

# RGB-D 이미지에서 인체 영역 검출을 위한 프레임워크

홍성진<sup>†</sup>, 김명규<sup>\*\*</sup>

## A Framework for Human Body Parts Detection in RGB-D Image

Sungjin Hong<sup>†</sup>, Myounggyu Kim<sup>\*\*</sup>

### ABSTRACT

This paper propose a framework for human body parts in RGB-D image. We conduct tasks of obtaining person area, finding candidate areas and local detection in order to detect hand, foot and head which have features of long accumulative geodesic distance. A person area is obtained with background subtraction and noise removal by using depth image which is robust to illumination change. Finding candidate areas performs construction of graph model which allows us to measure accumulative geodesic distance for the candidates. Instead of raw depth map, our approach constructs graph model with segmented regions by quadtree structure to improve searching time for the candidates. Local detection uses HOG based SVM for each parts, and head is detected for the first time. To minimize false detections for hand and foot parts, the candidates are classified with upper or lower body using the head position and properties of geodesic distance. Then, detect hand and foot with the local detectors. We evaluate our algorithm with datasets collected Kinect v2 sensor, and our approach shows good performance for head, hand and foot detection.

**Key words:** Body Parts Detection, Quadtree Structure, Graph Model, Geodesic Distance

### 1. 서 론

영상기반 사람의 신체 영역 검출은 객체 추적, 포즈 획득, 제스처 인식 그리고 HCI 어플리케이션에 활용될 수 있는 기술로 오래 전부터 컴퓨터 비전과 패턴 인식 분야에서 연구되어왔다. 초기 2D영상에서 관심 영역의 특징 정보를 추출하고, AdaBoost, SVM, GMM와 같은 학습 모델을 구성해 얼굴, 눈, 손, 발과

같은 다양한 신체 영역들의 검출하는 방법들이 연구되었다. 최근에는 Kinect나 RealSense와 같은 RGB-D 카메라의 발달로 깊이와 색상 정보를 활용한 방법들이 제안되고 있다.

깊이 데이터를 활용한 인체영역 검출 방법으로 Shotton [1]은 픽셀 단위의 접근법을 제안했다. 각 픽셀에 대한 레이블링을 수행하고, 군집화된 정보를 기반으로 인체 영역 검출을 실시했다. 이 방법은 학

\* Corresponding Author: Sungjin Hong, Address: (34129) Electronics and Telecommunications Research Institute 218 Gajeong-ro, Yuseong-gu, Daejeon, Korea, TEL : +82-42-860-1396, FAX : +82-42-860-1051, E-mail : sjhong0117@etri.re.kr

Receipt date : Oct. 4, 2016, Revision date : Nov. 14, 2016  
Approval date : Nov. 23, 2016

<sup>†</sup> Sportainment Research Section, Creative Contents Research Division, SW·Contents Research Laboratory, Electronics and Telecommunications Research Institute

<sup>\*\*</sup> Sportainment Research Section, Creative Contents Research Division, SW·Contents Research Laboratory, Electronics and Telecommunications Research Institute (E-mail : mgkim@etri.re.kr)

\* This research project was supported by the Sports Promotion Fund of Seoul Olympic Sports Promotion Foundation from Ministry of Culture, Sports and Tourism (s072016122016)

습 모델을 생성하면 실시간으로 인체 영역을 검출할 수 있는 장점을 갖는다. 하지만 학습 모델을 얻기 위해 많은 양의 데이터가 필요하고, 카메라의 시점이 변경되거나 학습되지 않은 자세가 나타났을 때 급격한 성능 저하 문제가 발생한다.

또 다른 접근법중 하나는 Plagemann [2]가 제안한 누적 측지거리 활용한 방법이다. 인체의 중심으로부터 손, 발, 머리 부분의 누적 측지거리가 크게 나타나는 특성을 활용한다. 인체 영역을 그래프 모델로 구성하고, 다익스트라 알고리즘을 이용한 탐색으로 누적 측지거리가 가장 크게 나타나는 극점의 위치들을 획득한다. 해당 극점의 영역은 사전에 추출한 지역 형태 특징 매칭을 통해 손, 발, 머리인지 결정된다. 하지만 잘못된 그래프 구성은 극점 탐색 성능 저하로 부정확한 신체 영역을 검출 문제를 야기한다. Schwarz [3]는 문제를 해결하기 위해 배경이 제거된 깊이 데이터를 사용해 그래프를 구성하고, 신체 영역 검출 실패 시 유틸리티 플로우 정보를 기반으로 그래프를 수정해 정확도를 향상시켰다. 하지만 두 방법 모두 그래프 탐색의 속도가 느린 문제점을 갖고 있다. Baak [4]는 이 방법을 해결하기 위해 최적의 다익스트라 알고리즘 탐색 기법을 제안했다. 첫 번째 극점 탐색 시 획득한 누적측지 거리정보를 저장하고 두 번째 극점 탐색부터는 이전 누적측지 거리 정보를 활용해 거리 정보가 변화되는 노드에 대해서만 다익스트라 알고리즘 탐색을 실시한다. 이로써 두 번째 극점부터의 탐색 시간을 비약적으로 향상시켰다. 하지만 첫 번째 노드의 탐색 시간은 줄일 수 없었으며 그래프의 사이즈가 커질수록 속도가 저하되는 문제가 있어 이를 해결할 필요성이 있다.

