

SPAD과 CNN의 특성을 반영한 ToF 센서와 스테레오 카메라 융합 시스템

Fusion System of Time-of-Flight Sensor and Stereo Cameras Considering Single Photon Avalanche Diode and Convolutional Neural Network

김 동 엽[†] · 이 재 민¹ · 전 세 웅²

Dong Yeop Kim[†], Jae Min Lee¹, Sewoong Jun²

Abstract: 3D depth perception has played an important role in robotics, and many sensory methods have also proposed for it. As a photodetector for 3D sensing, single photon avalanche diode (SPAD) is suggested due to sensitivity and accuracy. We have researched for applying a SPAD chip in our fusion system of time-of-flight (ToF) sensor and stereo camera. Our goal is to upsample of SPAD resolution using RGB stereo camera. Currently, we have 64 x 32 resolution SPAD ToF Sensor, even though there are higher resolution depth sensors such as Kinect V2 and Cube-Eye. This may be a weak point of our system, however we exploit this gap using a transition of idea. A convolution neural network (CNN) is designed to upsample our low resolution depth map using the data of the higher resolution depth as label data. Then, the upsampled depth data using CNN and stereo camera depth data are fused using semi-global matching (SGM) algorithm. We proposed simplified fusion method created for the embedded system.

Keywords: Sensor Fusion, Time-of-Flight, Stereo camera, Single Photon Avalanche Diodes, Convolution Neural Network

1. 서 론

로봇이 3차원 공간지능을 갖기 시작하면서 보다 많은 서비스들을 제공할 수 있게 되었다. 드론이 자동으로 이착륙하고 목표점까지 비행^[1]을 하고, 이송로봇이 물건을 전달^[2]하는 데에 이러한 공간지능이 필수적이다. 스마트 공장에서 쓰이는

협업로봇이 3차원 공간정보를 활용하여 펜스를 제거하려는 연구가 진행되고 있고, 최근에는 인공피부 수준의 비접촉식 충돌감지도 개발되고 있다^[3]. 자율주행자동차는 주행을 위한 외부 3차원 환경정보^[4] 뿐만 아니라, 탑승자 모니터링을 위한 내부 3차원 환경정보^[5]도 활용되고 있다.

이러한 3차원 공간지능을 위한 센서로서, 반송시간을 측정하는 Time-of-Flight (ToF) 센서, 카메라의 양안시차(disparity)를 활용하는 스테레오 카메라, 구조광 (structured light), 적외선 패턴을 투광하여 패턴의 왜곡을 감지하는 라이트 코딩 (light coding) 등이 있다^[6]. 최근에는 센서칩 설계 기술과 광학계 기술의 발전으로 ToF 방식의 3D 센서가 널리 보급되고 있다. 스파드(Single Photon Avalanche Diode, SPAD)는 ToF 센서칩의 한 종류로서, 민감도(sensitivity)와 정확도(accuracy)가 높은 광검출기(photodetector)이다^[6]. 본 논문에서는 SPAD 센서칩을 적용하여, 고가의 외부 광학필터 없이 운용이 가능하면서 20,000 lux 이상의 야외에서 감지가 가능한 3D 센서를 개

Received : Sep. 6. 2018; Revised : Oct. 29. 2018; Accepted : Nov. 14. 2018

※ This project was funded by the Ministry of Trade, Industry & Energy (Project No. 10050534, Development of low-cost wide-FoV 3D environmental modeling sensor for indoor and outdoor robot application based on the fusion of ToF and stereo-vision sensor)

1. Researcher, Intelligent Robotics Research Center, Korea Electronics Technology Institute (KETI), Bucheon-si, Republic of Korea (qhrhtr2000@keti.re.kr)

2. Managerial Researcher, Intelligent Robotics Research Center, Korea Electronics Technology Institute (KETI), Bucheon-si, Republic of Korea (daniel@keti.re.kr)

† Senior Researcher, Corresponding author: Intelligent Robotics Research Center, Korea Electronics Technology Institute (KETI), Bucheon-si, Republic of Korea (sword32@keti.re.kr)

발하는 것을 목표로 하였다. 특히, 로봇 적용을 위한 저가형 3차원 실내의 공간 인식 센서를 목표로 SPAD의 성능을 활용하였다.

현재까지 연구로 확보된 SPAD 센서칩의 해상도는 64×32 로서, Kinect V2 (512×424)나 Cube-Eye (320×240) 대비 상대적으로 작다. 본 논문에서는 이를 극복하기 위하여 컨벌루션 뉴럴 네트워크(Convolution Neural Network, CNN)를 활용하여 고해상도 ToF 센서 수준의 해상도로 업샘플링(upsampling)하는 네트워크를 적용하였다. 이와 같이 센서의 업샘플링에 딥러닝(deep learning)을 적용하는 접근은 로봇용 3D 센서 뿐만 아니라 열영상 센서의 해상도 증대⁷⁾, MRI 의료영상^{8,9)} 등 다양하게 연구되고 있다. 센서의 해상도가 적을수록 센서의 가격 및 생산비용, 사용비용이 감소하기 때문에, 이를 후처리를 통해서 해상도를 증대하려는 연구는 오랫동안 계속되어 최근 딥러닝 기술에서도 널리 연구되고 있다⁷⁻⁹⁾.

