안했다(그림 1) [9]. 일반적인 형판 정합을 개선하기 위해, 각 제스처가 갖는 내/외접원 반지름의 비율에 기반하여 손가락을 편 그룹과 손가락을 접은 그룹으로 우선 분류하고 다시 각 그룹 내에서 형판 정합을 실시하여

91.15%의 인식률을 기록했다.

8.85%의 오류를 분석한 결과, 다음과 같은 원인들이 드러났다. 우선 Major/Minor 그룹 오분류가 1.1%인데 오분류의 원인이 부실한 이진화를 통한 손영역 검출에 있다. 이진화에 실패시 그림 2에서 보듯 내/외접원의 정확한 비율 계산에 실패해 오분류된다. 둘째로, 중국식 숫자 제스처 인식의 독특한 문제인 Minor 그룹에서 클래스 7과 10의 인식 오류가 2.7%에 달한다.

이를 개선하기 위해 다음과 같은 개선안을 제안한다. 첫째, 기존의 영상 처리 방식을 대대적으로 개선한다. 손 영역 검출에 사용하던 최소 교차 엔트로피 전역 이진화를 Watershed 방법으로 대체하고, 손 최장너비 기반 손바닥 검출[10]을 타원 추정 손바닥 영역 검출[11]로 대체한다.

둘째, Major/Minor 그룹 분류 특징으로 기존 내/외접원 지름비율에 손 면적을 추가해 오분류를 방지한다.

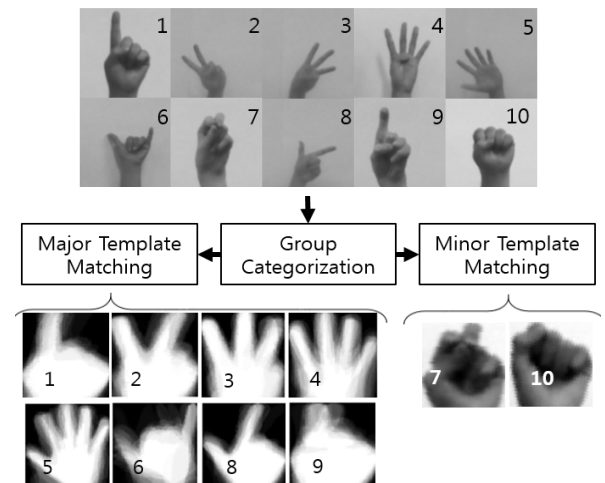


그림 1. 중국 숫자 손 제스처와 다단 형판 정합  
Fig. 1. Chinese number gestures and multi-stage template matching



그림 2. 부실 이진화  
Fig. 2. Bad binarization

셋째, 클래스 7과 10의 구분을 위해 우리는 피사체의 형태나 위치의 소소한 변화에 둔감하되, 영상의 기하학적 정보를 최대한 보장할 수 있는 방법으로서 HoG(Histogram of Oriented Gradients) 기술자(Descriptor)로 손을 표현 하고 이들의 막대그래프를 정합한다[12][14]. 마지막으로 제한된 양의 데이터를 사용하므로 4겹 교차검증을 실시해 엄격한 성능검증을 실시한다.

### III. 개선된 다단 형판 정합

#### 3.1 영상 처리 개선

그림 3은 다단 형판 정합을 위한 처리 과정을 설명한다. 형판 정합은 다량의 영상에 대해 특징점 추출 없이 피사체의 형태를 활용하는 경우 효과적이며 영상의 해상도 및 샘플링 포인트, 피사체의 크기 비율과 회전 정도를 동일화시켜야 한다. 기존 방법의 자세한 설명은 [9]를 참고하기 바라며 이곳에서는 개선된 사항만을 다룬다.

• 손 영역 검출(Hand Segmentation, HS)

손 영역 검출을 위해 기존 연구에서 사용하던 최소 교차 엔트로피 문턱값을 이용한 전역 이진화 대신 영상 분할(Segmentation)에서 흔히 사용되는 Watershed 방법으로 전역 이진화를 실시한다.

