JDCS
디지털콘텐츠학회논문지
Journal of Digital Contents Society
Vol. 21, No. 11, pp. 2011-2019, Nov. 2020

Check for updates

비디오 영상을 이용한 3차원 재구성 및 객체 인식 모델 개발

이 나 혁[1] · 이 경 택[2] · 박 영 섭[3] · 서 상 현[4] · 이 태 민[5*]
[1]중앙대학교 소프트웨어대학 소프트웨어학부 학사과정
[2]한국전자기술연구원 센터장
[3](주)이노시뮬레이션 수석연구원
[4]중앙대학교 예술공학대학 컴퓨터예술학부 교수
[5*]중앙대학교 다빈치SW교육원 특임교수

Development of 3D Reconstruction and Object Recognition Model using Video

Nahyuk Lee[1] · Kyungtaek Lee[2] · Youngsup Park[3] · Sanghyun Seo[4] · Taemin Lee[5*]
[1]Bachelor's Course, Computer Science & Engineering, ChungAng University, Seoul 06974, Korea
[2]Director, Korea Electronics Technology Institiue, Seongnam-Si, Gyeonggi-do 13509, Korea
[3]Senior Researcher, Innosimulation, Seoul 03925, Korea
[4]Professor, School of Computer Art, College of Art & Technology, Chung-Ang University, Anseong-Si, Gyeonggi-Do 17546, Korea
[5*]Special Affair Professor, Davinci SoftWare Institue, ChungAng University, Seoul 06974, Korea

[요 약]

콘텐츠 산업들이 발전함에 따라 이용되는 콘텐츠들이 2차원에서 3차원으로 범위가 넓어지고, 전문가들 뿐 아니라 일반 사용자들도 이 콘텐츠를 만들고 사용하고자 하였다. 하지만 3차원 정보들을 다루는 작업은 고도의 기술과 시간이 많이 필요하다. 따라서 본 연구에서는 SfM을 이용한 간단한 3차원 재구성 방법을 제시한다. 사용자가 간단한 방법으로 비디오 영상을 제작하면 이를 토대로 데이터셋을 증강하여 늘린 후에 늘어난 데이터를 이용하여 3차원 재구성을 진행한다. 또한 CNN 학습을 통해서 3차원 객체를 인식하는 모델도 제작한다. 즉 우리는 하나의 데이터셋과 하나의 정보 추출과정으로 서로 다른 두 가지 결과를 제작하였다.

[Abstract]

As the content industries developed, the scope of content used expanded from two to three dimensions, and not only experts but also ordinary users wanted to create and use this content. But handling three-dimensional information requires a lot of technology and time. Therefore, this study presents a simple three-dimensional reconstruction method using SfM. When users produce video images in a simple way, datasets are augmented and increased based on this, and then carry out three-dimensional reconstruction using the increased data. It also produces models that recognize three-dimensional objects through CNN learning. In other words, we produced two different results, one dataset and one information extraction process.

색인어 : 합성곱 신경망, 3차원 재구성, 객체 인식 모델, 이미지 증강, SURF
Key word : CNN, 3D Reconstruction, Object Recognition Model, Image Augmentation, SURF

http://dx.doi.org/10.9728/dcs.2020.21.11.2011

Received 18 October 2020; **Revised** 02 November 2020
Accepted 02 November 2020

***Corresponding Author; Taemin Lee**

Tel: +82-2-824-3018
E-mail: kevinlee@cglab.cau.ac.kr

Ⅰ. 서 론

기존 2D 콘텐츠 중심의 산업은 기술의 발달로 관찰자에게 보다 더 사실적이고 생생한 입체감을 줄 수 있는 3D 콘텐츠 창작에 집중하게 되었다. 3D 콘텐츠에 집중하게 되면서, 전문가들뿐 아니라 일반 사용자들도 이 정보들을 사용하고자 하는 욕구들이 늘어나기 시작하였다. 또한 콘텐츠 산업은 이제 더 이상 전문가가 제작한 콘텐츠를 단순히 시청하거나 체험하는 것에 국한되지 않는다. 사용자가 직접 만들고 공유할 수 있는 콘텐츠가 주를 이루기 시작하였다.

하지만 3D 객체를 컴퓨터 화면 속으로 옮기는 것은 쉬운 일이 아니다. 아티스트가 아닌 이상 사실적으로 묘사를 하기 힘들뿐더러 직접 모델링을 하는 것은 고도의 기술과 시간을 필요로 한다. 특히 콘텐츠 창작가와 같은 비전문가에게는 매우 어려운 주제이다. 그래서 본 연구에서는 SfM(Struture from Motion)을 이용한 3차원 재구성(3D Reconstruction) 방법을 제시한다. 이는 물체를 단순한 방법으로 촬영하면 그에 대한 3차원 객체를 획득할 수 있게 한다. 이를 통해 비전문가 또한 쉽게 3차원 객체를 획득할 수 있도록 도울 수 있다. 이렇게 얻어진 3차원 정보들은 다양한 곳에서 사용될 수 있다[1][2].

3D 콘텐츠 제작을 위한 3차원 객체를 획득했다면 이를 응용하여 보다 더 완성도 높은 콘텐츠를 만들도록 할 수 있다. 현 콘텐츠 시장은 인터랙티브한 콘텐츠가 가장 주목을 받고 있는 추세인데, 컴퓨터 비전과 인공지능 기술 발달에 따라 물체를 인식하고 구별하여 이를 콘텐츠에 활용하는 빈도가 급격히 높아지고 있다. 하지만 객체 인식기를 제작하는 과정은 굉장히 복잡하며 그 원리를 이해하고자 한다면 꽤 어려운 일이 된다. 따라서 우리는 비전문가를 위해 객체 인식기를 자동으로 제작해주는 연구를 진행하였다. 다만 객체 인식기의 경우 다양한 환경에서 촬영을 한 이미지 데이터 셋을 입력하는 것이 일반적인 방법인데, 사용의 편의를 제공하기 위해 우리는 3차원 재구성(3D Reconstruction)에 활용한 데이터 셋을 활용하여 3차원 재구성과 객체 인식 모델 개발을 동시에 수행한다.

