3차원 물체 복원을 위한 SFM 기술의 성능평가

Performance Evaluation of Structure-from-Motion Tools for 3D Object Reconstruction

최성록† · 최준혁[1] · 김기범[1] · 최중용[2] · 김현주[2]
Sunglok Choi†, Jun Hyeok Choi[1], Ki Beom Kim[1], Joongyong Choi[2], Hyunjoo Kim[2]

Abstract: Structure-from-motion (SFM) is an essential technology for visual 3D reconstruction of objects and spaces. This paper contains accuracy comparison of two popular SFM tools (VisualSFM and COLMAP). Similar to visual SLAM, SFM generates 3D camera poses and 3D points of observed objects and spaces. The public visual dataset, IVL-SYNTHSFM-v2, was used to calculate accuracy of two tools. Accuracy was measured by positional and rotational differences of the estimated camera pose to its ground truth given in the dataset. As a result, we found that COLMAP had better accuracy, and we could derive good insight and discussion for better 3D object reconstruction.

Keywords: Structure-from-Motion, Bundle Adjustment, 3D Reconstruction, Pose Estimation

1. 서 론

Structure-from-motion (이하 SFM)은 시점이 다른 다수의 영상을 이용하여 카메라의 3차원 자세와 영상 속의 물체 또는 공간의 3차원 좌표를 동시에 추정하는 기술이다. SFM 기술은 visual SLAM과 유사한 문제로, 일반적으로 다수의 영상 사이의 시간적 순서가 없는 일반적인 문제[1]로 여겨진다. 영상 사이의 관계 정보가 없기 때문에 매칭 관계를 찾는 과정이 포함된 SFM은 visual SFM과 달리 실시간 동작보다는 정확성이 중요한 척도이다. 본 논문에서는 공개되어있고 널리 알려진 SFM 기술인 VisualSFM[2]과 COLMAP[3]을 물체의 3차원 복원에 적용하고, 그 정확도를 평가한다.

본 논문에서는 SFM에서 추정한 카메라의 3차원 자세와 해당 카메라 시점의 참값을 비교하여 SFM 기술의 정확도를 평가한다. SFM을 통해 영상이 촬영된 카메라의 3차원 시점, 즉 자세 정보와 물체 표면에서 자동으로 추출된 특징점의 3차원 위치를 알 수 있는데, 특징점 추출 알고리즘에 따라 각기 다른 위치에서 특징점이 추출되기 때문에 특징점의 3차원 위치를 이용한 단순한 정확도 비교는 힘들다. 본 연구에서는 카메라 시점의 참값이 제공되는 IVL-SYNTHSFM-v2 데이터셋을 이용한다.

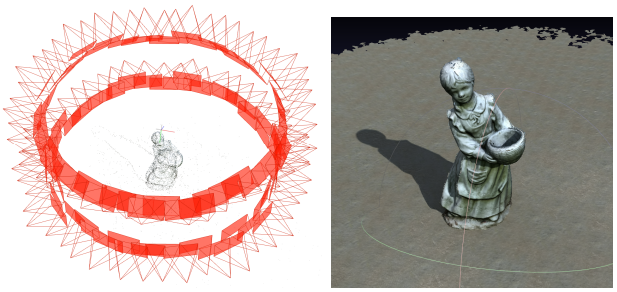
※ 본 연구는 문화재청 및 국립문화재연구원의 2021년도 '문화유산 스마트 보존·활용 기술 개발'사업으로 수행되었습니다. (과제명: 초고해상도 기가픽셀 3D 데이터 생성 기술 개발, 과제번호: 2021A02P02-001)

1. Master Student, Computer Science and Engineering Department, SeoulTech, Seoul, Korea
2. Senior Researcher, Content Research Division, ETRI, Daejeon, Korea
† Assistant Professor, Corresponding author: Computer Science and Engineering Department, SeoulTech, Seoul, Korea (sunglok@seoultech.ac.kr)

2. Structure-from-Motion 도구 소개

VisualSFM[2]은 Waymo에 근무하고 있는 Changchang Wu가 만든 SFM 도구로 간단하고 사용하기 쉬운 GUI를 가졌고, 실행 파일을 다운받아 복잡한 설정없이 사용할 수 있게 되어 있다. VisualSFM은 병렬처리와 GPU를 활용하여 고속으로 동작하는 SFM 기법이다. GUI를 포함하여 전체 SFM의 동작에 해당하는 소스코드는 공개되지 않았지만, SiftGPU[4]나 pba[5]와 같은 병렬처리를 이용한 핵심 기술은 오픈소스로 공개되어 있다. VisualSFM은 nvm 파일 포맷 형태로 SFM 결과를 저장하기 때문에 이를 다른 응용 프로그램에서 가공하여 사용할 수 있다.

COLMAP[3]은 SFM과 MVS(multi-view stereo)를 모두 포함하는 최신 기술이다. VisualSFM과 마찬가지로 incremental SFM 기술이 구현되어 있고, 자체 MVS 기능을 통해 별도의 프로그램이나 라이브러리 없이 dense reconstruction이 가능하다. COLMAP은 VisualSFM과 마찬가지로 다수의 CPU 코어와 GPU를 효율적으로 사용하여 빠르고 정확하게 SFM을 수행한다. COLMAP도 사용하기 쉬운 GUI가 탑재된 바이너리 파일을



[Fig 1.] SFM (Left) and dense recon (Right) by COLMAP @ Status data

제공하고 있고, VisualSFM과는 달리 모든 소스코드를 BSD 라이센스로 공개하고 있다.