- ① th(threshold, 임계값) 설정, pixelNo 설정
- ② 수평/수직방향 sobel filtering으로 경사도 크기(Gradients magnitude) 계산
- ③ th 값보다 작은 영역을 표식(Marker)으로 선정, th 값보다 큰 윤곽선 검출, 윤곽선 내 화소수를 pixelNo에 저장
- ④ 윤곽선 범위 내, 표식을 종자(Seed)로 삼아 Watershed 적용
- ⑤ 만약 pixelNo에 변화가 없으면 탈출, 변화가 있으면  $th = th+1$ 하고 ③으로 회귀
- ⑥ 연결성분 찾기(Connected component labelling)를 통해 이진화 오류를 보정

• 손바닥 영역 검출(Palm Segmentation, PS)

기존 연구에서 우리는 이진화 오류에 대비해 손 최장너비 기반 손바닥 검출 방법을 사용했는데 이때 손바닥 영역을 원으로 추정한다. 그러나 다양한 손 특성을 분석해보면 손바닥 형태는 원보다는 타원에 더 가깝다[11]. 이진화 손 영역의 중앙점을 기준으로 30도 각도 간격마다 간격 내에 놓인 윤곽선과 중앙점 간 최소점(그림 4의 붉은 점)을 찾는다.

$$ax^2 + 2bxy + cy^2 + 2dx + 2fy + g = 0 \quad (1)$$

이 점들의 위치로 최소 제곱법 근사를 시도하여 수식 (1)에서 주어지는 타원 일반식의  $a, b, c, d, e, f, g$ 를 추정한다[13]. 마지막으로 추정한 타원의 장축 길이에 근거하여 손목을 제거한다. 그림 4는 타원 장축 길이의 60% 위치를 분홍색 점으로 표시하는데 인식률 최적 위치를 찾기 위해 탐욕탐색(Greedy search)한다.

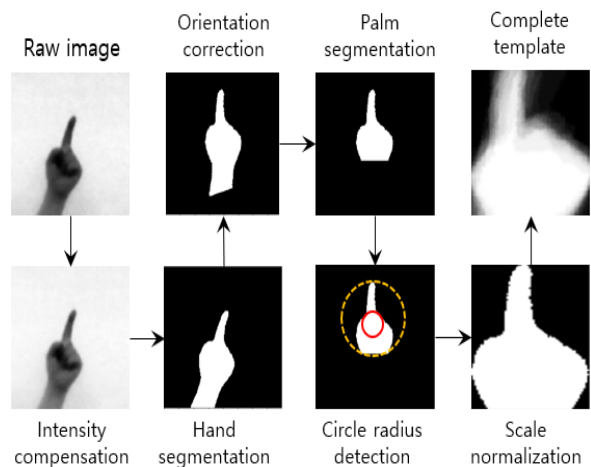


그림 3. 형판 제작을 위한 처리 과정  
Fig. 3. Processing for template implementation

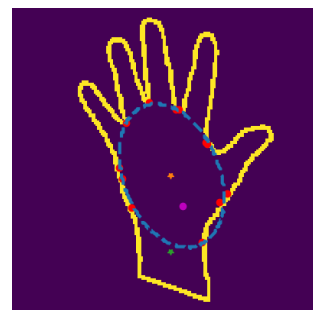
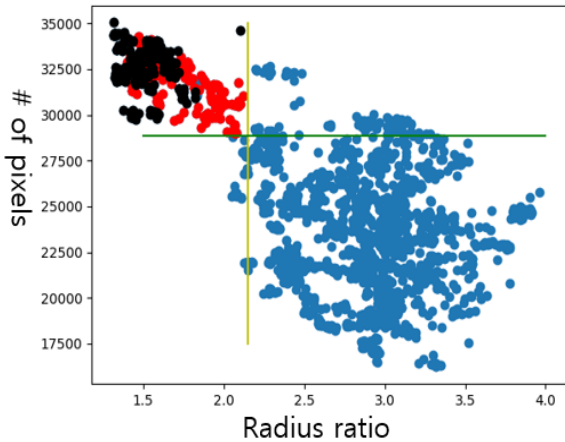


