

특집논문 (Special Paper)

방송공학회논문지 제27권 제1호, 2022년 1월 (JBE Vol.27, No.1, January 2022)

<https://doi.org/10.5909/JBE.2022.27.1.20>

ISSN 2287-9137 (Online) ISSN 1226-7953 (Print)

Synthetic fisheye 이미지를 이용한 360° 파노라마 이미지 스티칭

권혁준^{a)}, 조동현^{a)*}

Panorama Image Stitching Using Sythetic Fisheye Image

Hyeok-Joon Kweon^{a)} and Donghyeon Cho^{a)*}

요 약

최근 VR (Virtual Reality) 기술이 주목받기 시작하면서 생동감 넘치는 VR 콘텐츠를 볼 수 있는 360° 파노라마 영상이 많은 관심을 받고 있다. 이미지 스티칭 기술은 360° 파노라마 영상을 제작하는데 주요한 기술로서 많은 연구가 활발하게 이루어지고 있다. 일반적인 스티칭 알고리즘은 특징점 기반 이미지 스티칭을 기반으로 한다. 하지만 기존의 특징점 기반 이미지 스티칭 방법들은 특징점에 크게 영향을 받는다는 문제가 존재한다. 이러한 문제를 해결하기 위해서 최근에는 딥러닝 기반의 이미지 스티칭 기술들이 연구되고 있지만 이미지 간의 겹치는 영역이 거의 없거나 큰 시차가 존재할 때 여전히 많은 문제점이 존재한다. 또한 실제 환경에서는 라벨링된 정답 파노라마 영상을 얻을 수 없으므로 완전한 지도학습에 한계가 존재한다. 따라서 자율주행분야에 많이 이용되는 칼라(Carla) 시뮬레이터를 통해 카메라 센터가 다른 3개의 fisheye 이미지와 그에 대응되는 정답 영상을 제작하였다. 우리는 제작한 fisheye 영상으로 360° 파노라마 영상을 만드는 이미지 스티칭 모델을 제안한다. 최종 실험 결과로는 실제 환경과 비슷하게 구성된 가상의 데이터 세트로 다양한 환경과 큰 시차에도 강인한 스티칭 결과를 검증한다.

Abstract

Recently, as VR (Virtual Reality) technology has been in the spotlight, 360° panoramic images that can view lively VR contents are attracting a lot of attention. Image stitching technology is a major technology for producing 360° panorama images, and many studies are being actively conducted. Typical stitching algorithms are based on feature point-based image stitching. However, conventional feature point-based image stitching methods have a problem that stitching results are intensely affected by feature points. To solve this problem, deep learning-based image stitching technologies have recently been studied, but there are still many problems when there are few overlapping areas between images or large parallax. In addition, there is a limit to complete supervised learning because labeled ground-truth panorama images cannot be obtained in a real environment. Therefore, we produced three fisheye images with different camera centers and corresponding ground truth image through carla simulator that is widely used in the autonomous driving field. We propose image stitching model that creates a 360° panorama image with the produced fisheye image. The final experimental results are virtual datasets configured similar to the actual environment, verifying stitching results that are strong against various environments and large parallax.

Keyword : 360-panorama, Fisheye, Stitching, Carla simulator

1. 서론

현대 기술의 발전과 더불어 이미지 스티칭 기술도 많은 발전이 이루어지고 있다. 이미지 스티칭 기술은 시야각이 좁은 여러 영상을 합성하여 시야각이 넓은 큰 이미지 영상을 만드는 기술이다. 이미지 스티칭 기술은 많은 분야에서 이용되는데 위성사진 이미지나 디지털 방식의 이미지 안정화에도 활용된다. 최근 가상환경에서 360°의 시야각을 가지는 입체적인 영상이 사람들의 많은 관심을 받고 있는데 360°의 시야각을 가지는 영상을 만드는데 이미지 스티칭 기술이 사용된다.

일반적인 스티칭 방법은 특징점 기반 스티칭 방법을 이용하고 있는데 그 과정은 다음과 같다. 스티칭 할 이미지들의 겹치는 영역에서 특징점 검출 알고리즘(SIFT, SURF, FAST, Harris Corner Detector)으로 특징점을 찾아 추출한다^{[1][5]}. 추출된 특징점들은 특징점을 비교하여 강한 특징점끼리 일대일로 매칭시킨 후 매칭시킨 특징점 간의 변환 관계를 이용하여 호모그래피(Homography)를 추정한다^[6]. 추정한 호모그래피를 이용해 타겟 이미지를 와핑하여 기준 이미지에 스티칭한다. 마지막으로 경계선 부분을 부드럽게 만들어 주는 블렌딩 과정을 거쳐 최종 파노라마 영상을 만들게 된다^{[7][8]}.

하지만 특징점 기반 이미지 스티칭 기술은 특징점에 크게 영향을 받는다. 특히나 겹치는 영역이 적어서 특징점을 거의 찾지 못하거나 강한 특징점을 찾지 못하면 스티칭 된 이미지가 크게 왜곡될 수 있다. 또한 서로 다른 카메라 센터에서 찍었기 때문에 각각의 이미지의 같은 장면에서 다른 깊이(Depth) 수준을 가지게 된다. 즉 겹치는 장면에서 시차 차이가 발생하는데 이를 단일 호모그래피만을 이용해서 이미지 스티칭을 하게 되면 고스팅(Ghosting)이 발생할 수 있다.

따라서 이러한 시차 왜곡을 많이 완화하는 알고리즘들이 활발하게 연구되고 있다. 먼저 단일 호모그래피만을 추정하는 것이 아닌 다양한 영역의 다중 호모그래피를 추정하는 방식이 있다^{[9][13]}. 다양한 영역별로 각각 호모그래피를 추정하기 때문에 다양한 깊이를 고려한 이미지 스티칭을 수행할 수 있으나 이미지의 전경에서 큰 시차가 발생할 때 시차왜곡이 발생해 잘못된 스티칭 결과가 나오는 한계가 존재한다. 다른 방식으로서는 이미지들 간의 최적의 경계선(Seam)을 추정하는 방식이 있다^{[14][17]}. 최적의 경계선 추정은 특정 영역을 경계선에서 배제하여서 그 영역에서 시차 왜곡이 발생하지 않게 만든다. 따라서 특정 영역에서 시차에 따른 불일치, 고스팅을 크게 줄일 수 있으나 특정 영역에서의 물체의 모양, 크기, 배치에 따라서 제대로 된 경계선을 추정하지 못해 시차 왜곡이 발생할 수 있다. 또한 스티칭된 결과 이미지가 시각적으로 자연스러운 이미지처럼 보이지만 기하학적으로 잘못된 결과가 나올 수 있다.

