

(별지 서식 7)

## 저작권양도서 (Copyright Transfer Form)

소속 : 인하대학교 정보통신공학과

성명 : 김유진, 신재이

학번 : 12161711, 12171798

논문제목 : AR/DR in images using camera pose estimation and inpainting

본인은 상기 논문을 2020 학년도 1 학기 정보통신프로젝트 최종 보고서 겸 결과 논문으로 제출하고자 합니다. 본 논문의 내용은 저자가 직접 연구한 결과인 것과 이전에 출판된 적이 없음을 확인합니다. 또한 공저자와 더불어 인하대학교 정보통신공학부에서 발간하는 논문집에 본 논문을 수록하는 것을 허락하며 제반 저작권을 정보통신공학부에 양도합니다.

2020 년 7 월 10 일

주저자 : 김유진, 신재이 (김유진), (신재이)

정보통신공학과장 귀하

# 카메라 포즈 추정과 인페인팅을 이용한 AR/DR

김유진, 신재이

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

Yujin Kim, Jael Shin

**요약 :** 본 논문에서는 장소와 상황에 어울리는 가상의 사물을 합성하는 Augmented Reality(AR)와 지우고 싶은 사람을 제거하는 Diminished Reality(DR)의 융합을 제안한다. AR을 구현하기 위한 카메라 포즈는 Colmap을 사용해 취득한다. 가상 사물의 합성을 자연스럽게 입체감 있게 표현하기 위해 Phong reflection model을 이용한다. DR을 위해 사용하기 쉬운 자동 human inpainting tool을 구현했다. 이미지에서 Mask-RCNN 모델에 의해 사람이 감지된다. 모델 인페인팅 단계에서, 사용자가 지정한 사람을 Deepfill v2 기반 재학습 모델에 의해 인페인팅한다. 모델 인페인팅 결과의 정제를 원할 시 호모그래피 행렬을 사용한 인페인팅 방법을 제시한다. 이 방법은 SIFT 기반으로 취득한 매칭의 수가 가장 많은 배경 이미지와의 호모그래피 행렬을 얻어 정제하고 싶은 영역을 인페인팅한다. 제안하는 기법으로 한 장소에서 AR과 DR을 같이 적용한 이미지를 얻을 수 있다.

**키워드 :** AR, DR, 카메라 포즈, 모델 인페인팅, SIFT, 호모그래피 인페인팅

**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. By using the Phong reflection model, we express the synthesis of virtual objects naturally and deeply. 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 Deepfill v2 based re-trained model. If we want to refine the model inpainting result, we suggest inpainting method using homography matrix. This method takes homography matrix with background image which has the largest number of matches acquired on a SIFT basis and fill the area that we want to refine. With the proposed method, an image of AR and DR can be obtained in one place.

**Keyword :** AR, DR, camera pose, model inpainting, SIFT, homography inpainting

## 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. 그래서 사람이 없는 곳에서 사진을 찍기 위해 줄을 서는 경우를 경험한 적이 있을 것이다. 이를 위해, 소품 대신 장소에 어울리는 사물을 합성하고 관심 없는 사람을 지우는 아이디어를 고안했다. 아이디어의 구현으로 특정장소와 상황에 맞는 가상 사물을 추가하는 Augmented Reality와 관심 없는 사람을 제거하는 Diminished Reality를 융합하였다.

본 논문에서는 Augmented reality와 Diminished reality의 기존연구에 대해 소개하고 AR과 DR의 전체 과정에 대한 흐름도를 제시한다. AR은 Colmap[2]을 사용해 camera pose를 취득한다. 취득한 camera pose와 phong reflection model을 이용해 장소에 어울리는 가상 사물을 합성한다. DR은 이미지에서

사람을 감지한 후 Deepfill v2[4]를 기반으로 하는 Model inpainting과 SIFT 알고리즘과 homography 행렬을 사용하는 Homography inpainting 방법을 제시한다. 훈련한 모델로 사람 영역을 inpainting한 후 사람의 그림자, 손, 발 등 정제하고 싶은 영역을 선택 후 배경 이미지와 본 이미지 간의 homography 행렬을 구해 inpainting을 실시한다.