하지만 딥러닝 방법을 활용한 단일 센서의 업샘플링의 경우에는 학습 데이터(training dataset)의 해상도를 넘을 수 없다는 구조적인 한계가 있다. 이 문제를 해결하기 위해서 오랫동안 이종센서(heterogeneous sensor)를 융합하는 알고리즘들이 제안되어왔다. [10]에서는 구속 최적화기법(constrained optimization)을 활용하여 176×144 해상도의 ToF 센서(MESA SR4000)를 1280×960 RGB카메라(Point Grey社. Flea)와 융합하였다. [11]에서는 200×200 해상도의 ToF 센서(PMD社. CamCube 2.0)를 두 개의 1312×1082 해상도의 카메라를 사용하여 1082×1082 해상도로 업샘플링했다.

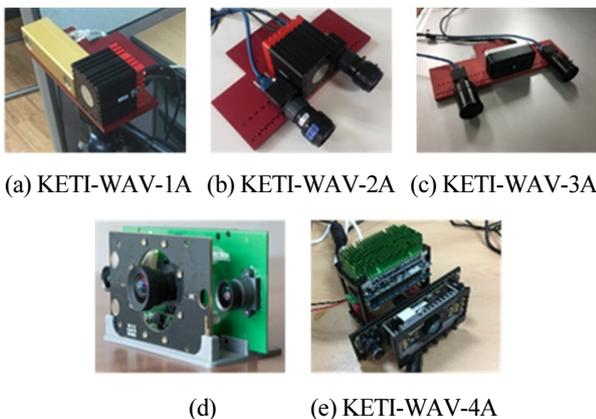
본 논문에서는 [Fig. 1]과 같이 [11]에서 제안하는 ToF 센서와 스테레오 카메라를 융합하는 방법론을 따랐다. [11]에서는 기존 스테레오 카메라 알고리즘인 SSD cost를 사용하는 블록 매칭(block-matching) 알고리즘을 바탕으로, 각 센서의 특성을 반영하여 ToF 센서와 카메라를 융합하는 방법을 사용하였다. 융합을 위한 구체적인 방법으로, ToF의 IR 강도(intensity,

C^{TOF}), ToF의 depth gradient ($C^{\nabla TOFD}$), 레퍼런스 카메라의 horizontal gradient ($C^{\nabla xL}$), 두 센서의 겹침여부(occlusion, C^{LOc})로 센서 충실도(fidelity)를 계산한 후, 스테레오 매칭을 위한 검색영역(joint per pixel search range map)을 추정한다¹¹⁾.

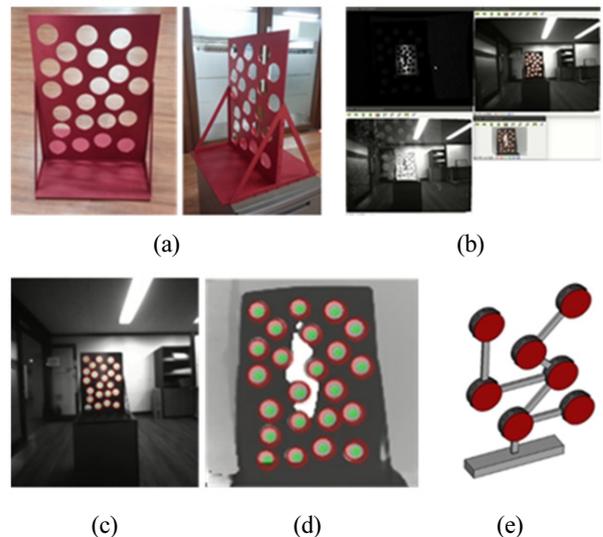
최근에는 이와 같은 ToF 센서 - 스테레오 카메라 융합에 딥러닝 방법을 적용하는 연구가 활발하게 진행되고 있다. 본 논문에서는 CNN을 ToF 업샘플링에 직접적으로 활용하는 것과 달리, [12]는 CNN을 활용하여 ToF 센서와 스테레오 카메라의 신뢰도 지도(Confidence map)를 추론(inference)한다. 이 신뢰도(confidence)를 이웃 픽셀로부터의 유사성(plausibility)를 융합하는 가중치(weight)로 사용하였다.

센서융합의 선결조건인 ToF 센서와 스테레오 카메라를 캘리브레이션함에 있어서, 전체 장비의 가격과 소요시간에 대해서 많은 고려를 하였다. 왜냐하면, 본 논문에서 제안하는 센서 모듈이 실제 생산에 들어갔을 때의 생산단가에 직결되는 문제이기 때문이다. 이러한 관점에서 첫 번째로 시도한 것은 [10]에서 사용한 캘리브레이션 방법이다.

[Fig. 2]의 (a)는 제작한 캘리브레이션 지그이고, [Fig. 2]의 (b)는 데이터를 취득하는 GUI 화면이다. [Fig. 2]의 (c)와 (d)에서 보듯이 카메라와 ToF 깊이(depth) 이미지에서 허프 원 알고리즘(Hough Circle algorithm)으로 캘리브레이션 지그에 있는 원형 구멍의 중심을 측정했다. 이 원형구멍의 중심을 특징점으로 해서 [13]에서 제안하는 방법으로 캘리브레이션을 수행했다. ToF 센서 특성 상 물체의 외곽선에서 발산 픽셀(flying pixel)이 발생해서 에러에 취약하기 때문에¹¹⁾, Hough Circle 알고리즘의 투표하는 특성(voting)으로 이 에러의 영향을 축소하였다¹⁴⁾.