우리의 연구는 다음과 같은 공헌도를 가지고 있다. 첫째, 3차원 재구성 작업과 객체 인식 모델은 각각을 따로 수행할 경우 많은 시간이 소요된다. 이를 하나의 플로우로 통일시킴으로써, 시간을 단축할 수 있다. 둘째, 3D 개념의 이해 없이 구현하기 힘든 주제를 비전문가도 쉽게 활용할 수 있도록 도울 수 있다. 일반적으로 물체를 3차원으로 구성하기 위해서는 3D Max, Fusion, MAYA와 같은 전문가들이 사용하는 툴을 사용해야 한다. 하지만 비디오를 찍은 것만으로도 3차원 모델을 제작해 낼 수 있기 때문에, 사용자들에게 3D를 쉽게 접근할 수 있도록 해준다. 셋째, 서로 다른 두 가지 데이터 셋을 통일 시켰다. 3차원 재구성과 객체 인식 모델은 초기에 입력하는 데이터 셋의 특징이 원래 다르다. 3차원 재구성의 경우 단일 환경에서 촬영된 데이터 셋이어야 하지만 객체 인식기의 경우 성능 향상을 위해 다양한 환경에서 촬영된 데이터 셋을 활용하는 것이 좋다. 하지만 우리는 단일 인풋을 활용하여 두 작업을 동시에 수행하였다. 따라서 이 논문에서는 이 두 과정에 대한 데이터 셋을 단일화하고 그 수행과정을 자동화하는 연구를 진행하였다.

Ⅱ. 관련연구

SfM(Structure from Motion) 기반의 3차원 재구성 기법은 이미 많이 연구 개발이 진행된 상태로, 주로 SIFT와 SURF와 같은 이미지 특징점(Feature) 검출기를 이용하여 특징점을 찾아내고 이를 매칭함으로써 카메라의 움직임과 위치를 검출해내는 방식을 이용한다[3].

이러한 움직임 내 특징점을 매칭하는 과정에서 두 연속되는 이미지 간의 유사도가 60% 이상이 되어야 비교적 정확한 결과를 추출할 수 있다.[4]. 하지만 객체 인식을 위한 수많은 데이터 셋은 이를 만족하지 않는데, 그 이유는 사용자가 원하는 어떠한 특정 객체가 아닌 보편적인 물체에 대한 인식을 주된 목적으로 하기 때문이다.

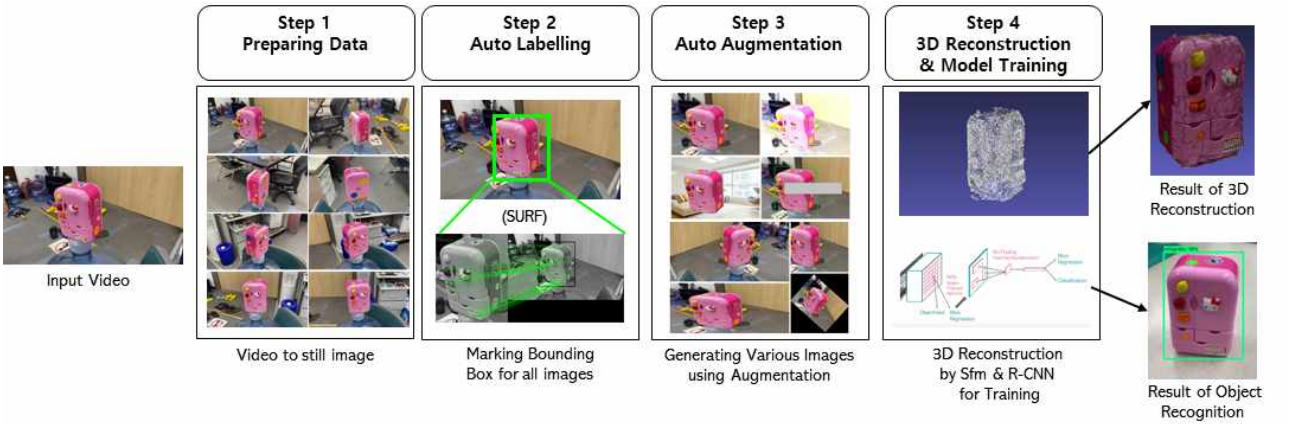

그림 1. 시스템 흐름도 (비디오가 입력으로 들어오면 4가지 스텝을 통하여 3차원 재구성과, 객체 인식 모델을 진행한다.)
Fig. 1. System Flow (When the video enters the input, three-dimensional reconstruction and object recognition models are carried out through 4 steps)

3차원 재구성을 위해 수집한 데이터셋을 그대로 객체 인식모델 제작을 위한 학습 데이터로 활용하게 되면 이미지 데이터셋의 목적이 다르기 때문에 데이터의 양이나 정보가 부족할 수 있다는 문제점이 생기게 된다. 또한 3차원 재구성을 위한 데이터셋으로 학습된 객체 인식 모델은 밝기, 물체와 카메라 간의 거리 등의 다양한 환경 변수들에 대한 대응을 하지 못한다. 선행 연구[5]에 따르면, 이미지 분류를 위한 데이터 증강 기법을 이용하였을 때 상대적으로 부족한 이미지 데이터 개수에 대한 데이터 증강을 이용하면 정확도 등의 그 학습 효과를 극대화시킬 수 있다.

따라서 본 연구에서는 하나의 오브젝트를 3차원 재구성할 때 필요한 데이터셋을 구축한다. 그 이 후에 이를 인식하는 높은 정확도의 학습 모델을 생성하기 위해서 데이터를 증강시킨다. 또한 이러한 과정을 자동화하여 그 결과물을 동시에 추출하는 것을 목적으로 하였다.