본 논문에서는 RGB-D 이미지에서 손, 발, 머리를 빠르게 검출하는 프레임워크를 제안한다. 그래프의 사이즈가 커지는 상황에서도 빠르게 극점 탐색을 실시하고, 순차적인 지역 특징자 매칭을 통해 손, 발, 머리를 검출한다. 초기 RGB-D 입력 이미지에서 탐색 속도 향상과 주변 정보에 의한 검출 오류를 줄이기 위해 배경 분리를 실시하고, 노이즈를 제거해 사담 영역을 획득한다. 획득한 인체 영역은 깊이 정보와 색상 정보를 기반으로 쿼드트리 구조로 분할한다. 쿼드트리 기반 그래프 구성으로 기존 대비 그래프의 사이즈를 효과적으로 줄이고 빠르게 극점 탐색을 수행한다. 탐색된 극점 영역은 지역 검출기를 활용해

머리 위치를 결정하고, 누적 측지거리 경로 정보를 기반으로 극점 영역이 상체에 속하는지 하체에 속하는지 판단한다. 이후 HOG 기반 서포트 벡터 머신 검출기를 활용해 손과 발을 검출한다. 제안한 알고리즘의 성능은 [2]와 [4]의 방법과 비교해 극점 탐색의 처리 속도를 측정했고, 손, 발, 머리 검출의 정확도를 측정했다. 실험 결과, 제안한 방법이 기존 방법 대비 속도적인 측면에서 향상이 있었음을 확인할 수 있었다.

본 논문의 구성은 다음과 같다. 2장은 본 논문의 기본 이론이 되는 다익스트라 알고리즘 기반 누적 측지거리 극점 탐색 방법, 이미지 쿼드트리 분할, HOG와 서포트 벡터 머신에 대해서 서술한다. 3장은 제안하는 손, 발, 머리 검출 프레임워크에 대해서 설명한다. 제안하는 알고리즘의 전체적인 개요, 배경 제거 및 사람 영역 획득, 손, 발, 머리 후보영역 탐색 및 검출 순서로 상세하게 설명한다. 그리고 4장에서는 제안한 방법의 성능을 평가하고, 5장에서 결론을 맺는다.

## 2. 이 론

### 2.1 다익스트라 알고리즘 기반 누적 측지거리 극점 탐색 방법

다익스트라 알고리즘은 정점과 간선들로 구성된 그래프에서 주어진 두 정점간의 최단거리 문제를 푸는 알고리즘이다. 매 순간 도달 가능한 정점들 중 가장 적은 비용을 가진 경로를 선택하면서 시작 정점에서 다른 정점까지 도달하는 최단거리를 구한다. 다익스트라 알고리즘의 시간복잡도는  $O(|E| + |V| \log |V|)$ 로  $|V|$ 와  $|E|$ 는 각각 정점의 개수와 간선의 개수를 나타낸다[5]. 인체의 깊이맵 데이터에서 손, 발, 머리 영역은 몸의 중심으로부터 누적 측지 거리가 다른 신체 부분에 비해 멀리 나타나는 특성을 갖는다. 다익스트라 알고리즘 기반 손, 발, 머리 후보 영역 검출 방법[2]은 다음과 같다.

### 2.2 이미지 쿼드트리 분할

쿼드트리는 부모노드가 4개의 자식노드를 갖는 트리 구조를 말한다. 쿼드트리를 이용한 이미지 분할은 유사한 속성을 가진 픽셀들을 군집화시켜 이미지를 표현하는 픽셀수를 효과적으로 줄인다. 최악의 경

```

Pseudocode 1
input : undirected weighted graph  $G$ , start vertex  $V_s$ , the number of end points  $n$ 
output : accumulative geodesic end points  $E = \{e_k | k = 1, \dots, n\}$ 

function FindExtrema ( $G, V_s, n$ )
 $G_1 = G$ 
for  $i = 1$  to  $i \leq n$ 
    (1) set  $V_s$  as start vertex, and search  $G_i$  with Dijkstra's algorithm
    (2) set  $e_i$  to the longest accumulative geodesic end point of (1)
    (3) Update  $G_i$  to  $G_{i+1}$  by adding zero weighted edge between  $V_s$  and  $e_i$ 
    
```

우 쿼드트리 of 모든 리프 노드들이 하나의 픽셀로 구성되어 군집화의 효과를 얻을 수 없지만 일반적으로 인접한 픽셀들 간의 유사한 속성으로 군집화가 이뤄져 이미지의 크기를 줄이고 연산량을 효과적으로 줄일 수 있다[9,10]. Fig. 1은 쿼드트리로 분할된 이미지를 나타낸다. 최상위 노드로 설정된 이미지는 최대 깊이가 3인 쿼드트리로 구성됐고, 이미지는 10개의 분할된 영역으로 표현된다.