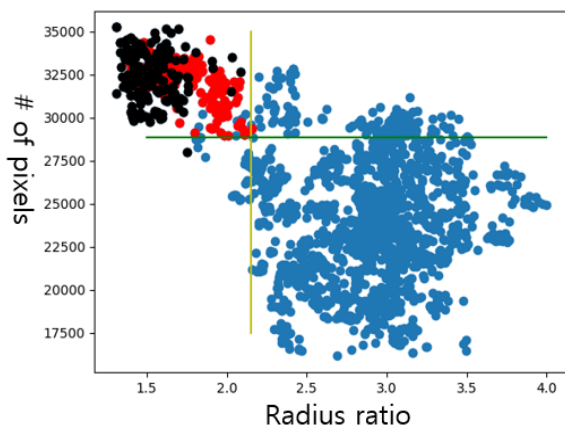
그림 4. 타원 추정 기반 손목 제거  
Fig. 4. Palm area detection by ellipse estimation

3.2 Major/Minor 그룹분류를 위한 손 면적 추가

기존 연구에서 우리는 손가락을 모두 편 Major 그룹과 손가락을 모두 접은 Minor 그룹을 분류하기 위해 손 윤곽선에 그린 내/외접원의 지름 비율을 사용하였다. 그러나 이 비율만의 단독 사용으로는 완벽한 분류가 어려워 우리는 검출된 손바닥 영역의 넓이를 추가적인 특징으로 제안한다. 그림 3에서 보듯 손의 방향과 크기를 보정 후 손바닥에 해당하는 화소의 수를 세어 넓이를 추정한다. 이때, 손의 크기가 정규화된 상황에서 손가락을 펼친 제스처는 손가락을 접은 제스처에 비해 손바닥 영역의 화소 수가 더 작다.



(a) 개선된 손 영역 검출  
(a) Improved hand segmentation



(b) 기존 이진화 적용  
(b) Conventional hand segmentation

그림 5. 원 비율과 손 면적으로 표현한 클래스 분포  
Fig. 5. Class sample distribution represented by radius ratio and no. of pixels

이 결과를 그림 5가 보여주는데, 파란 점은 Major 그룹에 해당하는 클래스 샘플이고 검은 점은 클래스 9, 붉은 점은 클래스 7에 해당하는 샘플이다. 그림 5(b)는 영상처리를 기존 방법으로 처리한 경우이며 5(a)가 개선된 방법을 적용한 결과이다. 5(b)에 비해 5(a)에서 양 그룹이 깨끗이 분리되는 것을 확인할 수 있다.

3.3 Minor 그룹 내 HoG 정합

기존에 우리는 Minor 그룹 내 정합을 위해 방향과 크기가 정규화된 그레이 영상간 형상 정합을 시도하였다. 본 논문에서는 이를 개선해 그레이 영상을 HoG 기술자로 표현 후 이를 정합하는 방법을 제안한다. 형판 정합은 원래 영상의 기하학적 정보를 그대로 유지해야 정합이 되는 반면, 막대그래프 정합은 형태나 위치 변화에는 둔감하나 대상의 기하학적 정보를 버리고 단지 분포 정보만을 기억한다. HoG는 형판 정합과 막대그래프 정합의 중간 단계에 있으며 블록 단위로는 기하학적 정보를 유지하되, 각 블록 내부의 분포 정보를 사용함으로써 지역적인 변화에는 강인한 특성을 갖는다.

구체적인 HoG 기술자의 구현과정을 설명한다.

- ① 수평/ 수직방향 Sobel 필터를 적용하여 구한 화소별 경사도(Gradients),  $G_x$ ,  $G_y$ 를 계산
- ② 식 (2), (3)를 이용, 화소별 크기와 방향 계산[12]

$$\text{magnitude} = \sqrt{G_x^2 + G_y^2} \tag{2}$$

$$\text{orientation} = \tan^{-1} \left( \frac{G_y}{G_x} \right) \tag{3}$$

- ③ 각 영상에 대해, 영상의 크기를 고려해 적당한 크기의 셀로 분할
- ④ 경사도 방향의 각 셀에서 방향별 화소를 분류
- ⑤ 경사도 크기의 각 셀에서 위 화소에 해당하는 크기를 합산하여 막대그래프 생성
- ⑥ 막대 그래프 계수값을 셀의 기하학적 위치에 기준하여 연결시켜 하나의 벡터 생성

