

Multi View System 에서 Depth Map Fusion 을 위한 개선된 기법

정우경, 김해광, 한종기

세종대학교

trancis314@sju.ac.kr, hkkim@sejong.ac.kr, hjk@sejong.edu

Improved Method for Depth Map Fusion in Multi View System

Woo-Kyung Jung, Haekwang Kim, Jong-Ki Han

Sejong University

요 약

실감 미디어에 대한 수요가 증가함에 따라 고품질의 실감 미디어에 대한 중요성이 증가하고 있다. 이러한 실감미디어를 제작하기 위해 사용되는 일반적인 기법 중 하나인 Multi View Stereo 는 깊이 영상 추정 및 해당 깊이 영상을 이용하여 3 차원에 point cloud 를 생성하는 fusion 과정을 거치게 된다. 본 논문에서는 다중 시점 영상의 깊이 영상을 정합하는 fusion 과정을 개선하기 위한 방법을 제안한다. 제안하는 방법에서는 깊이 영상, 색상정보를 이용하여 기존 시점의 depth map 을 이용한 fusion 과정을 거친다. 실험을 통하여 제안한 알고리즘을 이용한 결과가 기존보다 개선됨을 보인다.

1. 서론¹

하드웨어 발전 및 5G 네트워크가 상용화됨에 따라 고품질, 대용량의 멀티미디어 수요가 증가하고 있다. 특히 실감 미디어가 주요 콘텐츠로 주목받고 있다.

이러한 실감미디어를 제작하기 위한 주요 방법으로는 Multi View Stereo 이 있다. 해당 알고리즘은 영상들의 자세정보를 이용하여 깊이 영상을 생성하고, 이를 이용해 3 차원 공간상에 point cloud 를 생성하는 fusion 과정을 거친다.

본 논문에서는 Multi View Stereo 를 통해 복원된 point cloud 의 품질을 개선하기 위한 Fusion 알고리즘을 제안한다. 제안하는 방법에서는, 기존 깊이 영상 및 이웃 깊이 영상에 해당하는 confidence map, 그리고 색상정보를 이용하여 위치정보가 개선된 Point 를 생성하여 오류가 최소화하는 것을

목표로 한다.

본 논문의 구성은 다음과 같다. 2 절에서는 Multi View Stereo 의 과정 중 하나인 Fusion 에 대해 살펴본 후, 문제점에 대하여 토의한다. 3 절에서는 제안하는 기법에 대하여 설명한다. 4 절에서는 본 논문에서 제안하는 알고리즘을 적용하여 성능을 확인한다. 마지막으로 5 절에서는 본 논문에 대한 결론을 맺는다.

2. 기존 알고리즘 및 문제점

Multi View Stereo 는 다 시점의 영상을 이용하여 카메라들의 위치를 추정하고, 해당 정보를 이용하여 깊이 영상을 추정한다. 이후 추정된 깊이 영상 및 위치 정보를 이용하여 3 차원 공간상에 point cloud 를 복원하는 것을 목표로 한다. [1]

¹ 연락 저자: 한종기

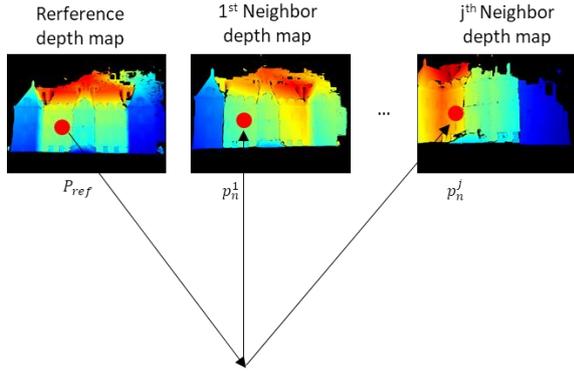


그림 1. 참조 시점의 깊이 영상으로부터 생성된 P_{ref} 와 이웃 시점으로 투영된 결과인 P_n 들

이중 point cloud 를 생성하기 위해서 여러 개의 깊이 영상들을 정합하여 3 차원 공간상에 point 를 생성하는 과정을 fusion 이라 한다.

[2]에서는 참조 카메라의 깊이 영상에 존재하는 각각의 픽셀마다 해당 픽셀의 point P_{ref} 와 이에 해당하는 이웃 카메라의 픽셀의 point P_n 의 위치를 고려하여 각 픽셀의 깊이 정보의 유효성을 판단한다.

P_n 이 P_{ref} 보다 더 깊은 깊이 정보를 가졌을 경우, [2]은 이를 가려진 point 라고 판단하고 P_n 에 해당하는 픽셀의 깊이 정보를 제거한다. 만약 P_n 과 P_{ref} 과의 깊이 값의 차이가 threshold 보다 작을 경우, 해당 두 포인트가 중복되는 point 로 판단하여 P_{ref} 를 제거한다. 위 두가지 경우에 해당하지 않는 경우에는 깊이정보를 제거하지 않고 다음 P_{ref} 에 대해 위 과정을 반복한다. 이 과정을 통해 중복, 혹은 가려지는 point 에 대한 처리가 가능하지만, 중복되는 point 를 제거하는 과정에서 P_{ref} 에 대한 개선과정 없이 그대로 이용하게 되어 이웃 시점의 결과가 전혀 반영되지 않는다는 문제가 발생한다.

이를 개선하기 위하여 본 논문에서는 기준 깊이 영상 및 이웃 깊이 영상에 해당하는 confidence map, 그리고 색상정보를 이용하여 위치정보가 개선된 Point 를 생성하여 오류가 최소화하는 알고리즘을 제안한다.

