

주행 로봇을 위한 단일 카메라 영상에서 손든 자세 검출 알고리즘

Hand Raising Pose Detection in the Images of a Single Camera for Mobile Robot

권 기 일[†]

Gi-Il Kwon[†]

Abstract This paper proposes a novel method for detection of hand raising poses from images acquired from a single camera attached to a mobile robot that navigates unknown dynamic environments. Due to unconstrained illumination, a high level of variance in human appearances and unpredictable backgrounds, detecting hand raising gestures from an image acquired from a camera attached to a mobile robot is very challenging. The proposed method first detects faces to determine the region of interest (ROI), and in this ROI, we detect hands by using a HOG-based hand detector. By using the color distribution of the face region, we evaluate each candidate in the detected hand region. To deal with cases of failure in face detection, we also use a HOG-based hand raising pose detector. Unlike other hand raising pose detector systems, we evaluate our algorithm with images acquired from the camera and images obtained from the Internet that contain unknown backgrounds and unconstrained illumination. The level of variance in hand raising poses in these images is very high. Our experiment results show that the proposed method robustly detects hand raising poses in complex backgrounds and unknown lighting conditions.

Keywords: Pose Detection, Mobile Robot, Appearance-based Detection, Single Camera

1. 서 론

로봇과 사용자의 상호작용을 위해서 손을 이용한 동작 인식에 관한 연구는 오래 전부터 수행 되어 왔다. 본 논문은 주행 로봇에서 사람과 상호작용을 하기 위한 주행 로봇에 달린 카메라를 이용한 손을 든 자세 검출 알고리즘을 제안한다. 카메라 센서는 깊이 센서에 비해 조명이나 배경변화에 약한 단점을 지니고 있다. 하지만 깊이 센서에 비해 먼 거리의 객체 검출이 가능하고 카메라 센서의 이미지의 경우 객체의 모양, 색상, 텍스처 등 보다 많은 정보를 제공한다. 손을 드는 동작은 사용자에게 직관적이고 기본적인 동작으로 이를 통해 로봇과 상호작용이 가능하다.

카메라 센서를 이용한 기존의 동작 검출 알고리즘은 환경에 제약이 강하고 강인하지 못하다. 하지만 주행 로봇과 사용자의 상호작용을 위해서는 다양한 환경에서 강인하게 동작하는 동작 검출 알고리즘이 필요하다. 기존의 연구들은 고정된 카메라를 가정하여 배경 차감 기법을 이용해서 사용자의 팔 영역을 분리해주어 동작을 검출했다.^[1-3] 하지만 주행 로봇에 부착된 카메라 센서를 이용해서 주행로봇이 움직이는 상황에서 손을 든 동작 검출을 할 경우 배경을 예측 불가능하고 움직이고 있기에 배경 차감기법을 이용한 배경 추출이 불가능하다. 또한 기존의 기법들은 로봇이 움직임에 따른 조명 변화를 고려하지 않았다. 그리고 강의실이나 교실과 같은 단순한 배경을 가진 이미지 안에 사용자의 손든 동작을 검출 해왔다.^[1-5] 하지만 주행 로봇이 동작하는 실제 환경은 복잡한 배경을 가지고 조명은 로봇의 움직임에 따라 변하는 동적인 환경이다.

Received : Apr. 9. 2015; Reviewed: May 21. 2015; Accepted : Jun. 17. 2015

[†] Corresponding author: Integrated Control Research Team, National Fusion Research Institute(NFRI) 169-148 GWAHAK-RO, YUSEONG-GU, DAEJEON 305-806, KOREA (Eoeun-dong), (getupforone@gmail.com)