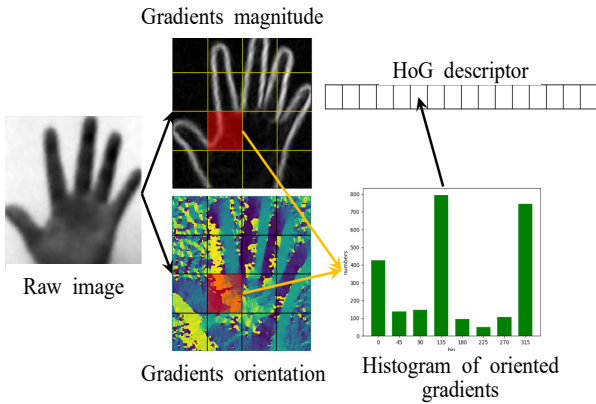


그림 6. HoG 기술자

Fig. 6. Histogram of oriented gradients descriptor

① ~ ⑥ 과정을 그림 6에서 설명하는데 그림 6은 50×50 크기로 나눈 셀에 대해 8가지 방향(45°간격)으로 구한 막대그래프 예시이다. 셀의 개수가 총 16개고 셀당 8개의 막대 그래프 계수가 나오므로 최종 벡터의 크기는 16×8= 128이다.

#### IV. 실험 및 결과

##### 4.1 실험 설계

실험 영상은 10개의 클래스에 대해 피험자 17명으로부터 획득한 총 2000장의 중국 숫자 손 제스처이다. 그림 7에서 보듯, 200×200의 그레이 영상이며 손의 크기와 방향, 조명, 손목의 노출 정도도 제각각이다. 모든 실험은 적용되는 알고리즘만 달리할 뿐 동일한 데이터에 근거해 진행된다.

우선 기존 방법과 새롭게 개선된 방법들을 추가할 때마다 달라지는 성능을 비교한다. 영상처리 방법을 단계별로 적용하고 다음에 손 영역 면적을 이용해 양 그룹을 분리한다. 마지막으로 Minor 그룹 내에 HoG 정합을 적용한다. 이때, 우리는 셀의 크기, 방향의 개수를 다양화한다. 셀의 크기는 2×2, 4×4, 5×5, 10×10, 20×20, 50×50, 100×100, 200×200로, 방향의 개수는 4방향, 8방향, 16방향으로 구분한다.

제안하는 방법에서 조명은 보상되기 때문에 조명 정규화를 위해 사용하는 블록 크기는 1×1로 고정한다. 최종적으로 기존에 발표된 중국식 손 숫자 제스처 인식 방법론과 성능을 비교한다. 모든 성능 비교는 4겹 교차검증을 통해 진행된다.



그림 7. 실험에 사용된 영상 예시

Fig. 7. Gesture images

##### 4.2 실험 결과 및 분석

제안한 개선안의 최종적인 성능은 93.65%를 기록했다. 우리가 제안하는 방법의 성능을 검증하기 위해 기존 다단 형판 정합에 사용된 방법과 비교한 결과를 표 1에서 제시한다. 다단 형판 정합의 처리 과정을 HS, PS, Major/Minor 그룹 분류를 위해 추가된 손면적(Area), Minor 그룹내 HoG 정합(Hog)로 분류 후, 3장에서 제안된 알고리즘이 적용된 경우를 ‘O’로, 기존 [9]에 적용된 방법을 사용한 경우를 ‘X’로 표시한다. 가령, 표 1에서 실험(Ex) 1은 모든 과정이 ‘X’이므로 기존 다단 형판 정합과 동일하다. 실험 9는 기존 다단 형판 정합 알고리즘에서 Major/Minor 그룹 분리 방법과 Minor 그룹내 형판 정합 알고리즘을 우리가 제안한 방법으로 대체한 경우이다.

표 1. 개선 알고리즘 성능 비교

Table 1. Performance comparison of improved algorithms

Ex	HS	PS	Area	HoG	Recog. rate(%)
1	X	X	X	X	91.15
2	O	X	X	X	89.35
3	X	O	X	X	90.15
4	O	O	X	X	92.50
5	X	O	O	X	90.35
6	X	X	O	X	91.40
7	O	O	O	X	92.55
8	O	O	O	O	93.65
9	X	X	O	O	92.80

우선, 우리가 제안한 4가지 개선에서 가장 효과적인 것은 Minor 그룹 내 HoG정합으로 보인다. 실험 6과 9, 실험 7과 8을 비교해보면 1.1%, 1.4%의 성능향상이 나타난다. 반면에, 개선된 손 영역 검출과 손바닥 검출을 단독으로 사용한 경우, 기존 방법에 비해 도리어 성능이 하락하는 것을 실험 1, 2, 3