[Fig. 1] Fusion module of ToF sensor and stereo camera



[Fig. 2] (a) Calibration jig (b) Calibration GUI (c) Reference camera image (d) ToF depth image (e) Calibration jig suggested in¹⁵⁾

문제는 ToF 센서의 낮은 해상도 때문에, [Fig. 2]의 (d)와 같이 ToF depth 이미지에서 원이 찾아지는 영역이 적다. 다시 말하자면, 제작된 캘리브레이션 지그의 원형 구멍 크기가 유효한 센서모듈과 캘리브레이션 지그 사이의 거리가 제한된다는 것이다. 센서 감지거리가 길수록 다양한 크기의 캘리브레이션 지그가 필요하고, 낮은 해상도의 ToF 센서 때문에 원형구멍의 크기도 커져야 한다. [Fig. 2]의 (e)는 이러한 한계를 극복하기 위하여 [Fig. 2]의 (a)의 음각이 아닌 양각방식의 캘리브레이션 지그를 설계한 것이다^{15,16}.

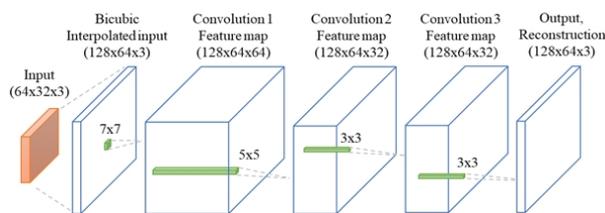
본 논문에서는 ToF 센서와 스테레오 카메라를 융합하기 위한 알고리즘을 제안한다. 2절에서는 SPAD의 특성을 활용하기 위한 CNN기반의 업샘플링에 대해서 설명하고, 3절에서는 이중센서 융합을 위한 알고리즘을 설명한다. 4절에서는 실험을 통하여 본 논문에서 제안하는 알고리즘의 적합성을 보인다.

2. CNN을 사용한 ToF 해상도 증대

1절에서 언급했듯이, 우리는 기존 Depth 센서보다 광특성이 훌륭한 SPAD를 활용한 로봇용 저가 3차원 센서를 개발하는 것을 목표로 하고 있다. 문제는 현재 확보한 SPAD의 현재 해상도가 64×32 라는 점이다. 이는 최근 시판(off the shelf)되고 있는 Depth 센서들의 해상도 보다 적은 편이다.

이를 해결하기 위하여, 본 논문에서는 상기 한계를 역으로 이용하는 접근을 제안하였다. 고해상도의 영상을 CNN으로 업샘플링하기 위해서 고해상도의 영상을 레이블(label)로서 학습하는데^{7,9}, SPAD를 업샘플링하기 위해서 필요한 고해상도 영상을 기존 Depth 센서의 데이터를 활용하는 것이다. 상용 320×240 의 해상도의 Depth 센서인 미래컴퍼니社의 MDC600SB를¹⁷ SPAD의 해상도에 맞춰서 서브샘플링(subsampling)하여 학습 데이터(training data)의 저해상도 영상을 생성했다.

본 논문에서 제안하는 CNN은 [Fig. 3]에서 표현된 것과 같이 4개의 Convolution layer로 이루어져있다. [7]에서와 같이 SPAD의 출력데이터를 CNN의 성능이 제일 좋은 배율(scale factor) $\times 2$ 를 적용하여, 64×32 의 두 배인 128×64 로 업샘플링하였다. 일반적인 CNN과 달리, [Fig. 3]의 네트워크는 업샘

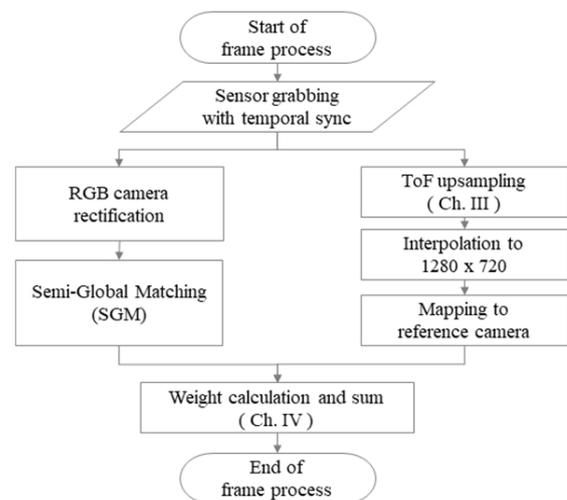


[Fig. 3] Convolution Neural Network for upsampling of Single Photon Avalanche Diodes data

플링이 목적이므로 특징점(feature)의 위치(shift)를 보상하기 위한 풀링 레이어(pooling layer)는 생략되었다. ¹⁷과 같이, 본 논문에서도 저가형 로봇용 센서모듈을 목표로 하기 때문에 컨벌루션 레이어(Convolution layer)의 개수를 제한하여 계산 부담이 적게 설계하였다. Convolution layer 1에서는 $7 \times 7 \times 3$ 크기의 filter를 64개 적용하였다. Convolution layer 2에서는 $5 \times 5 \times 64$ 크기, 그리고 Convolution layer 3과 아웃풋 레이어(output layer)에서는 $3 \times 3 \times 32$ 크기를 적용하였다. 이렇게 설정함으로써, 패딩(padding)의 크기를 레이어(layer)를 지날 때마다 3에서부터 1까지 줄일 수 있었다¹⁷.