## II. 기존 연구

### 2.1. AR

Augmented reality는 Marker-based와 Markerless로 나뉜다. Marker-based AR은 장치 카메라가 주변을 시각적으로 인식하고 장치의 소프트웨어가 특정 마커를 인식해 가상의 사물을 합성한다. 이 방법은 안정적으로 작동하나 마커 이미지가 필요하다는 큰 단점을 가지고 있으며 흑백 대비가 강해야 하고 빛을 반사하면 안되는 등 마커의 제약 조건이 존재한다.

Makerless AR은 SLAM(Simultaneous Localization And Mapping)을 활용해 주변을 기억하고 대상의 현재위치를 실시간으로 동기화한다. 3D 콘텐츠를 장면에 오버레이하고 공간의 특정 지점에 고정시킬 수 있다. 도보 방향 레이블, 가상 관광 안내, 지역 정보 등 다양한 응용 분야에 유용하다. 이 방법은 3D 공간에서 현재 위치의 카메라 포즈를 취득해야 하는 본 논문에서 필요하다.

## 2.2. DR

Diminished reality 즉 inpainting은 크게 두 가지 방법으로 나뉜다. 가장 고전적인 방법인 sequential based[1] 방법의 경우, patch 단위로 이미지를 분석해 가장 비슷한 patch를 찾아와 채워 넣는 patch based 방법과 가려진 부분의 외곽에서부터 내부까지 이미지 내용 자체를 전파시켜서 채우는 diffusion based 방법이 있다. 최근 들어서 neural network를 기반으로 하는 CNN, GAN based 방법이 생겨나기 시작하였고, GAN의 결점을 보완한 Wasserstein distance를 사용한 WGAN을 본 연구에서 사용한다.

## III. 제안하는 기법 및 기술

### 3.1. 흐름도

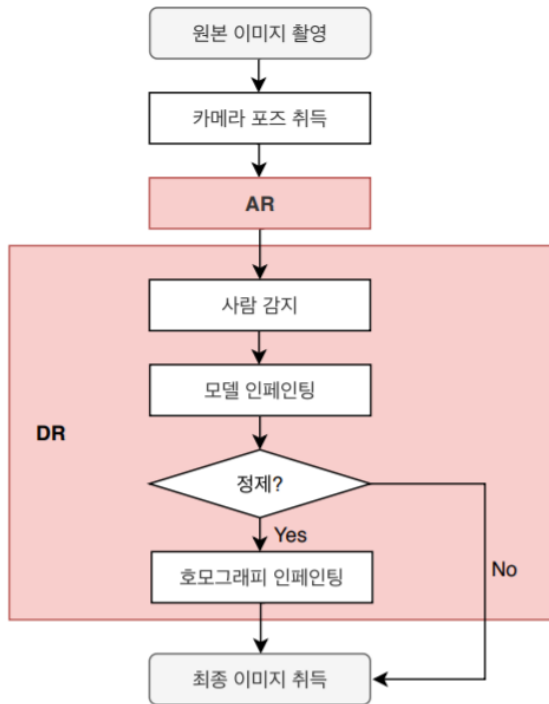


그림 1. 흐름도

그림 1은 전체 순서의 흐름을 나타낸다. Colmap 프로그램을 사용한 camera pose 취득을 위해 한 장소에서 최소 10장의 원본 이미지를 촬영한다. Camera pose를 취득한 후, AR을 구현한다. DR 과정에서는 AR 결과 이미지에서 사람을 감지하고 학습 모델을 사용해 사람 영역을 채운다. 모델 인페인팅의 결과에 대한 정제를 원한다면 homography 행렬을 이용해 원하는 영역을 인페인팅 할 수 있다.

### 3.2. AR

다양한 장소에서의 실험을 위해 표 1과 같이 장소를 선정해 영상을 취득하였다. 한 장소에서 다양한 각도에서 취득하였다.

