

# 카메라 포즈 추정과 인페인팅을 이용한 AR/DR

김유진, 신재이, 박인규  
인하대학교 정보통신공학과

e-mail : kimyujin1088@gmail.com, jaeshin1@gmail.com, pik@inha.ac.kr

## AR/DR in images using camera pose estimation and inpainting

Yujin Kim, Jael Shin, and In Kyu Park

School of Information and Communication Engineering  
Inha University

### Abstract

In this paper, we propose the fusion of Augmented Reality(AR) which puts virtual objects suitable for places and situations, and Diminished Reality(DR) which removes the person who we want to erase. Camera pose for AR implementation is acquired using Colmap. We designed an easy-to-use automatic human inpainting tool for DR. People in image are detected by the Mask-RCNN model. In the model inpainting phase, the user-specified person area is filled by the re-trained model based on Deepfill v2. In the refine inpainting phase, we suggest inpainting method using homography matrix. With the proposed method, the result image of AR and DR can be obtained in one place.

### I. 서론

관광지나 추억의 장소와 같은 특별한 장소에서 기억에 남는 사진을 촬영하기 위해 여러 소품을 사용하기도 한다. 또한, 사진을 촬영 후, 다른 사람이 등장해 사진을 다시 촬영하는 경우가 많다. 이를 위해, 소품 대신 장소에 어울리는 사물을 합성하고 관심 없는 사람을 지우는 아이디어를 고안했다. 특정장소와 상황에 맞는 가상 사물을 추가하는 Augmented Reality와 관심 없는 사람을 제거하는 Diminished Reality를 융합하였다.

본 논문에서는 AR과 DR의 전체 과정에 대한 흐름도를 제시한다. AR은 Colmap[4]을 사용해 camera pose를 취득한다. DR은 Deepfill v2[2]를 기반으로 하는 model inpainting과 SIFT 매칭 및 homography 행렬을 사용하는 homography inpainting 방법을 제시한다.

### II. 본론

#### 2.1. 흐름도

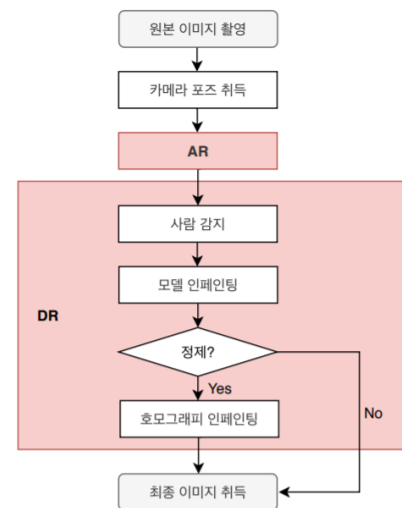


그림 1. 흐름도

#### 2.2. AR

Camera pose를 취득하기 위해 카메라 정보 없이 이미지만으로 카메라 포즈를 구할 수 있는 Colmap을 사용한다. 3D reconstruction의 결과물 중 이미지 정보의 사원수와 translate 요소를 이용해 Rotation 행렬과 Translation 행렬을 계산한다. 아래 수식(1)에 따라 카메라 내부 행렬과 Rotation 행렬과 Translation 행렬을 이용해 camera pose를 구한다. Camera pose 행렬은 카메라 좌표계에서 월드 좌표계로 변환되기 때문에 전치행렬로 변환해 적용한다.

$$P = KR[I|t] \quad (1)$$

AR에 사실감과 음영을 더하기 위해 Phong reflection model을 이용해 조명 효과를 구현한다. 광원의 ambient, diffuse, specular 성분과 위치를 설정하여 obj를 더 입체적이고 깊이 있게 표현한다.