3. ToF 센서 - 스테레오 카메라 융합

본 논문에서 제안하는 센서융합 방법을 설명하기에 앞서서 각 센서의 감지범위에 대해서 언급을 하겠다. 첫째, 스테레오 카메라는 [Fig. 1]의 (e)에서 볼 수 있듯이 왼쪽 카메라와 오른쪽 카메라가 있는 공간적 스테레오 기법(spatial stereo)을 적용하였다. 이 방법의 장점은 스테레오 카메라의 disparity의 거리 단위(meter)로 변환이 용이하고, 시간적 스테레오 기법(temporal stereo, 또는 moving stereo) 대비 센서모듈의 위치가 변하지 않아도 depth 정보를 취득할 수 있다는 점이다. 반면에, 양 카메라의 시야가 겹쳐지지 않는 영역에서는 거리가 측정되지 않는다. [Fig. 1]의 (e)의 구성에서는 센서모듈의 바로 앞부터, 약 1 m까지의 거리가 이러한 음영영역인 센서 최소치(z_{ToF}^{min} , range min)에 해당한다. 둘째, ToF 센서는 발광부가 점멸하는 속도에 의해서 결정되는 최대감지 범위의 제한이 있다. Light coding 방식의 경우는 IR 패턴을 인식하는 센서의 해상도에 의해 최대감지 범위의 제한이 있다. 본 논문에서는 6 m의 거리를 ToF 센서의 최대치(z_{ToF}^{max} , range max)로 설정하였다.



[Fig. 4] Flow chart for fusion of ToF sensor and stereo camera

이를 바탕으로, 본 논문에서 제안하는 ToF 센서와 스테레오 카메라를 융합하는 방법이 [Fig. 4]이다. 첫째, 스테레오 카메라는 Semi-Global Matching (SGM) 알고리즘을 활용하여 1280×720 의 depth map으로 변환한다^[18]. 둘째, ToF 센서는 64×32 의 SPAD 데이터를 2절에서 제안하는 CNN 기반의 방법으로 128×64 로 upsampling을 한 후에 1280×720 으로 바이큐빅 보간법(Bicubic interpolation)을 수행한다. 두 센서 데이터를 하나의 좌표계로 합치기 위해 1절에서 제안하는 방법으로 얻어진 이중센서간 캘리브레이션을 적용하여 ToF 데이터를 기준 카메라(Reference camera)인 왼쪽 카메라 좌표계로 변환한다. 이 단계에서, 스테레오 카메라와 ToF 센서 데이터 모두 동일한 방향을 가리키는 각 픽셀에 depth 값이 저장된 depth map 형태로 되어있다.

이 두 depth를 하나의 값으로 합치기 위하여 아래 식과 같은 형태로 가중치 합계(weight sum)을 각 픽셀별로 수행한다.

$$z_{output}(i, j) = \alpha z_{str}(i, j) + \beta z_{ToF}(i, j) \quad (1)$$

$z_{output}(i, j)$ 는 i 행, j 열의 depth 최종결과이다. $z_{str}(i, j)$ 는 SGM의 결과로 기준 카메라 좌표계(reference camera coordiante)로 되어있는 depth 값이다. $z_{ToF}(i, j)$ 도 마찬가지로 기준 카메라 좌표계로 되어있는 ToF depth 데이터이다. α 와 β 는 각각의 depth에 해당하는 가중치(weight)로 다음과 같은 알고리즘으로 결정된다.

[Algorithm 1]의 라인 1은 해당 픽셀의 depth가 ToF가 유효한 범위를 초과(z_{ToF}^{max} , range max)하였기 때문에 스테레오 카메라 depth를 활용하는 경우를 의미하며, 반대로 라인 4는 너무

[Algorithm 1] Weight calculation for fusion of ToF sensor and stereo camera

Algorithm 1 Weight Calculation

```

1: if  $z_{ToF}(i, j) > z_{ToF}^{max}$  then
2:    $\alpha = 1$  // use only stereo camera depth
3:    $\beta = 0$ 
4: else if  $z_{ToF}(i, j) < z_{ToF}^{min}$  then
5:    $\alpha = 0$ 
6:    $\beta = 1$  // use only ToF depth
7: else if  $z_{ToF}(i, j)$  is not available then
8:    $\alpha = 1$  // use only stereo camera depth
9:    $\beta = 0$ 
10: else
11:    $\alpha = 0.15$ 
12:    $\beta = 0.85$ 
10: end if

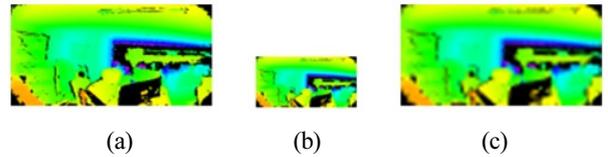
```