3. 제안하는 알고리즘

fusion 과정에서 현재 깊이 영상으로 생성된 point P_{ref} 의 3 차원 좌표에 대해, 이웃 시점들로 projection 하여 해당 위치의 point P_n 의 깊이 영상에 해당하는 confidence map 및 색상정보를 이용하여 P_{ref}' 으로 갱신한다.

$$P_{ref}' = \sum_Q \frac{w(P_{ref}, Q)(Q)}{\sum_Q w(P_{ref}, Q)} \quad (1)$$

식(1)을 통해 P_{ref} 및 P_n 들로부터 갱신된 point P_{ref}' 를 유도해낸다. 이때 Q 는 P_{ref} 및 P_n 을 포함하는 point 집합을 의미한다.

$$w(P_{ref}, Q) = \prod_{j=1}^2 \exp\left(-\frac{c_j(P_{ref}, Q)}{\sigma_j^2}\right) \quad (2)$$

여기서 가중치 w 는 두가지 특성 c_1, c_2 에 의해 유도되며 식 (2)로 표현된다. 각 특성은 matching confidence, hue correlation 부터 유도된다.

$$c_1 = 1 + \frac{1 + (\text{confidence}_{P_{ref}} - \text{confidence}_{P_n})}{2} \quad (3)$$

식 (3)에서 보여주는 첫번째 특성은 matching confidence 이다. 이는 Multi View Stereo 과정에서 깊이 영상을 추정하기 위한 패치 단위 매칭 과정에서 계산되는 매칭 정확도를 의미한다. 이 논문에서는 ZNCC 를 이용하였으며, [0, 1] 사이의 값을 가진다. [3] 따라서 첫번째 특성 c_1 은 보다 정확히 추정된 깊이 영상에 의해 생성된 point 에 대하여 더 작은 값을 가지므로 수식 (1)에서 더 높은 신뢰도를 반영하게 된다.

$$c_2 = 1 + |H(P_{ref}) - H(P_n)| \quad (4)$$

두번째 특성은 두 point 간의 hue correlation 이다. hue component H는 depth map 의 연관된 input texture 이미지의 RGB 성분을 이용하여 계산이 가능하며, [0, 1] 사이의 값이다. [4] Hue 는 명암에 관계없이 texture 영상의 순수한 색상정보만을 나타내기 때문에, P_{ref} 와 P_n 사이의 색상정보의 correlation 을 나타내기에는 적합하다. Hue component 를 이용한 특성 c_2 는 식 (4)와 같이 표현되며, texture 이미지에서 P_{ref} 와 P_n 에 해당하는 픽셀의 색상이 유사할수록 작은 값을 가진다.

4. 실험 결과

본 논문에서는 제안된 알고리즘을 평가하기 위하여 [ETH3D]에서 제안하는 평가 기준인 F1 점수 및 [5]에서 제공하는 데이터셋 중 고해상도-다 시점 훈련 데이터셋 중 일부 데이터

셋(delivery_area, kicker, relief, playground)으로 수행되었다.

F1 점수는 completeness 와 Accuracy 이 두 가지 요소의 조화 평균값으로, 각 영상 세트의 ground truth point cloud 와 비교하여 계산된다. 이를 통하여 복원된 Point cloud 의 완성도 및 정확도를 평가할 수 있다.

Tolerance	1	2	5	10	20	50	100
[2]	0.09	0.224	0.678	1.563	3.400	8.848	18.198
Proposed Method	0.11	0.248	0.714	1.611	3.511	9.327	18.990

표 1. 기존 방법[2] 및 제안한 방법의 F1 점수

식(2)에서 이용되는 스케일 파라미터 σ_1, σ_2 는 각각 1 로 설정하였다. F1 점수를 측정하기 위한 Tolerance 는 각각 1,2,5,10,20,50,100(cm)에서 계산되었다.

실험결과 제안한 방법이 기존 방법보다 더 높은 F1 점수를 달성함을 통해 성능이 향상됨을 확인할 수 있다.

5. 결론

본 논문에서는 깊이 영상 Fusion 을 개선하기 위한 새로운 방법을 제안하였다. 제안하는 방법은 depth map 의 confidence 와 해당하는 texture image 의 색상정보를 이용하여 depth map fusion 과정에서 reference point 를 개선하는 알고리즘이다. 4 절에서 실험을 통하여 제안한 알고리즘을 이용한 point cloud 결과가 기존 알고리즘을 이용한 point cloud 결과보다 개선됨을 보였다.

ACKNOWLEDGMENT

This work was partly supported by the National Research Foundation of Korea (NRF) under Grant (No. 2019R1F1A1046236) and partly by Institute for Information & communications Technology Promotion (IITP) under Grant (2017-0-00486) funded by the Korea government (MSIT).

References

- [1] Theo Moons; Luc Van Gool; Maarten Vergauwen, 3D Reconstruction from Multiple Images: Part 1: Principles , now, 2010.
- [2] S. Shen, "Accurate Multiple View 3D Reconstruction Using Patch-Based Stereo for Large-Scale Scenes," in IEEE Transactions on Image Processing, vol. 22, no. 5, pp. 1901-1914, May 2013, doi: 10.1109/TIP.2013.2237921.
- [3] Aschwanden, P., and W. Guggenbuhl. "Experimental results from a comparative study on correlation-type registration algorithms." Robust computer vision (1992): 268-289
- [4] Alvy Ray Smith. 1978. Color gamut transform pairs. SIGGRAPH Comput. Graph. 12, 3 (August 1978), 12-19. DOI:https://doi.org/10.1145/965139.807361
- [5] T. Sch T., J. L. Sch. L. Sch, S. Galliani, T. Sattler, K. Schindler, M. Pollefeys, A. Geiger, "A Multi-View Stereo Benchmark with High-Resolution Images and Multi-Camera Videos", Conference on Computer Vision and Pattern Recognition (CVPR), 2017.