최근 영상처리 분야에서 딥러닝 기반 알고리즘들이 큰 성과를 보이고 있는데, 이미지 스티칭 분야도 마찬가지로 딥러닝 모델을 기반으로 한 연구들이 활발히 이루어지고 성능 또한 기존 알고리즘을 상회하는 결과를 보여주고 있다. 딥러닝은 이미지의 특징 추출에서도 우수한 성능을 보여주고 있는데 몇몇 연구들은 기존의 특징점 추출 알고리즘을 이용하지 않고 딥러닝 기반 특징점 추출 방식을 통해 이미지 스티칭 모델을 제안하였다^{[18][19]}. 또한 딥러닝을 이용해 특징점 추출뿐만 아니라 중간 간 고정된 시점의 영상을 정합하는 딥러닝 기반 이미지 스티칭 모델이 제안되었다^{[20][22]}. 최근에는 고정된 시점에서의 영상을 정합하는 것이 아닌 다양한 시점에서 영상을 정합하는 연구가 제안되었다^[23]. 하지만 이러한 연구들에는 큰 한계가 존재하는데 실제 환경에서는 정확한 파노라마 영상을 찍을 수 없으므로 라벨링 된 완벽한 정답 영상으로 학습이 가능하지 않다. 따라서 완전한 지도학습을 수행하지 못하기에 실제 파노라마 영상과 스티칭된 파노라마 영상과는 다소 차이가 발생할 수 있다.

따라서 본 논문에서는 완전한 지도학습이 가능한 가상환경을 이용하여 실제 환경과 유사한 다양한 날씨, 다양한 물체가 존재하는 가상환경에서의 카메라 세팅을 제안하였다. 카메라 센터가 각각 다른 3개의 360° 큐브 맵 영상을 시물레이터를 통해 만들었다. 만든 3개의 360° 큐브 맵 이미지

a) 충남대학교 전자공학과(Department of Electronics Engineering, Chungnam National University)

‡ Corresponding Author : 조동현(Donghyeon Cho)

E-mail: cdh12242@cnu.ac.kr

Tel: +82-42-821-7699

ORCID:https://orcid.org/0000-0002-2184-921X

※ This work was supported by Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No. 2018-0-00207, Immersive Media Research Laboratory) and the National Research Foundation of Korea (NRF) grant funded by the Korea Government (MSIT) under Grant 2020R1F1A1065573.

• Manuscript received November 23, 2021; Revised December 27, 2021; Accepted December 27, 2021.

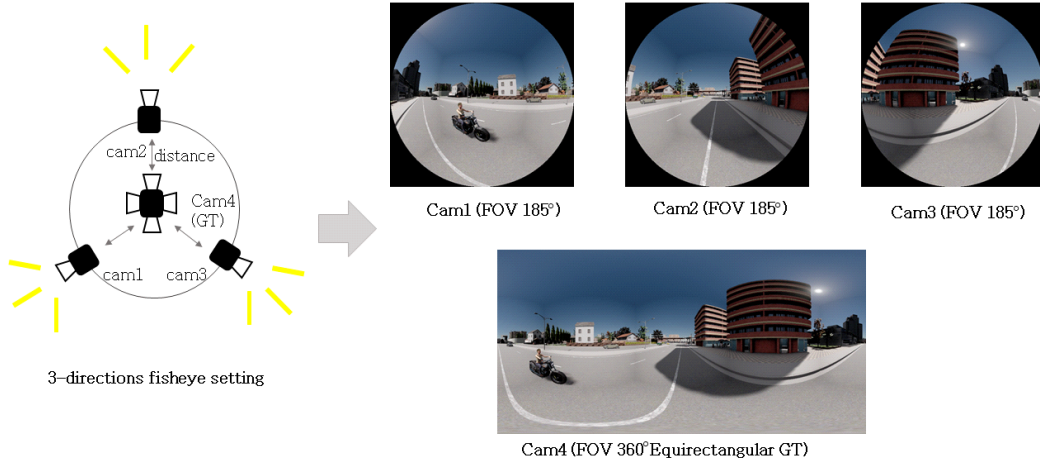


그림 1. 3방향 카메라 세팅
Fig. 1. 3 - directions camera setting

를 fisheye projection을 통해 그림1에서 보이듯이 3개의 185°의 시야각을 가지는 fisheye 영상을 제작하였다. 마찬가지로 3개의 카메라 센터의 중심에 정답 영상인 360° 파노라마 이미지를 제작하였다. 제작한 fisheye 영상과 정답 영상으로 본 논문은 멀티 호모그래피를 추정하여 다양한 깊이 에 따른 이미지 스티칭이 가능하고 완전한 지도학습을 통해 정확한 파노라마 영상을 만드는 모델을 제안하였다.

본 논문의 구성은 다음과 같다. 2장에서는 우리가 제안하는 이미지 스티칭 모델을 들어가기 전의 fisheye 영상과 정답 영상을 제작하는데 필요한 관련 연구에 관해 설명하고, 3장에서는 본 논문이 제안하는 딥러닝 기반 이미지 스티칭 기술에 관해서 설명하고, 4장에서는 우리의 모델의 실험 결과에 관해 설명한다. 마지막으로 5장에서는 본 논문의 결론을 서술한다.

