360° 영상 응용을 위한 벤치마크 데이터 생성 연구

이종성, *이의진
서울과학기술대학교
syniez@g.seoultech.ac.kr, *yeejinlee@seoultech.ac.kr

Benchmark Dataset Generation for 360-degree Image Applications

Jongsung Lee, *Yeejin Lee
Seoul National University of Science and Technology

요 약

최근 가상현실 및 증강 현실에 대한 관심도가 높아지면서, 깊이 추정, 객체 인식, 영상 분할 등의 다양한 컴퓨터 비전 알고리즘을 360° 영상에 적용하는 연구가 활발히 진행되고 있다. 이 중, 다수의 RGB 카메라를 활용하여 3 차원 정보를 추출하는 깊이 추정 기술은 보다 나은 몰입감을 제공하기 위한 핵심 기술이다. 그러나 깊이 추정 알고리즘의 객관적 성능 평가를 위한 정제된 360° 영상 데이터셋은 극히 부족하며, 이로 인하여 관련 분야 연구에 한계가 있다. 따라서 본 논문에서는 객관적인 알고리즘 성능 평가가 가능하며, 정제된 360° 동영상 데이터셋을 제안하고, 추후 다양한 360° 영상 응용 알고리즘 개발에 활용하고자 한다.

1. 서론

최근 가상 현실(Virtual Reality), 증강 현실(Augmented Reality) 및 혼합 현실(Mixed Reality)에 관한 관심과 수요가 증가하면서 360° 영상의 수요 또한 증가하였다. 360° 영상에 대한 깊이 추정(Depth Estimation), 시점 합성(View Synthesis), 객체 인식(Object Recognition), 영상 분할(Semantic Segmentation) 등 여러 분야의 연구가 진행되고 있지만, 연구를 위한 콘텐츠 생성을 위해서는 고가의 장비가 필요하며 이로 인하여 콘텐츠 수급에 어려움을 겪고 있다. 따라서 저렴한 다수의 카메라를 활용하여 획득한 영상을 사용하여 360° 영상 콘텐츠를 수급하려는 노력을 하고 있다. 최근에는 넓은 시야각(Field of View, FoV)을 가진 영상을 적은 수의 카메라를 사용하여 획득하기 위하여 비교적 저렴한 360°카메라를 활용하는 연구도 활발히 진행되고 있다.

특히, 다수의 RGB 카메라를 활용하여 3 차원 정보를 추출하는 깊이 추정 기술은 보다 나은 몰입감 있는 가상, 증강, 및 혼합 현실의 콘텐츠를 제공하기 위한 핵심 기술이다. 서로 다른 시점을 가진 영상을 활용한 깊이 추정은 중첩된 영역을 가진 여러 장의 영상을 획득한 후, 기준 영상 대한 시점이 다른 영상의 위치 관계를 파악하여 정렬(Registration)하고, 정렬한 영상의 각 화소(Pixel)를 중심으로 유사한 영역을 찾아 시점이 다른 영상 간의 시차(Parallax)에 의한 디스패리티(Disparity)를 계산한 후, 계산한 시차를 깊이로 변환하여 깊이 맵(Depth Map)을 생성하는 과정으로 이루어진다. 현재까지 제안된 대부분의 디스패리티 추정 알고리즘들은 완벽하게 정렬된 영상들을 입력으로 하여 수평 혹은 수직 한 방향으로만 유사 영역을 탐색하므로 카메라 간의 정확한 위치 관계 파악 및 영상을 오차 없이 정렬하는 과정이 선행되어야만 한다. 다수의 영상들이 촬영된 시점을 파악하고 정렬하기 위해서는 영상마다

(a) (b)