표 1. 영상 취득 대상 장소

장소	특징
하이테크 건물 고층부 앞	일반 빌딩 앞
인경호	자연 환경
비룡탑	반복되는 패턴
비행기	공원
본관 앞	반복되는 패턴

Camera pose를 취득하기 위해 Colmap 소프트웨어를 사용한다. 한 장소에서 카메라 내부 파라미터, 외부 파라미터 정보 없이 이미지만으로 카메라 포즈를 구할 수 있는 장점이 있다. Colmap은 정렬되지 않은 이미지 셋에서 3D reconstruction을 하는 Incremental Structure-from-motion 방법을 사용한다. 특징점 추출과 매칭 후 검증을 하는 대응점 탐색 단계를 거치고 초기화, 이미지 등록, 삼각 측량, 번들 조정, 아웃라이어 필터링이 있는 재구성 단계를 거쳐 reconstruction이 완료된다. 위 단계의 3D reconstruction의 결과물은 카메라 내부 파라미터, 이미지 정보, 특징점 정보이다. 그 중 카메라 정보와 이미지 정보를 이용해 각 이미지에 대한 camera pose를 계산한다. 사원수와 translate 요소를 이용해 Rotation 행렬과 Translation 행렬을 계산한다. 수식(1)에 따라 camera pose를 구한다. camera pose 행렬은 카메라 좌표계에서 월드 좌표계로 변환되기 때문에 월드 좌표계에서 카메라 좌표계로 변환하는 행렬을 만들기 위해 전치 행렬로 변환해 적용한다.

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

AR에 사실감과 음영을 더하기 위해 Phong reflection model을 이용해 조명 효과를 구현한다. Phong model은 표면 위 점들의 지역 조명의 모델이며 반사되는 빛에 의한 간접광인 ambient 성분, 난반사 조명 성분인 diffuse 성분, 하이라이트를 표현하는데 쓰이는 specular 성분으로 이루어진다. 광원의 위 세 성분과 위치를 설정하여 obj를 더 입체적이고 깊이 있게 표현한다.

### 3.3. DR

DR의 전체 과정은 사람 감지, model inpainting, homography inpainting으로 나누어진다.

#### 3.3.1 사람 감지

사람 감지를 위해 Mask-RCNN[3]을 사용한다. Mask-RCNN는 Faster-RCNN에 segmentation mask branch가 추가된 구조로 detection 성능이 우수하다. 감지되는 하나의 클래스 안에 여러 instance들의 segmentation에 취득에 용이하다. 모든 사람을 감지하고 segmentation을 취득해야 하는 해당 연구에 적합한 방법이다.

#### 3.3.2 Model inpainting

Model inpainting에 사용되는 model은 Deepfill version2 pretrained model을 기반으로 생성한다. Deepfill[4] model은 GAN에서 discriminator가 온전한 역할을 다 하지 못하여 model이 최적점에 도달하지 못하는 경우를 보완하기 위해 Wasserstein distance를 이용한 WGAN을 기반으로 한다. 학습을 위해 취득한 데이터는 약 1,650개이며, augmentation을 거쳐 최종적으로 약 13,170개의 데이터를 생성한다. Data augmentation을 통해 model의 예측범위를 넓혀줄 수 있는 장점이 있다. Places2라는 실내, 실외 등 전반적인 장소에 대한 dataset을 학습시킨 pretrained model에 취득한 data를 더해 재학습을 진행하였다. Epoch당 54분, 총 157epoch까지 학습시켜 141시간이 소요되었다.

#### 3.3.3 Homography inpainting

Trained model inpainting의 정확도가 낮을 시에 두 평면 간의 변환관계인 homography 행렬[5]을 이용한 inpainting을 적용한다.