II. 관련 연구

1. Projection

1.1 ERP (Equirectangular Projection)

칼라 가상환경에서 정답인 360° 파노라마 영상을 얻기 위해서는 90°의 시야각을 가진 6방향의 큐브맵을 이용해 360° 파노라마 영상을 얻을 수 있다. 그 과정은 다음과 같다. 파노라마 이미지의 각 픽셀에 대해서 구의 표면에 해당되는 구면 좌표와의 관계를 구해야 한다. 그림2에서 보듯이 2-D 360° 파노라마 이미지는 3-D 구면 좌표를 평평하게 편 이미지이므로 각 픽셀에 대해서 서로 연관 지을 수 있다. 그 후 그림 3처럼 360° 큐브 맵을 구의 표면에 내접시킨다. 구의 그 지점에서 색상을 얻기 위해서 구의 중심에서 극좌

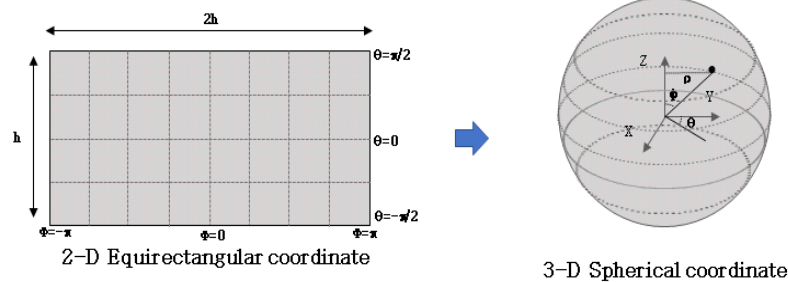


그림 2. 2-D equirectangular 영상과 3-D 구면 좌표와의 관계
Fig. 2. The relationship between 2-D equirectangular image and 3-D spherical coordinate

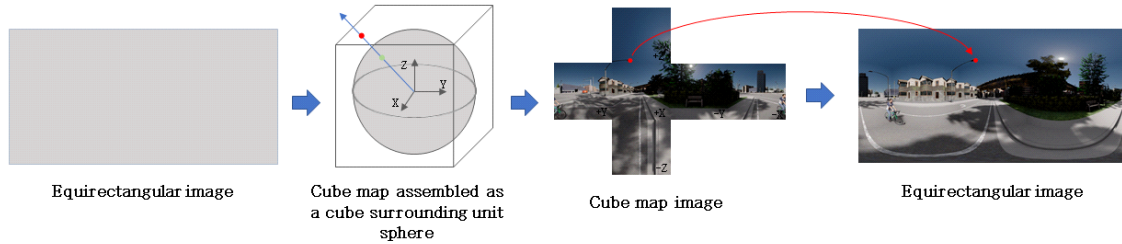


그림 3. 등장방형도법 과정
Fig. 3. Equirectangular projection procedure

표를 통해 광선을 투사하고 그것이 정육면체 표면에 닿는 위치를 계산한다. 이 정보를 사용하여 큐브 맵의 지점에서 색상을 얻고 처음에 계산한 구면 좌표와의 관계를 이용해서 360° 파노라마 이미지에서의 색상을 구한다.

1.2 Fisheye projection

실제 환경에서 어안렌즈로 촬영된 fisheye 영상을 가상환경에서 얻기 위해서는 ERP와 유사한 과정을 거쳐야 한다. 우리는 입력 fisheye 영상을 얻기 위해서 앞선 선행연구들과 같이 6방향의 큐브 맵 영상을 사용하여 360°의 큐브 맵을 만들었다^{[24][25]}. Fisheye 영상의 픽셀에 상응하는 구면 좌표를 구하기 위해서 우리는 Scaramuzza^[26]가 제안한 카메라 모델을 사용하였다. 모델은 전 방향 카메라 모델을 사용하고 영상촬영 시스템을 소형 시스템으로 취급합니다. 3-D 세계 좌표계를 2-D 영상과 연관시키려면 카메라 외적 및 고유 파라미터를 얻어야 한다. 세계 좌표계는 외적 매개변수를 사용하여 카메라 좌표로 변환되고 카메라 좌표는 내장 파라미터를 사용하여 영상 평면에 매핑된다. 제안된 카메라 모델은 전 방향 카메라의 내부파라미터를 계산할 수 있으므로 fisheye의 2-D 영상에 상응하는 3-D 구면 좌표와의 관계를 알 수 있다. 2-D fisheye 영상의 픽

셀을 상응하는 3-D 구면 좌표의 각 픽셀에 대해서 연관 지은 후 그림 4과 같이 큐브 맵을 구의 표면에 내접시킨다. 그 후 구의 중심에서 극좌표를 향해 광선을 투사하고 큐브 맵 표면의 투과된 픽셀의 색상을 통해 앞서 계산한 3-D 구면 좌표의 관계를 이용해 fisheye 영상의 픽셀값을 구할 수 있게 된다.

1.3 Fisheye unwarping

보통 fisheye 영상은 광각 렌즈에 의해 생성된 선과 왜곡 효과가 있으므로 유용한 이미지를 생성하기 위해 이미지를 펴준다. 우리의 모델에서는 3개의 fisheye 영상을 와핑하여 더 완벽한 파노라마 영상을 만들기 위해서 fisheye 영상을 ERP 형태로 펴는 과정을 수행하게 된다. Fisheye 영상을 펴주는 과정은 그림 5에서 보듯이 크게 2가지로 나뉘는데 먼저 2-D fisheye 이미지를 3-D 구면 좌표계로 투영시키게 된다. 수식으로 살펴보면 fisheye 영상의 임의의 한점 (x,y) 에서 구면 좌표 $P(\cos\phi\sin\theta, \cos\phi\cos\theta, \sin\phi)$ 로 투영시키게 된다. 여기서 fisheye 영상으로부터 (θ, ϕ) 를 구할 수 있다. Fisheye 영상의 시야각이 f , 이미지의 높이와 넓이를 각각 W, H 라고 하면 다음식을 얻을 수 있다.

