

다양한 환경에서 강건한 RGB-Depth-Thermal 카메라 기반의 차량 탑승자 점유 검출

Robust Vehicle Occupant Detection based on RGB-Depth-Thermal Camera

송창호¹ · 김승훈[†]
Changho Song¹, Seung-Hun Kim[†]

Abstract: Recently, the safety in vehicle also has become a hot topic as self-driving car is developed. In passive safety systems such as airbags and seat belts, the system is being changed into an active system that actively grasps the status and behavior of the passengers including the driver to mitigate the risk. Furthermore, it is expected that it will be possible to provide customized services such as seat deformation, air conditioning operation and D.W.D (Distraction While Driving) warning suitable for the passenger by using occupant information. In this paper, we propose robust vehicle occupant detection algorithm based on RGB-Depth-Thermal camera for obtaining the passengers information. The RGB-Depth-Thermal camera sensor system was configured to be robust against various environment. Also, one of the deep learning algorithms, OpenPose, was used for occupant detection. This algorithm is advantageous not only for RGB image but also for thermal image even using existing learned model. The algorithm will be supplemented to acquire high level information such as passenger attitude detection and face recognition mentioned in the introduction and provide customized active convenience service.

Keywords: Smart car, RGB-Depth-Thermal camera, OpenPose, Occupant detection system

1. 서 론

최근 차량 자율 주행 기술이 발달 되면서 차량 안전 기술이 화두가 되고 있다. 에어백이나 안전벨트와 같은 수동안전시스템에서, 운전자를 포함한 탑승자의 상태 및 행동을 파악하여 위험을 완화 시켜주는 능동형시스템으로 변화되고 있다. 이러한 능동형시스템 개발을 위해서는 탑승자의 상태 및 행동을 파악하는 센서 시스템 개발과 지능형 알고리즘 개발을 위한 데이터베이스 수집이 필수적이다.

차량 탑승자의 상태를 파악할 수 있는 센서로는 대표적으로 가시 카메라(RGB 카메라)를 예로 들 수 있다. RGB 카메라는 저조도 환경이나 터널 입구 등의 광량 변화가 큰 환경에서 정보손실이 크기 때문에, 최근에는 다중 스펙트럼 카메라, 열화상 카메라(Thermal 카메라) 등을 이용하여 다양한 환경에서도 가시 정보를 얻고 이를 활용하는 연구가 진행되고 있다^[1,2]. 본 논문에서는 RGB 카메라뿐만 아니라 다양한 조도 환경에 강건한 Thermal 카메라와 앞좌석, 뒷좌석 탑승자 구별, 탑승자의 체적 분석 등을 위한 Depth 카메라를 포함하여 RGB-Depth-Thermal 카메라를 구성하였다.

RGB-Depth-Thermal 센서로 얻을 수 있는 정보는 저수준(Low level) 정보와 고수준(High level) 정보로 나눌 수 있다. 저수준 정보는 RGB-Depth-Thermal 영상 자체에서 획득할 수 있는 정보를 의미하는데 Depth 영상을 통해 앞좌석 뒷좌석 구분, 사람의 체적 등과 Thermal 영상을 통해 사람의 체온, 차량 내온도 등을 예로 들 수 있다. 고수준 정보는 RGB-Depth-Thermal

Received : Jan. 18. 2018; Revised : Feb. 9. 2018; Accepted : Feb. 9. 2018

* This project was funded by Ministry of Trade, Industry and Energy and is currently supported by the publication grant (10077617, Development of Personalized Active Comfort System (Seat, A/C, D.W.D warning) based on Multi Bandwidth IR Sensor Module)

1. Intelligent Robotics Research Center, Korea Electronics Technology Institute(KETI), Bucheon, Korea (sweetu@keti.re.kr)

† Corresponding author: Intelligent Robotics Research Center, Korea Electronics Technology Institute(KETI), Bucheon, Korea (ksh1018@keti.re.kr)

영상에 기계학습 알고리즘 등을 통해 얻는 탑승자 점유 유무 검출, 탑승자 인식, 탑승자 자세를 예로 들 수 있다. 더 나아가, 획득한 저수준 및 고수준 정보를 활용하여 탑승자에게 맞는 시트 변형, 공조 조작, DWD (Distraction While Driving) 경고 등의 맞춤형 서비스를 제공할 수 있을 것이라 예상된다.

