

실외 경비 환경에서 강인한 객체 검출 및 추적을 위한 실외 멀티 모달 센서 기반 학습용 데이터베이스 구축

노동기[†], 양원근^{**}, 엄태영^{***}, 이재광^{****}, 김형록^{*****}, 백승민^{*****}

Multi Modal Sensor Training Dataset for the Robust Object Detection and Tracking in Outdoor Surveillance (MMO (Multi Modal Outdoor) Dataset)

DongKi Noh[†], Wonkeun Yang^{**}, Teayoung Uhm^{***}, Jaekwang Lee^{****},
Hyoung-Rock Kim^{*****}, SeungMin Baek^{*****}

ABSTRACT

Dataset is getting more import to develop a learning based algorithm. Quality of the algorithm definitely depends on dataset. So we introduce new dataset over 200 thousands images which are fully labeled multi modal sensor data. Proposed dataset was designed and constructed for researchers who want to develop detection, tracking, and action classification in outdoor environment for surveillance scenarios. The dataset includes various images and multi modal sensor data under different weather and lighting condition. Therefor, we hope it will be very helpful to develop more robust algorithm for systems equipped with difference kinds of sensors in outdoor application. Case studies with the proposed dataset are also discussed in this paper.

Key words: Outdoor Surveillance, Action Classification, Object Detection & Tracking, Mobile Robot, Multi Modal Sensor Dataset

1. 서 론

과거에는 공장 자동화 로봇이 로봇 시장의 주를 이루었다면 최근에 가정용 또는 상업용 서비스 로봇 시장이 점차 확대되고 있다. 서비스 로봇 시장도 과거 가정용 청소 로봇에 국한되어 있었다면 현재는

안내, 서빙 등을 포함한 다양한 분야로 확대되고 있다. 적용되는 환경도 제한된 실내에서 대면적 광역 실외 공간으로 점차 확대되어가고 있다. 실외 서비스 로봇의 경우는 실외 배송, 실외 경비 등과 관련된 기초적인 연구 및 실증 테스트가 진행되고 있다.

본 논문의 연구도 실외 경비 로봇에 관한 연구로

※ Corresponding Author: SeungMin Baek, Address: (06772) Yangjae 2(i)-dong, Yangjae-daero 11-gil, Seoul, Korea, TEL: +82-2-6912-6401, E-mail: seungmin2.baek@lge.com

Receipt date: Jul. 10, 2020, Approval date: Jul. 28, 2020

[†] Advanced Robotics Lab. LG Electronics Inc.
(E-mail: dongki.noh@lge.com)

^{**} Advanced Robotics Lab. LG Electronics Inc.
(E-mail: wonkeun.yang@lge.com)

^{***} Korea Institute of Robotics & Technology Convergence
(E-mail: uty@kiro.re.kr)

^{****} Advanced Robotics Lab. LG Electronics Inc.
(E-mail: jaekwang3d.lee@lge.com)

^{*****} Advanced Robotics Lab. LG Electronics Inc.
(E-mail: hyoungrock.kim@lge.com)

^{*****} Advanced Robotics Lab. LG Electronics Inc.

※ This research was supported by Institute for Information & Communication Technology Promotion (IITP) grant funded by the Korea government (MSIT) (2017-0-00306)

경비 시 대처해야 하는 다양한 이상 상황을 인식하는 알고리즘을 개발하는 과제이다. 본 연구는 정보통신기획평가원(ИTP)에서 지원받아 2017년부터 진행되고 있고 해당 과제 진행 시 수집한 학습 및 평가용 데이터베이스를 본 논문을 통해 공개하고자 한다. 참고로 본 연구는 악천후 등의 다양한 상황에서 동작 가능한 실외 경비 로봇의 정보분석 플랫폼을 개발하는 연구이다. 실외 로봇 애플리케이션의 경우 실내보다 강인하고 안정적인 알고리즘이 필요하다. 이를 위해서는 조명, 날씨 등의 환경 변화가 다양하게 발생할 수 있으므로 각 상황에서도 안정적이고 강건하게 동작할 수 있는 학습 방법이 필수적이다. 따라서 다양한 상황에서 동일한 성능을 내려면 상황이 바뀌더라도 주변 상황을 인식할 수 있는 센서 시스템과 이에 맞는 알고리즘이 주요한 요소이다. 예를 들어 밤, 낮에 모두 사람을 검출해야 한다면 RGB 카메라만으로 해당 임무(Task)를 수행하기 힘들고, 열 영상 카메라(Thermal Camera)나 나이트 비전 카메라(Night Vision Camera) 데이터가 추가되어야 한다. 따라서 다양한 센서 데이터를 획득할 수 있는 센서 시스템과 이를 활용할 수 있는 알고리즘이 필요하다.

본 연구는 포항 안전로봇 실증단지과 광주광역시 나노 산단 일대에서 진행되었으며, 이 지역에서 다양한 환경과 경비 시나리오에 대해 멀티모달 센서 데이터를 수집, 정리하였고, 알고리즘 개발과 실증을 진행하였다. 특히 일반적인 상황의 데이터셋 뿐만 아니라 야간이나 우천 상황에서의 데이터를 다수 확보하였다. 대상은 주로 무인 경비 시나리오와 연관성이 큰 사람과 차량에 대한 데이터셋을 확보하였고, 전체 데이터셋 크기는 Labeling 된 데이터 기준으로 약 30만 장 이상을 확보하였다. 해당 데이터를 활용하여 일반 환경이 아닌 다양한 악천후 환경에서 사람과 차량의 검출 및 추적 알고리즘을 개발 및 테스트하여 유용한 결과를 얻는 것이 가능하다.

본 논문은 다음과 같이 구성되어 있다. 다음 장에서 데이터셋의 구성과 특징에 관해서 설명하고, 이 데이터를 획득한 센서에 대해 간단히 기술한다. 그리고 데이터셋을 활용하여 개발하고 검증할 수 있는 다양한 알고리즘과 이에 관해 서술한다. 마지막으로 동 데이터셋으로 개발된 알고리즘을 무인 경비 정보 분석 플랫폼에 활용된 예로 결론을 맺는다.