2.3 Histogram of Oriented Gradients (HOG)

HOG는 Dalal [6]가 제안하는 지역적 특징 추출 방법으로 보행자 검출 목적으로 활용했다. 컴퓨터 비전과 영상처리 분야에서는 객체의 특정 영역이나 전체 영역에 대한 HOG 특징을 추출하고 조합해 객체 검출에 활용한다[7,8,12]. HOG는 지역적 특징추출을 위해 셀(cell)이라는 최소단위로 구분한다. 각 셀들은 여러 개의 픽셀들로 구성되고, 여러 개의 셀들이 모여 하나의 블록(block)을 구성한다. 블록과 셀의 크

기는 조절가능하고, 블록들은 인접한 블록들과 중첩하여 구성 가능하다. 셀의 형태는 사각형과 log-polar 방식으로 선택 가능하다. 미분 방향에 대한 공간은 0~180도 혹은 0~360도에 대해 일정한 크기를 가진 빈(bin)들로 구성된다. 각 픽셀에서 측정된 미분 방향은 셀 단위로 가중치(미분의 크기)를 갖은 투표로 누적되어 히스토그램으로 표현되고, 블록 단위로 셀의 히스토그램들을 1차원으로 나열하여 특징 벡터로 활용한다.

2.4 서포트 벡터 머신

서포트 벡터 머신은 분류와 회귀 분석에 활용되는 지도학습 모델이다. n 종류의 학습 데이터가 주어졌을 때 각 종류를 구분할 수 있는 결정 경계면인 n-1 개의 초평면을 설정하는데, 초평면으로부터 가장 가까운 위치에 존재하는 샘플인 서포트 벡터와의 여백 (margin)을 최대화시켜 일반화 능력을 극대화한다. 이진분류인 경우 결정 초평면을  $w \cdot x + b = 0$  으로 표

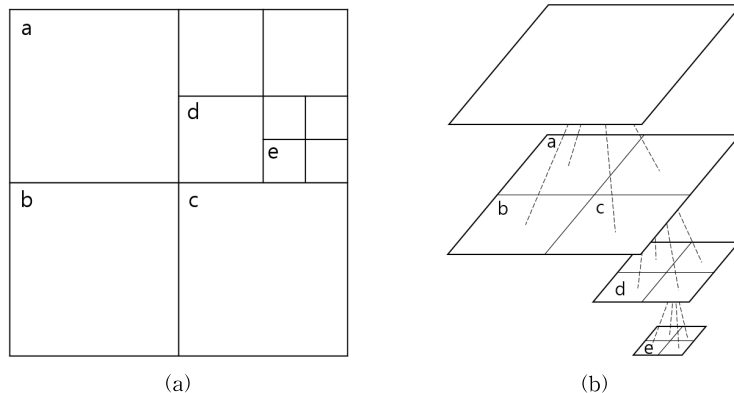


Fig. 1. Quadtree structure representation of an image : (a) the segmented image and (b) the quadtree structure of the image.

현했을 때 여백의 크기는  $\frac{2}{|w|}$ 이다. 최적의 초평면을 구하는 작업은  $\frac{2}{|w|}$ 를 최대화하는  $w$ 와  $b$ 를 얻는 것으로 라그랑제 승수(lagrange multiplier)를 도입하여 KKT 조건을 이용하고, wolfe 듀얼 문제로 변형하여 계산한다[11].

초기 vopnik의 서포트 벡터 머신은 선형 분류 문제를 해결하기 위한 방법으로 제안되었고, 이후 특징 공간을 변형시키는 커널이 제안되어 비선형 분류기로 확장되었다. 특징 변환 커널로는 대표적으로 다항식(polynomial) 커널, RBF(radial basis function) 커널, 하이퍼볼릭 탄젠트(hyperbolic tangent) 커널이 있다.

### 3. 제안한 방법

#### 3.1 제안한 알고리즘의 개요

본 논문에서는 RGB-D 이미지에서 손, 발 머리를 검출하는 프레임워크를 제안한다. 사람의 손, 발, 머리 부분은 신체의 다른 부분에 비해 동작의 변화가 크게 나타나 많은 정보를 담고 있어 제스처 인식이나 포즈 맵핑, 동작 검색에서 주요 특징으로 활용된다. Fig. 2는 제안하는 프레임워크의 대략적인 흐름도를 보여준다.