## 2.3. DR

DR의 전체 과정은 사람 감지, model inpainting, homography inpainting으로 나누어진다. 사람 감지를 위해 Mask-RCNN[1]을 사용한다. Mask-RCNN은 감지되는 하나의 클래스 안에 여러 instance들의 segmentation에 취득이 용이하고 다른 CNN 기반 방법에 비해 detection 성능이 우수해 해당 연구에 적합한 방법이다. Model inpainting에 사용되는 model은 Deepfill version2 pretrained model을 기반으로 생성한다. Deepfill[2] model은 기존 GAN에서 discriminator를 원인으로 model이 최적점에 도달하지 못하는 경우를 방지하기 위해 WGAN을 기반으로 한다. Data augmentation을 통해 최종적으로 13,170개의 데이터를 생성한다. 이는 model의 예측범위를 넓혀줄 수 있는 장점이 있다. Places2 dataset을 학습시킨 pretrained model에 취득한 데이터를 더해 재학습을 진행하였다. Epoch당 54분, 총 157 epoch까지 학습시켜 141시간이 소요되었다. Trained model inpainting의 정확도가 낮을 시에 homography 행렬[3]을 이용한 inpainting을 적용한다.

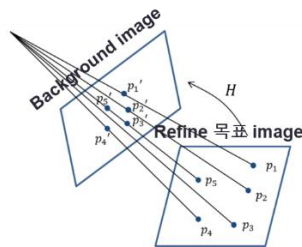


그림 2. Homography 개념 설명

그림 2에서 구해진  $H$ 를 이용한 inpainting의 의미는 refine 목표 영역에 있는 pixel들을  $H$ 를 적용해 얻은 background의 pixel로 채우는 것을 의미한다.



그림 3. Background 선택 과정

그림 3과 같이 background를 선정하는 과정이 필요하다. 해당 장소에서 찍은 후보 background 이미지 30장과 refine 목표 이미지를 SIFT 알고리즘을 이용해

good features의 개수가 가장 많은 이미지를 background 이미지로 선택하고 이 이미지와의  $H$  행렬을 inpainting에 적용한다.

## III. 구현

### 3.1. AR

AR에 대한 실험 결과이다. 그림 4는 인하대학교 캠퍼스의 비행기를 배경으로 한 이미지이다. 이미지에 따라 camera pose가 달라져도 같은 곳에 가상 사물이 로드된다.



그림 4. AR 구현 사진

그림 5은 조명 효과에 따른 전과 후 이미지이다. 상단에 헬리콥터 obj를 로드시키고 조명 효과를 적용함으로써, obj가 현실적이고, 깊이 있게 표현된다.



그림 5. 조명 효과 적용 전과 후

Camera pose의 정확성을 측정하기 위해 표 1과 같이 카메라의 초점 거리에 대한 오차율을 계산하였다. 스펙이 다른 카메라에서 5개의 장소 당 오차율을 계산했다. 해상도에 따라 각 장소에서 오차율이 달라진다. 고해상도에서 초점 거리에 대한 오차율이 증가하는 것은 촬영 환경의 영향으로 판단된다. 10장의 사진 모두 같은 광원의 세기, 위치를 가지고 있을 경우 오차율이 감소한다.

표 1. 카메라 초점 거리에 대한 오차율

해상도 장소	960*720	4032*3024
하이테크	0.113%	0.389%
인경호	0.42%	2.155%
비룡담	0.471%	2.939%
비행기	2.781%	0.467%
본관건물	0.241%	0.992%
평균	0.805%	1.388%