2. 데이터셋 구성 및 특징

2.1 시스템 구성

본 논문에서 제안하는 데이터셋은 실외 무인 경비 로봇을 위한 지능형 정보분석 시스템을 기반으로 취득되었다. 시스템은 모바일 로봇인 이동형 에이전트와 고정형 에이전트로 구성되어 있고 에이전트에서 들어오는 정보를 지능분석 플랫폼에서 해석하여 이상 상황을 인식하게 된다. 지능분석 플랫폼의 모듈 구성은 다음과 같다. 우선 고정형/이동형 에이전트와 경비 대상이 되는 사람 또는 차량 등의 객체를 검출 및 추적하는 모듈이 있다. 그리고 에이전트의 센서 데이터와 객체 검출 및 추적 모듈의 결과를 활용하여 확률 기반의 환경지도를 만들고 이를 이용한 이상 상황 분석 및 인식 모듈이 있다. 마지막으로 각 에이전트를 모니터링 할 수 있는 관제 시스템으로 구성되어 있다. 전체 시스템의 개괄적인 구성도는 Fig. 1과 같다. 경비 환경(Security Environments)에 있는 멀티 에이전트로부터(Multi-Agents) 취득된 멀티 모달 센서 데이터가(Multi Modal Sensor Data) 클라우드 기반의 지능분석 플랫폼으로 입력되고 이를 플랫폼에서 분석하여 경비 영역에서의(Security Environments) 이상 유/무를 판단하게 된다.

2.2 데이터셋 구성

2.1 장에서 설명한 에이전트에 부착된 센서는 다음과 같다. 이동형 에이전트인 로봇과 고정형 에이전트에는 Fig. 2와 같은 모듈이 부착되어 있고 이동형 에이전트의 경우는 Fig. 2의 모듈과 3D LiDAR 센서가 부착되어 있다.

이동형 에이전트는 진동 영향을 최소화하기 위해 흡진기(Damper) 모듈 하단에 설치하였고, LiDAR 센서의 경우는 로봇과 강체(Rigid Body)로 결합하였다. [1] 로봇의 형태는 센서 모듈과 상관없이 변형할 수 있으나, 본 논문의 부록(Appendix)에 공개하는 데이터셋을 획득한 로봇 형태를 첨부하였다. 데이터셋의 구성은 Fig. 3과 같이 8가지로 구성되어 있지만, 본 논문을 통해 공개되는 데이터셋은 객체에 대한 검출 및 추적 알고리즘과 관련된 것이므로, GPS, 3D LiDAR 데이터는 제외하였다. 다만, 객체 검출의 보조 수단으로 사용할 수 있도록 데이터셋 생성 시 취득한 LiDAR 데이터에 대해서는 함께 제공한다. 각

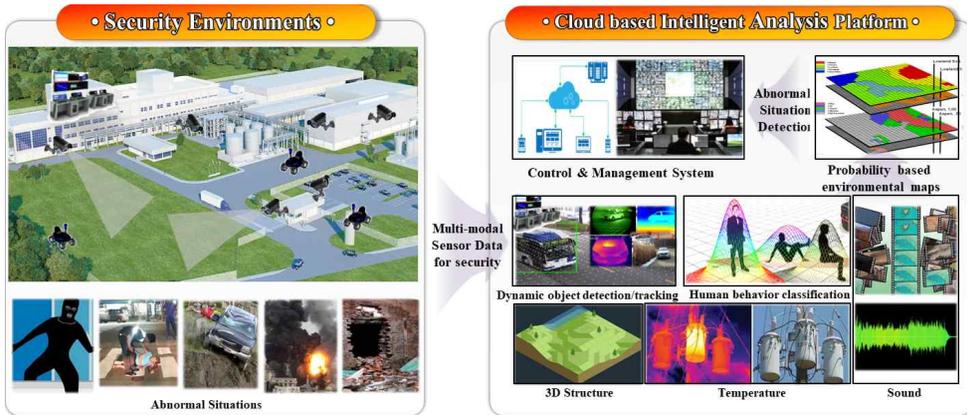


Fig. 1. Multi-modal sensor based intelligent systems for outdoor security robot.

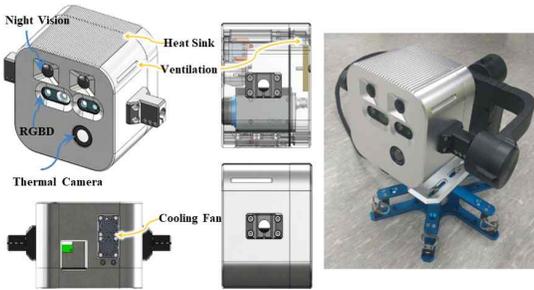


Fig. 2. Configuration of Multi Modal Sensor Module.

센서의 데이터는 독일의 도로 표지판 데이터셋과 (GTSRB) 같이 개별적인 Raw 데이터 형태로 구성되어 있고, rosbag 형태로 제공하지 않는다. [2] 각 센서 간 데이터 동기화 정보는 파일명으로 제공한다

다. 파일명이 같으면 동일 시간에 획득한 데이터로 간주한다.

결론적으로 데이터셋은 RGB 영상 데이터, Depth 데이터, IR 영상 데이터, 열(Thermal) 영상 데이터, 나이트 비전 카메라(Night Vision Camera) 데이터, 3D LiDAR 데이터로 구성되어 있다. 각 데이터의 획득 시간은 동기화되어 있다. 인텔의 D435 센서를 활용하여 RGB, IR, Depth 영상을 획득하였고, 영상 획득 시 센서에서 획득한 IMU 데이터도 본 데이터셋에서 기본적으로 제공된다. Fig. 4는 센서별 측정된 데이터의 예시를 보여주고 있다.