3. SFM 성능평가 실험

IVL-SYNTHSFM-v2 데이터셋은 Blender를 이용 가상의 물체 모델에 대해 렌더링한 RGB 영상, 즉 합성 영상을 이용한 데이터셋이다. 렌더링 프로그램을 이용한 합성 영상이기 때문에 물체의 형태 뿐만 아니라 카메라의 자세를 정확하게 알 수 있고, 정확한 참값을 이용한 정확한 카메라 자세의 평가가 가능하다. IVL-SYNTHSFM-v2 데이터셋은 5종류의 물체에 대해 다양한 조명과 노이즈 조건에 따라 1920x1080 (HD) 해상도의 100장의 영상으로 구성되어 있다.

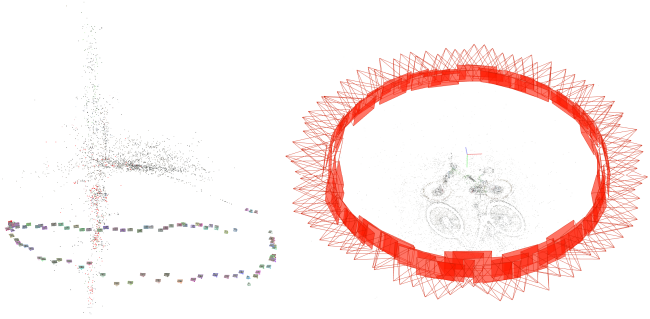
카메라 자세의 정확도는 절대궤적오차(absolute trajectory error; ATE) 관점에서 에러 함수를 평가 척도로 사용한다. 주어진 참값의 카메라 자세와 추정된 카메라 자세의 원점은 다를 수 있어서 첫 번째 영상의 자세를 전역 좌표계(global coordinate)의 원점과 좌표계로 설정하였다. 또 단안 영상만 이용하는 경우 복원의 스케일이 정해지지 않는 문제(scale ambiguity)가 있고, 첫 두 영상 사이의 스케일을 이용하여 추정된 카메라 궤적 전체의 스케일을 맞추었다. 정합된 두 궤적 사이의 각 영상에 대한 위치오차(position error) 와 방향각오차(orientation error) 를 아래와 같이 각각 계산한다.

$$e_p = \| \boldsymbol{p}_{GT} - \boldsymbol{p} \|_2 \qquad e_o = AxisAngle(R_{GT}^{\top} R)$$

위치오차는 두 자세 사이의 two-norm, 즉 유클리디안거리

[Table 1] Camera Pose Accuracy @ IVL-SYNTHSFM-v2

Datasets	VisualSFM		COLMAP	
	Position [m] Mean (STD)	Orientation [deg] Mean (STD)	Position [m] Mean (STD)	Orientation [deg] Mean (STD)
Statue	0.213 (0.1)	0.154 (0.1)	0.146 (0.1)	0.094 (0.0)
EmpireVase	0.031 (0.0)	0.138 (0.1)	0.011 (0.0)	0.187 (0.1)
Bicycle	21.401 (8.6)	101.341 (46.5)	0.139 (0.1)	0.131 (0.0)
Hydrant	0.251 (0.4)	2.664 (5.5)	0.002 (0.0)	0.030 (0.0)
Jeep	1.459 (2.5)	3.575 (5.7)	0.008 (0.0)	0.035 (0.0)



[Fig. 2] VisualSFM (left) and COLMAP (right) @ Bicycle data

(단위: m)로 정의하였고, 방향각오차는 두 자세 사이의 방향각이 나타내는 3차원 회전행렬의 차이(transpose)에 대한 axis-angle representation으로 표현된 자세의 각도의 크기(scale; 단위: deg)로 정의하였다.

IVL-SYNTHSFM-v2 데이터셋에서 두 가지 SFM 도구의 카메라 자세 정확도를 [Table 1]에 나타내었다. 전체적으로 COLMAP이 훨씬 정확한 결과를 나타내었다. 두 방법 모두 Statue와 Bicycle 데이터에 대해 다른 데이터보다 다소 높은 위치오차를 가졌고, 특히 VisualSFM의 경우 Bicycle 데이터에서 엉망진창의 결과가 나왔다. 실제 두 도구의 Bicycle 결과를 가시화하였을 때, VisualSFM의 결과에서 카메라의 위치가 크게 찌그러진 것을 확인할 수 있었다. Bicycle 데이터는 매우 얇은 프레임을 갖는 물체는 SIFT descriptor를 위한 안정적인 패치(patch) 획득이 어렵기 때문에, 이러한 형태의 물체의 3차원 복원은 다소 난이도가 높은 상황이라고 생각된다.

4. 결 론

본 논문에서는 두 가지 SFM 도구를 3차원 물체 복원에 적용하여 복원 카메라 자세의 정확도를 비교하였다. VisualSFM보다 COLMAP이 정확한 카메라 자세 추정이 가능함을 확인하였다. 또 다양한 형태의 물체 종류에 대한 실험을 통해 현재의 특징점 기반의 SFM은 Bicycle 데이터의 예와 같이 가느다란 형태의 물체의 정확한 3차원 복원이 어려움을 확인하였다.

References

[1] 최성록, 최중용, 김현주, Visual SLAM을 통해 살펴본SLAM 기술의 변화와 흐름, 로봇과 인간, 18권 4호, 2021년

[2] C. Wu, “VisualSFM: A Visual Structure from Motion System”, 2011. [Online] http://ccwu.me/vsfm/

[3] J. Schonberger et al., “Structure-from-Motion Revisited”, CVPR, 2016. [Online] https://demuc.de/colmap/

[4] C. Wu, “SiftGPU”. [Online] https://github.com/pitzer/SiftGPU

[5] C. Wu et al., “Multicore Bundle Adjustment”, CVPR, 2011. [Online] http://grail.cs.washington.edu/projects/mcba/