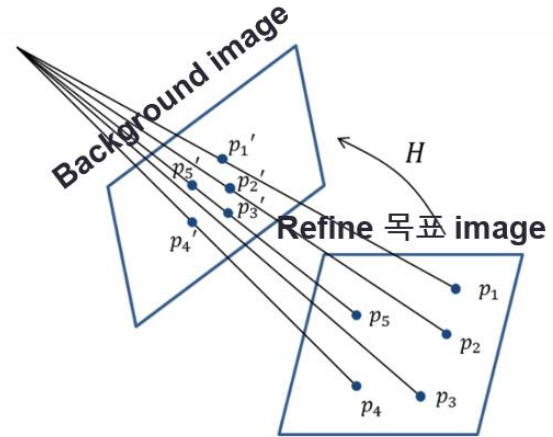


그림 2. Homography 개념

Homography는 그림 2와 같이 한 평면을 다른평면에 투영할 시, 투영되는 대응점 사이에 성립하는 일정한 변환관계를 의미한다. H를 이용한 inpainting의 의미는 refine 목표 영역에 있는 pixel를 H를 적용해 얻은 background의 pixel로 채우는 것을 의미한다.



그림 3. background 선택 과정

그림 3과 같이 어떤 background와 H를 구할지 선택하는 과정이 전제로 이루어져야 한다. 30장의 후보background 이미지들과 refine목표 이미지를 SIFT 매칭을 통해 good features의 개수가 가장 많은 이미지를 best background 이미지로 선택한다.

## IV. 실험 결과

### 4.1. 활용한 소프트웨어 및 실험 장비

본 논문의 실험은 1080-GPU를 장착한 컴퓨터에서 Ubuntu 18.04, Cuda 10.2, Python 3.7 환경에서 진행하였다. 또한, AR 구현을 위해 Colmap 소프트웨어를 사용하였다.

### 4.2. AR

AR에 대한 실험 결과이다. 그림 4는 인하대 캠퍼스의 비행기를 배경으로 한 이미지이다. 두 이미지에 대한 camera pose를 구해 비행기 obj를 띄웠다. 이미지에

따라 camera pose가 달라져도 같은 곳에 obj가 로드되는 것을 확인할 수 있다.



그림 4. 비행기에서의 AR 구현 사진

그림 5는 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 카메라의 초점 거리에 대한 오차율을 계산하였다. 측정된 카메라 초점 거리는 카메라 정보를 담고 있는 Colmap의 결과물인 cameras.txt에서 얻을 수 있다. 스펙이 다른 카메라에서 5개의 장소 당 오차율을 계산했다. 표 2의 결과에 따르면 하이테크 건물에서 오차율이 가장 적으며 해상도에 따라 각 장소에서 오차율이 달라진다. 960\*720 해상도의 이미지 오차율 평균은 0.805%이며 4032\*3024 해상도의 이미지 오차율 평균은 1.388%이다. 고해상도에서 초점 거리에 대한 오차율이 증가한 것은 촬영 환경의 영향으로 판단된다. 10장의 사진 모두 같은 광원의 세기, 위치를 가지고 있을 경우 오차율이 감소한다.

표 2. 카메라 초점 거리에 대한 오차율

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

#### 4.3. DR

DR의 모든 과정은 그림 8과 같이 UI 상에서 수행된다. UI 상에서는 3개의 버튼이 사용 가능하다. Open Image를 통해 inpainting할 사진을 load하고 Mask-RCNN을 기반, 자동으로 사람을 감지한다. 사용자의 선택으로 inpainting할 사람의 bounding box를 지정하고 model Inpainting 버튼을 통해 지정한 사람을 train한 model을 사용하여 inpainting한다. UI상에서 mouse input을 통해 정제하고자 하는 영역을 사용자가 선택할 수 있다. Refine inpainting 버튼을 통해 배경 이미지와의 homography 행렬을 사용하여 inpainting한 최종 이미지를 얻게 된다.



그림 6. 그림자 영역 refine 결과



그림 7. 사람 외곽 영역 refine 결과

그림 6과 그림 7에서 Model inpainting 후 남은 사람의 외곽부분이나 그림자 영역을 homography inpainting을 통해 refine할 수 있다.

표 3은 inpainting이 적용된 영역을 비교한 PSNR 값을 취득한 결과이다. 모든 dataset에서 평균적으로 30dB 이상의 값이고, Homography inpainting이 Model inpainting보다 더 좋은 성능을 보인다.