본 논문에서는 맞춤형 서비스를 위해 차량내 센서 시스템 구성 및 고수준 정보 획득을 위한 센서시스템을 구성한다. 자세히 말하면, 다양한 상황에 강건하도록 RGB-Depth-Thermal 카메라를 구성하고, 영상처리 알고리즘을 적용할 수 있도록 카메라 캘리브레이션을 수행했다.

한편, 최근 깊은 학습을 이용한 객체 검출 알고리즘이 사람 수준보다 더 뛰어남을 증명했다^[3]. 이러한 알고리즘은 대부분 RGB 영상에 대해서 학습을 하고 객체를 검출 및 성능 평가가 이루어 졌는데, 저조도 환경이나 광량 변화가 큰 환경에서 검출 성능을 기대하기는 여전히 어렵다. 따라서, 다양한 환경에서 객체 검출이 가능 하려면 RGB 영상뿐만 아니라 다중 스펙트럼 영상, Thermal 영상 등의 데이터베이스를 구축하고, 학습 및 성능 평가가 이루어져야 한다. 하지만, RGB가 아닌 다중 스펙트럼 영상이나 Thermal 영상은 대역별 특성이 다양하기 때문에 학습 데이터베이스를 구성하기 까다로운 문제점이 있다.

본 논문에서는 새로운 데이터베이스를 수집하고 깊은 학습 모델을 개발하는 것이 아닌, 기존 RGB 영상으로 학습된 모델을 이용하여 Thermal 영상에도 적용하는 방향으로 연구를 진행하였다. 특히, 입력 영상으로부터 다양한 정보(예를 들어 앞서 언급했던 탑승자 검출, 탑승자 인식, 탑승자 자세 정보 등)를 얻을 수 있는 end-to-end 형태의 깊은 학습 알고리즘을 고려했는데, 사람을 검출하고 자세와 얼굴 윤곽 등을 얻을 수 있는 OpenPose 알고리즘^[4]을 활용하는 방향으로 연구를 진행했다.

기존 자세 측정 알고리즘은 사람을 검출한 후 관절이 예측되는 부분을 찾아가는 Top-down 방식^[9]과 관절이 예측되는 부분을 먼저 찾고 각 사람들에 해당하는 관절위치로 재구성하는 Bottom-up 방식^[10]으로 나눌 수 있는데, 과거에는 Top-down 방식에 비해 Bottom-up 방식의 알고리즘 성능이 낮은 편이였지만, 컴퓨터 비전 분야 깊은 학습 연구가 발전하면서 Bottom-up 방식의 알고리즘 성능이 더 앞서게 되었다. 본 논문에서는 자세 측정을 위한 깊은 학습 알고리즘 중에 하나인 Z. Cao 가 제안한 자세 측정 알고리즘^[4]을 활용하여 탑승자 점유 유무 검출하는 알고리즘을 제안한다.

2. 본 론

2.1 RGB-Depth-Thermal 카메라 구성

본 논문에서는 다양한 상황에 강건하도록 [Fig. 1]과 같은

RGB-Depth-Thermal 카메라를 구성했다.

[Fig. 1]의 검은색 원 형태의 카메라(왼쪽)는 FLIR사의 Boson 모델로 Thermal 카메라이며, 진한 파랑 직사각형 형태의 카메라(오른쪽)은 Intel사의 R200 모델로 RGB-Depth 카메라이다. 두 개의 카메라로부터 얻은 RGB-Depth-Thermal 영상들의 정합 또는 좌표 정보 공유를 위해서는 카메라 파라미터를 얻어야 한다^[5]. RGB-Depth의 경우, Intel R200 모델의 SDK를 이용하여 카메라 파라미터를 구한다. 따라서, RGB와 Thermal 또는 Depth와 Thermal 간의 카메라 캘리브레이션을 수행하면 RGB-Depth-Thermal 영상들의 정합이 가능하다.