또한 그림 1과 같이 사용자가 착용한 옷에 따라, 사용자의 손드는 습관에 따라 사용자들의 손든 모습은 다양하다. 이러한 다양한 손든 모습을 일반성을 유지하며 학습하기 위해서는 매우 많은 학습 데이터가 필요하며 학습하기 어렵다.^[6] 이러한 환경에서 강인하게 손드는 사람을 검출하기 위해서 손든 사람의 자세나 옷에 상관없이 손든 동작을 검출하는 손 검출기술 기반 손드는 동작 검출 기법과 손든 사람 상반신 모습을 학습하여 검출하는 두 가지 방법을 제안하고 이를 병합하는 방법을 제안했다. 복잡한 배경과 조명 변화에 강인하게 손을 검출하기 위해서 HOG(Histogram of Oriented Gradient) 특징과 선형 SVM을 이용한 손바닥 검출기와 사용자 손든 모습 검출기를 학습했다. HOG 특징을 사용하여 객체를 검출할 경우 조명 변화와 배경 변화에 강인하게 객체 검출이 가능하다.^[7,8] 손을 든 영역을 검출하기 위해서 본 논문에서는 먼저 얼굴영역을 검출했다. 얼굴검출 후 오차요인을 줄이고자 검출된 얼굴 영역 주변 부분을 관심영역으로 설정하여 이 관심영역 안에서 손바닥 영역을 검출했다. 그리고 얼굴 영역에서 얻어진 살색상 분포를 이용하여 검출된 손 영역을 검증했다. 이렇게 구해진 손 위치가 얼굴 위치 보다 위에 있을 경우 손을 든 동작으로 인식했다. 그리고 얼굴이 검출되지 않았을 경우나 손바닥이 검출되지 않았을 경우를 고려하여 HOG특징과 선형 SVM을 이용한 손을 든 사람 상반신 모습을 학습한 검출기를 개발했다. 손든 사람 상반신모습의 경우 손든 모습은 다양할 수 있으나 검출 대상을 팔을 곧게 펴고 위로 곧 바르게 손을 든 모습으로 제한하여 학습했다. 그리고 이 두 가지 기법을 응용하여 손든 사람 모습 검출하는 기법을 제안했다.



Fig. 1. Hand Raising Pose Dataset, RYH DB

2장에서는 본 논문에서 제안한 손든 사람 검출 알고리즘에 대해 설명하고 3장에서는 손 검출기술 기반 손드는 동작 검출 기법을 설명하고 4장에서는 손든 사람 상반신 모습을 검출하는 기법에 대해 설명했다. 5장에서는 웹 캠을 이용해서 얻어진 이미지와 인터넷 이미지를 이용해 자체 제작한 데이터를 이용한 손든 사람 동작 검출기를 평가한다. 6장에서는 이 논문에 결론을 내리고 향후 연구 방향에 대해 제안한다.

2. 손든 사람 검출 알고리즘

2.1 손든 사람 검출 알고리즘

손든 사람 검출 알고리즘은 주행로봇에 부착된 카메라 센서를 이용해서 얻어진 영상에서 손든 사용자를 검출하는 알고리즘이다. 카메라에서 영상이 들어오면 이영상을 이용하여 두 가지 기법을 이용하여 손든 사람을 검출했다. 먼저 사용자의 손 영역과 얼굴영역을 검출 후 얼굴의 위치와 손의 위치의 기하학적 위치를 이용하여 손을 들었는지 판단하는 알고리즘이다. 그리고 사용자의 손든 상반신 모습을 검출하는 알고리즘이 있다. 손 검출기 기반 방식은 이미지에서 얼굴 영역을 검출하여 손이 검출될 만한 검출 후보 영역을 얼굴영역 윗 부분 영역으로 설정한다. 그리고 얼굴 영역에서 살색 색상 분포를 구하고 이미지에서 살색 영역을 구한다.^[9] 얼굴이 검출되지 않았을 때는 화면의 윗 부분의 70%를 관심영역으로 설정했다. 그리고 기존에 정의 되어있던 살색 정의를 이용하여 살색 영역을 구한다.^[10] 영상에서 손 후보영역을 검출한다. 각 손 후보 영역은 식 (1)과 같이 손 영역에 살색 색상영역이 일정 이상 포함되어 있는지 판단하여 일정 이상 포함되어있으면 손 영역으로 판단한다. i 번째 손 후보 영역에서 살색 픽셀의 수 $N_{skin,i}$ 와 전체의 픽셀 수 $N_{tot,i}$ 의 비중을 보아 일정 임계값 이상이면 손 후보영역을 손 영역으로 보아 i 번째 손 후보영역 H_i 에 손이 존재하는 것으로 판단하여 1 값을 할당해주었다.

$$H_i = \begin{cases} 1, & \frac{N_{skin,i}}{N_{tot,i}} > \text{threshold} \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

알고리즘에 정확도를 높이기 위하여 손 검출기 기반 방식

과 동시에 손든 상반신을 검출하는 기법을 통해서 손을 든 모습을 검출했다.

