

## Learning 3D Representation for Visual Perception

2017320114 컴퓨터학과 최재원

이번 수업은 김승룡 교수님의 Learning 3D representation for visual perception이라는 주제를 중심으로 AI, 딥러닝, 컴퓨터 비전에 대해서 배울 수 있는 시간을 가졌다. 특히 그 중에서도 3D 컴퓨터 비전이라는 것은 무엇일까에 대해서 심층적으로 알 수 있었다. 현재 우리가 풀어야 될 문제들과 해결해야 되는 문제들은 다 3D 공간상에 존재를 하는데, 우리가 가지고 있는 영상들, 또는 이미지들은 2D 이미지이기 때문에 이러한 2차원 이미지들로부터 다시 한번 3차원 정보를 복원해내는 것은 정말로 중요한 기술 중 하나다. 따라서 3D 컴퓨터 비전은 현재 매우 중요한 분야이다.

3차원 정보를 얻어내는 것은 크게 두가지가 있다. Active sensing과 Passive Sensing이다. Active sensing은 Raider 혹은 Kinect처럼 어떠한 하드웨어를 가지고 3차원을 복원해 내려고 하는 연구들을 말하고 Passive Sensing은 2D로 찍힌 많은 영상들을 가지고 3차원을 복원해내고자 하는 3D reconstruction이라고 할 수 있다. 인간의 눈은 왼쪽과 오른쪽이 보이는 영상이 다르다(Disparity). 즉 가깝고 먼 정도의 따라서 보이는 view 자체가 달라진다. 따라서 이러한 human eye system에 착안해서 영상을 다양한 viewpoint에서 찍어서 그러한 영상들의 정보를 가지고 3차원을 복원해보자라는 것이 Stereo 3D reconstruction이라고 할 수 있다. Stereo Matching이란 left image와 right image를 주고 한 점 마다 내 search range를 기반으로 matching cost를 만들어서 어떤 픽셀이 가장 나랑 매칭이 잘 되어있는가, 즉 disparity의 확률을 예측을 하고 실제로 이 점에서의 깊이 정보가 얼마큼인지를 알아내는 (Aggregation) 과정이다. 또한 교수님께서 Kitti라는 벤치마크에서 이러한 stereo matching의 depth 정도를 경쟁하는 표를 보고 3차원을 복원해 내는 것은 어렵고 현재 매우 각광 받는 기술이라는 것을 알게 되었다.

이전까지는 2 view를 기준으로 찾는 것이었다면 이번에는 multi camera를 이용해서 3D를 찾는 연구로 확장이 된다. 대표적인 연구 분야가 Structure From Motion(Multi View

Stereo)라고 하는 분야라고 하는데 다양한 각도에서 촬영된 사진들을 stereo matching 등을 통해서 또 하나의 3차원으로 복원해 내는 것이라고 할 수 있다. 예시로 구글 3D 지도 등이 있다. Multi View Stereo는 내 reference view와 다른 view의 영상들을 가지고 각각을 stereo라고 가정을 하고 probability cost를 가지게 해서 딥러닝 네트워크를 이용해서 Multi View Depths를 찾아내는 것이다. Multi View Stereo의 핵심은 동일 Scene을 수 많은 view에서 찍어내고 그 다른 view에서 찍힌 영상들 간의 대응점을 찾는 문제라고 볼 수 있다. 현재는 딥러닝을 이용해서 이러한 문제를 풀지만 예전에는 optimization based, learning-based 등 여러가지 방법들이 있었다.

앞선 두 가지 큰 기술 외에도 여러가지 기술들이 있다. Photometric Stereo란 내가 reconstruct 하고 싶은 scene을 다양한 조명 환경에서 찍는 것이다. 조명에 따른 그림자가 달라지기 때문에 그 정보로부터 3차원을 복원할 수 있다. 이것에 대한 예시로 오바마 전 미국 대통령의 석고를 찍어내는 영상을 보게 되었는데, 흥미롭게 느껴졌다. 3차원은 센서를 통해서도 얻어낼 수 있다. 우리가 아는 자율주행 차는 LiDar 센서를 통해 얻는 경우가 많다. 하지만 이 LiDar 센서의 단점은 이 센싱의 범위가 너무 sparse하다는 것이다. 따라서 많은 연구들은 이 LiDar 센서와 카메라 센서들을 같이 활용해서 초정밀 3차원을 복원해낸다.

가장 최근의 3D reconstruction은 Monocular 3D reconstruction을 사용한다. 이 기술은 monocular 즉 single image에서 어떤 knowledge를 통해서 학습을 잘 시켜 놓으면 single image를 가지고도 어떤 깊이 정보를 얻어낼 수 있지 않을까라고 하는 것이 이 기술의 핵심이고 현재 3D reconstruction이 가야하는 방향이기도 하다.

이번 강의를 통해 3D 복원 기술에 조금이나마 알 수 있었다. 또한 아직 기술의 발전이 진행중이라는 것에도 매우 흥미를 가지게 되었고 기회가 된다면 이 분야로서의 진로도 생각을 해볼 것 같다. 평소에 컴퓨터 비전에 관심이 많았는데 3D reconstruction에 대한 high level의 지식을 습득한 것에 대해 매우 만족하고 뿌듯한 강의였다.