Ⅲ. 본 론

본 연구는 단일 데이터셋만을 이용하여서 주어진 물체에 대한 3차원 재구성(3D Reconstruction)과 그에 대한 객체 인식 모델을 개발하였다. [그림 1]은 우리의 전체시스템 흐름을 도식화한 것이다. 시스템 흐름도에서 볼 수 있듯이, 우리의 연구는 크게 4단계로 구분할 수 있다. 첫 번째, 데이터 수집(Preparing Data) 단계에서는 사용자에게 입력받은 동영상을 각 프레임 단위로 분해하여 데이터를 저장한다. 이는 이후에 이미지 증강등을 통해서 객체 인식 모델 제작을 위한 학습 데이터로 활용된다. 두 번째 단계는 자동 레이블링(Auto Labelling)단계이다. 이 단계에서는 데이터 수집단계에서 수집된 이미지 프레임들에 자동으로 Bounding Box를 찾아주는 작업을 한다. 뿐만 아니라 세 번째 단계인 자동 증강(Auto Augmentation) 단계에서는 1~2 단계를 통해 수집된 데이터 뿐만 아니라, 더 다양한 학습 데이터를 수집하기 위해서, 기존의 영상에서 사용된 정보들을 다른 형태로 증강시켜 학습 데이터셋을 증가 시켜주는 작업을 한다. 이렇게 수집된 이미지 데이터셋들은 마지막 단계에서 3차원 재구성을 하고, 객체 인식 모델을 학습한다.

3-1 데이터 수집(Preparing Data)

사용자로부터 데이터를 입력받는 과정이다. 3차원 재구성과 객체 인식 모델을 제작할 대상 객체를 주위로 하여 360도 돌면서 그 주위를 촬영한다. 그리함으로써 물체가 바닥에 맞닿는 면을 제외한 다른 모든 부분이 동영상 내에 모두 포함될 수 있도록 한다. 이렇게 촬영된 미가공 동영상(Raw Video)로부터 각 프레임을 이미지 데이터로 추출한다.

추출된 미가공 이미지(Raw Image) 데이터셋은 3차원 재구성을 위한 데이터셋으로써 활용된다. 연속된 이미지 데이터를 활용하여 깊이(Depth) 값을 계산할 수 있다. 또한, 이러한 미가공 이미지 데이터셋은 이후 진행할 이미지 증강(Image Augmentation) 등의 가공을 통해 객체 인식 모델 제작을 위한 학습 데이터로 활용한다.

그림 2. 이미지 증강을 위한 데이터 셋 예시
Fig. 2. Example of dataset for image augmentation

[그림 2]는 연구를 위해 수집되는 영상의 예시를 보여준다. 본 연구에서 사용된 이미지 영상은 입력된 미가공 동영상에서 592개의 프레임을 추출하여 증강용 영상으로 사용되었다.

3-2 자동 레이블링(Auto Labelling)

객체를 검출(Object Detection)하기 위해 등장한 CNN 모델인 R-CNN(Regions with CNN)[6]은 이미지 내에 있는 객체들을 찾아내어 각각이 어떤 클래스에 속하는지에 대한 예측을 수행한다. R-CNN 신경망 학습을 위해선 객체의 클래스와 위치를 이 Bounding Box를 이용하여 나타내주어야 한다. 하지만 수 천, 수 만에 달하는 이미지 데이터 셋에 대해 한 장씩 객체에 대한 레이블링을 해주는 것은 매우 많은 시간과 비용이 소모된다.

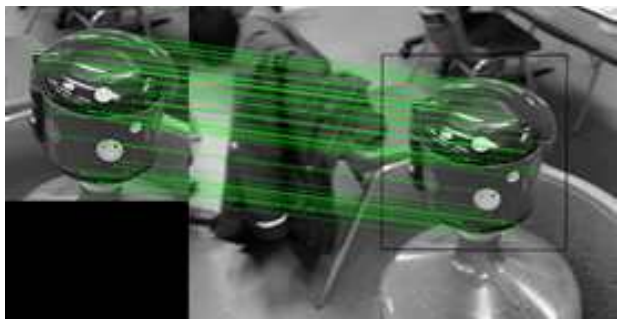

그림 3. SURF를 이용하여 다음 프레임에서 레이블링 한 결과 예시(특징점이 많이 매칭된 것을 볼 수 있다.)
Fig. 3. An example of labeling in the following frame using SURF(We can see many matched features.)

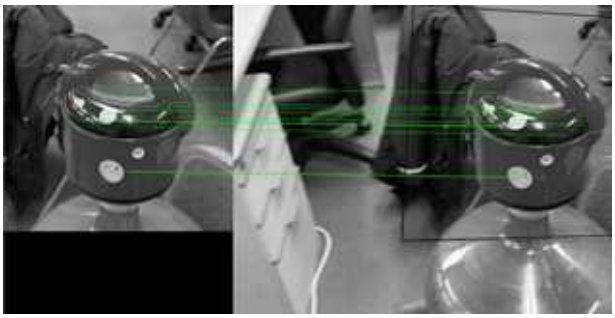

그림 4. 특징점이 적어서 자동 레이블링이 제대로 되지 않은 결과 (이 경우, 반자동화를 진행한다.)

Fig. 4. Result of poor auto-labeling due to few feature points (In this case, semi-automation proceeds).

따라서 본 연구에서는 레이블링 과정을 자동화하기 위한 방법을 제시한다.