우선, 배경 정보를 알고 있다고 가정한다. 입력되는 RGB-D 이미지에서 배경을 제거해 전경을 획득한다. 전경 영역에서 군집화된 영역의 크기를 측정하고, 일정 크기 이상인 경우 사람이라고 판단한다. 다음은 사람이라고 판단된 영역 내에서 손, 발, 머리 후보영역인 극점을 빠르게 찾는 작업을 수행한다. 인체를 쿼드트리 구조로 분할해 그래프로 구성하고, 다익스트라 알고리즘을 이용해 몸의 중심으로부터 가장 멀리 떨어진 극점들을 탐색한다. 마지막으로 극점 영역에 대해 HOG 특징벡터를 이용한 검출기로 머리

를 검출하고, 측지거리 속성을 기반으로 상체와 하체를 구분한 뒤 손과 발을 검출한다.

#### 3.2 배경 제거 및 사람 영역 획득

입력되는 RGB-D 이미지에서 사람 영역 획득을 위한 전처리 작업을 수행한다. 관심영역만 추출함으로써 연산량을 줄이고, 주변 정보로 인해 발생하는 오류를 최소화한다. 먼저, 배경 깊이 영상을 저장해 활용한다. 깊이 영상은 색상 영상에 비해 조명의 변화에 강한 특성이므로 안정적인 관심영역 획득이 가능하다. RGB-D 이미지가 입력되었을 때 배경 영상과 차 영상을 계산해 전경을 획득한다. 이후 외각선 검출 알고리즘을 이용해 군집화된 영역들의 크기를 추정하고, 일정크기이상인 경우 사람이라고 판단한다.

#### 3.3 손, 발, 머리 후보영역 검출

인체 영역에서 손, 발, 머리가 나타나는 극점 영역을 빠르게 탐색하는 방법을 제안한다. 쿼드트리 분할 기법을 적용해 인체를 표현하는 그래프  $G$ 의 정점과

```

Pseudocode 2
input : image  $I = \{depth, color\}$ , init tree depth
 $treeDepth = 0$ 
output : decomposed image with quadtree structure
function Quadtree ( $I, treeDepth$ )

    if( $treeDepth \geq Max_d$ )
        return  $node(I)$ 
    else
         $Std_d =$  standard deviation of  $I_{depth}$ 
         $Std_c =$  standard deviation of  $I_{color}$ 
        if ( $Std_d > T_d \parallel Std_c > T_c$ )
             $I_{sub} =$  split  $I$  as four sub images
            Quadtree ( $I_{sub}, treeDepth + 1$ )
    
```

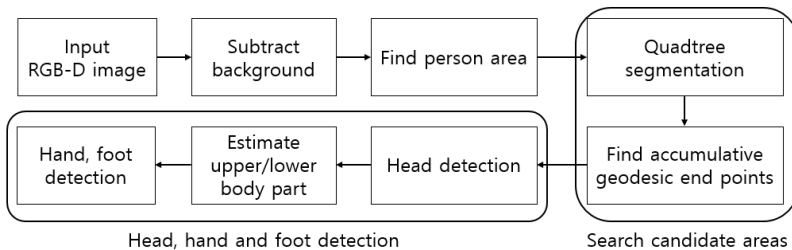


Fig. 2. Overview of a proposed framework for head, hand and foot detection in RGB-D image.

간선의 개수를 효과적으로 줄임으로써 인체 영역의 크기가 커지는 상황에서도 다익스트라 알고리즘을 이용한 극점 탐색 시간을 보장한다.

입력되는 이미지  $I = \{depth, color\}$ 는 깊이 정보와 색상 정보를 기반으로 쿼드트리 구조로 분할한다. 트리를 구성하는 노드는 깊이와 색상의 분산값  $T_d$ 와  $T_c$ 를 기준으로 분할되고, 트리의 최대 깊이는  $Max_d$ 로 제한된다.

Fig. 3은  $T_d=8$ ,  $T_c=5$  그리고  $Max_d=8$ 로 설정했을 때의 쿼드트리 분할 영상이다. 배경이 제거된 전경 영상에서 색상 정보만 활용한 경우, 깊이 정보만

활용한 경우 그리고 색상과 깊이 정보를 모두 활용한 경우의 분할 결과를 보여준다. 색상 정보만 활용한 경우 공간 속성 고려하지 못해 신체의 일부가 배경 영역과 혼동하거나 인접하지 않은 픽셀들이 하나의 노드로 구성되는 오류가 발생한다. 깊이 정보만 활용한 경우 공간 속성을 반영하지만 팔이 몸에 밀착되는 상황과 같이 공간 특징이 제대로 나타나지 않는 경우 다른 신체부위임에도 불구하고 하나의 노드로 구성되는 문제가 발생한다. 하지만 깊이 정보와 색상 정보를 동시에 활용한 경우 상호보완적 관계로 공간적인 특성을 고려하고 공간적인 특성이 뚜렷하진 않은