가까워서 스테레오 카메라의 데이터를 사용할 수 없는 경우이다. 라인 7은 ToF의 가려짐 영역이나 depth의 경계에서 ToF flying 픽셀이 발생하는 경우를 가리킨다^[11]. 라인 11과 12에서 ToF 센서와 RGB 스테레오 카메라가 정상적인 상황일 때 융합을 하는 비율을 고정하고 있다. 많은 연구들에서 해당 비율을 상황에 맞게 (adaptive) 조정하는 방법들을 제시하고 있으나, 본 논문에서는 본 센서모듈을 로봇에 적용한다는 점과 CNN에 소요되는 시간을 고려하여 이를 적용하지 않는 것으로 결정하였다.

4. 실험

4.1 CNN을 사용한 ToF 해상도 증대

2절에서 제안한 대로, SPAD에서 얻어진 64×32 의 입력영상을 업샘플링하기 위해 레퍼런스 ToF 센서인 미래컴퍼니사 MDC600SB를^[17] 활용하여, 1,600장의 320×240 고해상도 ToF 센서 데이터를 취득하였다. 이를 1,400장의 학습 데이터셋(training set)과 200장의 test set으로 분리하였다. 학습 데이터셋의 양을 늘리기 위하여(data enrich) 좌우반전(horizontal flip)을 하여 2,800장의 데이터를 생성하였다. 본 논문에서는 모바일 로봇의 3차원 센서를 목표로 하기 때문에, 상하반전(vertical flip) 데이터, 즉 depth map에서 바닥이 영상의 윗면에 있는 영상은 무의미하다고 판단하여 배제하였다. [Fig. 5]는 CNN 학습을 위해 준비한 테스트 데이터셋(test set) 이미지 중 하나이다. [Fig. 5]의 (a)는 CNN 학습에서 레이블(label)로 사용되는 128×64 이미지이다. [Fig. 5]의 (b)는 SPAD와 동일한 해상도인 64×32 이미지이고, 이를 bicubic interpolation한 이미지가 [Fig. 5]의 (c)이다.

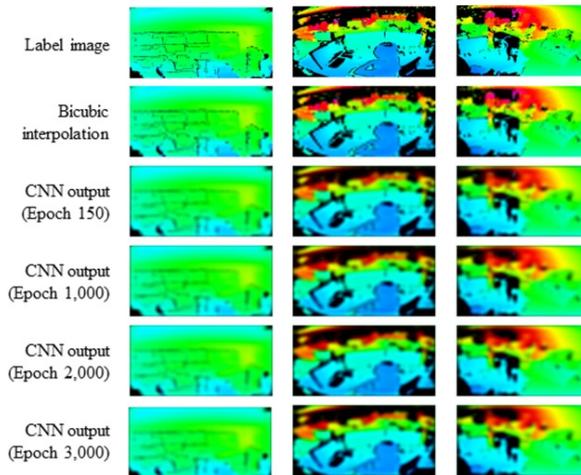


[Fig. 5] (a) label image (128×64). (b) 64×32 image. (c) CNN input image upsampled with bicubic interpolation (128×64)

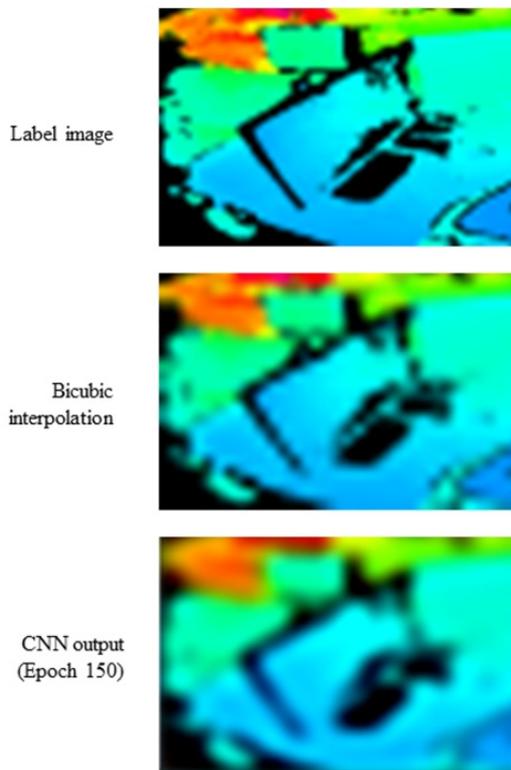
[Table 1] Evaluation result of PSNR (dB)

	Bicubic	Epoch 150	Epoch 1,000	Epoch 2,000	Epoch 3,000
Min.	68.0073	64.8614	64.8877	64.8790	64.8601
Med.	70.1132	66.3620	66.4878	66.5230	66.4929
Avg.	70.4750	66.9537	67.0514	67.0728	67.0446
Max.	75.0917	71.3673	71.6846	71.8192	71.7972

Anaconda 5.2 for Windows, Python 3.6, CUDA 9.0, TensorFlow 1.10, NVIDIA GTX1080 (dual SLI) 환경에서 2절에서 제안한 CNN을 구현하였다. 학습(Training)을 위한 파라미터(parameter)로, 러닝 레이트(learning rate)는 0.001을 사용하였고 배치 크기(batch size)는 128이었다. 위에서 취득한 2,800장의 학습 데이터셋(training set)으로 150 epoch, 1,000 epoch, 2,000 epoch 3,000 epoch을 수행한 결과가 [Table 1]이다. 성능비교를 위하여 [7]과 동일하게 PSNR을 측정하였다.