2.2 손 검출 알고리즘

영상에서 손 검출을 위해서 본 논문에서는 손을 들었을 때 손을 펴서 손바닥을 보여주는 편 손 자세로 제한했다.

손을 편 자세 검출기는 그림 3과 같이 손을 강체로 가정하여 64x64 크기의 손 이미지를 HOG 특징을 선형 SVM 을 이용하여 검출기를 학습했다. 검출기를 학습하기 위해서 RYH라는 자체 제작된 손 이미지 데이터베이스를 사용했다. 이 이미지들은 인터넷에서 수집된 다양한 배경과 다양한 조명환경에서 손을 든 이미지 중에서 손을 편 손 이지만 편집하여 제작된 데이터 베이스다. Positive 샘플 4000장과 negative 샘플 16000 장을 사용하여 학습했다. 손 검출은 슬라이딩 윈도우 방식을 이용해서 검출했다. 그림 3은 손 검출 결과를 보여주고 있다. 검은 사각형은 얼굴검출결과이고 하늘색 선은 ROI를 나타낸다. 얼굴 검출 영역 아래 경계선에서 위로 1/3되는 지점 이상의 영역을 ROI로

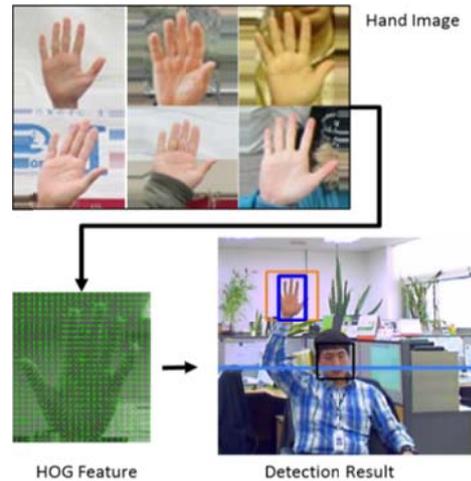


Fig. 3. Open palm detection

설정했다. 주황색은 손 검출기의 검출 결과를 나타내고 파란색은 검출기에서 검출된 손 후보 영역 안에 얼굴색상 분포로 구해진 살색 영역을 나타낸 것이다. 손 검출기는 640x480 해상도의 이미지에서 1~3미터 정도 떨어진 손을 검출 가능하다. 카메라의 해상도가 높아지면 더 먼 거리도 가능하다.

2.3 얼굴 색상 기반 살색 영역 검출

영상에서 손 후보 영역을 검증하기 위해 본 논문에서는 살색 영역을 얼굴 색상 분포 정보를 이용하여 검출했다.^[8] 색상 정보는 조명환경에 따라 변화가 심하다. 조명의 변화에 따라 카메라를 통해 얻어지는 이미지의 색상정보는 변화한다. 그렇기 때문에 기존에 정의 되어있는 색상정보를 가지고 살색 영역을 검출할 경우 살색 영역이 잘 검출되지 않는다. 이와 같은 문제점을 해결하고자 본 논문에서는 얼굴영역에서 검출된 살색 분포를 이용하여 살색영역을 검출했다.^[9] 조명 변화에 따른 사용자의 얼굴 영역의 색상 변화와 손 영역의 색상변화가 유사하다는 가정을 가지고 사용자의 살색 영역을 검출했다. 일단 카메라에서 들어온 영상에서 Normalized-RGB 이미지를 구한다. Normalized-RGB는 그림 4(ب)와 같고 아래와 같은 식으로 구한다.