입력받은 동영상으로부터 이미지 프레임 데이터를 얻었을 때, 사용자가 첫 프레임에 대하여 한 번의 레이블링을 직접 수행한다. 이 Bounding Box는 첫 프레임 내 객체에 대한 레이블 역할도 수행하지만, 곧 다음 프레임에서 새로운 Bounding Box를 그리기 위해 이미지 매칭을 수행하는 구역이 되기도 한다. Bounding Box 범위만큼 크롭(Crop)된 범위가 다음 프레임에서 매칭되는 범위를 SIFT[7], SURF[8] 등의 이미지 매칭 알고리즘을 통해 계산하고, 곧 그를 레이블로서 활용하여 연속적으로 계속하여 동작을 수행하게 된다. 우리의 결과에서는 SURF가 더 높은 속도와 정확도를 갖는다. 따라서 본 연구에서는 SURF를 이용한 자동레이블링을 진행하였다. [그림 3]은 SURF 매칭을 이용하여 레이블링을 자동화한 예시를 보여준다.

하지만 그 과정 중간에 프레임 간 특징점 매칭 개수가 적어 정상적으로 Bounding Box를 추정하지 못하게 되는 경우가 존재할 수 있다. Bounding Box의 넓이를 계산하여 사전에 설정한 오차를 넘어설 경우 새롭게 사용자가 Bounding Box를 그릴 수 있도록 하여 이미지 매칭이 정상적으로 이루어질 수 있도록 반자동화 할 수 있다. [그림 4]는 프레임간 특징점 매칭 개수가 적어서, Bounding Box의 결과가 정확하지 않는 예시를 보여준다. 이 경우에는 사용자가 직접 Bounding Box를 수정하여 준다.

이렇게 하면 손쉽게 객체 인식 모델 학습을 위한 이미지 데이터셋의 레이블링 작업을 마칠 수 있다. 이는 단일 객체에 대한 레이블링이라는 점에서 가능한 방식으로, 특히 3차원 재구성(3D Reconstruction)을 위한 360도 회전 동영상을 기반으로 하여 가능한 방식이라고 할 수 있다.

3-3 자동 증강(Auto Augmentation)

이전 과정까지의 이미지 데이터 셋만으로는 성능이 좋은 객체 인식기를 제작할 수 없다. 단 하나의 환경에서만 촬영된 이미지로 구성되어 있기 때문에 장소, 특히 밝기와 객체 크기 등에 있어 다양한 환경에서의 인식 성능이 매우 떨어지게 된다. 하나의 독립 환경에서 생성되어 다양한 환경 변수의 변화에 대응하지 못하기 때문에 우리는 이미지 데이터의 증강 작업을 실시하였다. 이미지 증강(Image Augmentation)은 부족한 데이터 수를 보완하기 위하여 실시하는 작업으로, 이 연구에서는 증강된 이미지가 원본 이미지와 레이블이 동일한 증강과 동일하지 않은 증강으로 나누어 작업을 실시하였다. 즉, 이미지 내에서의 객체의 Bounding Box 크기와 위치가 증강 후에도 동일한 경우가 바로 원본 이미지와 레이블이 동일한 경우를 의미하는 것이다.

1) 원본 이미지와 레이블이 동일한 증강

원본 이미지와 레이블이 동일한 증강은 아래와 같이 크게 세 가지 방식으로 진행하였다. [그림 5]은 그 결과로 나온 이미지들의 예시를 보여준다.

1-1) 배경 교체

원본 이미지 데이터 셋의 경우 한 공간에서 촬영된 동영상을 기반으로 하기 때문에 이 데이터 셋만을 이용해선 다양한 환경에서 작동하는 객체 인식기를 제작하기 힘들다. 따라서 원하는 객체를 제외한 배경을 교체하는 작업을 진행하였다. 먼저, 배경 제거에는 Grabcut 알고리즘이 사용되었다. 전경에 해당되는 이미지는 기존 이미지의 레이블에 포함되어있는 Bounding Box를 활용하여 지정하였다. 그리하여 배경이 제거되어 추출된 전경에 무작위 배경 이미지를 합성시켜 교체를 진행하였다.

1-2) 색상 및 밝기 증강

원본 이미지 데이터 셋의 경우 다양한 촬영의 경우에서 환경 변수를 모두 고려하지 못한다. 이를테면 만약 원본 이미지 데이터 셋이 어두운 곳에서 촬영이 되었다면, 그를 통해서만 학습된 객체 인식기는 밝은 환경에서의 성능이 떨어지게 된다. 또한, 카메라의 종류나 성능에 따라 저장되는 이미지의 색상이 다를 수 있다. 이런 다양한 환경 변수에 대응하기 위하여 밝기와 색상을 일부 조정하는 증강 작업을 진행하였다.

그림 5. 원본 이미지와 동일한 레이블링(왼쪽 위 : 원본, 오른쪽 위 : 배경 교체, 왼쪽 아래 : 색상 및 밝기 증감, 오른쪽 아래 : 무작위 임의 구역 제거)

Fig. 5. Same labeling as the original image (left-top : source image , right-top : background replacement, left-bottom : color and brightness increments, right- bottom : remove random area)

1-3) 무작위 임의 구역 제거

객체 인식 도중 특정한 물체 혹은 손가락 등으로 객체가 가려지는 경우가 발생할 수 있다. 객체의 일부가 임의로 제거되더라도 인식을 성공적으로 수행할 수 있도록 무작위로 객체 내 임의 구역을 제거하는 작업을 진행하였다.

2) 원본 이미지와 레이블이 동일하지 않은 증강

원본 이미지와 레이블이 동일하지 않은 증강은 아래와 같이 크게 네 가지 방식으로 진행하였다. [그림 6]은 그 결과로 나온 이미지들의 예시를 보여준다. 이미지의 검은 배경은 원본 사이즈와의 비율 비교를 위하여 시각화한 것이고, 실제로 학습이 진행될 때는 이 부분이 제거되어 진행되었다.

2-1) 반전

이미지를 상하 반전, 좌우 반전, 상하좌우 반전 시키는 증강을 진행하였다. 이 경우 Bounding Box의 좌표는 원점 중심을 기준으로 하여 값을 반전시켜주었다.

2-2) 회전

이미지를 회전시키는 증강을 진행하였다. 회전 정도를 바꾸어 가며 다양한 이미지를 제작하였고, 그에 따라 Bouding Box도 회전되었다.