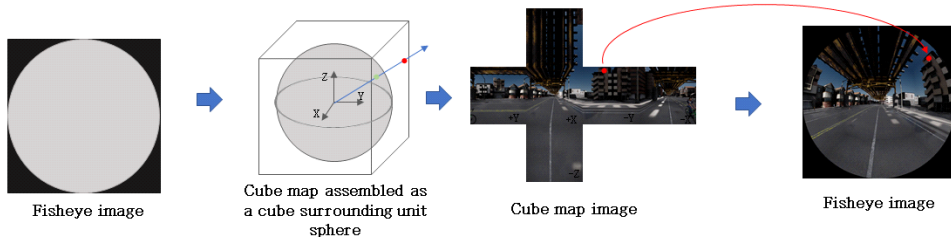


그림 4. Fisheye 프로젝션 과정
Fig. 4. Fisheye projection procedure

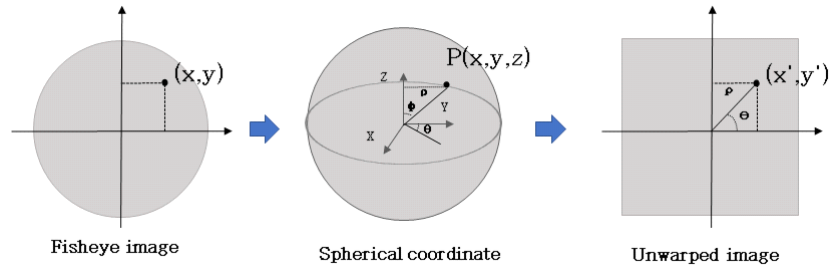


그림 5. Fisheye unwarping 과정
Fig. 5. Fisheye unwarping procedure

$$\theta = f \frac{x}{W} - 0.5, \phi = f \frac{y}{H}, \rho = \frac{H}{f} \tan^{-1} \frac{\sqrt{x^2 + y^2}}{y},$$

구한 (θ, ϕ) 로 구면좌표계의 점 $P(x, y, z)$ 을 구한다. 계산된 구면좌표계의 점을 Equirectangular projection 시키면 최종적으로 다음과 같은 2-D 좌표점을 얻을 수 있다.

$$x' = 0.5W + \rho \cos \theta, y' = 0.5H + \rho \sin \theta, \theta = \tan^{-1} \left(\frac{z}{x} \right),$$

위 과정을 통해 퍼진 fisheye 영상을 우리 모델의 입력 영상으로 사용한다.

III. Fisheye 영상을 이용한 360 파노라마 이미지 스티칭

우리가 제안한 이미지 스티칭 모델은 기존 이미지 스티칭 모델과는 달리 정답 영상을 제작하였기 때문에 전처리를

통해 퍼진 fisheye 영상을 입력으로 완전한 지도학습을 통해 다중 호모그래피, 이동 맵(Displacement map), 가중치 맵(weight map)을 한 번에 추정한다. 또한 기존의 단일 호모그래피를 추정하는 것이 아닌 다중 호모그래피를 추정하고 이동 맵을 이용해 와핑 맵을 미세 조정 후 이미지를 와핑한다. 그 후 가중치 맵을 이용해 와핑된 이미지를 합산하기 때문에 다양한 깊이를 고려한 이미지 스티칭을 수행할 수 있다. 이번 장에서는 우리가 제안한 모델의 구조에 대해서 소개하고, 딥러닝 모델의 학습 과정을 설명한다. 그림6은 우리의 딥러닝 기반 이미지 스티칭 모델의 파이프라인을 간략히 보여준다.