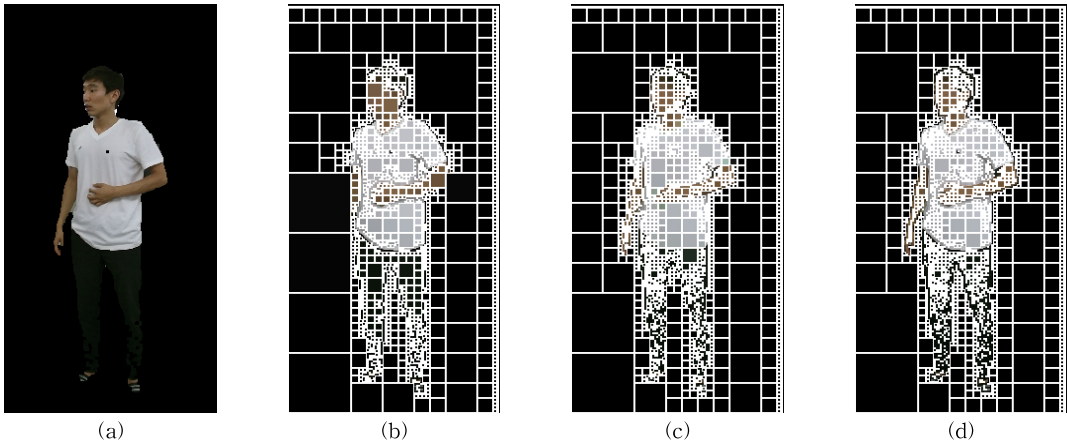


Fig. 3. Comparison of quadtree based image segmentation : (a) a foreground image, (b) color-based segmentation, (c) depth-based segmentation, and (d) color and depth-based segmentation.

```

Pseudocode 3
input : undirected weighted graph  $G$ , start vertex  $v_{start}$ , the number of end points  $n$ 
output : accumulative geodesic end points  $E$ 

function FindExtrema( $G, v_{start}, n$ )
 $s_0 = v_{start}$ 
 $Geo = empty$ 
for  $i = 1$  to  $i \leq n$ 
    if  $i \leq 1$ 
        (1) set  $s_i$  as start vertex, and search  $G_i$  with dijkstra algorithm
        (2) save the accumulative geodesic distances of (1) to  $Geo_i$ 
        (3) set  $e_i$  to the longest accumulative geodesic end point of  $Geo_i$ 
        (3) Update  $G_i$  to  $G_{i+1}$  by adding zero weighted edge between  $V_s$  and  $e_i$ 
    else
        (4) set  $s_i$  as start vertex, and partially search  $G_i$  whose vertices are nearer to  $S_i$  than to  $S_{i-1}$ 
        (5) update  $Geo_{i-1}$  to  $Geo_i$  using result of (4)
        (6) set  $e_i$  to the longest accumulative geodesic end point of  $Geo_i$ 

 $s_{i+1} = e_i$ 
    
```

부분은 색상 정보로 구분함으로써 다른 신체 영역이 하나의 노드로 구성되는 오류를 줄일 수 있다.

분할된 쿼드트리 영역은 그래프  $G$ 의 정점  $v_x \in G$ 로 설정되고, 인접한 정점들간의 유클리디안 거리를 가중치로 갖는 간선  $e_x \in G$ 으로 연결하여 무향성 연결형 가중치 그래프를 생성한다. 다익스트라 알고리즘 탐색을 위한 시작 위치  $v_{start}$ 는 몸의 중심으로 인체 영역 깊이 정보의 평균값으로 설정한다. [4]가 제안한 방법을 활용해 그래프  $G$ 에 대해서 극점  $E = \{e_k | k = 1, \dots, n\}$  탐색을 실시한다.

Fig. 4는 10개의 극점 탐색 결과를 보여준다. 빨간 점은 극점 탐색의 시작 위치를 나타내고, 초록점은 탐색된 극점을 나타낸다. 탐색된 극점들 가운데 관심 영역인 손, 발, 머리 부분이 포함된 것을 확인할 수 있다.

### 3.4 손, 발, 머리 검출

3.3에서 탐색한 극점 위치에서 일정한 크기의 탐

색 영역을 설정한 뒤 슬라이딩 윈도우 기법과 멀티스케일 기법을 적용해 손, 발, 머리를 검출한다. 검출기는 HOG 지역 특징 추출자를 이용하고, 서포트 벡터 머신을 이용해 학습한다. 또한 낮은 해상도의 이미지에서 유사한 특징으로 오검출되는 문제를 해결하기 위해 강인한 특징이 나타나는 머리를 검출하고, 극점 탐색 영역을 상/하체로 구분한 뒤 각 영역에 대해서 손과 발을 검출한다. Fig. 5는 극점 영역에서 손, 발, 머리를 검출하기 위한 흐름도 보여준다.