[Fig. 6] CNN output of test set with epoch 150, 1000, 2000, and 3000



[Fig. 7] A comparison of plausibility with neighbor depth pixel.

[Fig. 6]는 테스트 데이터셋(Test set) 결과를 일부 발췌한 결과이다. CNN 전후의 PSNR의 개선 정도는 평균 3.5 dB 내외로 [7]과 동등한 수준의 성능을 확인하였다. 학습 이폭(Training epoch)이 늘어난다고 해서 성능의 개선은 미미하므로 150 epoch이면 충분한 성능을 얻을 수 있었다.

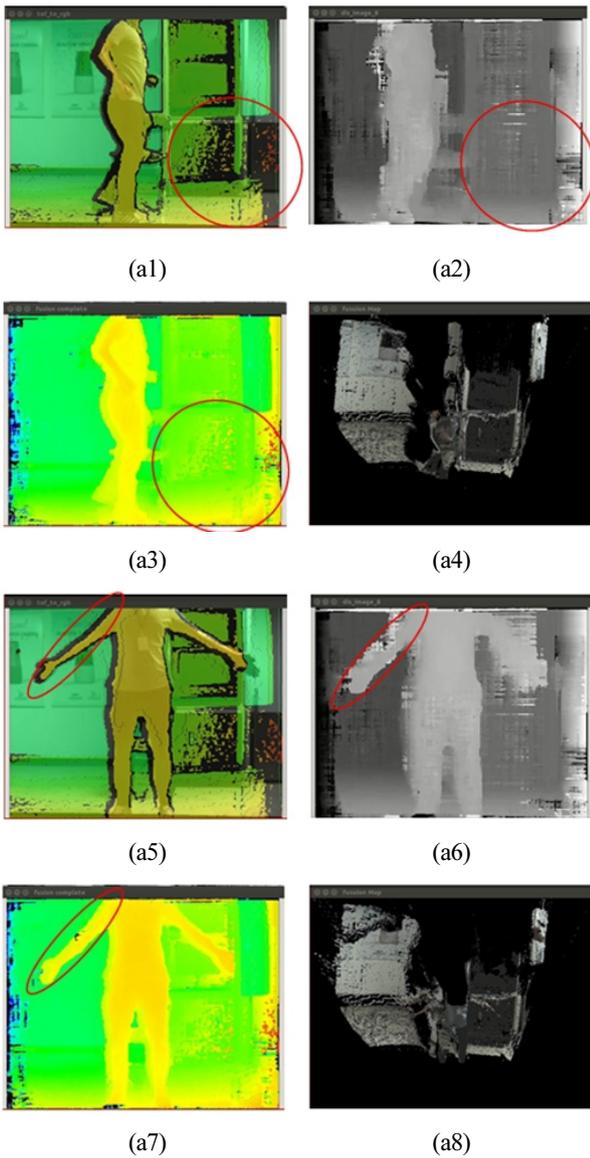
[Fig. 7]은 CNN을 활용한 업샘플링 결과와 bicubic interpolation 결과를 비교하기 위해 [Fig. 6]의 두 번째 열 영상에서 모니터 부분만 확대한 것이다. 레이블 영상(label image)는 320×240 을 128×64 로 축소된 원본 영상이고, 다른 두 영상은 64×32 를 128×64 로 확대한 영상이다. Bicubic interpolation의 경우 영상의 선명도가 상대적으로 높으나 이웃 depth 픽셀과의 유사성이 훼손되어 있음을 알 수 있다. 모니터의 왼쪽 가장자리를 보면, 물결이 치는 듯한 노이즈가 있음을 볼 수 있는데, 이것을 depth로 변환하면 모니터 평판이 평평하지 않는 것처럼 표현되는 노이즈인 것이다. 반면에 CNN 결과는 영상의 선명도가 상대적으로 낮지만 이웃 depth 픽셀과의 유사성이 높기 때문에 depth 센서의 데이터로서 더 적합하다. 이와 관련하여, 단순히 Bicubic interpolation을 10배하는 것보다, 2배의 CNN upsampling과 5배의 Bicubic interpolation을 하는 것이 더 유리하는 결론을 얻었다.

4.2 ToF 센서 - 스테레오 카메라 융합

[Fig. 8]은 본 논문에서 제안하는 융합 방법으로 얻어진 결과 영상이다. [Fig. 8]의 (a1)과 (b1)은 1절에서 제안한 방식으로 캘리브레이션을 수행하여 ToF 데이터를 흑백의 RGB 영상에 overlay한 것이다. 이를 통하여 ToF 데이터 캘리브레이션이 유효함을 확인할 수 있다. 참고로, (a1)과 (b1)은 융합 결과가 아니라 ToF 데이터를 출력한 영상이기 때문에, ToF의 단점인 빈 영역(hole)이 두드러진다. 이를 극복하기 위하여 RGB 스테레오와 융합을 하는 것이다. [Fig. 8]의 (a2)와 (b2)는 SGM의 결과이고 이것을 융합 알고리즘의 입력 데이터로 사용한다. [Fig. 8]의 (a3)과 (b3)은 본 논문에서 제안한 알고리즘을 통해서 융합한 결과이다. [Fig. 8]의 (a3)과 (b3)을 3차원으로 출력한 영상이 (a4)와 (b4)이다.