2.1.1 RGB-Thermal 카메라 캘리브레이션

일반적으로 카메라 캘리브레이션은 미리 정한 크기의 체스 보드를 촬영 후 획득한 영상으로부터 특징점을 추출하고 월드 좌표계와 비교하여 카메라 파라미터를 얻는 것을 의미한다. Thermal 영상은 RGB 영상과 특성이 다르기 때문에 앞서 말한 일반적인 카메라 캘리브레이션을 수행 할 수 없으며 [Fig. 2]와 같은 특수한 카메라 캘리브레이션 보드를 통해 카메라 캘리브레이션이 가능하다.

[Fig. 2]의 (a)는 마스크 기반의 보드^[6]로 기존의 가열된 종 이를 이용하는 방법에 비해 선명한 체스 보드 영상을 얻을 수 있는 장점이 있지만, 마스크의 두께에 따라 왜곡이 발생하는 단점이 있다. [Fig. 2]의 (b)는 구리와 고방사성 잉크 기반의 보



[Fig. 1] RGB-Depth-Thermal camera. The black circular (Left) and deep blue rectangular (Right) camera are the FLIR Boson Thermal and Intel R200 RGB-Depth camera, respectively^[5]



(a) Mask-based board^[6]



(b) Copper and high-emissive ink-based board^[7]



(c) Copper line-based board^[11]



(d) LED-based board^[5]

[Fig. 2] The comparison of RGB-Thermal camera calibration boards^[5]

드^[7]로 열방사가 잘 되도록 주변 온도를 높여야 하며, 적당한 온도를 맞추기 어려운 단점이 있다. (c)는 구리 격자 선 기반의 보드^[1]로, 균일한 열 분포를 갖기 때문에 격자가 선명하게 보인다는 장점이 있지만, 기존 체스 보드 검출 알고리즘을 이용할 수 없다는 단점이 있다. (d)는 본 논문에서 사용한 LED 기반의 보드^[5]이다. 체스 보드 교점에 LED를 장착한 보드로 주변 환경에 강건하며, 기존 체스 보드 검출 알고리즘을 활용할 수 있다는 장점이 있다. 3장에서 [Fig. 2]의 (c)와 (d)의 카메라 캘리브레이션 결과 비교실험을 진행했으며, 본 논문에서 사용한 LED 기반의 보드가 효과적임을 증명했다.

2.1.2 RGB-Thermal 영상 정합

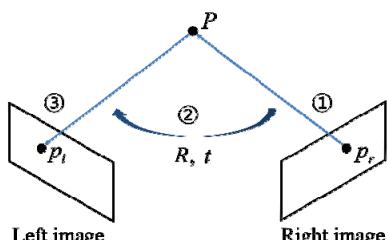
RGB-Thermal 영상을 정합하는 과정은 [Fig. 3]의 ①~③의 순서대로 진행된다^[5]. 먼저 왼쪽 영상을 RGB 영상, 오른쪽 영상을 Thermal 영상이라 했을 때, 오른쪽 영상의 좌표 한 점(p_r)을 월드 좌표계 점(P)으로 투영한다. 월드 좌표계로 변환하는 과정에서 Depth 정보를 이용한다. 그 다음, 카메라 캘리브레이션을 통해 얻은 파라미터와 R, t (Rotation and translation matrix)를 이용하여 월드 좌표계로 투영된 점(P)의 좌표계를 왼쪽 영상의 좌표계로 변환 한 후, 왼쪽 영상의 점(p_l)로 투영하면 정합이 완료된다.

2.2 OpenPose를 이용한 좌석 점유 판단

2.2.1 사람 검출

본 논문에서는 사람 검출을 위해 자세 측정 라이브러리인 OpenPose 라이브러리를 이용했으며, OpenPose 라이브러리 중에서도 2016 COCO Keypoints challenge에서 우승한 Z. Cao가 알고리즘^[4]을 이용하여 사람을 검출한다. 이 알고리즘은 Bottom-up 방식이기 때문에 기학습된 모델을 이용하여 Thermal 영상에도 적용 가능하다는 장점이 있다. 이 알고리즘의 내용은 다음과 같다.