의 비교를 통해 확인할 수 있다. 또한 그룹 분류를 위해 면적을 추가하는 경우는 실험 1과 6, 실험 4와 7의 비교를 통해 0.25%, 0.05%의 인식률 향상을 기록한다. 둘째, 손 영역 검출과 손바닥 영역 검출은 상호간에 밀접한 관련을 갖는다. 제안한 개선안이 단독으로 쓰일 경우 성능하락을 보이거나 함께 사용할 경우 기존 방법보다 1.35%의 향상을 보인다 (실험 1과 4). 또한 그룹 분류를 위한 면적 추가 역시 그림 5에서 보듯 손 영역 검출의 성능에 종속적이다.

따라서 이진화 성능의 검증은 이들 모두가 사용된 실험 1과 7의 비교를 통해 가능하며 1.45% 향상된 성능으로 볼 때, 적용한 이진화 방법의 상대적인 우수성을 추정할 수 있다. Minor 그룹 내 HoG정합에서 다뤄지는 셀의 크기는 4×4에서 20×20에서 가장 높은 성능 향상을 보이고 있다. 기대와는 달리 방향의 개수는 4나 8이 16방향보다 더 높은 성능을 기록한다.

표 2. 클래스별 인식율

Table 2. Recognition rates in each class

Class	ConvexHull+SVM[3](%)	Multi-state template matching[9](%)	Proposed(%)
C1	64.00	93.00	90.00
C2	96.00	90.00	89.00
C3	94.00	100.00	100.00
C4	85.00	98.50	94.00
C5	99.00	100.00	99.50
C6	84.50	83.00	85.50
C7	50.50	81.00	94.50
C8	95.00	99.50	100.00
C9	59.00	74.50	89.00
C10	72.50	92.00	95.00
Total	79.95	91.15	93.65

제안한 개선안을 통해 가장 향상되는 클래스(Class)는 7과 9로 드러났다. 표 2에서 기존 방법론과 진행한 클래스별 성능 비교를 보여주는데, 클래스 7은 13.50%, 클래스 9는 14.50%의 성능 향상을 보인다. 중국식 손 숫자 제스처에서 흔히 발생하는 오류를 표 3에서 정리하는데, 클래스 7과 10의 혼동이 Minor 그룹 내 HoG 정합에 의해 극복되어 클래스 7의 인식률 향상을 야기하고, 그룹 분류를 위해 추가된 면적으로 인해 클래스 9가 Minor 그룹으

로 오분류되는 경우가 줄어들었으므로 인식률의 향상이 나타나는 것으로 분석된다. 클래스 6과 9의 인식률 향상에도 불구하고 Major 그룹 내 오류는 1.40% 증가되었는데 이는 개선된 이진화의 결과물이 손가락 폭을 더 좁게 검출함으로써 형판 정합의 단순한 화소 합산이 클래스 2, 3, 4, 5 간 오류를 증가시키기 때문이다. 오류를 합산시 2.50%의 향상이나 본 논문이 제안하는 것은 손 영역 검출, 손바닥 영역 검출, Major/Minor 그룹 분류를 위한 면적 추가, Minor 그룹내 HoG 정합을 통한 각 해당 오류의 개선이며, 표 3을 통해 이를 확인할 수 있다.

표 3. 오류 분석

Table 3. Error analysis

Error	Multi-state template matching[9](%)	Proposed (%)
Group categorization fails	1.95	0.05
Errors in class 6	1.70	1.40
Errors in class 9	2.55	1.10
Errors in minor group	2.65	0.50
Errors in major group excluding class 6 and class 9	1.95	3.35
Total	8.85	6.35

## V. 결 론

한 손으로 10개의 수를 표현하는 중국식 숫자 제스처 인식은 제스처 간 형태적 유사성으로 인해 5, 6개 클래스를 인식하는 기존 방법으로는 처리가 힘들다. 본 논문은 이를 위해 제안된 다단 형판 정합의 성능을 개선하기 위해 손 영역 검출과 손바닥 영역 검출을 개선하고 Major/Minor 그룹 분류의 정확도를 향상시키기 위해 기존에 사용하던 내/외접 원 지름 비율에 손바닥 넓이를 추가한다. 끝으로 Minor 그룹 내에서 자연적 유사도로 구분이 난해한 클래스 7과 10을 분류하기 위해 HoG 정합을 대안으로 제시한다. 피험자 17명으로부터 획득한 10개의 클래스, 2000장의 영상으로 검증한 성능은 평균 인식률은 93.65%이며, 이는 기존에 보고된 인식률 91.15% 보다 2.50% 향상된 기록이다.