2.3 데이터셋에 제공되는 정보(Annotation)

기본적으로 모든 데이터는 멀티모달 센서 데이터

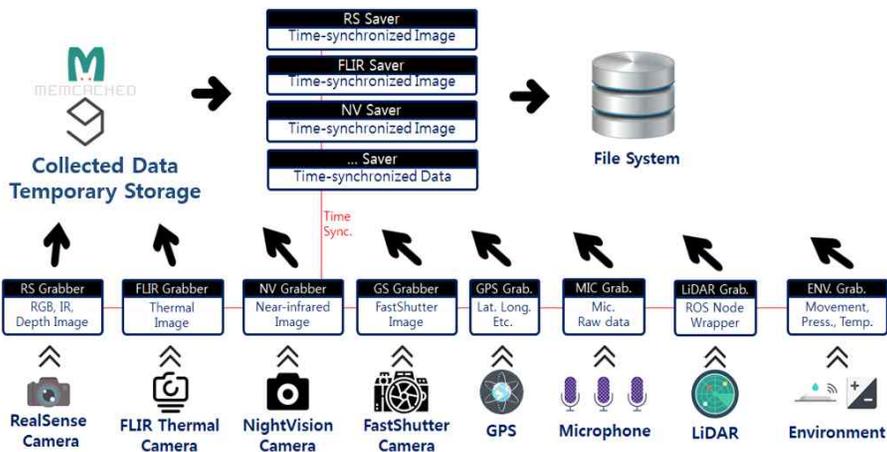


Fig. 3. Description of Synchronized Dataset: Realsense (RGB, Depth, IR), Thermal Camera, Night Vision Camera, 3D LiDAR, GPS

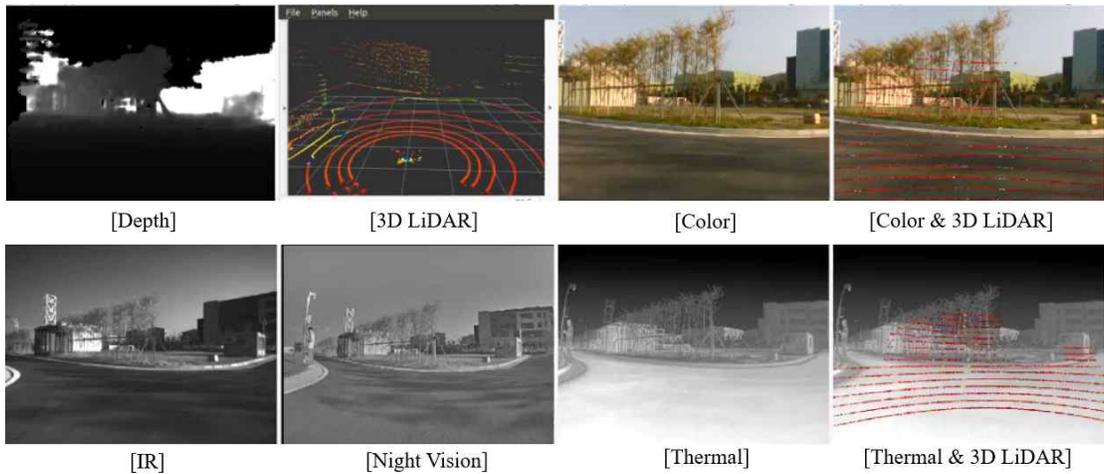


Fig. 4. Example of Dataset.

이고 Labeling이 되어있다. 각 센서 간 기구학적 관계(Geometric)는 별도의 파일로 제공되고, 센서의 종류에 대해서는 2.2장에서 상세히 설명하였다. 제공되는 데이터셋 중 객체 검출(Detection) 및 추적(Tracking) 데이터셋은 다수의 객체가 있는 데이터셋이고, 행동 분류용 데이터셋은 단일 객체에 대해서 제공된다. 경비 로봇의 경비 상황에서 볼 수 있는 다양한 상황에 대해 데이터셋을 구성하였고, 특히 행동 분류에 대해서는 정상과 이상 상황에 대해 15개의 (Table 1 참조) 시나리오로 분류하여 데이터를 수집하고 정리하였다. 멀티 모달 센서 데이터의 특징점을 활용하기 위하여 일반적인 환경 데이터 외에도 밤, 우천, 폭우 등의 데이터를 추가로 구성하였다. 데이터셋에 등장하는 사람은 남녀 전문 배우 각각 2명을 포함하여 10명 이상의 인원으로 구성하였다. 설명한 내용에 대한 정보는 아래 Fig. 4와 같이 데이터별로 XML 파일 형태로 Labeling 되어 제공된다. 일반적인 DB와의 차이점은 같은 장면(Scene)에 대해 멀티 모달 센서 데이터를 제공한다는 점과 다양한 환경 변화(e.g. XML의 environment 탭)에 대한 데이터셋을 제공한다는 점이다. 앞의 Fig. 5로 첨부한 XML을 보면 <subfolder> 항에 사용한 센서가 표기되고 <environment> 항에 센서 모듈이 고정 및 이동 여부에 따라 고정형(FIX)과 이동형(MV)으로 분류되어 있다. 이 항에 날씨(e.g. Sunny)와 촬영장소(e.g. Pohang(포항))가 표기되고 마지막으로 검출(Detection) 및 추적용인지 행동(Action) 분류용인지에 대

한 정보가 포함된다. <size> 항에서 센서의 해상도에 정보가 표기되고 <object>항에서 Annotation된 객체의 정보가 표기된다. 참고로 행동 분류 DB의 경우는 별도의 경계(Bounding Box)가 없다. 추가로 단순한 경계 정보(Bounding Box)뿐만 아니라, 행동 분류 데이터셋의 경우 Mask 이미지를 제공함으로써 Pixel Labeling 정보도 함께 제공한다.