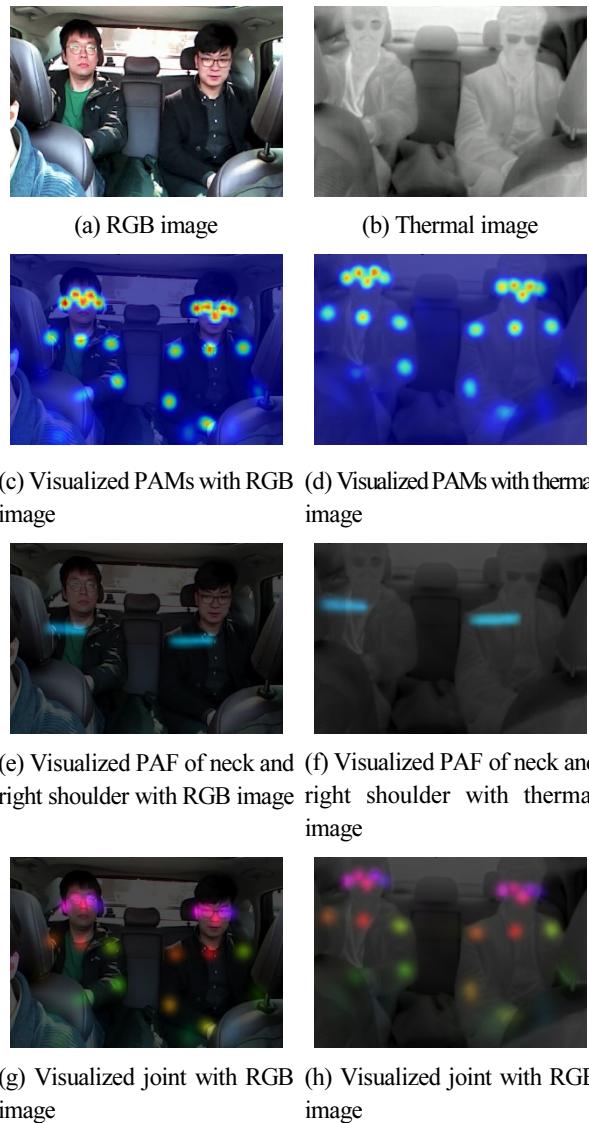
먼저, 입력 영상을 2분기 다중 스테이지 CNN(Convolutional neural network) 모델을 통과시켜 관절이 있을 것 같은 위치를



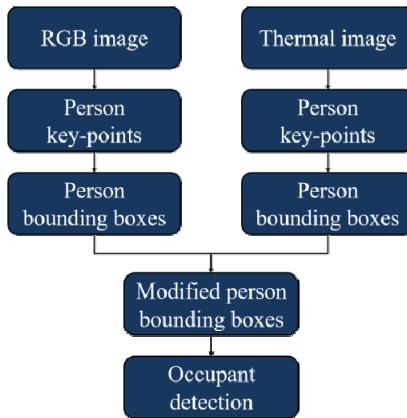
[Fig. 3] The registration of left-right image using depth information, rotation matrix and translation matrix^[5]

표현하는 Part Confidence Maps(PAMs)와 관절과 관절 사이 연결 관계를 표현하는 Part Affinity Fields(PAFs)를 얻는다. 그리고, 관절이 어떤 사람의 것인지 판단하기 위해 NP-Hard 매칭 문제를 Greedy 방법으로 풀어 빠르면서도 효과적으로 해결한다^[4]. [Fig. 4]은 RGB 및 Thermal 입력 영상에 대한 PAMs, PAFs, 추정된 관절의 시각화를 보인다.

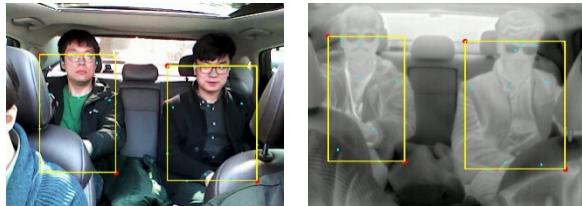
[Fig. 5]은 제안하는 탑승자 점유 유무 검출 알고리즘의 블록 다이어그램을 보인다. 먼저 RGB, Thermal 각 영상에 Z. Cao가 제안한 알고리즘^[4]을 적용하여 사람 관절 포인트들(Person key-points)을 얻는다. 그 다음, 각 포인트들의 (x, y) 좌표의 최대값과 최소값을 구하여 사람 바운딩 박스(Person bounding boxes)를 구한다. [Fig. 6]의 (a)와 (b)에서 노란색 박스는 RGB