[Fig. 8]의 (a1) 우측 하단의 ToF 센서 난반사 영역이 [Fig. 8]의 (a3)의 우측 하단에서 스테레오 카메라를 활용하여 depth 데이터를 얻어질 수 있음을 확인할 수 있다. [Fig. 8]의 (b1)에서 사람의 팔 주변의 ToF 가려짐 영역이 [Fig. 8] (b2)의 스테레오 카메라 결과와 융합하여 (b3)에서와 같이 극복되었다.

[Fig. 9]는 제안하는 캘리브레이션 방법론의 유효성을 알기 위하여 SSIM (Structural Similarity Index)를 계산하기 위한 입력 영상이다. 측정의 정확도를 위하여 SGBM 결과 영상과 ToF 센서 영상에서 전경(foreground)만 남긴 이진화 영상(binary

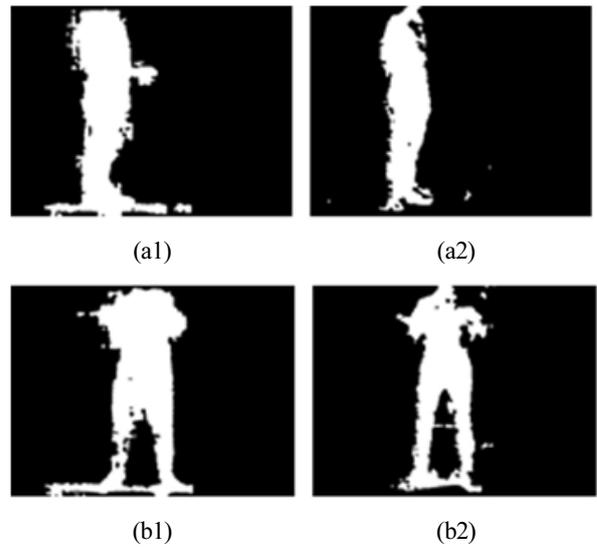


[Fig. 8] Output of fusion for ToF sensor and stereo camera.

image)을 생성하여 입력하였다. [Fig. 9]의 (a1)과 (b1)은 SGBM에 대한 전경 영상이고, (a2)와 (b2)는 ToF 센서에 대한 전경 영상이다. 이러한 영상의 469 set에 대하여 측정 결과, SSIM의 평균은 0.8599, 최대값 0.8880, 최소값은 0.7775였다.

5. 결 론

ToF 센서와 RGB 스테레오 카메라는 상호 보완적인 센서로서, 본 논문에서는 이를 융합하여 보다 다양한 환경에서 대응이 가능한 3차원 센서를 제안하였다. 특히, 본 논문에서는 64×32 의 저해상도 SPAD의 한계를 극복하기 위하여 상용 레퍼런스 ToF 센서를 활용하여 128×64 로 upsampling하는



[Fig. 9] Input images for SSIM comparison.

CNN을 적용하였다. 이 결과를 바탕으로 ToF 센서와 스테레오 카메라를 융합하는 알고리즘을 제안하여 SPAD 센서를 저가형 로봇용 3D 센서모듈에 적용하기 위한 연구를 수행하였다. 실험을 통하여 제안하는 CNN 결과의 PSNR로 유효성을 검증하였으며, 융합 결과로 각 센서의 한계를 극복할 수 있음을 확인하였다.

향후 연구로서, 본 논문이 제안하는 방법이 20 m의 장거리, 90도의 광시야각에서도 적용이 가능하고, 생산성을 확보하기 위한 연구를 진행할 예정이다. 이를 위하여, 현재 2~3m를 타겟으로 하는 고정 융합비율이 아니라 상황에 따라 적응적으로 (adaptive) 융합비율이 변하는 연구를 진행할 계획이다. 뿐만 아니라, [19]에서 제안된 출력 직전에 디컨벌루션(Deconvolution)을 이용한 업샘플링을 로봇을 위한 저가형 3차원 센서에 맞게 최적화하는 연구를 통해 본 논문이 제안하는 CNN의 성능을 개선하고자 한다.

References

- [1] S.-H. Kim, D.-K. Lee, J.-H. Cheon, S.-J. Kim, and K.-H. Yu, "Design and Flight Tests of a Drone for Delivery Service," *Journal of Institute of Control, Robotics and System*, vol. 22, no. 3, pp. 204-209, Mar., 2016.
- [2] S. Cho, D. Lee, Y. Jung, U. Lee, and D. H. Shim, "Development of a Cooperative Heterogeneous Unmanned System for Delivery Services," *Journal of Institute of Control, Robotics and System*, vol. 20, no. 12, pp. 1181-1188, Dec., 2014.
- [3] V. Magnanimo, S. Walther, L. Tecchia, C. Natale, and T. Guhl, "Safeguarding a mobile manipulator using dynamic safety fields," *2016 IEEE/RSJ International Conference on Intelligent*

Robots and Systems (IROS), Daejeon, South Korea, pp. 2972-2977, 2016.