2-3) 잘라내기

원본 이미지의 크기를 변경하되, 왜곡시키지 않고 잘라내어 증강을 실시하였다. 이 경우, Bounding Box 영역을 침범하여 잘라내었는지 침범하지 않고 잘라내었는지를 고려하여 레이블을 다시 지정해주었다.

2-4) 왜곡

원본 이미지의 크기를 변경하며 동시에 왜곡시키는 증강을 실시하였다. 이미지 크기가 변한 비율을 계산하여 동일하게 Bounding Box의 크기를 조정하였다.

3-4 3차원 재구성과 모델 학습 (3D Reconstruction and Model Training)

완성된 3차원 재구성을 위한 이미지 데이터 셋, 그리고 학습용 이미지 데이터 셋을 통해 각각 원래 목적의 작업을 수행한다. 3차원 재구성의 경우 SfM(Structure from Motion) 기법을 이용하여 두 연속 2차원 프레임 간의 3차원 구조를 계산한다. 계산된 깊이 맵(Depth Map)을 예측한 뒤, 추출된 Point Cloud에 대한 메쉬(Mesh)화와 텍스처 매핑(Texture Mapping)을 수행하여 객체에 대한 최종 모델을 추출한다. 3차원 재구성을 위한 단계는 [표 1]에 나와 있는 순서대로 진행하였다[9][10].

또한 학습의 경우 R-CNN 계열 중 빠른 속도와 높은 성능을 자랑하는 Faster R-CNN을 이용한다. 이외 신경망은 레이블링 과정이 보다 복잡하여 자동화에 적합하지 않거나 성능이 떨어진다는 단점이 존재한다. 학습용 데이터와 검증용 데이터를 7:3 비율로 분할하여 학습을 수행하였다. 모델 학습을 위한 R-CNN 구조는 [그림 7]과 같다.

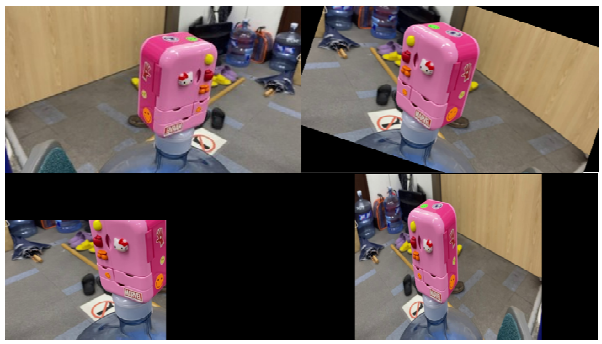

그림 6. 원본 이미지와 동일하지 않은 레이블링(왼쪽 위 : 반전, 오른쪽 위 : 회전, 왼쪽 아래 : 잘라내기, 오른쪽 아래 : 왜곡)

Fig. 6. Labeling not identical to the original image (top left : invert, top right : rotation, bottom left : cut, bottom right : distortion)

표 1. 3차원 재구성을 위한 구현 단계

Table 1. Implementation Steps for 3D Reconstruction

level	state	contents
1	CameraInit	Initialize camera
2	FeatureExtraction	Extract features in an image using SURF algorihtm
3	ImageMatching	Find an image looking at the same area of the scene
4	FeatureMatching	Match feature points between candidate images
5	StrurctureFromMotion	Estimate 3D structures from 2D image sequences using Sfm algorithm
6	PrepareDenseScene	Prepare process for Depth-map calculation
7	DepthMap	Create Map by searching for the depth value of each pixel for all frames
8	DepthMapFilter	Maintain consistency with Depth-map calculated independently
9	Meshing	Generate mesh of point cloud to create geomnetric surfaces
10	MeshFiltering	Simplify Mesh to reduce unnecessary advantages
11	Texturing	Map a calculated UV map to the generated Mesh

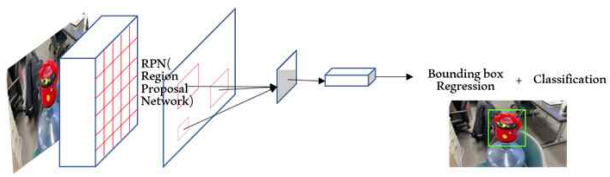

그림 7. 객체 인식을 위한 Faster R-CNN 학습 모델

Fig. 7. Faster R-CNN Learning Model for Object Recognition

Ⅳ. 실험결과

4-1 구현 환경

본 연구에는 Intel(R) Core(TM) i7-6700 CPU @ 3.40GHz, RAM 16GB와 NVIDEA RTX 2060 SUPER GPU 환경에서 실행되었다. Python 3.6.5.를 이용하여 프로그램 되었으며 CUDA 10.1 라이브러리와 OpenCV-contrib 3.4.2.16, Numpy 1.18.0 Tensorflow 1.15.0 패키지[11]가 사용되었다. 프레임 워크는 AliceVision 2.3.0 for meshroom이고 Meshroom software 2020.1.1.[12]을 사용하였다.

학습 모델 예측은 epoch=10000으로 세팅하여 진행하였으며, 레이블의 개수가 적어서 한번의 학습당 약 25분의 시간이 소요되었다. 3차원 재구성의 경우, 예제에서 사용된 냉장고를 재구성하는 데 걸리는 시간은 약 23분이고, 각 11단계에서 소요되는 시간은 [표 2]와 같다.