표 3. 각 Dataset에 따른 PSNR

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



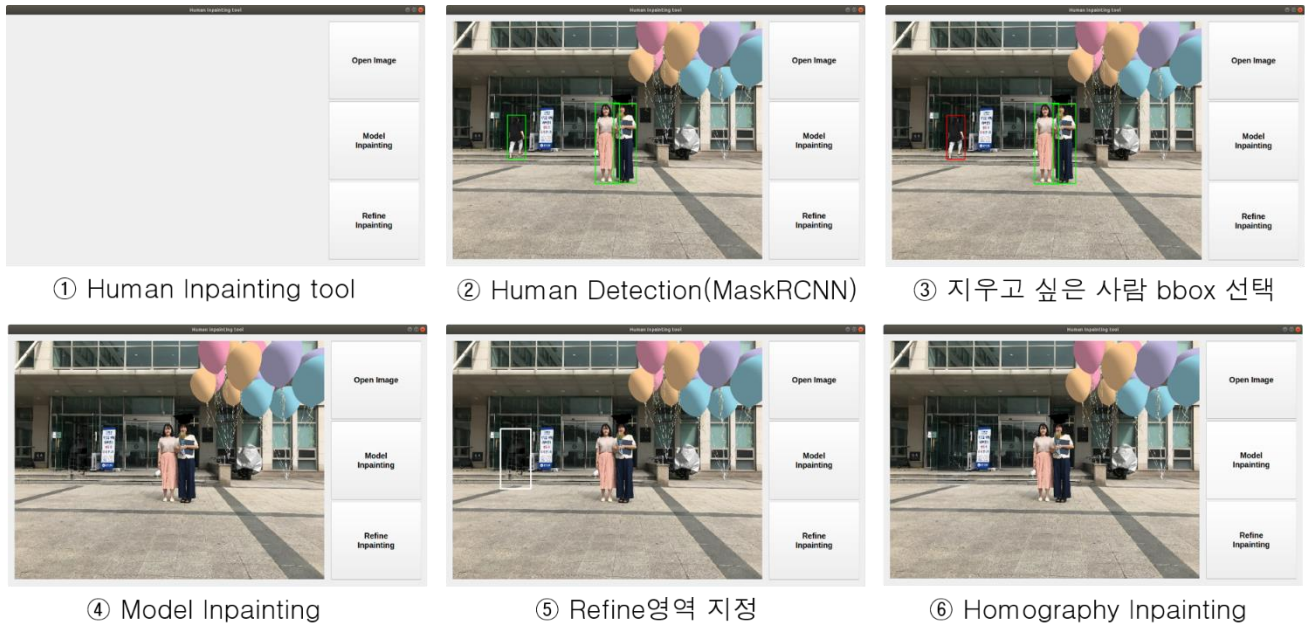


그림 8. Human inpainting tool(UI)

## V. 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. 데이터 개수 제약의 한계를 벗어난다면, model inpainting과 homography inpainting의 정확도를 향상시킬 수 있을 것이라 기대한다.

## 참고문헌

- [1] R. T. Pushpalwar and S. H. Bhandari, "Image inpainting approaches - review," Proc. of IEEE International Conference on Advanced Computing, February 2016.
- [2] J. L. Schönberger and J. Frahm, "Structure-from-motion revisited," Proc. of IEEE Conference on Computer Vision and Pattern Recognition, June 2016.

- [3] K. He et al., "Mask r-cnn," Proc. of IEEE International Conference on Computer Vision, October 2017.
- [4] 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.
- [5] D. DeTone, T. Malisiewicz and A. Rabinovich, "Deep image homography estimation," Proc. of IEEE Conference on Computer Vision and Pattern Recognition, June 2016.



김 유 진 (인하대 정보통신공학과4)

2016년 ~ 현재 인하대학교 정보통신공학 학부과정 재학중.

2020년 8월 졸업 예정.

관심분야는 컴퓨터 비전, 딥러닝.



신 재 이 (인하대 정보통신공학과4)

2017년 ~ 현재 인하대학교 정보통신공학 학부과정 재학중.

2021년 2월 졸업 예정.

관심분야는 컴퓨터비전, 알고리즘.