촬영된 시나리오는 다음 Table 1과 같다. 무인 경비 로봇 입장에서의 정상 상황과 이상 상황으로 분류하고 각 상황에 대한 행동(Action Type)을 Table 1과 같이 정의하였다. 각 시나리오에 대해 15회 이상 촬영을 진행하였다.

Table 1에서 보듯 다양한 이상 상황들이 배우들에 의해 연출되어 있어 경비 로봇의 이상 동작 감지 알고리즘 개발에 매우 유용한 데이터이다. 데이터별 통계적 수치는 다음과 같다(Fig. 6).

3. 데이터셋을 활용한 알고리즘 분석

3.1 행동 분류(Action Classification) 데이터셋

행동 분류 데이터셋의 경우 경계 정보(Bounding Box) 대신 Mask 이미지를 제공함으로써 픽셀(Pixel) 기반의 Segmentation 정보를 제공한다. 주어진 데이터셋을 활용하여 행동 분류 관련 알고리즘 개발 및 성능 평가를 진행할 수 있다(Fig. 7참조).

일반적으로 기존의 행동 분류 알고리즘 및 성능을 평가하는 데이터셋은 UCF 11 데이터셋을 사용하여

Table 1. Scenario and Action Type

No.	Situation	Action Type	Example	No.	Situation	Action Type	Example
1	Normal	lie down on ground		9	Abnormal	sitting on a chair → fall down	
2	Normal	standing		10	Abnormal	sitting on ground → fall down	
3	Normal	walking		11	Abnormal	standing → fall down	
4	Normal	running		12	Abnormal	walking → fall down	
5	Normal	sitting on a chair		13	Abnormal	running → fall down	
6	Normal	sitting on ground		14	Abnormal	convulsions	
7	Normal	stand → sit on ground → stand		15	Abnormal	drunken person	
8	Normal	stand → sit on a chair → stand		16	Abnormal	etc.	others

```

<?xml version="1.0" encoding="UTF-8"?>
<Annotation>
  <folder>MV12_0604_d</folder>
  <subfolder>thermal</subfolder>
  <filename>MV12_THERMAL_190604_215959_08.png</filename>
  <environment>
    <camerastate>MV</camerastate>
    <weather>Sunny</weather>
    <venue>Po-hang</venue>
    <purpose>Detection</purpose>
  </environment>
  <size>
    <width>640</width>
    <height>512</height>
    <depth>3</depth>
  </size>
  <object>
    <name>Human</name>
    <pose>sit</pose>
    <bndbox>
      <xmin>334</xmin>
      <ymin>251</ymin>
      <xmax>349</xmax>
      <ymax>271</ymax>
    </bndbox>
  </object>
  <object>
    <name>Human</name>
    <pose>sit</pose>
    <bndbox>
      <xmin>275</xmin>
      <ymin>252</ymin>
      <xmax>289</xmax>
      <ymax>272</ymax>
    </bndbox>
  </object>
  <object>
    <name>Human</name>
    <pose>sit</pose>
    <bndbox>
      <xmin>255</xmin>
      <ymin>253</ymin>
      <xmax>268</xmax>
      <ymax>273</ymax>
    </bndbox>
  </object>
  <object>
    <name>Human</name>
    <pose>stand</pose>
    <bndbox>
      <xmin>237</xmin>
      <ymin>251</ymin>
      <xmax>250</xmax>
      <ymax>279</ymax>
    </bndbox>
  </object>
  <object>
    <name>Human</name>
    <pose>stand</pose>
    <bndbox>
      <xmin>220</xmin>
      <ymin>252</ymin>
      <xmax>233</xmax>
      <ymax>280</ymax>
    </bndbox>
  </object>
  <object>
    <name>Human</name>
    <pose>sit</pose>
    <bndbox>
      <xmin>200</xmin>
      <ymin>251</ymin>
      <xmax>218</xmax>
      <ymax>267</ymax>
    </bndbox>
  </object>
</Annotation>

```

```

<?xml version="1.0" encoding="UTF-8"?>
<Annotation>
  <folder>MV06_0927</folder>
  <subfolder>rgbdepth\color</subfolder>
  <filename>FX01_RGB_190927_103423_03.png</filename>
  <environment>
    <camerastate>MV</camerastate>
    <weather>Sunny</weather>
    <venue>Po-hang</venue>
    <purpose>Action</purpose>
  </environment>
  <size>
    <width>640</width>
    <height>480</height>
    <depth>3</depth>
  </size>
  <object>
    <name>Human</name>
    <pose>stand</pose>
    <status>normal</status>
    <masking>FX01_MASK_190927_103423_03.png</masking>
  </object>
</Annotation>

```

Fig. 5. XML : Annotation of multi modal sensor data for detection and tracking.

개발 또는 실증한다. 해당 데이터셋은 Fig. 8과 같이 11가지 사람 행동에 대한 비디오로 제공된다. UCF 11의 경우는 RGB 비디오 만을 제공하므로 시공간 기반의 인식기를 개발/실증하기는 좋으나 악천후나 야간 환경에서의 행동 분류기를 개발하는 데이터셋으로는 부적절하다[3,4].

현재 UCF 11 데이터셋을 활용하여 Fig. 9과 같이 VGG 기반의 LSTM 네트워크를 사용하는 경우 UCF 11 데이터셋에서 2019년도 기준으로 약 95.6% 정도의 정확도를 가지는 행동 분류기를 설계할 수 있다.

UCF 11 기반의 행동 분류 연구는 LSTM 등의 시공간의(Temporal & Spatial) 데이터를 활용하는 인식기의 개발에는 의미가 있을 수 있으나, 무인 경비 로봇이 적용되는 실제 환경에는 부적합한 부분이 많

다. 경비 상황은 주로 야간 환경이 많으며, 침입자는 우천 환경을 선택하기 쉽다. 따라서 일반적인 환경을 포함하여 야간, 우천 등의 환경에서 취득한 데이터셋이 무인 경비 로봇의 알고리즘 개발하는 연구자들에게 필요하다. 따라서 본 논문에서 제안하는 “MMO Dataset”은 Fig. 10과 같이 다양한 악천후 환경에 대한 행동 분류 연구용 데이터셋을 제안한다.