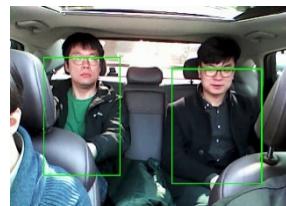
[Fig. 4] The visualization of PAMs, PAF (case of neck and right shoulder), and estimated joint



[Fig. 5] Block diagram of occupant detection algorithm



(a) Person detection in RGB image (b) Person detection in thermal image

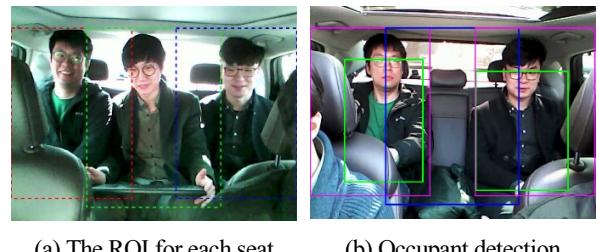
[Fig. 6] The visualization of person detection. (Cyan dots: person key-points, Red dots: maximum or minimum points of (x, y) at person key-points, Yellow boxes: person bounding boxes, Green boxes: modified person bounding boxes)

와 Thermal 영상에서 사람 검출 결과를 의미한다.

마지막으로, 앞서 구한 사람 바운딩 박스의 위치 정보는 2.1.2 절에서 언급한 정합 방법으로 바운딩 박스 위치 정보를 정합 시킨 후, 비최대 억제(Non-Maximum Suppression, NMS)^[8] 알고리즘을 통해 수정된 사람 바운딩 박스(Modified person bounding boxes)를 구하여 사람 검출을 완료한다. [Fig. 6]의 (c)는 수정된 사람 바운딩 박스를 RGB 영상에 표시한 결과를 보인다.

2.2.2 좌석 점유 판단

RGB-Depth-Thermal 카메라는 고정되어 있다고 가정하고, [Fig. 7]의 (a)와 같은 좌석 별 관심영역(Region of interest, ROI)



[Fig. 7] The ROI for each seat and the results of occupant detection (Red, green, and blue dashed lines: the ROI of each seat, Blue boxes: non-occupied seat, Magenta boxes: occupied seats)

$$\text{IoROI} = \frac{\text{Area of Overlap}}{\text{Area of ROI}}$$

[Fig. 8] Intersection over ROI

을 설정한다. 좌석에 앉은 사람의 행동 반경을 고려하여 사용자가 설정한다.

그 다음 ROI 영역과 수정된 사람 바운딩 박스 영역간의 오버랩 영역 비율(Intersection over ROI, IoROI)을 구하여 점유 유무를 판단한다. IoROI는 [Fig. 8]과 같이 구하며, IoROI 값이 T 를 초과하면 점유 했다고 판단한다.

3. 실험

3.1 RGB-Thermal 카메라 캘리브레이션 실험 결과

본 논문에서는 주변 환경에 강건하며 기존 채스보드 검출 알고리즘을 이용 가능한 [Fig. 2]의 (d)인 LED기반의 보드^[5]를 사용하였다. [Table 1]은 [Fig. 2]의 (c)인 구리선 기반의 보드^[1]와 LED기반의 보드^[5]에 따른 카메라 캘리브레이션 후 평균 재투영 에러(Mean reprojection error)를 보인다. 평균 재투영 에러는 측정한 카메라 파라미터를 이용하여 월드 좌표계 포인트를 카메라 좌표계로 투영 시킨 위치의 좌표와 입력 영상에서 실제로 관측된 위치의 좌표의 평균 에러를 의미한다. 본 논문에서 사용한 방법의 평균 재투영 에러가 낮게 나온 것을 확인했다.