그림 1. (a) Matterport3D dataset, (b) Stanford 2D-3D dataset

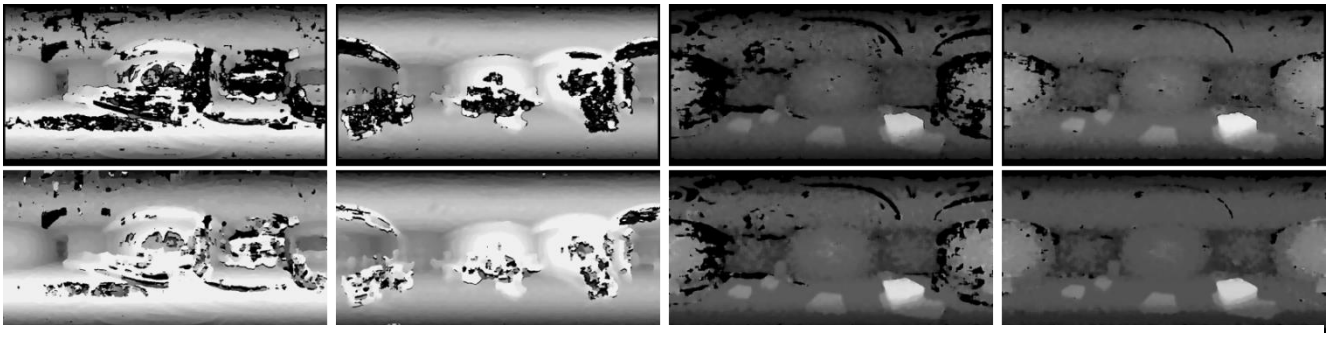
특징점(Feature Descriptor)을 추출하고, 추출된 특징점 중 영상 간에 동일하게 추출된 점들을 대응(Feature Matching)하여 대응된 특징점들의 영상 내 위치 관계 변화를 계산한 후, 획득한 영상들을 수평 혹은 수직 방향으로 재배열한다.

그러나 360° 영상의 경우, 구형 영상을 2 차원으로 투영하는 과정에서 위도에 따라 기하적 왜곡 정도가 다르며, 극점으로 갈수록 왜곡이 심해진다. 이러한 기하학적인 특성으로 인하여 다수의 360° 영상을 사용하여 깊이 추정을 수행할 경우, 영상 정렬을 위한 특징점 추출 및 특징점 매칭의 정확도가 떨어지며, 추정된 깊이 맵의 정확도 또한 보장할 수 없다. 뿐만 아니라, 실제 촬영한 영상을 활용하여 깊이 맵을 추정했다 할지라도, 영상 촬영 시 실측된 깊이 정보(Ground Truth)를 얻기 어려워 생성된 깊이 맵의 정확도의 객관적인 평가가 어렵다. 따라서 생성된 깊이 맵의 정확도와 사용된 알고리즘의 객관적인 성능 평가를 위해서는 그라운드 트루스(Ground Truth)가 존재하는 완벽하게 정렬된 데이터셋이 필요하다.

가장 대표적인 360° 영상 데이터셋으로는 Matterport3D [1]와 Stanford 2D-3D [3]등이 있다. 그러나 두 데이터셋 모두 데이터셋을 만들 때 실제 환경에서 촬영된 영상을 3 차원 좌표계로 투영한 다음, 다시 2 차원 영상 좌표계로 재투영하는 방식으로 제작되었기 때문에 실제 영상처럼 보인다는 장점이 있지만 여러 번 좌표계를 변환하며 그림 1 과 같이 물체 일부분이 비어 있거나 원본 영상과는 다르게 투영되는 등의 오차가 발생하여 영상의 품질이 떨어질 수 있다. 이러한 오차들은 깊이 추정이나 시점 합성 등의 분야에서 결과 영상이 원본 영상과 달라지는 문제를 일으킬 수 있다. 따라서 본 논문에서는 위에 기술된 현재 360° 영상 데이터셋의 한계점을 극복하고자 객관적인 알고리즘의 성능 평가가 가능하며 그라운드 트루스가 존재하는 360° 영상 벤치마크 데이터셋을 제안하며, 추후 360° 영상 응용 알고리즘 개발에 활용하고자 한다.