1. 전처리

입력 영상은 그림7과 같이 일련의 전처리 과정을 거치게 되는데 과정은 다음과 같다. 먼저 칼라 시뮬레이터를 통해

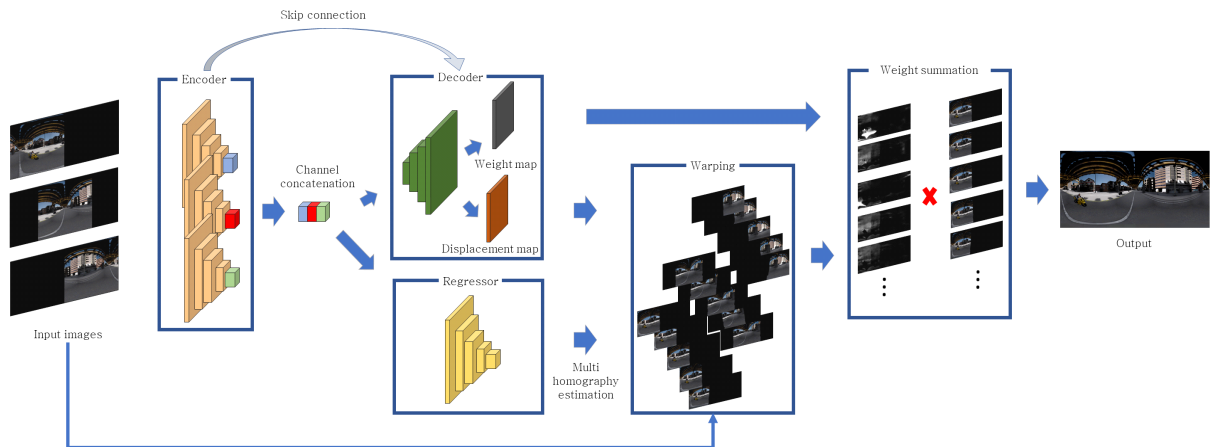


그림 6. 스티칭 모델의 전반적인 파이프라인
Fig. 6. Overall pipeline of our stitching model

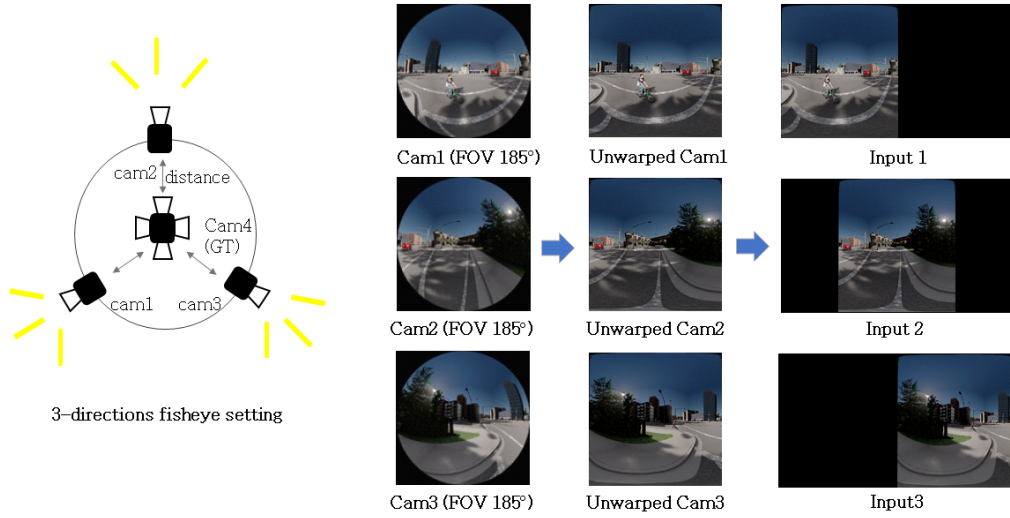


그림 7. 전처리 과정
Fig. 7. Preprocessing procedure

만들어진 카메라 센터가 다른 3개의 큐브 맵 이미지를 fisheye projection을 통해 3개의 fisheye 영상을 만들어 주었다. 그 후 fisheye 영상을 피는 과정을 통해 3개의 이미지를 구성한다. 마지막으로 정답 파노라마 영상과의 크기를 맞추주기 위하여 3개의 퍼진 이미지를 출력 영상 크기에 맞게 패딩을 진행하게 된다.

2. 다중 호모그래피 이미지 스티칭 네트워크

우리의 이미지 스티칭 네트워크는 크게 인코더(encoder), 디코더(decoder), 회귀 모듈(regressor) 3가지 구조로 되어 있다. 먼저 인코더에서는 전처리를 거친 3개의 전처리를 거친 fisheye 영상을 입력으로 받아 3개의 특징 맵을 추출하게 된다. 추출된 3개의 특징 맵을 채널 축으로 합쳐 각각 디코더, 회귀 모듈의 입력으로 넣어 주게 된다. 회귀 모듈에서는 다중 호모그래피를 추정하는 과정을 거치게 된다. 본 논문에서 하나의 입력 영상마다 5개의 호모그래피를 추정하였다. 회귀 모듈에서 추정한 호모그래피를 이용해 글로벌 와핑 맵을 구한다. 디코더의 입력으로 들어온 합쳐진 특징 맵은 연속적인 컨볼루션 층과 업 샘플링 과정을 거친다. 디코더의 마지막 층에서 특징 맵은 두 컨볼루션 블록의 입력으로 들어가 최종적으로 이동 맵과 가중치 맵을 출력하

게 된다. 이동 맵은 입력 영상을 와핑할 시 다중 호모그래피를 이용해 만든 글로벌 와핑 맵을 세밀하게 조정하는 역할을 하며, 가중치 맵은 입력 영상을 다중 호모그래피와 이동 맵을 이용해 와핑 한 후 가중치 별로 와핑 이미지들을 합산한다. 최종적으로 회귀 모듈의 출력인 다중 호모그래피를 이용해 글로벌 와핑 맵을 구하고 디코더의 출력인 이동 맵과의 합을 통하여 최종 다중 와핑 맵을 구한다. 구한 다중 와핑 맵으로 입력 영상을 와핑하고 가중치 맵을 이용해 가중치 별로 와핑된 이미지를 합산해 최종 파노라마 영상을 출력한다.

3. 학습

우리의 제안하는 이미지 스티칭 모델은 인코더, 디코더, 회귀 모듈을 한 번에 학습시키는 end-to-end 구조이다. 목적함수로 우리는 L1 함수와 두 이미지의 휘도, 대비, 및 구조를 비교하는 SSIM (Structural Similarity Index Map)을 사용하였다. 우리의 이미지 스티칭 모델의 목적함수는 다음과 같다.

$$L = \beta L_{PL1} + (1 - \beta) L_{SSIM},$$

$$L_{L1} = E[\|P_{gt} - P_{out}\|_1],$$

$$L_{SSIM} = E[\|1 - SSIM(P_{gt} - P_{out}\|_1)],$$

P_{gt} , P_{out} 는 각각 정답 이미지와 본 모델의 출력이며 하이퍼파라미터인 β 는 두 목적함수의 균형을 위한 가중치 계수이며 본 논문에서는 0.4로 고정하였다.

또한 이미지 레벨에서의 차이를 구하는 L1 함수를 이미지의 특징 맵 레벨에서 차이를 비교하는 지각손실 함수로 대체하는 실험도 진행하였다. 지각손실함수는 학습된 VGG-16 모델을 이용하여 각 컨볼루션 층의 출력인 특징 맵들의 L1 차이를 사용하였다. 지각손실함수의 목적함수는 다음과 같다.

$$L_{PL1} = E[\sum_{j=1}^5 \|\Psi_j(P_{gt}) - \Psi_j(P_{out})\|_1],$$

ψ_j 는 VGG-16의 j 번째 컨볼루션 층의 출력인 특징 맵이다.

IV. 실험 결과

1. 데이터 세트 및 실험 환경

무료 공개 소스인 칼라 시뮬레이터를 이용하여서 그림 8에서 보이듯이 다양한 날씨, 다양한 물체를 담고 있는 3개의 512×512 해상도의 입력 영상과 1024×512 해상도의 정답 영상을 한 세트로 하는 총 40,000 데이터 세트를 학습으로 사용하였으며 5,000 데이터 세트를 테스트 이미지로 사용하였다. 본 논문의 구현 환경은 우분투 18.04의

PyTorch이며 GPU로는 2개의 A6000을 사용하였다. 제안된 모델의 옵티마이저(optimizer)로는 학습률(learning rate) 4×10^{-3} 의 Adam을 사용하였으며, 총 20 에폭(epoch)을 진행하였다.