우선, 학습한 검출기를 이용해 머리 영역을 검출한다. 인체를 구성하는 그래프  $G$ 에서 검출한 머리 위치를 정점  $v_{head}$ 이라고 할 때 몸의 중심 정점  $v_{center}$ 을 기준으로 그래프  $G$ 를 구성하는 노드들의 상체와 하체 소속 여부는 다음과 같이 추정된다.

$$Thres = Dist(Pos(v_{head}), Pos(v_{center}))$$

$$Label(v_i) = \begin{cases} +1(upper) & \text{if } Dist(Pos(v_{head}), Pos(v_i)) \leq Thres \\ -1(lower) & \text{else} \end{cases} \quad (1)$$

$Dist$ 는 유클리디안 거리, 그리고  $Pos$ 는 해당 정점

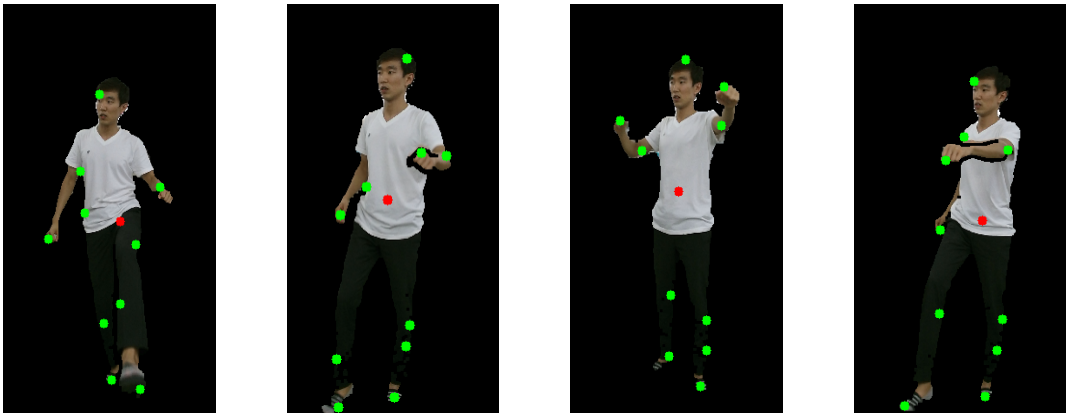


Fig. 4. Results of searching accumulative geometric end points in quadtree structured graph model ( $|E|=10$ ; red dot is  $v_{start}$  and greens dots are  $E$ ).

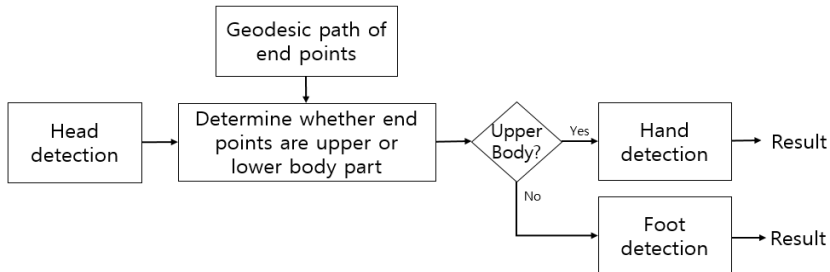


Fig. 5. Flowchart of head, hand and foot detection for geodesic end points.

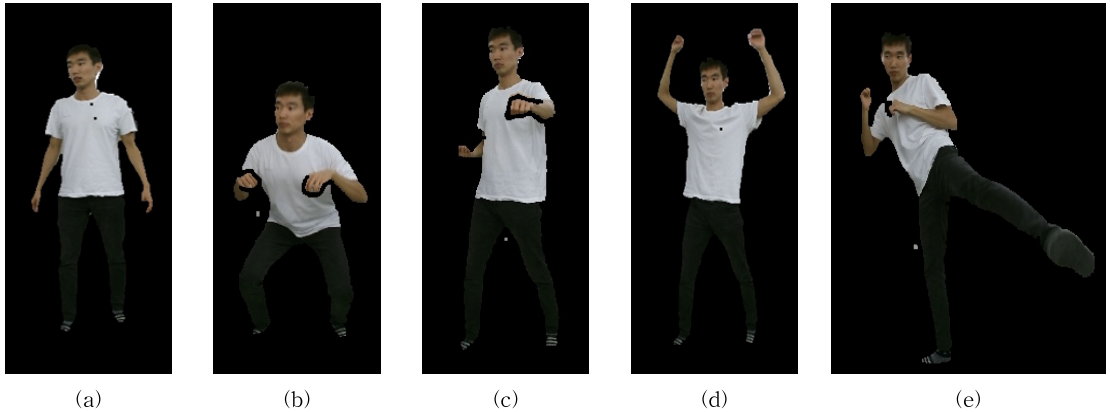


Fig. 6. Some samples of the experiment dataset with various poses : (a) standing, (b) sitting, (c) punching, (d) waving hands, and (e) kicking.