[4] J.-H. Moon, H.-D. Choi, N.-H. Park, C.-H. Kim, Y.-W. Park, and E.-T. Kim, "Database based global positioning system correction," *Journal of Korea Robotics Society*, vol. 7, No. 3, pp. 205-215, Sep., 2012.

[5] C. Song and S.-H. Kim, "Robust vehicle occupant detection based on rgb-depth-thermal camera," *Journal of Korea Robotics Society*, vol. 13, No.1, pp. 31-37, Mar., 2018.

[6] D. Shin, B. Park, Y. Chae, and I. Yun, "Structure variation effects on device reliability of single photon avalanche diodes," *Microelectronics Reliability*, vol. 76-77, pp. 610-613, Sep., 2014.

[7] Y. Choi, N. Kim, S. Hwang, and I. S. Kwon, "Thermal image enhancement using convolutional neural network," *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Daejeon, South Korea, pp. 223-230, 2016.

[8] S. Wang, Z. Su, L. Ying, X. Peng, S. Zhu, F. Liang, D. Feng, and D. Liang, "Accelerating magnetic resonance imaging via deep learning," *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)*, Prague, Czech Republic, pp. 514-517, 2016.

[9] B. Zhu, J. Z. Liu, S. F. Cauley, B. R. Rosen, and M. S. Rosen, "Image reconstruction by domain-transform manifold learning," *Nature*, vol. 555, pp. 487-492, Mar., 2018.

[10] J. Park, H. Kim, Y.-W. Tai, M. S. Brown, and I. Kwon, "High quality depth map upsampling for 3d-tof cameras," *2011 International Conference on Computer Vision*, Barcelona, Spain, pp. 1623-1630, 2011.

[11] R. Nair, F. Lenzen, S. Meister, H. Schäfer, C. Garbe, and D. Kondermann, "High accuracy tof and stereo sensor fusion at interactive rates," *European Conference on Computer Vision*, pp. 1-11, 2012.

[12] G. Agresti, L. Minto, G. Marlin, and P. Zanuttigh, "Deep learning for confidence information in stereo and TOF data fusion," *2017 IEEE International Conference on Computer Vision Workshops (ICCVW)*, Venice, Italy, pp. 697-705, 2017.

[13] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 22, No. 11, pp. 1330-1334, Dec., 2000.

[14] J. M. Lee, D. Y. Kim, I. Kim, S. Jun, and C.-W. Park, "A study on heterogeneous sensor calibration between ToF sensor and RGB camera," *Korean Society for Precision Engineering Spring Conference*, Jeju, Korea, pp. 108-109, 2016.

[15] D. Y. Kim, Y.-O. Kim, C.-W. Park, J.-M. Lee, and S. Jun, "System and method for concurrent calibration of camera and depth sensor," Korea Patent, Application No. 10-2015-0095091, Jul., 3, 2015. DOI: 10.8080/1020150095091.

[16] D. Y. Kim, C.-W. Park, J.-M. Lee, Y.-O. Kim, I. Kim, S. Jun, and J.-H. Hwang, "System and method for 3-dimensional calibration of camera and depth sensor, and calibration device and

calibration jig thereof," Korea Patent, Registration No. 10-1803705, Nov., 2017. DOI: 10.8080/1020160057145.

[17] http://www.cube-eye.co.kr/en/#/spec/product_MDC600S.html, [Online] Accessed: Jul. 24, 2018.

[18] D. Hernandez-Juarez, A. Chacón, A. Espinosa, D. Vázquez, J. C. Moure, and A. M. López, "Embedded real-time stereo estimation via semi-global matching on the GPU," *Procedia Computer Science*, Vol. 80, pp. 143-153, Jun., 2016.

[19] C. Dong, C. C. Loy, and X. Tang, "Accelerating the super-resolution convolutional neural network," *European Conference on Computer Vision*, Amsterdam, The Netherlands, pp. 391-407, 2016.



김 동 엽

2008 연세대학교 전기전자공학부(학사)
 2010 연세대학교 전기전자공학과(석사)
 2010~현재 전자부품연구원 지능로보틱스
 연구센터 전임연구원/선임연구원
 2017~현재 연세대학교 전기전자공학과
 박사과정

관심분야: 지능시스템, SLAM, Sensor Fusion, 3D Reconstruction, Computational Intelligence, Machine Learning, Deep Learning, Virtual Reality



이 재 민

2015 명지대학교 전기공학과(학사)
 2015~현재 전자부품연구원 지능로보틱스
 연구센터 연구원

관심분야: 임베디드 제어 및 응용, 센서융합



전 세 응

2003 광운대학교 정보제어공학과(학사)
 2006 연세대학교 전자공학과(석사)
 2006~현재 전자부품연구원 지능로보틱스
 연구센터 책임연구원
 2012~현재 연세대학교 전기전자공학과
 박사과정

관심분야: Artificial Intelligence, Robot Sensor Fusion