표 2. 3차원 재구성에서 소요된 시간 (s : 초, m : 분)
Table 2. Spending time for 3D Reconstruction(s : second, m : minute)

level	stage	time
1	CameraInit	0.3s
2	FeatureExtraction	32s
3	ImageMatching	0.3s
4	FeatureMatching	3m 28.3s
5	StrurctureFromMotion	1m 23.8s
6	PrepareDenseScene	19.4s
7	DepthMap	10m 5.2s
8	DepthMapFilter	2m 11.1s
9	Meshing	3m 24.7s
10	MeshFiltering	6.1s
11	Texturing	1m 37.8s
Total		23m 8.0s

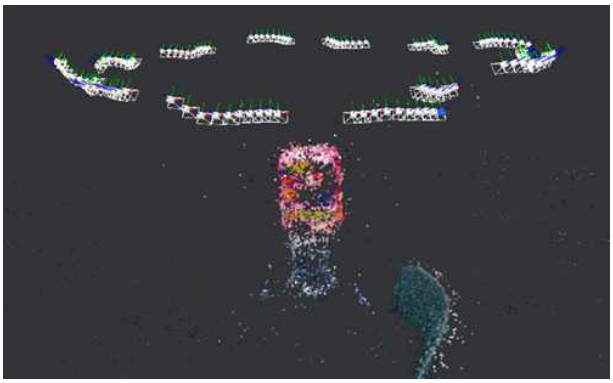

그림 8. 프레임 이미지를 통한 카메라 위치/방향 예측 결과
Fig. 8. Results of camera position/direction prediction via frame image

4-2 결과

1) 3차원 재구성 결과

비디오에서 분리된 프레임중 120장을 랜덤으로 선택하여 3차원 재구성을 진행하였다. [그림 8]은 사진을 통하여, 카메라의 위치와 방향을 예측한 결과이다. 약 600장의 프레임에서 5분의 1인 120장을 랜덤으로 추출하였기 때문에, 360도를 기준으로 물체를 둘러 싸는 다양한 각도에서 카메라의 위치가 추출되는 것을 확인할 수 있다. 이를 통하여 3차원을 재구성하기 때문에 우리는 정확한 3차원 정보를 구할 수 있게 된다.

프레임 이미지에서 얻어진 정보들을 이용하여, 우리는 포인트 클라우드를 추출하고 포인트 클라우드 기반으로 메쉬화 시켜 결과를 만든다. 그 과정의 예시는 [그림 9]에서 볼 수 있다. [그림 9]에서 얻어진 메쉬 모델을 이용하여 우리는 다양한 각도에서의 3차원 모델을 제작할 수 있다. 본 연구에는 냉장고로 실험하였고, 그 결과는 [그림 10]에서 볼 수 있다.

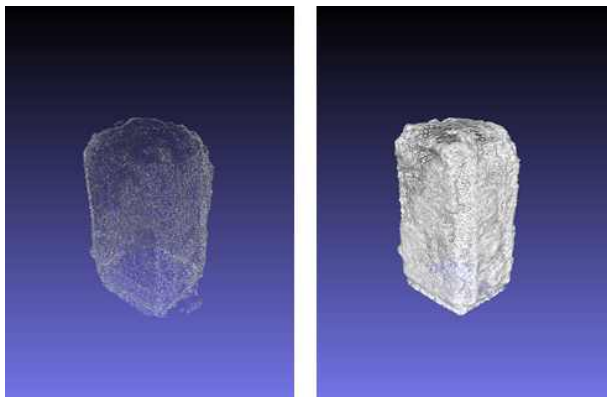

그림 9. 3차원 재구성을 위한 포인트 클라우드(왼)와 메쉬모델(오) (포인트 클라우드를 통해서 메쉬 모델이 제작되고 메쉬 모델을 통해서 3차원으로 재구성된 모델을 제작한다.)
Fig. 9. Point cloud (L) and mesh model (R) for 3D reconstruction (Point cloud creates a mesh model and a three-dimensional reconstruction through a mesh model).

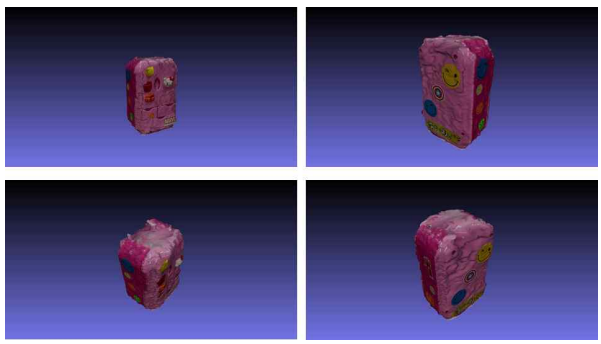

그림 10. 냉장고 3차원 재구성 결과 (오른쪽 상단에 있는 이미지가 [그림 9]의 메쉬모델을 통해 생성된 결과이다.)
Fig. 10. Result of refrigerator 3D reconstruction results (The image on the upper right is produced by the mesh model in Fig 9.).

2) 객체 인식 모델

임의의 배경에 놓여져 있는 냉장고를 실시간을 캡처했을 때 이 물체를 냉장고로 인지할 수 있는지에 대하여 실험하였다. 처음 입력된 비디오에서 추출된 프레임만 가지고 Faster R-CNN으로 학습한 모델과 자동 이미지 증강을 통해 더 많은 데이터를 가지고 학습한 모델을 비교하여, 이미지 증강의 필요성에 대해서도 확인해 보았다. [그림 11]은 두 가지 환경(비디오 프레임만 가지고 학습한 모델과 이미지 증강을 통한 입력 이미지 개수 확장한 모델)에서의 냉장고를 인식하는 정확도를 비교한 것이다. 그림에서 볼 수 있듯이, 정면이나 후면의 경우에는 두 환경의 차이가 크게 없는 것을 볼 수 있지만, 멀리서 캡처되거나, 위/아래와 같은 캡처 이미지가 약간의 왜곡이 있는 경우에는 입력 이미지를 늘려서 학습한 모델이 매우 높은 정확도를 갖는 것을 볼 수 있었다.

