가상 상품 배치를 위한 파이프라인

문승기 [1], 이정은 [1], 이창엽 [1], 최희진 [2], 박용현 [2], 김현지*
고려대학교 [1], SK 텔레콤 [2,*]

{moon44432[1], esilver[1], ckd248[1]}@korea.ac.kr, {astehelen[2], calm.ardent[2], hyeonji*}@sk.com

A Pipeline for Virtual Product Placement

Seunggi Mun[1], Jeongeun Lee[1], Changyeop Lee[1], Huijin Choi[2], Yonghyeon Park[2], Hyeonji Kim*
Korea University[1], SK Telecom[2,*]

요 약

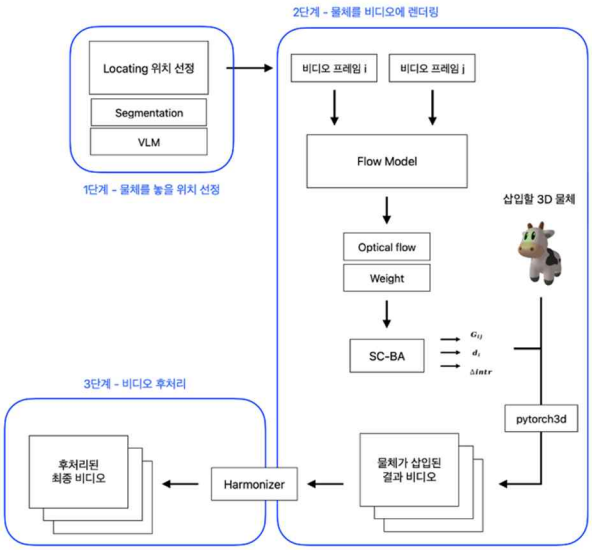
최근 촬영이 완료된 영상에 제품을 삽입하여 광고하는 가상 제품 간접 광고 (VPP; Virtual Product Placement)가 주목 받고 있다. 본 논문에서는 기존 VPP 과정을 자동화하고 제작 비용을 절감할 수 있도록, 영상과 제품 모델링 파일을 입력으로 받아 자동으로 영상 속에 물체를 삽입하고 렌더링할 수 있는 AI 기반의 VPP 생성 기술을 제안한다. 이를 위해 제품 배치 위치 탐색, 제품 렌더링 및 영상 후처리 과정을 포함하는 총 세 단계의 기술을 제시한다.

Ⅰ. 서론

가상 제품 간접 광고(VPP; Virtual Product Placement)는 촬영이 완료된 디지털 콘텐츠 내에 제품을 자연스럽게 합성하여 광고 효과를 창출하는 기술이다. 기존의 제품 배치 작업은 특수효과, 영상편집 등의 전문 도구에 대한 숙련도가 요구되는 과정으로 비용과 시간이 많이 소요되며, 비효율적이다.

본 연구에서는 이러한 문제를 극복하기 위해 AI 기반의 VPP 제작 파이프라인을 제안한다. 제안된 기술은 영상과 제품의 3D 모델링 파일만으로 자동으로 제품을 동영상에 삽입하고 렌더링할 수 있도록, 세 단계로 각 과정으로 구체화한다.

본 논문에서는 파이프라인의 각 단계를 설명하고, 이를 통해 AI 기반 VPP 제작 도구가 어떻게 기존의 복잡한 제품 배치 과정을 자동화하고 효율성을 향상시킬 수 있는지 논의한다.



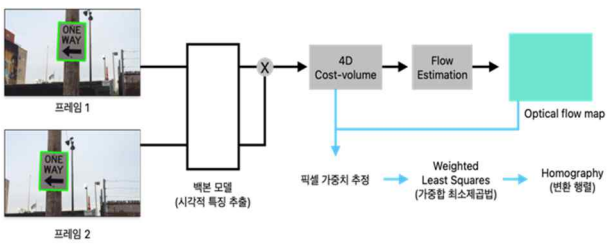
[그림 1] 전체 파이프라인 구조

Ⅱ. 본론

본 연구는 AI 를 통한 VPP 제작을 세 단계로 구체화한다. 제안한 파이프라인의 구조도는 그림 1 에 제시되어 있다. (1) 동영상에서 물체를 배치할 적절한 위치를 탐색 (2) 탐색된 위치에 물체를 렌더링하는 과정 (3) 삽입된 물체가 자연스럽게 합성되도록 후처리를 진행하는 과정이다. 첫번째 단계인 물체 배치 위치 탐색은, GPT-4o 같은 Visual-language 모델과 SAM 모델의 segmentation 결과를 활용할 수 있다. 본 연구에서는 두번째와 세번째 과정인 물체 배치와 후처리를 주로 다룬다.