2. 모델 분석

우리는 칼라 시뮬레이터의 정답 영상 카메라와 입력 영상 카메라의 거리를 각각 0.2, 0.4로 나누어 실험을 진행하였다. 또한 목적함수를 L1, SSIM, 지각손실함수를 각각 사용한 360° 파노라마 이미지 스티칭 결과를 비교하였다. 입력 영상을 우리의 모델에 넣었을 때 출력되는 360° 파노라마 영상과 정답 파노라마 영상을 PSNR (Peak Signal-to-noise ratio)과 SSIM을 이용해 평가하였다. 표1에서 확인할 수 있듯이 입력 영상과 정답 영상과의 적절한 거리에서는 비슷한 스티칭 성능이 나왔다. 하지만 목적함수에 따라서 큰 변화폭이 존재하였다. 먼저 목적함수로 지각손실함수와 SSIM을 사용하였을 때 다른 목적함수를 사용하였을 경우보다 가장 낮은 PSNR 성능이 나왔다. 그에 비해서 L1만을 사용하였을 경우는 L1, SSIM을 사용하였을 때보다 PSNR은 약간 높으나 그에 비해 SSIM은 낮은 것을 보여준다. 이 결과로 정답 영상으로 완전한 지도학습 시에는 PSNR 측면에서는 L1 함수가 유용한 것을 보여주고 SSIM 측면에서는

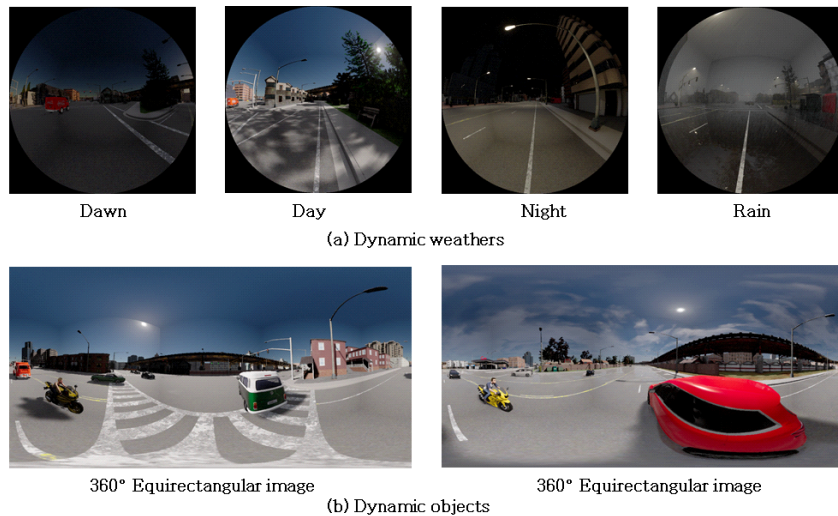


그림 8. 데이터 세트 예시들
Fig. 8. Dataset examples

표 1. 정답 영상과 입력 영상의 거리(0.2, 0.4)에 따른 정량적 결과

Table 1. Quantitative results according to the distance (0.2, 0.4) between the Ground-truth image and the input images

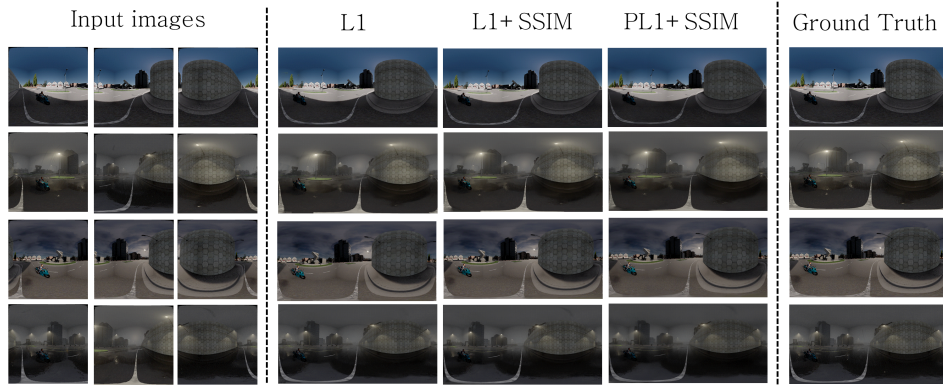
Distance (0.2) (PSNR/SSIM)			Distance (0.4) (PSNR/SSIM)		
L1	L1+SSIM	PL1+SSIM	L1	L1+SSIM	PL1+SSIM
32.56/0.8591	32.33/0.8713	31.07/0.8442	32.76/0.8567	32.62/0.8763	30.91/0.8554

L1, SSIM 함수를 모두 사용하는 것이 유용한 것을 알 수 있다. 또한 현존하는 알고리즘은 완전한 지도학습이 아니기 때문에 무 참조(No-Reference) 이미지 영상 평가인 NIQE^[27] (Natural Image Quality Evaluator), MetaQA^[28] (Meta-learning for No-Reference Image Quality Assessment)를 이용해 APAP^[11]와의 정성적 비교 평가를 하였다. 표 2에서 확인 할 수 있듯이 APAP에 의한 스티칭 결과 영상보다 우리의 모델의 스티칭 결과 영상 품질이 더 좋은

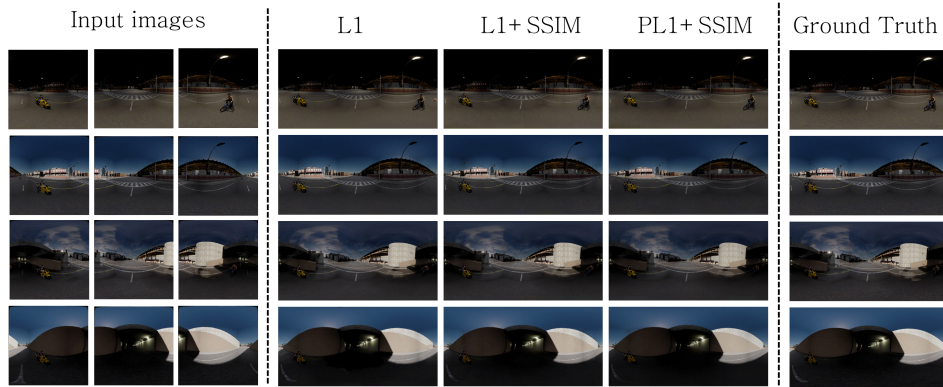
표 2. 정답 영상과 입력 영상의 거리(0.2, 0.4)에 따른 무 참조 이미지 영상 평가