Ⅴ. 결 론

본 논문에서는 비디오 기반의 단일 데이터셋을 이용하여 하여 3차원 재구성과 객체 인식기 제작을 동시에 수행하는 연구를 제시하였다. 주어진 비디오를 프레임 단위로 분할하여, 자동 레이블링을 하고, 적은 데이터양을 늘리기 위해 자동으로 이미지를 증강시킨다. 이는 3차원 재구성이나 모델 예측을 위해 필요한 정보양을 늘려줌으로써, 결과의 정확도를 향상 시키는데 도움을 주었다. 마지막으로 3차원 재구성 데이터셋을 통하여 객체 인식 모델을 제작하기 위해 여러 환경 변수를 고려한 데이터 증강을 수행하였다. 이를 통해 두 작업을 수행하는 데에 있어 시간 및 비용 비효율성을 줄이고 비전문가 또한 쉽게 원하는 3D 오브젝트와 이를 활용한 실시간 객체 인식 도구를 활용할 수 있도록 하였다.

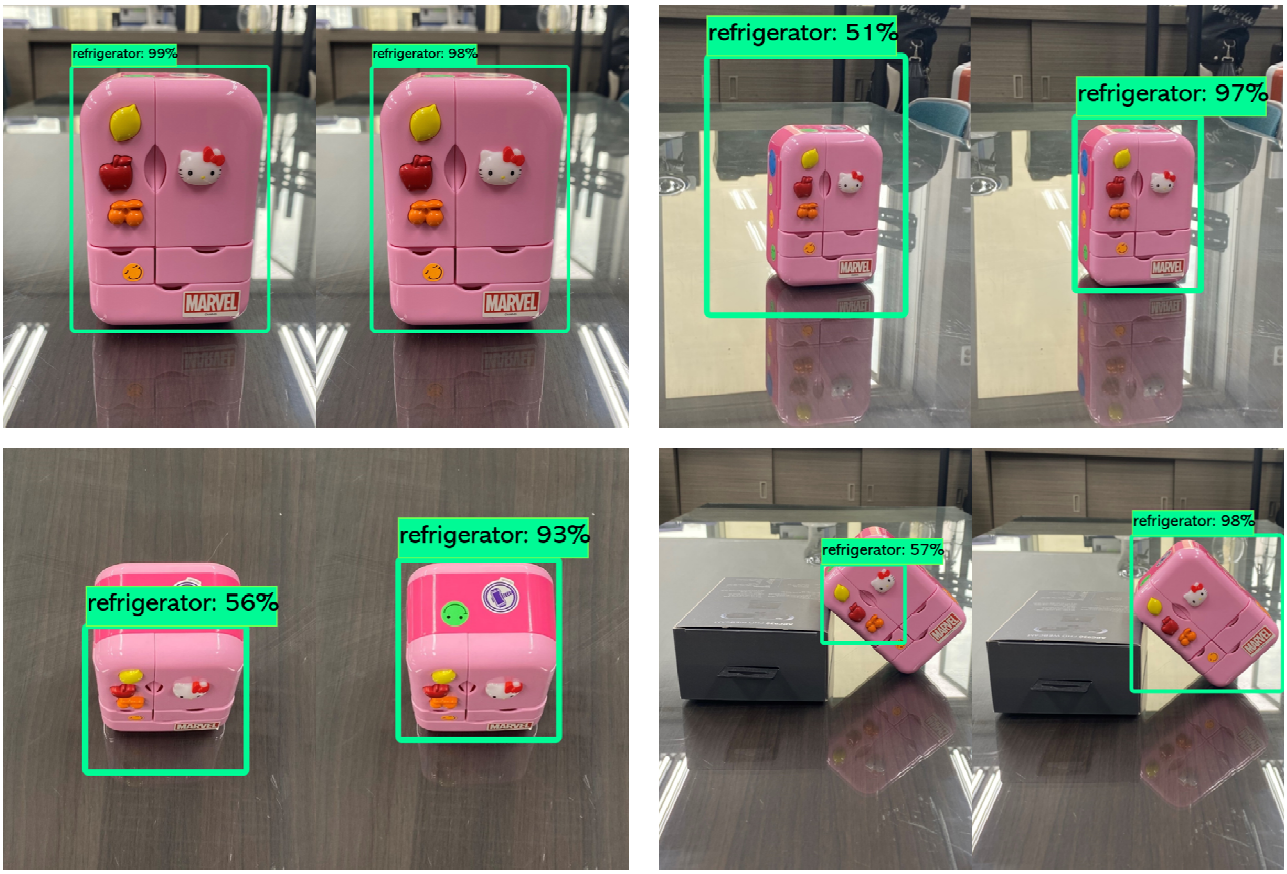

그림 11. 객체 인식 모델 결과 : 왼쪽은 이미지 증강 없이 프레임만을 이용한 학습 모델 예측 결과이고, 오른쪽은 이미지 증강을 통해 학습 데이터셋을 증가시켜 학습한 모델 예측 결과이다. (왼쪽 상단은 냉장고의 정면을 가까이 찍은 영상이다. 결과에서처럼 두 모델의 정확도에 차이가 거의 없다. 오른쪽 상단은 멀리서 정면을 찍어서 사용한 비교한 결과이다. 아래쪽 두 비교군은 이미지의 형태가 회전을 통해 왜곡되어 찍혔을 때 나타나는 결과 차이를 보여준다.)

Fig. 11. Object recognition model result: The left side is the result of prediction of learning model using frame only without image enhancement, and the right side is the result of prediction of model learned by increasing learning dataset through image enhancement. (Top left is a close-up shot of the front of the refrigerator. As in the results, there is little difference in the accuracy of the two models. The upper right hand is a comparison using a front-end shot from a distance. The bottom two comparators show the difference in results when the image is distorted by rotation.)

우리의 연구는 2D 비디오 셋을 이용하여 3차원 모델을 구성해내기 때문에, 특징을 정확하게 인지못하는 경우 3차원 모델의 결과가 좋지 못한 경우가 존재한다. 비디오를 찍을 때 3차원 정보를 수집할 수 있다면 더 높은 정확도와 품질의 3차원 재구성을 진행할 수 있을 것으로 예상된다. 이는 객체 인식 모델의 예측 정확도 향상에도 도움을 줄 수 있을 것으로 예상된다. 우리의 연구는 향후 콘텐츠 산업의 발전에 있어 기폭제 역할을 할 수 있을 것으로 예상하며, 3차원 공간 내 인식 객체 위 치 추정 기술에 응용될 수 있을 것으로 예상된다.