동일한 환경에 대해서도 행동 분류용 데이터셋은 Fig. 11과 같이 6종의 영상, 3D LiDAR 그리고 IMU 데이터로 구성되어 있다.

제안하는 데이터셋으로 행동 분류 및 사람을 검출한 결과 Fig. 12과 같이 야간 상황에서 RGB 뿐만 아니라 열영상(Thermal Image)을 활용할 수 있음을 확인하였다.

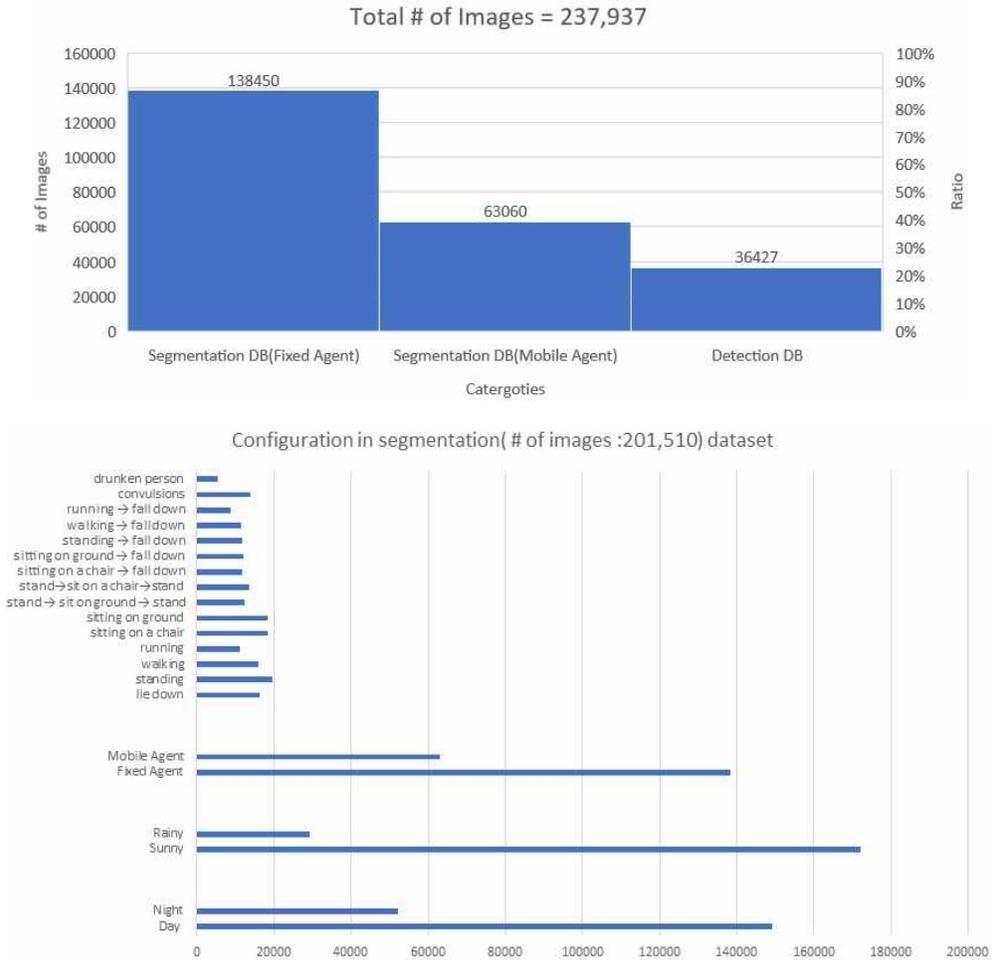


Fig. 6. Configuration Dataset.

3.2 객체 검출 및 추적 데이터셋

객체 검출 및 추적 관련 데이터셋은 다양한 분야에 걸쳐 공개되어 있다. 특히 차량에 대한 부분은 KITTI [5] 데이터셋과 Cityscapes Dataset [6] 이 유명하다. 연구용 데이터 셋으로는 MDP 주/야간, OTB

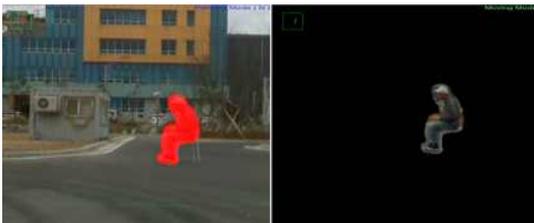


Fig. 7. Result of Pixel Labeling Red Color on the left image is pixel labeling.

데이터셋 등도 함께 활용하고 있다. 소개한 데이터셋들은 Reference 평가를 진행하기 위해 매우 좋은 데이터셋이고 OTB의 경우 Motion Blur에 대한 강인성을 확인하는 데 매우 유용하게 사용할 수 있는 장점이 있다. 결론적으로 해당 데이터셋들은 정량 평가를 위해 유용하게 사용할 수 있는 데이터셋이지만 현실적으로 무인 경비 로봇에 적용되었을 때 동일한 성능이 나온다는 것은 보장할 수 없다. 또한, 주로 RGB 영상과 스테레오(Stereo) 카메라 영상이 주를 이루고 있어 다양한 환경에서 강인하게 검출 및 추적 성능을 확보하는 것에 한계가 있다. 예를 들어 우천 상황이나 야간 상황에서는 RGB 외에도 다양한 특성을 가진 센서를 사용하는 것이 상식적으로 좋으나, 현재 공개된 KITTI[5], Cityscapes[6] 등의 데이터로는 다

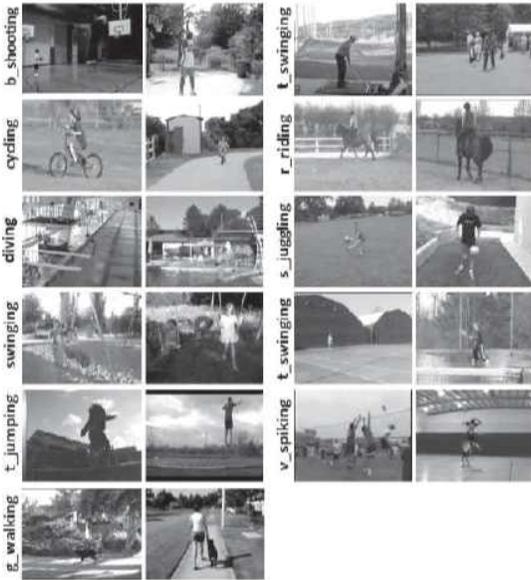


Fig. 8. UCF 11 Dataset.