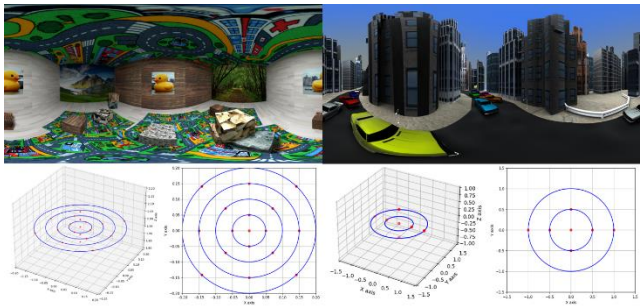
2. 제안하는 벤치마크 데이터셋

일반적으로 360° 영상 깊이 추정 알고리즘에서는 시차를 추정하기 위해 두 영상을 위아래로 정렬한다 [5], [6]. 이는 360° 영상에서 등방성으로(Isotropic) 존재하는 시차를 각 시차(Angle Disparity)로 변환함으로써 등방성이 아닌 한 방향으로만 시차가 발생하도록 하여 디스패리티 추정의 효율성을 높이기 위함이다. 디스패리티 추정은 각 화소를 중심으로 주변 영역과 가장 유사한 영역을 다른 시점의 영상에서 찾아 상대적인 좌표를 결정하며, 이때, 유사 영역 매칭 정확도 향상을 위해서는 시점이 다른 영상들이 오차 없이 위아래로 정렬되어 있어야 한다. 그러나 1 장에서 논의한 바와 같이 360° 영상의 경우, 영상의 고유한 기하학적인 특성으로 인하여 정렬을 위한 특징점 매칭의 정확도가 낮아진다. 따라서 제안하는 벤치마크 데이터셋에서는 다수의 카메라를 그림 2 와 같이 수평·수직 방향 모두에 배치하여 깊이 추정 알고리즘의 입력 영상을 오차가 존재할 수도 있는 영상 정렬 과정 없이 완벽하게 정렬된 상태로 제공하여 사용자의 응용 목적에 따라 카메라 위치를 선택할 수 있도록 한다. 또한, 자연스러운 동영상 연출을 위하여 움직이는 여러 개의 물체가 영상 내에 배치되어 있으며, 제공하는 데이터셋의 블렌더(Blender) 파일 내에서 사용자가 원하는 대로 객체를 추가할 수 있고 프레임 단위로 동영상의 길이도 조절 가능하다.



(a) (b) (c) (d)

그림 3. 첫째행: StereoBM, 둘째행: StereoSGBM, (a) MP3D, (b) Stanford 2D-3D, (c) 제안하는 데이터셋 (베이스라인:0.2m), (d) 제안하는 데이터셋 (베이스라인:0.05m)



(a) (b) (c) (d)

그림 2. (a) 실내 데이터셋 전체 카메라 배치도, (b) 실내 데이터셋 카메라 수평 배치도, (c) 야외 데이터셋 전체 카메라 배치도, (d) 야외 데이터셋 카메라 수평 배치도

제안하는 360° 동영상 데이터셋은 다양성 확보를 위하여 실내·실외 환경을 모두 고려하여 제작하였다. 실내 데이터셋은 Blender[2]를 사용하여 제작된 Room 데이터셋[4]을 다양한 분야에 응용할 수 있도록 보완하여 제공하였다. 그림 2 (a), (b)의 카메라 배치도와 같이, 실내 데이터셋의 경우에는 위아래로 카메라 2 대, 중앙 카메라를 기준으로 45°마다 카메라 2 대씩 배치하였다. 중앙 카메라와 같은 높이에 배치된 카메라들은 반지름이 0.05m, 0.1m, 0.15m, 0.2m 인 원 위에 배치하며, 중앙 카메라와 수직 방향의 카메라들은 각각 거리 차이가 0.05m 가 되도록 배치하였다.

야외 데이터셋은 실내 데이터셋보다 카메라와 물체의 거리, 물체 간의 거리를 멀게 배치하여 카메라 사이의 간격을 0.5m, 수평 방향으로 카메라 5 대, 수직 방향으로 카메라 3 대를 배치하였다. 야외 데이터셋의 경우 실제와 유사한 환경을 만들기 위해 다양한 높이의 빌딩을 배치하였으며 자동차와 펜스 같은 여러 가지 물체 배치, 배경에 색 변화를 적용하는 효과를 주었다. 또한, 정렬, 매칭과 같은 특징점 기반 알고리즘들이 문제없이 동작할 수 있도록 물체들의 표면에 텍스처가 포함되어 있다.

3. 데이터셋 평가

제작된 데이터셋의 평가를 위해 기존의 Matterport3D 와 Stanford2D-3D 데이터와 함께 동일한 조건에서 시차 추정을 하여 에러율을 계산하였다. 영상의 크기는 1024x512 로 고정하였고 시차 추정을 위해서 위아래로 정렬된 영상들을 시계방향으로 회전한 다음 OpenCV 의 StereoBM 알고리즘과 StereoSGBM 알고리즘을 사용하여 디스패리티를 계산하였다. 두 함수는 흑백의 좌, 우 영상을 입력으로 받아 블록 매칭 기법을 사용해 시차를 계산하며, 매칭되지 않는 부분은 음수를 반환한다. StereoSGBM 함수는 Semi-Global 블록 매칭 기법을 사용하며 StereoBM 함수 보다 더 넓은 영역에 대하여 유사도를 측정하므로 정확도가 높은 대신 알고리즘 수행 속도가 길다. 그림 3 은 제안하는 데이터셋과 비교 데이터셋을 사용하여 StereoBM 함수 및 StereoSGBM 함수를 적용하여 얻은 디스패리티 영상의 예시이다.