$$r = \frac{R}{R+G+B}, g = \frac{G}{R+G+B}, b = \frac{B}{R+G+B} \quad (2)$$

Normalized-RGB 3채널 중 살색을 잘 나타낼 수 있는 r, g 2채널을 가지고 색상 분포를 나타내는 color histogram을 그

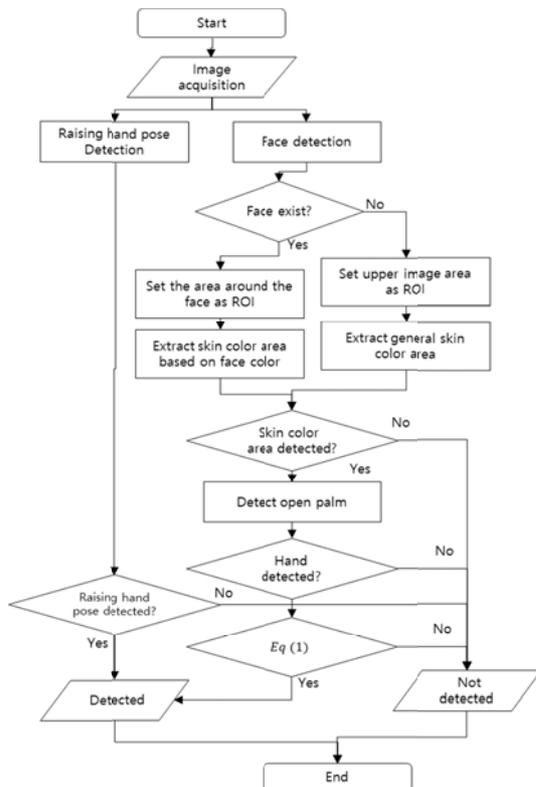


Fig. 2. Flow Diagram of proposed algorithm

림 4 (c)와 같이 구했다. 이중에서 검출된 얼굴 영역이 살색을 대표하는 영역이라 가정하여 얼굴 영역에서 살색 color histogram을 구했다. 그림 4 (c)를 보면 구해진 살색영역과 살색이 아닌 영역의 color histogram이 다른 것을 볼 수 있다. 그리고 Naïve bayes classifier를 구해진 color histogram에 적용해 주어서 살색 영역을 검출했다.^[7] 이를 위해 색상 c 가 살색일 확률 $p\left(\frac{c}{skin}\right)$ 와 살색이 아닐 확률 $p\left(\frac{c}{non\ skin}\right)$ 은 식 (3)과 같이 구한다.

$$p\left(\frac{c}{skin}\right) = \frac{s(c)}{T_s}, \quad p\left(\frac{c}{non\ skin}\right) = \frac{ns(c)}{T_{ns}} \quad (3)$$

여기서 $s(c)$ 는 살색 color histogram에서 색상 c 의 빈의 개수이고 $ns(c)$ 는 살색이 아닌 다른 색상 color histogram에서 색상 c 의 빈의 개수이다. T_s 는 살색 color histogram의 총 원소의 개수이고 T_{ns} 는 살색이 아닌 다른 색상 color histogram의 총 원소의 개수이다.

그리고 식 (4)와 같이 각 픽셀을 살색 픽셀과 살색이 아닌 픽셀로 구분했다.

$$\frac{p\left(\frac{c}{skin}\right)}{p\left(\frac{c}{non\ skin}\right)} > \theta, \quad 0 \leq \theta \leq 1 \quad (4)$$

여기서 θ 는 살색과 살색이 아닌 색상을 구분해줄 때 사용되는 임계 값이다. 그림 4 (d)와 같은 살색영역 이미지를 구했다.