3.2 실험 데이터베이스 수집 및 환경 별 분류

[Fig. 9]의 왼쪽 그림은 차량의 백미러에 장착된 RGB-Depth-Thermal 카메라, 오른쪽 그림은 RGB-Depth-Thermal 카메라를 이용하여 중형 차량 내부 촬영 데이터베이스를 얻는

[Table 1] Comparison of mean reprojection error between Copper line-based board and LED-based board^[5]

Method	RGB images	Thermal images
Copper line-based board ([Fig. 2] (c))	0.73 pixel	0.78 pixel
LED-based board ([Fig. 2] (d))	0.41 pixel	0.52 pixel



[Fig. 9] RGB-Depth-Thermal camera mounted on rearview mirror (Left) and collecting RGB-Depth-Thermal image database in vehicle (Right)

[Table 2] RGB-Depth-Thermal image database in vehicle with various environment

Name No.	Example of images				Number of frames	
		RGB	Depth	Thermal		
1	1_1_Day_Normal				2900	
	1_2_Day_FrontSideBackLight				1550	
	1_3_Day_TunnelEntrance				1300	
	1_4_Day_Tunnel				2700	
2	2_1_Rain_Normal				2000	
	2_2_Rain_TunnelEntrance				450	
	2_3_Rain_Tunnel				250	
3	Night	3_1_Night_Normal				1800

모습을 보인다.

본 논문에서는 데이터베이스 수집 후 [Table 2]와 같이 다양한 상황(낮, 비, 밤, 측후광, 터널 입출구 등)의 환경별로 분류를 했으며, 사용하는 RGB-Depth-Thermal 카메라의 화각이 좁은 관계로 뒷좌석에 대해서만 실험을 진행했다. 그리고 점유 유무 검출 성능 평가를 위한 점유 유무 실측 자료를 제작했다.

3.3 좌석 점유 유무 검출 성능 평가

실험은 Intel i5 dual core 2.2GHz, 16G RAM, Windows 10 환경에서 진행 되었으며, IoROI의 파라미터 T 값은 실험값으로 0.2로 설정했다. 식 (1)을 이용하여 각 좌석 별([Fig. 7]의 (a)에서 왼쪽부터 왼쪽, 가운데, 오른쪽)점유 유무 검출률을 계산했다.

$$\text{Detection rate} = \frac{\text{Successed frames}}{\text{Number of frames}} \quad (1)$$

[Table 3] The results of occupant detection rate for each seat

No.	Name	Left seat	Center seat	Right seat
1	1_1_Day_Normal	0.978	0.996	0.970
	1_2_Day_FrontSideBackLight	0.875	1.000	0.991
	1_3_Day_TunnelEntrance	0.978	0.874	0.948
	1_4_Day_Tunnel	0.993	0.915	0.993
2	2_1_Rain_Normal	0.957	0.978	0.984
	2_2_Rain_TunnelEntrance	1.000	0.993	1.000
	2_3_Rain_Tunnel	1.000	0.960	1.000
3	3_1_Night_Normal	0.747	0.999	0.995
Average detection rate		0.951	0.958	0.983

[Table 3]은 좌석별 점유 유무 검출률의 결과를 보인다. 오른쪽 좌석에 비해 왼쪽 좌석의 검출률이 낮게 나오는 경향이 있는데, 이는 RGB-Depth-Thermal 카메라가 한쪽으로 치우쳐 진 채로 촬영되어 폐색(Occlusion)되는 상황이 상대적으로 높게 나왔기 때문이다. 3_1_Night_Normal 데이터베이스의 왼쪽 좌석의 검출률이 다른 좌석에 비해 낮게 나온 이유는, 가운데 사람과 폐색이 발생하면서 검출 성능을 떨어뜨린 것으로 분석된다. 또한, 적은 수의 데이터베이스에 대해서만 실험을 진행한 점, 남성 성인에 대해서만 실험을 진행한 점으로 인해 전체적으로 검출률이 높게 나온 이유로 분석 된다.

4. 결 론

본 논문에서는 차량 탑승자 정보 획득을 위한 센서 시스템 구성과 탑승자 점유 유무 검출 알고리즘을 제안한다. 센서 시스템의 경우, 다양한 상황에 강건하도록 다중 카메라 시스템인 RGB-Depth-Thermal 카메라 센서 시스템을 구성하고, RGB-Depth-Thermal 영상 좌표 정보 공유 및 정합을 위해 캘리브레이션을 수행했다. RGB-Depth 카메라 캘리브레이션은 SDK를 이용하여 수행했으며, RGB-Thermal 카메라 캘리브레이션은 RGB-Thermal 영상의 특성이 고려된 LED 기반의 캘리브레이션 보드를 이용했으며, 기존 방법에 비해 효과적임을 증명했다. 그리고 Depth 정보와 카메라 캘리브레이션 결과를 이용하여 RGB-Depth-Thermal 영상의 좌표 정보 공유 및 정합을 수행했다. 탑승자 점유 검출을 위해서는 깊은 학습 알고리즘 중 하나인 자세 측정 알고리즘을 이용했다. 이 알고리즘은 Bottom-up 방식이기 때문에 기학습된 모델을 이용해도 RGB