Table 2. No-Reference Image Quality Assessment results according to the distance (0.2, 0.4) between the Ground-truth image and the input images

Distance	APAP		Ours	
	0.2	0.4	0.2	0.4
NIQE	8.01	8.897	5.73	5.596
MetalQA	0.32	0.31	0.41	0.43



(a)



(b)

그림 9. 목적함수에 따른 정성적 결과. (a) 정답 영상과 입력 영상 거리(0.2). (b) 정답 영상과 입력 영상 거리(0.4)

Fig. 9. Qualitative results according to the objective function and distance. (a) The distance (0.2) between the GT image and the input images. (b) The distance (0.4) between the GT image and the input images



그림 10. 스티칭이 실패한 사례
Fig. 10. Cases of stitching failure

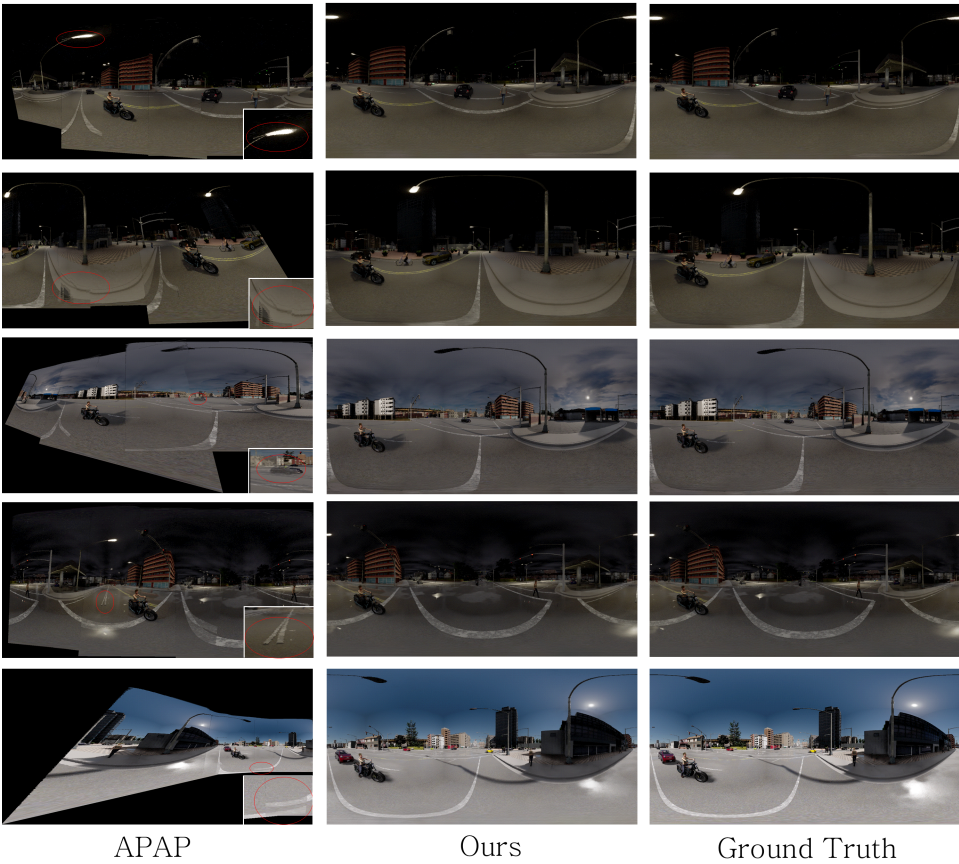


그림 11. 이미지 스티칭 품질의 시각적 비교
Fig. 11. Visual comparison of the image stitching quality

것을 보여준다. 그림 9는 목적함수에 따른 이미지 스티칭 결과를 정성적인 결과로 보여준다. 정답 영상과 입력 영상의 거리가 0.2일 때의 이미지 스티칭 결과는 대부분 우수한 성능을 자랑한다. 거리가 0.4의 경우에도 겹치는 영역에서 적은 시차가 발생할 경우에는 우수한 성능을 자랑한다. 하지만 겹치는 영역에서 큰 시차가 있는 물체가 있을 때 고스팅과 심한 왜곡이 일어나는 것을 그림 10에서 확인할 수 있다. 그림 10의 4가지 경우 전부 겹치는 영역에 보행자, 전봇대, 차량 같은 심한 시차가 발생한 물체들에서 모두 고스팅과 아티팩트(Artifacts)가 존재한다. 다음 실험으로는 현존하는 이미지 스티칭 알고리즘 중에 APAP와 우리의 모델과 정량적 비교 결과를 그림 11에서 확인 할 수 있다. APAP의 경우 시접선 부분에서 일치하지 않는 부분과 고스팅이 심하게 일어난다. 또한 실제 정답 영상과는 다소 차이가 발생하는 스티칭 결과가 나오지만, 우리의 모델은 정답 영상과 차이가 거의 없는 360° 파노라마 영상 스티칭 성능을 보여준다.

V. 결 론

본 논문에서는 현실 세계에서는 입력 영상과 라벨링 된 완벽한 정답 영상을 만드는 게 불가능하므로 우리는 가상 환경에서 이를 해결하고자 했다. 우리는 칼라 시뮬레이터를 이용한 실제 환경과 유사한 가상환경을 구성하고 3방향의 시야각이 185°인 fisheye 데이터 세트와 그에 대응되는 360° 파노라마 정답 영상을 제작하였다. 우리는 제작한 데이터 세트를 이용해 스티칭 과정을 수행해 360° 파노라마 영상을 만드는 모델을 구현하였다. 제안된 모델은 완전한 지도학습으로 다중 호모그래피, 가중치 맵, 이동 맵을 한번에 추정하여 어느 정도의 시차에도 강인한 이미지 스티칭을 수행하였다. 본 연구의 후속 연구로 가상환경에서 학습한 모델을 현실환경에서 완전한 지도학습을 하도록 도와주는 도메인 적응 연구를 진행해야 할 것이다.