시차 추정 성능 평가를 위해서는 StereoBM, StereoSGBM 함수의 출력으로 나온 결과 영상에서 음수인 픽셀의 개수를 측정하였다. 실험을 위한 베이스라인은 Matterport 와 Stanford 데이터셋에서 사용한 0.2m 와 동일하게 설정하였다. 또한, 다양한 영상셋의 성능도 검증하기 위하여 베이스라인을 0.05m 로도 설정하여 실험을 반복하였다.

표 1 에서 보는 바와 같이 Matterport 데이터의 경우 StereoBM 을 사용했을 때 전체 픽셀의 약 26.9%에 해당하는 픽셀이 매칭되지 않았고, Stanford 데이터의 경우 전체 픽셀의 약 19%가 매칭되지 않았다. 본 연구에서 제안한 데이터셋의 경우는 베이스라인이 0.2m 일 때 14.87%, 베이스라인이 0.05m 일 때 약 9.5% 매칭이 되지 않아 기존 데이터셋들 보다 시차 추정에 있어 오차가 더 적음을 확인할 수 있었다. StereoSGBM 알고리즘을 사용한 결과도 Matterport 데이터에서 약 7.33%, Stanford 데이터에서 약 5.35%의 화소값이 매칭되지 않았으며, 본 연구에서 제안한 베이스라인 0.2m 의 데이터에서는 약 3.52%, 베이스라인 0.05m 일 때 약 3.24%가 매칭되지 않아 StereoBM 알고리즘을 사용한 결과와 같이 오클루젼(Occlusion)영역이 감소하였다. 이와 같은 실험결과에 기반하여 본 논문에서 제안하는 데이터셋이 시차 추정에 효율적임을 확인할 수 있다.

4. 결론

본 논문에서는 다양한 분야의 360° 영상 연구에 필요한 벤치마크 데이터셋을 제안하였다. 제안하는 데이터셋은 기존의 데이터셋들 보다 사용자의 이용 목적에 따라 카메라 위치,

표 1. StereoBM 과 StereoSGBM 함수의 결과 영상 음수 픽셀 수, 전체 영상 중 음수 화소의 비율. 전체 화소의 수는 524288 임.

데이터셋	StereoBM	StereoSGBM
Matterport3D	141113 (26.91%)	38452 (7.33%)
Stanford2D-3D	99593 (18.99%)	28025 (5.36%)
Ours (0.2m)	77971 (14.87%)	18446 (3.52%)
Ours (0.05m)	49778 (9.49%)	16994 (3.24%)

카메라 간의 거리, 영상의 해상도, 프레임 수 등을 쉽게 조절할 수 있다. 특히, 제안하는 데이터셋의 그라운드 트루스 깊이를 활용하여 깊이 추정 알고리즘의 객관적인 성능 평가가 가능하므로 추후 360° 영상 응용의 다양한 분야에 활용 가능할 것으로 기대한다.

Acknowledgement

이 논문은 2021 년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (No.2016-0-00144, 시청자 이동형 자유시점 360VR 실감미디어 제공을 위한 시스템 설계 및 기반기술 연구).

참 고 문 헌

[1] Chang, Angel, et al. "Matterport3D: Learning from RGB-D Data in Indoor Environments." *International Conference on 3D Vision (3DV)*, Qingdao, China, Oct. 2017 , pp. 667-676.

[2] Community, B. O. (2018). *Blender - a 3D modelling and rendering package*. Stichting Blender Foundation, Amsterdam. Retrieved from http://www.blender.org

[3] Armeni, Iro, Alexander Sax, and Amir R. Zamir Silvio Savarese. "Joint 2D-3D-Semantic Data for Indoor Scene Understanding."

[4] Zhang, Zichao, et al. "Benefit of large field-of-view cameras for visual odometry." *IEEE International Conference on Robotics and Automation (ICRA)*, Stockholm, Sweden, May. 2016, pp. 582-588

[5] Wang, Ning-Hsu, et al. "360SD-Net: 360° Stereo Depth Estimation with Learnable Cost Volume." *IEEE International Conference on Robotics and Automation (ICRA)*. 2020.

[6] Pathak, Sarthak, et al. "Virtual reality with motion parallax by dense optical flow-based depth generation from two spherical images." *IEEE/SICE International Symposium on System Integration (SII)*. Taipei, Taiwan, Dec. 2017, pp. 887-892