### 3.2. DR

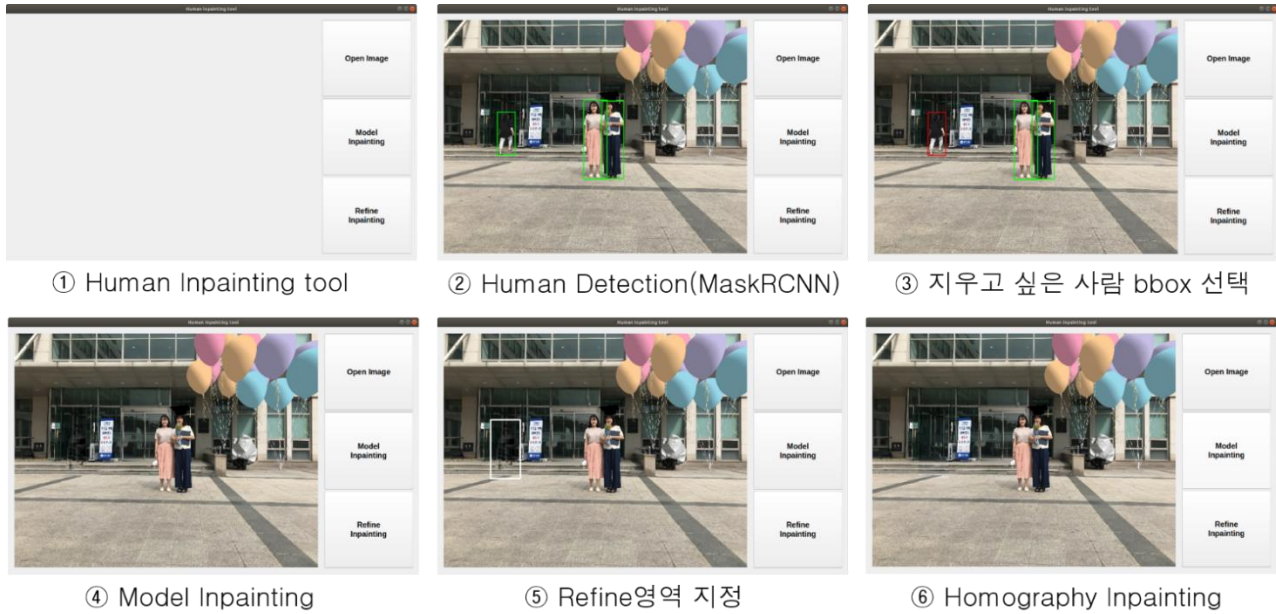


그림 6. Human inpainting tool(UI)

그림 6과 같이 DR의 모든 과정은 UI 상에서 수행된다. UI 상에서는 3개의 버튼이 사용 가능하다. Open Image를 통해 inpainting할 사진을 load하고 Mask-RCNN model이 자동으로 사람을 감지한다. 이 후, 사용자의 선택으로 inpainting할 사람의 bounding box를 지정한다. 지정한 사람을 Model Inpainting 버튼을 통해 학습한 model을 사용하여 inpainting한다. Refine Inpainting을 위해 UI 상에서 mouse input을 통해 정제하고자 하는 영역을 사용자가 선택할 수 있다. Homography를 사용한 Refine inpainting이 완료되면 최종 이미지를 취득할 수 있다.

표 2는 inpainting이 적용된 영역을 비교한 PSNR 값을 취득한 결과이다. 모든 dataset에서 평균적으로 30dB 이상의 값이 얻어지고, Homography inpainting이 Model inpainting보다 더 좋은 성능을 보인다.

표 2. 각 Dataset에 따른 PSNR

Inpainting	Dataset	PSNR(dB)
Model inpainting	Hightech	31.9
	Top	31.6
Homography inpainting	Hightech	41.4
	Top	44.4

### IV. 결론

본 논문에서는 Augmented Reality와 Diminished Reality의 개념을 한 데 적용하여, 영상의 원하는 것을 더하고 원하지 않는 것을 빼는 아이디어를 제시했다. 그림 1의 흐름도와 같이, 하나의 원본 이미지에 대해서 AR과 DR를 같이 적용한 최종 이미지를 얻을 수 있다. Human inpainting의 방법으로 WGAN based model에 이어 homography inpainting을 사용해 결과를 정제할 수 있는 장점이 있다. 사용자 친화적인 Human inpainting tool을 일반적인 image inpainting 분야에서 쉽게 이용 가능하다.

### 참고문헌

- [1] K. He et al., "Mask r-cnn," Proc. of IEEE International Conference on Computer Vision, October 2017.
- [2] J. Yu, Z. Lin, J. Yang, X. Shen, X. Lu and T. S. Huang, "Generative image inpainting with contextual attention," Proc. of IEEE Conference on Computer Vision and Pattern Recognition, June 2018.
- [3] D. DeTone, T. Malisiewicz and A. Rabinovich, "Deep image homography estimation," Proc. of IEEE Conference on Computer Vision and Pattern Recognition, June 2016.
- [4] J. L. Schönberger and J. Frahm, "Structure-from-motion revisited," Proc. of IEEE Conference on Computer Vision and Pattern Recognition, June 2016