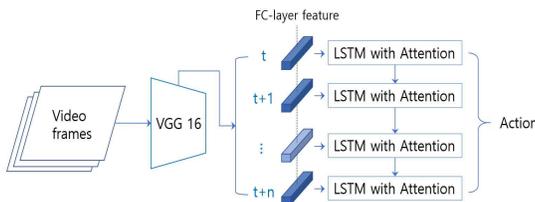


Fig. 9. VGG based LSTM Network.



Fig. 10. MMO Dataset on various environments for action classification

양한 악천후 시나리오를 테스트하기 어렵다. 본 논문에서 제안하는 데이터는 기본적으로 로봇 기준의 시점에서 촬영되었고, 다양한 상황과 환경에서 촬영되

었기 때문에 지능형 경비 로봇의 객체 검출 및 추적 알고리즘 개발에는 의미 있는 학습 성능을 확보할 수 있다.

3.2.1 객체 검출

획득한 멀티모달 데이터를 활용한 검출 성능을 검증하기 위해 Fig. 13과 같은 구조의 딥러닝 네트워크를 사용하였다. 제안된 네트워크는 RefineDet을 활용하여 RGB 외 영상을 사용할 수 있다[9]. 이 네트워크 구조를 활용하여 KITTI[5] 등의 기존 데이터셋에 대해서도 검출 성능을 확인하였다. 해당 네트워크는 이동 로봇에서 온라인으로 처리해야 하는 제약조건을 만족하기 위해 anchor refine module을 활용하여 고속화를 구현되었고, 백본(Backbone) 네트워크도 ResNet34를 활용하여 경량화하였다. 사전학습용 데이터로는 COCO[7]와 VOC[8] 데이터를 활용하였다.

멀티 모달 데이터 사용을 위해 서울대학교 MIPAL Lab.에서 제안된 RefineDet을 사용하여 이미 공개된 데이터셋에 대해서도 아래 Table 2과 같이 검출 성능에 대해 검증해보았다. 특히 Night 상황에서 Day와 큰 차이가 없음을 확인할 수 있었다.

본 논문에서 제안된 데이터셋에 대해서도 Fig. 12에서 제안된 알고리즘을 적용하여 Fig. 14과 같이 확인하였다. 다 개체(Multi-Object) 상황의 야간 영상에서도 무인 경비 대상인 사람에 대한 검출이 문제없이 됨을 확인할 수 있었다.

3.2.2 객체 추적

객체 추적은 객관적인 성능 평가에서는 OTB-100 데이터를 최근까지 많은 연구자가 활용하였다. 해당 데이터셋은 카메라가 흔들리는 상황도 연출되어 있어 모바일 로봇을 개발하는 연구자들에게도 의미가 있는 데이터셋으로 판단된다. 다음 장에 있는 Fig. 15와 같은 OTB-100 데이터셋에 대해 객체 추적 성능에 확인하였다.

Table 2. The performance for our algorithm using dataset

Dataset	Performance (mAP)
KITTI	56.52
MDP (Day)	56.01
MDP (Night)	55.24



Fig. 11. Multi modal dataset : (Top-Left) RGB(D435) (Top-middle) IR (Top-Right) Depth (Bottom-Left) Night Vision (Bottom-middle) Thermal (Bottom-Right) RGB(Global Shutter)

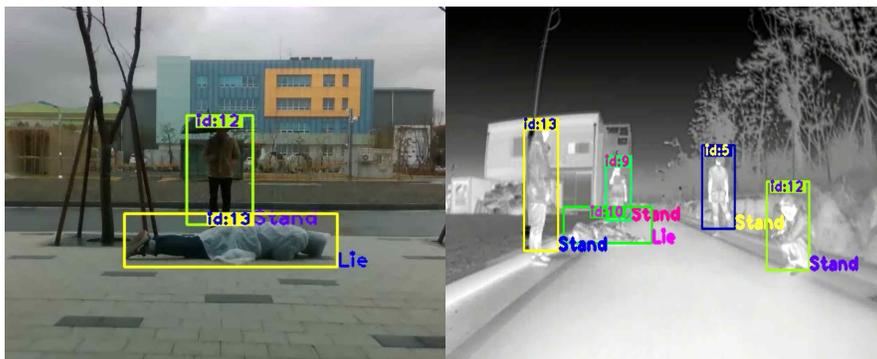


Fig. 12. Action classification using RGB with Depth(Left) and thermal Image(Right) (Left) Rain Condition, (Right) Night Condition,