의 3차원 좌표를 의미한다. 3.3에서 검출된 극점  $E_i$  이고,  $E_i$ 까지 도달하는 누적 축지거리 경로가  $Geo_{e_i}$  일 때 극점  $E_i$ 가 소속된 신체의 영역은 다음과 같이 결정된다.

$$Body(E_i) = \begin{cases} \sum_{i=1}^{Geo_{e_i}} Label(v_i) > 0, +1 \\ else, -1 \end{cases} \quad (2)$$

$Body(E_i)=1$  이면 해당 극점은 상체의 영역으로 판단하고 손 검출을 실시하고, 그렇지 않은 경우 발 검출을 실시한다.

#### 4. 실험 결과 및 고찰

제한한 알고리즘의 성능은 i5 3.1GHz dual core CPU, 8GB Ram의 일반적인 범용 PC에서 테스트하였다. RGB-D 데이터 수집을 위한 장치로 Kinect v2을 이용해 색상과 깊이 데이터를 획득해 실험 데이터로 활용했다. Fig. 6은 실험 데이터 샘플로 사람 영역만 추출한 이미지이다. 사용된 데이터는 서기, 앉기 이외에도 주먹 지르기, 양손 흔들기, 발차기 등의 자세를 포함한다.

Table 1은 약 1.7m 거리에서 촬영한 영상에서 인체 영역을 그래프로 표현했을 때 쿼드 트리 기반의 이미지 분할을 적용했을 때와 그렇지 않았을 때의 그래프 크기를 비교한 것이다. 제안하는 방법을 적용했을 때 기존 대비 그래프의 정점 사이즈는 약 63%, 간선 사이즈는 약 65% 줄어든 것을 확인할 수 있었다.

Table 1. Comparison of the graph size between proposed method and original

	Quadtree structured graph model	Depth map based graph model
The number of vertices	5759	15683
The number of edges	37083	105487

Fig. 7은 제안하는 방법과 [2]와 [4] 방법의 극점 탐색 속도를 비교한다. 10개의 극점을 탐색했을 때의 평균 소요 시간을 보여준다. 첫 번째 노드의 평균 탐색 속도는 제안하는 알고리즘의 경우 24.3ms인 반면 다른 두 알고리즘은 140.3ms로 6배 이상의 속도향상을 보였다. 또한, 두 번째 노드부터의 평균 탐색속도는 제안하는 방법이 2.3ms인 반면 [4]는 29.9ms, [2]은 193.4ms로 제안하는 방법이 두 알고리즘에 비해 각각 13배, 84배 빠른 성능을 보였다. 제안하는 알고리즘의 초기 쿼드트리 생성 속도 20ms를 고려하더라도 극점 탐색에서의 속도향상을 확인할 수 있었다.

제안하는 알고리즘의 극점 탐색 성능은 578장의 이미지를 이용해 측정했다. Fig. 6과 같이 다양한 자세를 포함하는 이미지에서 5개와 10개의 극점 탐색 시 검출하고자하는 손, 발, 머리 탐색 성공률을 측정했다. 손과 발 각각 1156개의 영역, 머리 578개의 영역에 대해 5개 극점 탐색 시 86.7%, 89.9%, 43.6%의 성공률을 보였고, 10개의 극점 탐색 시 99.6%, 96.9%, 99.3%의 성공률을 보였다. 인체의 포즈 변화로 인해

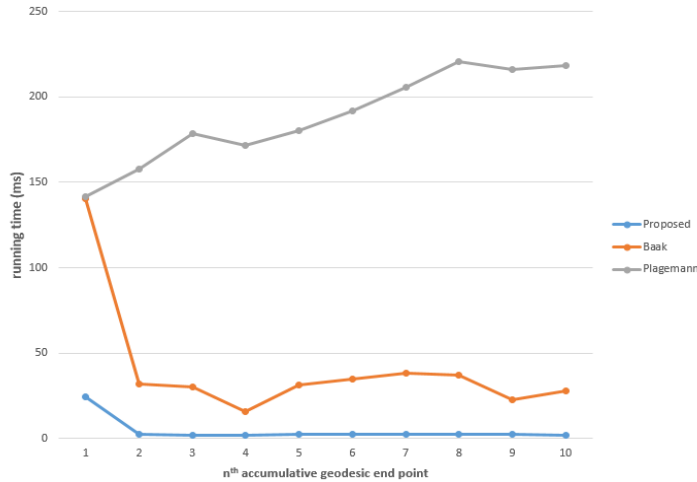


Fig. 7. Comparison of the average running time for finding  $n^{\text{th}}$  accumulative geodesic end point with proposed method and others.

그래프 구성이 달라지더라도 10개 극점 탐색 시 95% 이상의 확률로 손, 발, 머리 영역을 탐색할 수 있었다.