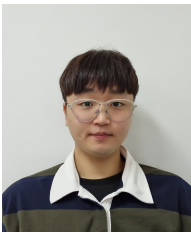
참 고 문 헌 (References)

- [1] Harris, C., & Stephens, M. A combined corner and edge detector. In *Alvey vision conference*, Vol. 15, No. 50, pp. 10-5244, August 1988.

- [2] Rublee, E., Rabaud, V., Konolige, K., & Bradski, G. ORB: An efficient alternative to SIFT or SURF. In *International conference on computer vision*, pp. 2564-2571, November 2011.
- [3] Bay, H., Tuytelaars, T., & Van Gool, L. Surf: Speeded up robust features. In *European conference on computer vision*, pp. 404-417, May 2006.
- [4] Rosten, E., & Drummond, T. Machine learning for high-speed corner detection. In *European conference on computer vision*, pp. 430-443, May 2006.
- [5] Lowe, D. G. Distinctive image features from scale-invariant keypoints. In *International journal of computer vision*, 60(2), 91-110, 2004.
- [6] Fischler, M. A., & Bolles, R. C. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. In *Communications of the ACM*, 24(6), 381-395, 1981.
- [7] Burt, Peter J., and Edward H. Adelson. "A multiresolution spline with application to image mosaics." In *ACM Transactions on Graphics (TOG)* 2.4, 217-236, 1983.
- [8] Pérez, P., Gangnet, M., & Blake, A. (2003). Poisson image editing. In *ACM SIGGRAPH*, pp. 313-318, 2003.
- [9] Gao, J., Kim, S. J., & Brown, M. S. Constructing image panoramas using dual-homography warping. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 49-56, June 2011.
- [10] Lin, W. Y., Liu, S., Matsushita, Y., Ng, T. T., & Cheong, L. F. Smoothly varying affine stitching. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 345-352, June 2011.
- [11] Zaragoza, J., Chin, T. J., Brown, M. S., & Suter, D. As-projective-as-possible image stitching with moving DLT. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2339-2346, 2013.
- [12] Chang, C. H., Sato, Y., & Chuang, Y. Y. Shape-preserving half-projective warps for image stitching. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3254-3261, 2014.
- [13] Lee, K. Y., & Sim, J. Y. Warping residual based image stitching for large parallax. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 8198-8206, 2020.
- [14] Liu, F., Gleicher, M., Jin, H., & Agarwala, A. Content-preserving warps for 3D video stabilization. In *ACM Transactions on Graphics (ToG)*, 28(3), 1-9, 2009.
- [15] Zhang, F., & Liu, F. Parallax-tolerant image stitching. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3262-3269, 2014.
- [16] Lin, K., Jiang, N., Cheong, L. F., Do, M., & Lu, J. Seagull: Seam-guided local alignment for parallax-tolerant image stitching. In *European conference on computer vision*, pp. 370-385, October 2016.
- [17] Hejazifar, H., & Khotanlou, H. Fast and robust seam estimation to seamless image stitching. In *Signal, Image and Video Processing*, 12(5), 885-893, 2018.
- [18] Hoang, V. D., Tran, D. P., Nhu, N. G., & Pham, V. H. Deep feature extraction for panoramic image stitching. In *Asian Conference on*

- Intelligent Information and Database Systems*, pp. 141-151, March 2020.
- [19] Shi, Z., Li, H., Cao, Q., Ren, H., & Fan, B. An image mosaic method based on convolutional neural network semantic features extraction. In *Journal of Signal Processing Systems*, 92(4), 435-444, 2020.
- [20] Wang, L., Yu, W., & Li, B. (2020, June). Multi-scenes image stitching based on autonomous driving. In *IEEE 4th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC)*, Vol. 1, pp. 694-698, June 2020.
- [21] Li, J., Zhao, Y., Ye, W., Yu, K., & Ge, S. Attentive Deep Stitching and Quality Assessment for 360° Omnidirectional Images. In *IEEE Journal of Selected Topics in Signal Processing*, 14(1), 209-221, 2019.
- [22] Song, D. Y., Um, G. M., Lee, H. K., & Cho, D. End-to-End Image Stitching Network via Multi-Homography Estimation. In *IEEE Signal Processing Letters*, 28, 763-767, 2021.
- [23] Nie, L., Lin, C., Liao, K., Liu, M., & Zhao, Y. A view-free image stitching network based on global homography. In *Journal of Visual Communication and Image Representation*, 73, 102950, 2020.
- [24] Sekkat, A. R., Dupuis, Y., Vasseur, P., & Honeine, P. The omniscap dataset. In *IEEE International Conference on Robotics and Automation (ICRA)*, pp. 1603-1608, May 2020.
- [25] Berenguel-Baeta, B., Bermudez-Cameo, J., & Guerrero, J. J. OmniSCV: An Omnidirectional Synthetic Image Generator for Computer Vision. In *Sensors*, 20(7), 2066, 2020.
- [26] Scaramuzza, D., Martinelli, A., & Siegwart, R. (2006, October). A toolbox for easily calibrating omnidirectional cameras. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 5695-5701, October 2006.
- [27] Mittal, Anish, Rajiv Soundararajan, and Alan C. Bovik. Making a “completely blind” image quality analyzer. In *IEEE Signal Processing Letters (SPL)*, pp. 209-212. 20.3 (2012)
- [28] Zhu, Hancheng, et al. MetalQA: Deep meta-learning for no-reference image quality assessment. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 2020.

저 자 소 개



권 혁 준

- 2021년 : 충남대학교 전자공학과 졸업(공학사)
- 2021년 ~ 현재 : 충남대학교 전자공학과 석사과정
- ORCID : <https://orcid.org/0000-0002-5788-2038>
- 주관심분야 : 영상처리, 이미지 스티칭



조 동 현

- 2019년 : KAIST 전기 및 전자공학부 졸업(공학박사)
- 2019년 ~ 현재 : 충남대학교 전자공학과 조교수
- ORCID : <https://orcid.org/0000-0002-2184-921X>
- 주관심분야 : 컴퓨터 비전, 딥러닝