감사의 글

이 논문은 과학기술정보통신부가 지원한 '5G 기반 VR·AR 디바이스 핵심기술개발 사업'으로 지원을 받아 수행된 연구 결과입니다. [과제명: 가상공간구성을 위한 5G 기반 3D 공간 스캔 디바이스 기술 개발 / 과제고유번호 : IITP202]

참고문헌

[1] M. Kim, H. Hong, "2D - 3D Conversion Method Based on Scene Space Reconstruction", *The journal of the Korea Contents Society*, 14(7), pp. 1-9, 2014.

[2] S. Jung, J. Lee, "User-friendly 3D Object Reconstruction Method based on Structured Light in Ubiquitous Environments" *The journal of the Korea Contents Society*, 13(11), pp. 523-532, 2013.

[3] M.J. Westoby, J. Brasington, N.F. Glasser, M. J.Hambrey, J.M. Reynolds "Structure-from-Motion' photogrammetry: A low-cost, effective tool for geoscience applications", *Geomorphology*, 179, pp. 300-314, 2012.

[4] Kraus, K., Photogrammetry: Geometry from Image and Laser Scans, *Walter de Gruyter*, pp. 459, 2007.

[5] Hiroshi Inoue, "Data Augmentation by Pairing Samples for Image Classification", *arXiv*, 1801.02929, 2018.

[6] S. Ren, K. He, R. Girshick and J. Sun, "Faster R-CNN : Towards Real-Time Object Detection with Region Proposal Networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(6), pp. 1137-1149, 2017, doi:10.1109/TPAMI.2016.25 77031.

[7] PC Ng, S. Henikoff, "SIFT: Predicting amino acid changes that affect protein function", *Nucleic Acids Res*., 31, pp.3812-3814, 2003.

[8] H. Bay, A. Ess, T. Tuytelaars and L. Van Gool, "Speeded-up robust features (SURF)", *Comput. Vis. Image Understanding*, 110(3), pp. 346-359, 2008.

[9] R. Otero, M. Delbracio, "Anatomy of the SIFT Method", *Image Processing On Line*, 4, pp. 370-396, 2014.

[10] Y. Li, S. Wnag, Q. Tian and X. Ding, "A survey of recent advances in visual feature detection", *NeuroComputer*, 149, pp. 736-751, 2015.

[11] Tensorflow, TensorFlow Model Garden, 2020, GitHub repository, https://github.com/charlespwd/project-title

[12] Alicevision, Meshroom, 2020, GitHub repository, https://github.com/alicevision/meshroom

이나혁(Nahyuk Lee)

2019년～현재 : 중앙대학교 학사 과정

※관심분야 : 컴퓨터비전(Computer Vision), 기계학습(Machine Learning)

이경택(Kyung Taek Lee)

1996년 : 인하대하교 대학원 (공학석사)
2008년 : 연세대학교 (공학박사-전기전자공학)

1996년～1998년: ㈜인켈
1998년～2001년: ㈜아이앤씨테크놀로지
2015년～2016년: Carnegie Mellon University, HCII(Human Computer Interaction Institute) Visiting Researcher
2006년～현 재: 한국전자기술연구원 콘텐츠응용연구센터 센터장
※관심분야 : XR(eXtended Reality), 디지털트윈, 3D 공간복원 등

박영섭(Youngsup Park)

1995년 : 대전대학교 공학사
2001년 : 중앙대학교 첨단영상대학원 영상공학과(공학석사-컴퓨터그래픽스및가상현실)
2006년 : 중앙대학교 일반대학원(공학박사-컴퓨터그래픽스)

1995년~1998년 : 와이즈컨트롤 대리
2007년~2008년 : 중앙대학교 정보통신연구소 연구교수
2009년~2015년 : 에이알비전 기술이사
2015년~2015년 : 성광유니텍 연구소장
2015년~현재 : 이노시뮬레이션 수석연구원
※관심분야 : XR(eXtended Reality), 디지털 트윈(Digital Twin), 딥러닝(Deep Learning)

서상현(Sanghyun Seo)

1998년 : 중앙대학교 컴퓨터공학과(공학사)
2000년 : 중앙대학교 첨단영상대학원 영상공학과(공학석사-컴퓨터그래픽스)
2010년 : 중앙대학교 첨단영상대학원 영상공학과(공학박사-컴퓨터그래픽스및가상환경)

2002년~2005년 : ㈜지노시스템, 다차원공간기술 연구소, 선임연구원
2010년~2011년 : 중앙대학교, 박사후연구원
2011년~2013년 : 프랑스 리옹 1대학, LIRIS 연구소, Post-Doc
2013년~2016년 : 한국전자통신연구원, 선임연구원
2016년~2019년 : 성결대학교 미디어소프트웨어학부 조교수
2019년~현재 : 중앙대학교 예술공학대학 컴퓨터예술학부 부교수
※관심분야 : 컴퓨터그래픽스(Computer Graphics), 비사실적 렌더링(Non-Photorealistic Rendering), 가상/증강현실(Virtucal Reality/Augmented Reality), 게임 기술(Game Technology)

이태민(Taemin Lee)

2011년 : 중앙대학교 컴퓨터공학과(공학사)
2013년 : 중앙대학교 일반대학원(공학석사-컴퓨터그래픽스)
2019년 : 중앙대학교 일반대학원(공학박사-컴퓨터그래픽스)

2019년~현재 : 중앙대학교 다빈치 SW교육원 특임교수
※관심분야 : 비사실적 렌더링(Non-Photorealistic Rendering), 색상 이론(Color Theory), 감성 컴퓨팅(emotional computing), 인공지능(Artificial Intelligence)