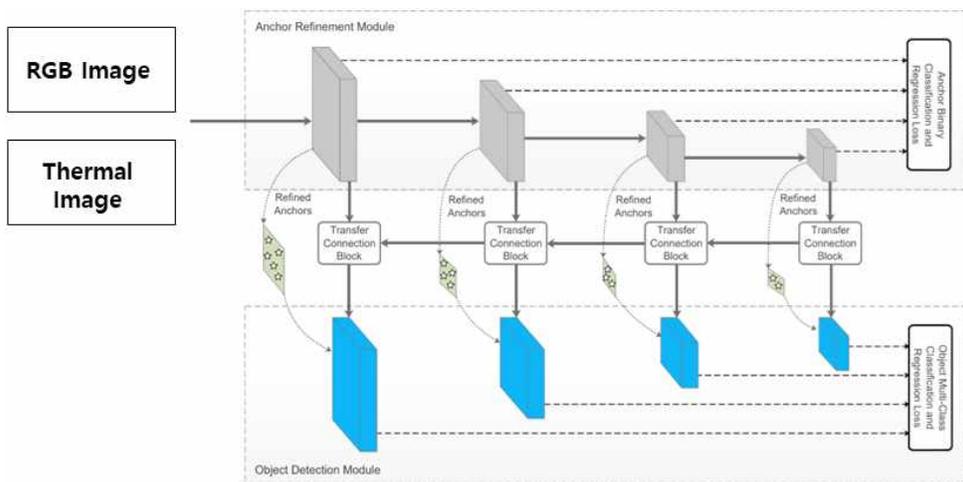


Fig. 13. RefineDet with anchor refine module to use multi modal data,



Fig. 14. Detection Result with multi modal sensor data in night.

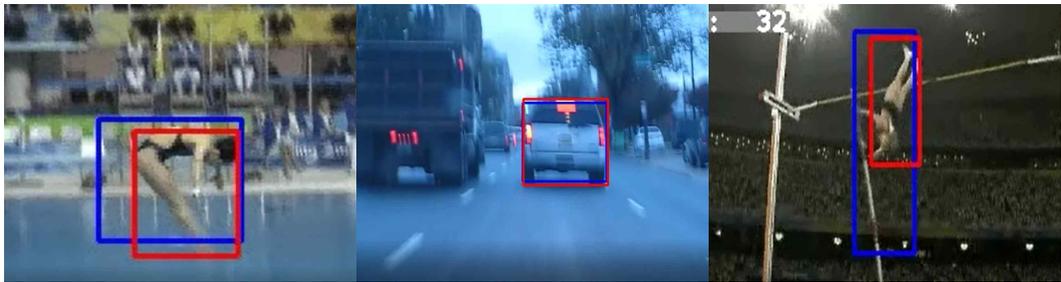


Fig. 15. Tracking Result for OTB 100 Dataset.

추적 성공률은 참값(Ground Truth)의 영역과 (Bounding Box) 객체 추적으로 나온 영역과(Bounding Box) 겹치지 않는 영역의 픽셀(Pixel) 수로 판단하였다. Fig. 16과 같이 겹치는 영역의 Pixel 차이가 기준인 20 Pixel 이하로 어긋날 경우만 추적 성공으로 판단하였고, OTB 100의 경우 약 87.5% 정확도를 확보할 수 있음을 확인하였다.

OTB 100의 경우는 RGB 기반의 실외 비디오 데이터므로 Motion Blur에 대한 알고리즘 강인성은 확인할 수 있었으나 실외 경비 상황과 관련된 알고리즘을 검증하는데 부족한 점이 있었다. 이에 본 논문에서

제안하는 멀티모달 데이터를 활용하여 Fig. 14의 기준으로 추적 성능을 확인하였다. 추적하는 방법은 객체 검출과 마찬가지로 서울대학교 PI Lab.에서 Tracking-by-Detection 개념으로 낮은 연산량으로 멀티모달 데이터를 사용하는 방법을 제안하였고 이를 확인해보았다. 본 논문에서는 추적(Tracking)을 진행하는 각 장면(Scene)별로 참값(Ground Truth)에 해당하는 경계 영역(Bounding Box)을 제공하므로 Fig. 17과 방법으로 성능을 확인할 수 있다.

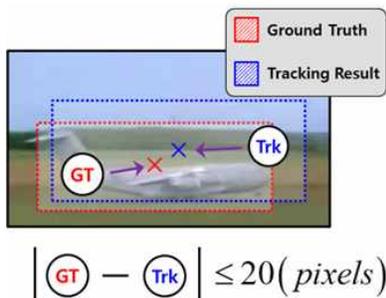


Fig. 16. Evaluation for tracking performance.



Fig. 17. Tracking result with multi modal data in night.

4. 결론 및 고찰

본 논문에서는 실외 무인 경비 로봇 플랫폼을 개발하는 과정에서 획득한 다양한 상황의 실외 데이터셋의 구성과 수집 방법에 관해 설명하고 이를 공개하고자 한다. 해당 데이터는 약 3년 동안 여러 가지 시행착오를 통해 구축한 상태이고, 공개되는 모든 데이터는 Annotation이 되어있다. 이미 공개된 데이터셋과 비교해보면 VOC[8]의 경우는 객체(Object)에 편중되어 있어 경비 시나리오와는 맞지 않는다. 물론 VOC[5] 데이터도 사전학습(Pre-training)용으로는 사용할 수 있다. 본 논문에서 제안한 해당 데이터셋은 KITTI[5], Cityscapes[6], COCO[7], 등과 같은 공개 DB와 비교 시 실외 경비에 특화된 다양한 형태의 데이터를 사용할 수 있다는 장점이 있으며, 실제 연구/개발되고 있는 딥러닝 네트워크를 활용하여 검증해본 결과 연구/개발에 문제없이 사용 가능할 것으로 검증되었다. 많은 시간과 자원을 투입하여 제안된 데이터셋을 구축하였으며, 본 과제의 알고리즘 개발 과정에서 데이터의 학습 성능은 충분히 검증되었다. 향후 실외 경비 상황에 대한 더 많은 시나리오에 대해 데이터셋을 확장하고 발전시킬 예정이다.

마지막으로 본 논문에서 제안한 데이터셋은 한국로봇융합연구원(KIRO)에서 관리하고 있고, 다운로드하는 부록(Appendix)에서 안내한 방법으로 할 수 있다.

Acknowledgments

3년 동안 알고리즘 개발과 함께 데이터 셋 구축에 힘쓴 ETRI, KIRO, 서울대 연구원들에게 감사하고, 본 논문의 데이터 셋 이름을 제안해준 카이스트 미래도시로봇 연구실 학생들에게도 감사하다.