\* 영상데이터를 제공한 충북대 유비쿼터스 게임 연구실 이상무씨에게 감사드립니다.

## References

- [1] Jeong-Hoon Shin, "Finger Counting Using Computer Vision", Conference of the Korean Institute of Communication Sciences, pp. 657-658, Jan. 2013.
- [2] Dong-Hyun Hwang and Kyung-Sik Jang, "Finger-Gesture Recognition Using Concentric-Circle Tracing Algorithm", Korea Institute of Information and Communication Engineering, Vol. 19, No. 12, pp. 2956-2962, Dec. 2015.
- [3] Sang-Mu Lee, Mi-Hye Kim, and Jong Gwan Lim, "Finger Counting Recognition Using Hand Skeleton Model", HCI, pp. 873-875, Feb. 2014.
- [4] Jaewan Park, Dae Hyen Song, and ChilWoo Lee, "Real-time Finger Gesture Recognition", HCI, pp. 847-850, Feb. 2008.
- [5] Do-Yeob Lee, Dong-Kyoo Shin, and Dong-Il Shin, "A Finger Counting Method for Gesture Recognition", Journal of Korean Society for internet Information, Vol. 17, No. 2, pp. 29-37, Apr. 2016.
- [6] Ji-Woo Song and Jeong-Su Oh, "Finger Detection using a Distance Graph", Journal of Korea Institute of Information and Communication Engineering, Vol. 20, No. 10, pp. 1967-1972, Oct. 2016.
- [7] Ping Lee, Hee-seong Lee, and Mi-hye Kim, "Classification Method of Chinese a Hand Numeric Representation to No.1 and No.9 Using Length Information", Korea Entertainment Industry Association, Vol. 6, No. 4, pp. 129-135, Dec. 2012.
- [8] Ping Lee, Mi-hye Kim, and Hee-seong Lee, "Recognition Method of Chinese Finger Number 2, 6, 8 Using Angle Information", Journal of Korea Game Society, Vol. 12, No. 6, pp. 121-130, Dec. 2012.
- [9] Dong Kyu Kim, Yong Hyeon Kim, and Jong Gwan Lim, "Multi-stage Template Matching for Hand Gesture Recognition", Journal of KIIT, Vol. 16, No. 5, pp. 15-21, May 2018.
- [10] Eun Chang Choi, Jun Yeon Kim, Jae Won Lee, and Jong Gwan Lim, "Palm Area Detection by Maximum Hand Width", Journal of KCA., Vol. 18, No. 4, pp. 398-405, Apr. 2018.
- [11] Giulio Marin, Marco Fraccaro, Mauro Donadeo, Fabio Dominio, and Pietro Zanuttigh, "Palm Area Detection for Reliable Hand Gesture Recognition", Proceedings of International Workshop on Multimedia Signal Processing, Sardinia Italy, pp. 120-121, Oct. 2013.
- [12] William T. Freeman and Michal Roth, "Orientation Histograms for Hand Gesture Recognition", International Workshop on Automatic Face and Gesture Recognition. Vol. 12, 1995.
- [13] Radim Halir and Jan Flusser, "Numerically stable direct least squares fitting of ellipses", In Proc. 6th International Conference in Central Europe on Computer Graphics and Visualization. WSCG, Vol. 98, pp. 125-132, 1998.
- [14] McConnell, Robert K. "Method of and apparatus for pattern recognition", U.S. Patent 4,567,610, issued January 28, 1986.

## 저자소개

임 종 관 (Jong Gwan Lim)



2016년 2월 : KAIST  
기계공학(공학박사)  
2017년 3월 ~ 현재 : 목원대학교  
지능로봇공학과 교수  
관심분야 : HRI, 영상/신호처리,  
머신러닝