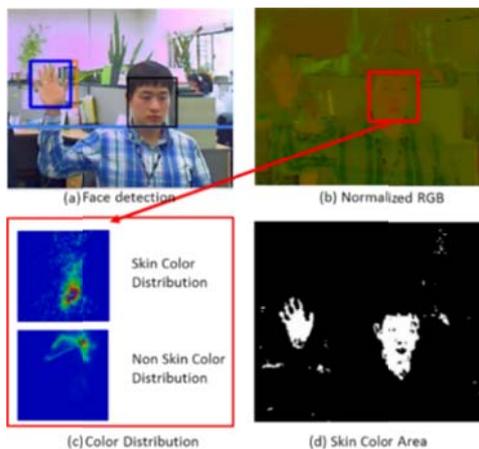


Fig. 4. Skin Color Detection Algorithm

2.4 손 든 상반신 검출 알고리즘

영상에서 손 든 상반신을 검출하기 위해서 본 논문에서는 손을 들었을 때 사람의 자세를 그림 6와 같이 손을 곧게 머리 위쪽으로 편 자세와 45° 정도 팔을 기울여서 든 자세로 제한했다. 이 두 자세에 대한 손든 자세 검출기를 HOG 특징과 선형 SVM을 사용하여 구현했다. 손을 곧게 편 자세는 128x80 크기의 이미지를 가지고 학습을 했고 손을 기울여서 든 자세는 128x128 크기의 이미지를 이용해서 검출기를 학습했다. 학습을 위해서 자체 제작된 RYH라는 손든 사람 이미지 데이터 베이스를 사용했다. 이 이미지들은 인터넷에서 수집된 다양한 배경과 다양한 조명 환경에서 손을 든 이미지만 편집하여 제작된 데이터 베이스다. 128x80 크기의 positive 샘플 2000장, negative 5000장, 128x128 크기의 positive 샘플 5000장, negative 10000장을 사용하여 두 종류의 검출기를 학습했다. 손 든 상반신 검출은 슬라이딩 윈도우 방식을 이용해서 검출했다. 손 검출 때와 달리 ROI는 전체 이미지 영역으로 설정했다. 그림 6와 그림 7는 손을 든 상반신 자세 검출 결과를 보여주고 있다. 그림 6의 노란 사각형은 128x80크기의 검출기 결과이고 빨간 사각형은 128x128크기의 검출기 결과이다. 손든 상반신 자세 검출기는 얼굴이나 손이 검출되지 않아도 손 든 자세 검출이 가능하기 때문에 손 검출기 기반의 방식에 비해 더 먼 거리인 1~4 미터에서 검출 가능하다.

3. 실험 및 결과

본 논문에서는 제안된 방법의 성능을 검증하기 위해 다음과 같이 실험을 진행했다. 실험에 사용된 컴퓨터의 CPU는 Intel i7 2.4GHz이고 메모리는 8GB이다. 카메라는 30 fps를 가지는 Logitech pro5000 웹 캠을 사용했다. 1미터 높이에 15도 정도 위로 향한 pan-tilt가 지원되지 않는 주행 로봇에 고정된 카메라를 가정했다. 주행 로봇의 움직임에 따라 조명과 배경이 바뀌고 다양한 카메라 관점을 가지게 된다. 그리고 움직임에 따른 흐려짐이 이미지에 존재한다. 이와 같은 환경에서 제안한 기법의 성능 검증을 위해서 두 가지 실험 데이터를 사용했다. 먼저 웹 캠을 이용하여 조명 환경, 배경, 카메라의 관점을 변화 시켜주며 손든 사람의 자세를 검출하는 실험을 진행했다. 이때 손든 포즈를

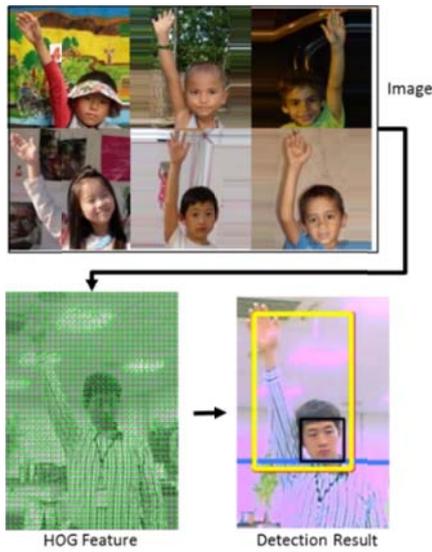


Fig. 5. Hand Raising Pose Detection Algorithm

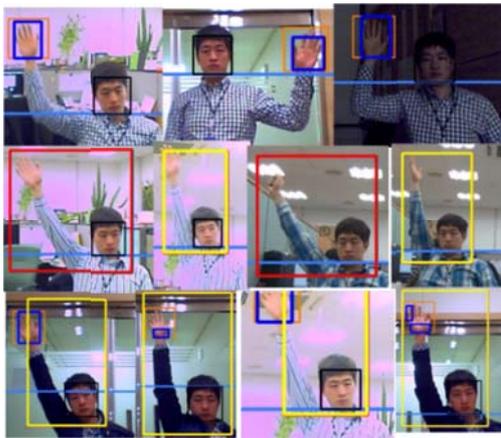


Fig. 6. Experiment result using webcam

바꾸며 173번 손을 드는 것을 웹 캠을 통해서 검출 했다. 그림 6 은 웹 캠을 이용하여 실험을 진행한 결과이다. 그리고 그림 7은 자체 제작된 손든 사람 이미지 데이터 베이스인 RYH DB로 실험한 결과 이다. RYH DB는 인터넷에서 수집된 다양한 배경과 다양한 조명환경에서 여러 사람이 손을 든 이미지만 편집하여 제작된 데이터 베이스다. RYH DB는 얼굴색상과 손색상이 비슷한 인종의 사진으로 구성되어 있다. 얼굴색상과 손색상이 다른 인종 사진은 제외되어 있다.