손, 발, 머리 검출성능을 평가를 위해 총 1768장의 학습 데이터를 사용했다. 각 이미지에서 손, 발, 머리의 이미지 패치를 추출하고, 각각 48×48, 48×48, 64×64 크기로 조절하여 학습 모델을 생성했다. 테스트 데이터는 1247장의 이미지를 사용했고, 인체에서 탐색한 10개의 극점 영역에 대해 손, 발, 머리 검출을 실시했다. 검출 성공률은 손의 경우 2494개 중 1818개를 검출해 72.89%의 성능을 보였고, 발은 2490개 중 2027개, 머리는 1247개 중 1083개로 각각 81.41%와 86.85%의 성능을 보였다. 10개의 영역에 대한 평균 검출 속도는 32ms이고, 제한하는 프레임워크의 모든 과정은 115ms가 걸렸다.

5. 결 론

본 논문에서는 RGB-D 이미지에서 손, 발, 머리를 검출하는 프레임워크를 제안하였다. 쿼드트리 기반의 극점 탐색은 인체를 표현하는 그래프의 사이즈를 효과적으로 줄여 빠르게 손, 발, 머리의 후보 영역을 탐색할 수 있었다. 또한 강인한 특징을 가진 머리를 검출하고, 극점 영역을 상체와 하체를 분류한 뒤 손과 발을 검출함으로써 인체의 많은 정보를 갖고 있는 손, 발, 머리를 안정적으로 검출할 수 있었다.

REFERENCE

[ 1 ] J. Shotton, R. Girshick, A. Fitzgibbon, T.

Sharp, M. Cook, M. Finocchio, et al., “Efficient Human Pose Estimation from Single Depth Images,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 35, No. 12, pp. 2821–2840, 2013.

[ 2 ] C. Plagemann, V. Ganapathi, D. Koller, and S. Thrun, “Real-time Identification and Localization,” *Proceeding of IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 3108–3133, 2010.

[ 3 ] L.A. Schwarz, A. Mkhitarian, D. Mateus, and N. Navab, “Estimating Human 3D Pose from Time-of-Flight Images Based on Geodesic Distances and Optical Flow,” *Proceeding of IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 700–706, 2011.

[ 4 ] A. Baak, M. Müller, G. Bharaj, H.P. Seidel, and S. Theobalt, “A Data-Driven Approach for Real-Time Full Body Pose Reconstruction from a Depth Camera,” *Proceeding of IEEE International Conference on Computer Vision*, pp. 71–98, 2011.

[ 5 ] Dijkstra’s Algorithm in Wikipedia, [https://en.wikipedia.org/wiki/Dijkstra%27s\\_algorithm](https://en.wikipedia.org/wiki/Dijkstra%27s_algorithm), (accessed Aug., 28, 2016).

[ 6 ] N. Dalal and B. Triggs, “Histograms of Oriented Gradients for Human Detection,” *Com-*



- puter Vision and Pattern Recognition*, Vol. 1, pp. 886-893, 2005.
- [7] P.F. Felzenszwalb, R.B. Girshick, D. McAllester, and D. Ramanan, "Object Detection with Discriminatively Trained Part Based Models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 32, No. 9, pp. 1627-1645, 2010.
- [8] G. Jiang, J. Cheng, J. Pang, and Y. Guo, "Real-time Hand Detection Based on Multi-stage HOG-SVM Classifier," *Proceeding of IEEE International Conference on Image Processing*, pp. 4108-4111, 2013.
- [9] C. Stoll, N. Hasler, J. Gall, H.P. Seidel, and C. Theobalt, "Fast Articulated Motion Tracking using a Sums of Gaussians Body Model," *Proceeding of International Conference on Computer Vision*, pp. 951-958, 2011.
- [10] Quadtree in Wikipedia, <https://en.wikipedia.org/wiki/Quadtree>, (accessed Sep., 1, 2016).
- [11] I.S. Oh, *Pattern Recognition*, Kyobobook, Seoul, 2008.
- [12] T.B. Nguyen, V.T. Nguyen, and S.T. Chung,, "A Real-time Pedestrian Detection based on AGMM and HOG for Embedded Surveillance," *Journal of Korea Multimedia Society*, Vol. 18, No. 11, pp. 1289-1301, 2015.



### 홍 성 진

2013년 인하대학교 컴퓨터정보공학과 학사와 2015년 인하대학교 컴퓨터·정보공학과 석사 졸업 후 현재 한국전자통신연구원 SW·콘텐츠연구소 차세대콘텐츠연구본부 연구원으로 재직 중이다. 주 관심 관심분야는 컴퓨터비전, 영상

처리, 머신러닝 등이다.



### 김 명 규

서울대학교 이학과와 미국 메릴랜드 주립대학교 이학박사 후 서울대학교 연구원을 거쳐 현재는 한국전자통신연구원 책임연구원으로 재직 중이다. 한국전자통신연구원에서는 기기 스포츠 시뮬레이터 개발 연구책임자와 무용

체험 시스템 개발 연구책임자였으며, 현재는 청소년용 VR스포츠 개발 연구책임자이다. 주 관심 연구 분야는 인체 모션 분석, 고속 객체 모션 분석, 물리 시뮬레이션과 웨어러블 센서 등이다.