부록(Appendix)

1. 데이터 셋 주소 : <https://kiro.myds.me:5001/sharing/1anwk0HxJ> (저자들에게 이메일로 용도, 이름(소속포함), 연락처를 보내주시면 해당 주소의 패스워드 전달 예정)
2. 주요 클래스 : 다양한 포즈와 상황의 사람(Human) 및 차량
3. Actor 수 : 전문 배우 성인 4명(남 2명, 여 2명)

의 10명 내외

4. 환경 : 밤, 낮, 우천, '20년 눈 오는 환경 추가 예정 (본문의 Fig. 10 참조)
5. 멀티 모달 센서 종류 : 영상 6종, 3D LiDAR, IMU Data
6. 행동 분류의 경우 시나리오당 15회 촬영(본문의 표 1 참조)
7. 총 데이터 수 : 237,937장 (Labeling 된 Data 기준, 본문의 Fig. 6 참조)
 - 이 중 36,427장 Detection Dataset
8. 센서 데이터 취득 로봇(이동형 에이전트, 아래 그림의 좌측) 및 고정형 에이전트(아래 그림의 우측 그림 내 붉은 원)



REFERENCE

- [1] T. Uhm, J. Park, J. Lee, G. Bae, and Y. Choi, "Multi-modal Sensor System and Database for Human Detection and Activity Learning of Robot in Outdoor," *Journal of Korea Multimedia Society*, Vol. 21, No. 12, pp. 1459-1466, 2018.
- [2] J. Stallkamp, M. Schlipsing, J. Salmen, and C. Igel, "Man vs. Computer: Benchmarking Machine Learning Algorithms for Traffic Sign Recognition," *Journal of Neural Networks*, Vol. 32, pp. 323-332, 2012.
- [3] M.D. Rodriguez, J. Ahmed, and M. Shah, "Action MACH: A Spatio-temporal Maximum Average Correlation Height Filter for Action Recognition," *Proceedings of the 2008 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1-8, 2008.
- [4] K. Soomro and A.R. Zamir, "Action Recognition in Realistic Sports Videos," *Computer Vision in Sports*, Springer International Pub-

- lishing in New York, NY 10036, 2014.
- [5] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, "Vision Meets Robotics: The KITTI Dataset," *International Journal of Robotics Research*, Vol. 32, No. 11, pp 1231-1237, 2013.
- [6] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, et al., "The Cityscapes Dataset for Semantic Urban Scene Understanding," *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3213-3223, 2016.
- [7] T.Y. Lin, M. Maire, S. Belongie, L. Bourdev, R. Girshick, J. Hays, et al., "Microsoft COCO: Common Objects in Context," *Proceedings of the 2014 European Conference on Computer Vision*, pp. 740-755, 2014.
- [8] M. Everingham, S.M.A. Eslami, L.V. Gool, C.K.I. Williams, J. Winn, and A. Zisserman, "The Pascal Visual Object Classes Challenge: A Retrospective," *International Journal of Computer Vision*, Vol 111, No. 1, pp. 98-136, 2015.
- [9] S. Zhang, L. Wen, X. Bian, Z. Lei, and S.Z. Li, "Single-shot Refinement Neural Network for Object Detection," *Proceedings of the 2018 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4203-4212, 2018.



노 동 기

2002년 고려대학교 전자전자전
과공학부 학사
2004년 서울대학교 전기컴퓨터
공학부 석사
2020년~현재 KAIST 전기및전
자공학부 박사과정

2004년~2006년 LG디스플레이 회로설계 주임연구원
2006년~2010년 삼성중공업 로봇시스템연구 책임연구원
2010년~현재 LG전자 로봇선행연구소 책임연구원
관심분야: Cognitive Robotics and Planning, Computer
Vision, Robot Vision, SLAM



이 재 광

2003년 홍익대학교 전기전자공학
부 학사
2005년 인하대학교 전기공학 석사
2007년 UST 컴퓨터소프트웨어
박사수료
2005년~2010년 ETRI-UST 연
수생

2010년~현재 LG전자 로봇선행연구소 책임연구원
관심분야: Computer Vision, Robot Vision, SLAM



양 원 준

2004년 인하대학교 전자공학과
학사
2006년 인하대학교 전자공학과
석사
2011년 인하대학교 전자공학과
박사

2012년~현재 LG전자 로봇선행연구소 책임연구원
관심분야: Robot Vision, Deep Learning



김 형 록

2000년 서울대학교 기계항공공학
부 학사
2002년 KAIST 기계공학 석사
2010년 KAIST 기계공학 박사
2010년~현재 LG전자 로봇선행
연구소 책임연구원

관심분야: Robot Navigation, Human-Robot Interaction



엄 태 영

2004년 경기대학교 전자공학과
학사
2006년 한양대학교 전자통신컴
퓨터공학과 석사
2014년 한양대학교 전자통신컴
퓨터공학과 박사

2014년 9월~2016년 4월 한국과학기술연구원 박사 후 과정
2016년 5월~2017년 4월 한국과학기술연구원 위촉연구원
2017년 5월 ~ 현재 한국로봇융합연구원 지능로봇연구
본부 선임연구원
관심분야: Computer Vision, Human-Robot Interaction,
AI, Autonomous Driving Car, AR



백 승 민

1996년 성균관대학교 전자공학과
학사
1998년 성균관대학교 전자공학과
석사
2002년 성균관대학교 전자공학과
박사

2003년~2004년 일본 나고야대학교 박사 후 과정
2004년~2008년 성균관대학교 지능시스템 연구센터 연
교수
2006년 프랑스 Blaise Pascal Univ. LASMEA 연구소 박
사 후 과정
2008년~2019년 LG전자 로봇선행연구소 연구위원
2020년~현재 LG전자 로봇선행연구소 소장(상무)
관심분야: Cognitive Robotics, Robot Vision, Learning