여러 사람의 손든 동작으로 이루어져 있는 데이터 베이스이기 때문에 웹 캠을 이용한 실험 보다 손을 든 사람의

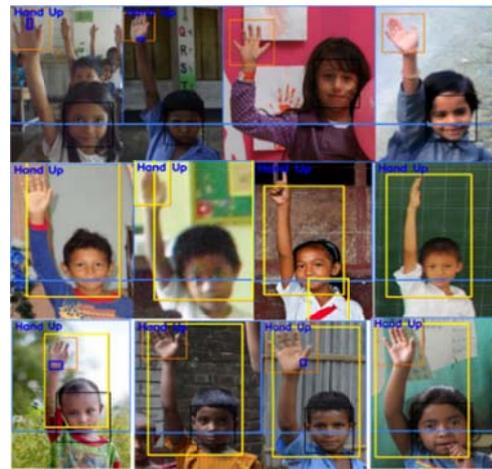


Fig. 7. Experiment result using RYH DB

모습의 변화가 크다. 그림 6과 그림 7에 첫째 줄 사진은 손 검출기 기반 검출 결과이고 둘째 줄은 손든 사람 상반신 모습 기반 검출 결과 이다. 마지막 줄은 손 검출기와 손든 사람 상반신 포즈 검출기를 동시에 사용한 결과이다. 그림 6과 그림 7을 보면 다양한 조명 환경과 배경 그리고 카메라 관점이 변하는 상황에서 강인하게 손든 사람을 검출하는 것을 볼 수 있다.

특히 웹 캠을 이용한 실험 영상은 손을 드는 움직임에 의해 이미지가 흐려지는 효과에도 불구하고 강인하게 검출하는 것을 볼 수 있다. 제안된 기법을 정량적으로 평가 하기 위해 정확도(precision)와 재현 율(recall)을 계산했다. 표 1와 표 2를 보면 첫째 열에 Hand 결과는 Nazare^[4]가 제안한 방식과 유사하게 손과 얼굴의 상대적인 위치만으로 손든 자세를 검출하는 기법이다. 다만 본 논문에서는 오 검출을 줄이기 위해 얼굴색상분포 기반 손 영역 검증 기법을 적용했고, 검출 율을 높이기 위해 손 검출을 위해 HOG 특징을 사용한 점이 다르다. 표 1과 표 2에 Upper Body 결과는 손든 상반신 자세를 검출한 결과이다. 그리고 마지막 줄에 Hand & Upper body결과는 앞에 두 방식을 결합한 방식이다. Upper body 방식과 Hand & Upper body 방식은 본 논문에서 제안한 방식이다. 웹 캠을 이용한 실험결과인 표 1에 보면 Hand & Upper body 방식이 Nazare^[4]가 제안한 방식보다 정확도는 74.6%로 약간 낮지만 재현 율은 85%로 기존 방식의 재현 율 71% 보다 좋은 결과이다. 또한 RYHDB를 이용하여 실험한 결과인 표 2에서도 제안한

Table 1. Detection experiment result using webcam

#1 Webcam		recall	precision
	Hand[4]	71%	76.5%
	Upper Body	62%	79.8%
	Hand & Upper body	85%	74.6%

Table 2. Detection experiment result using RYH DB

#2 RYH DB		recall	precision
	Hand[4]	72%	75%
	Upper Body	50%	32%
	Hand & Upper body	84%	38%

방식의 재현율이 84%로 기존의 방식보다 우수한 성능을 가진다. 이 같은 결과는 Upper body 방식은 Hand 방식에서 손이 검출되지 않은 경우나 얼굴이 검출되지 않는 경우에도 손든 상반신을 검출하게 되어 기존 기법에 비해 높은 재현율을 보여준다. 손든 상반신 검출 기법의 경우 손 이미지에 비해 검출 대상에 배경이 포함된 비율이 크다. 이로 인해 손든 사람 상반신 검출기는 손 검출기에 비해 배경 변화에 민감하다. 이러한 이유로 RYH DB의 경우 그림 8과 같이 배경이 복잡하기 때문에 표 2와 같이 Upper body 방식의 정확도가 32%로 기존 방식에 비해 낮고 이로 인해 Hand & Upper body 방식의 정확도도 38%로 낮은 것을 볼 수 있다. 하지만 그림 7과 같이 서비스 로봇이 작동하게 될 비교적 덜 복잡한 배경과 조명인 실내환경에서는 제안된 기법인 Upper body의 정확도가 79.8%로 기존 기법 정확도 76.5%보다 높았고 Hand & Upper body의 정확도가 74.6%로 기존 기법과 비슷하고 재현율은 85%호 기본 기법보다 더 높다.