영상뿐만 아니라 Thermal 영상에도 적용이 가능한 점이 장점이며 실험 결과를 통해 그 성능을 증명하였다. 향후 탑승자 자세 검출, 얼굴 인식 등의 정보를 획득하고 맞춤형 능동 편의 서비스를 제공할 수 있도록 알고리즘을 보완할 예정이다.

References

- [1] N. Kim, Y Chio, S Hwang, K Park, J. S. Yoon, and I. S. Kweon, “Geometrical calibration of multispectral calibration,” *2015 12th International Conference on Ubiquitous Robots and Ambient Intelligence*, Goyang, South Korea, pp. 384-385, 2015.
- [2] C. Palmero, A. Clapés, C. Bahnsen, A. Møgelmose, T. B. Moeslund, and S. Escalera, “Multi-modal RGB-depth-thermal human body segmentation,” *International Journal of Computer Vision*, vol.118, no. 2, pp. 217-239, June, 2016.
- [3] K. He, G. Gkioxari, P. Dollár, and R. Girshick, “Mask R-CNN,” *2017 IEEE International Conference on Computer Vision*, Venice, Italy, pp. 2980-2988, 2017.
- [4] Z. Cao, T. Simon, S. Wei and Y. Sheikh, “Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields,” *2017 IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, HI, USA, 2017, doi: 10.1109/CVPR.2017.143.
- [5] C. Song and S. Kim, “A study of an effective camera calibration and image registration method for RGB-Depth-Thermal image fusion,” *KSPS Autumn Conference*, Gwangju, Korea, pp. 236-237, 2017.
- [6] S. Vidas, R. Lakemond, S. Denman, C. Fookes, S. Sridharan, and T. Wakr, “A mask-based approach for the geometric calibration of thermal-infrared cameras,” *IEEE Transactions on Instrumentation and Measurement*, vol. 61, no. 6, pp. 1625-1635, 2012.
- [7] W. Usrine, F. Calado, G. Teixeira, H. Diniz, S. Silvino, and R. de Andrade, “Thermal/visible autonomous stereo visio system calibration methodology for non-controlled environments,” *11th International Conference on Quantitative InfraRed Thermography*, Naples, Italy, 2012, doi: 10.2161/qirt.2012.261.
- [8] R. Girshick, F. Iandola, T. Darrell, and J. Malik, “Deformable part models are convolutional neural networks,” *2015 IEEE Conference on Computer Vision and Pattern Recognition*, Boston, MA, USA, pp. 437-446, 2015.
- [9] A. Newell, K. Yang, and J. Deng, “Stacked hourglass networks for human pose estimation,” *European Conference on Computer Vision*, Amsterdam, The Netherlands, pp. 483-499, 2016.
- [10] L. Pishchulin, E. Insafutdinov, S. Tang, B. Andres, M. Andriluka, P. Gehler, and B. Schiele, “Deepcut: Joint subset partition and labeling for multi person pose estimation,” *2016 IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, NV, USA, 2016, doi: 10.1109/CVPR.2016.533.



송 창 호

2015 한국교통대학교 전자공학과 (공학사)
2017 중앙대학교 영상학과 (공학석사)
2017 ~ 현재 전자부품연구원지능로보틱스
연구센터

관심분야: Robot vision, Computer vision, Deep learning



김 승 훈

2002 한양대학교 전자컴퓨터공학과 (공학사)
2007 연세대학교 전기전자공학과(공학석사)
2007~2008 현대중공업 지능로보틱스연구실
2009 ~ 현재 전자부품연구원 지능로보틱스
연구센터

관심분야: Robot vision, Computer vision, SLAM, Sensor fusion