4. 결 론

본 논문에서는 주행 로봇에 부착된 카메라를 통해서 손과 얼굴을 검출한 뒤 두 위치를 비교하여 손든 모습을 검출하는 기법과 손든 사람의 상반신 모습을 학습한 뒤 검출하는 두 가지 기법을 제안했다. 자체 제작한 RYH DB를 통해서 손 검출기와 손든 사람 상반신 학습 기를 제작했다. 그리고 얼굴 색상 분포를 이용하여 손 영역을 검증했다. 이 두 가지 기법을 통해서 다양한 배경, 조명, 카메라 관점을 가진 이미지 데이터 베이스에서 기존의 기법에 비

해 높은 검출율을 보였다. 추후 DB를 조명환경과, 카메라 관점 변경, 배경의 복잡도에 따라 정량적으로 분류하고, 보다 정량적으로 알고리즘을 평가하겠다. 이를 통해 배경 변화에 강인하고 먼 거리에 있는 손든 사람을 검출하는 기법에 대해 연구할 계획이다.

References

- [1] Kim, DoHyung, et al., "Vision-based arm gesture recognition for a long-range human-robot interaction", *The Journal of Supercomputing*, vol. 65, issue 1, pp.336-352, July, 2013.
- [2] Chiang, Cheng-Chieh. "Automatic Raising Hand Detection in an Intelligent Classroom.", *IJEIR*, vol. 3, no.2, pp.151-155, March, 2014.
- [3] Xiaodong Duan; Hong Liu, "Detection of hand-raising gestures based on body silhouette analysis," *IEEE Robotics and Biomimetics*, Bangkok, Thailand, 2009, pp.1756-1761.
- [4] Nazaré, Tiago S., and Moacir Ponti. "Hand-raising gesture detection with Lienhart-Maydt method in videoconference and distance learning." *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*. Springer Berlin Heidelberg, 2013. 512-519.
- [5] Nyan Bo Bo; Van Hese, P.; Van Cauwelaert, D.; Veelaert, P.; Philips, W., "Detection of a hand-raising gesture by locating the arm," *IEEE Robotics and Biomimetics (ROBIO)*, Phuket, Thailand, 2011, pp.976-980.
- [6] Kim Juchang, Park Jeong-Woo, Kim Woo-Hyun, Lee Won-Hyong, Chung Myung-Jin, "Primitive Body Model Encoding and Selective Asynchronous Input-Parallel State Machine for Body Gesture Recognition", *Journal of Korea Robotics Society*, 2013, vol.8, no1., pp.238-246.
- [7] Jeongdae Kim1, Yongtae Do, "Human Detection in the Images of a Single Camera for a Corridor Navigation Robot", *Journal of Korea Robotics Society*, 2013, vol.8, no4., pp.1-7..
- [8] Dalal, N.; Triggs, B., "Histograms of oriented gradients for human detection," in *IEEE Computer Vision and Pattern Recognition*, 2005, vol.1, no., pp.886-893, June 2005.
- [9] P. Yogarajah, A. Cheddad, J. Condell, K. Curran and P. McKeivitt "A dynamic threshold approach for skin segmentation in color images", In *Proc. of IEEE Int. Conf. on Image Processing*, 2010, pp.2225,2228.
- [10] P. Kakumanu, S. Makrogiannis, N. Bourbakis., "A survey of skin-color modeling and detection methods", *Pattern Recognition*, vol. 40, issue 3, pp.1106-1122, March, 2007.



권 기 일

2006 성균관 대학교 기계공학부
(공학사)

2008 카이스트 로봇학제전공 (공학
석사)

2014 카이스트 로봇학제전공 (공학
박사)

2014~현재 국가핵융합 연구소 선임 기술원

관심분야: Computer Vision, Machine Learning, Robot