Ⅱ-1. 동영상 내 객체 배치

동영상 내 객체 배치는 선행 연구[1]를 참고하여, 크게 두 가지 기술로 이루어진다. 먼저 optical flow 추정을 기반으로 한 평면 트래킹 기술과, 동영상이 촬영된 카메라의 카메라 포즈를 추정하기 위한 카메라 포즈 추정 단계이다. 트래킹한 평면 위에 물체가 위치하게 되는데, 이는 WOFT(Planar Object Tracking via Weighted Optical Flow) 모델을 활용한다. 이후 카메라의 포즈 추정을 통해 프레임의 카메라 각도 및 위치 변화에 대응할 수 있게 된다.



[그림 2] WOFT 모델 구조

Ⅱ-1.1. 평면 트래킹 모델

동영상 내에 배치할 객체는 매 프레임에서 일관적인 위치를 유지해야 한다. 이를 optical flow 기반의 평면 트래킹 모델을 통해 구현한다.

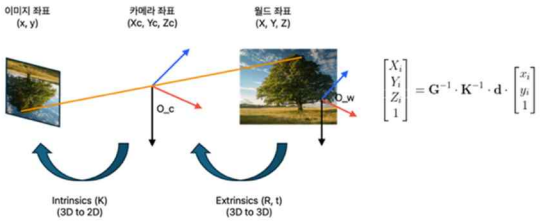
WOFT[2] 모델은 평면 객체 트래킹 벤치마크인 POT-280에서 우수한 성능을 기록한 모델로, 영상 내에서 특정 지점을 기준으로 그 주변의 특징을 분석하여 움직이는 평면의 위치와 방향을 파악한다. 예를 들어, 동영상에서 동일한 책상 표면을 추적하는 경우, WOFT 모델은 책상 표면 주변의 픽셀 이동 경로와 그 경로의 신뢰도를 고려하여 가중치를 부여함으로써 물체의 움직임을 보다 정확하게 추적한다.

또한 물체가 반사되거나 일부가 가려지는 경우, 크기가 변하는 경우, 모션 블러가 있는 경우에도 성공적으로 물체를 추적할 수 있다.

Ⅱ-1.2. 카메라 포즈 추정 모델

객체를 사실적으로 배치하기 위해서는 카메라 위치, 객체의 위치, 물체가 놓일 평면 등을 고려하여 객체를 렌더링해야 한다. 이를 위해 카메라 포즈 추정 모델을 활용한다.

DroidCalib[3]은 비디오로부터 depth map (d), 카메라 intrinsic (K), 카메라 포즈 (G)를 추론하는 모델이다. 객체를 놓을 평면의 픽셀 좌표를 2D-to-3D projection 을 통해 3 차원의 world 좌표로 변환한 후, world 좌표계에서의 카메라 위치, 객체 위치, 평면의 법선 벡터를 계산한다.



[그림 3] 좌표계 변환

DroidCalib 의 depth map 은 카메라의 움직임에 기반한 추론 결과이기 때문에, depth 경계가 모호하여 world 좌표가 부정확한 문제가 있다. 이를 해결하기 위해 Depth Anything v2[4]를 사용해 depth map 을 개별적으로 추론한다. Depth Anything v2 는 하나의 이미지만을 보고 depth 를 추론하는 모델로, 이 모델의 depth 추론 결과를 활용해 정확한 world 좌표를 계산함으로써 렌더링의 성능을 개선한다.

DroidCalib[3] Depth Anything v2[4]

[그림 4] Depth Map 추론 및 렌더링 성능 비교

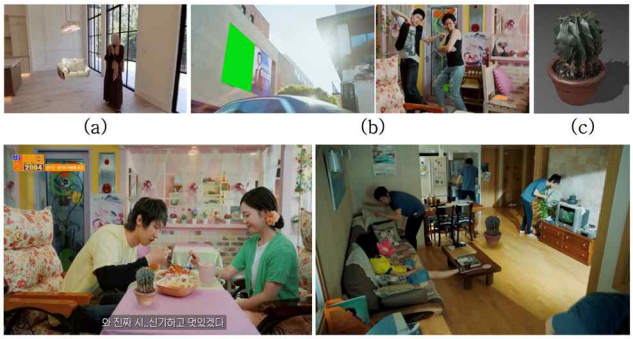
이후 Pytorch3D[5]를 활용해 3D 메쉬 객체를 평면의 법선 벡터 방향과 일치하게 회전시키고, 카메라 위치와 객체 위치를 바탕으로 2D 이미지를 렌더링하여 비디오와 합성한다.

Ⅱ-2. 동영상 후처리

동영상 후처리 단계에서는 Harmonizer[6]를 사용해 보다 자연스러운 비디오를 생성한다.

Harmonizer 는 동영상의 색상, 조명, 그림자 등 다양한 시각적 요소를 분석하여, 새로 삽입된 물체의 특성을 주변 환경과 일치시키는 역할을 한다. Harmonizer 는 iHarmony4 데이터셋을 활용하여 학습된 모델로, 개별 프레임 단위가 아닌 비디오 전체를 고려한 처리 방식을 통해 더욱 우수한 성능을 달성한다.

[그림 5] Harmonizer 적용 전/후 사진



(a) (b) (c) (d)

[그림 6] 기술 적용 사례

그림 6 은 본 논문에서 제안된 파이프라인을 실제 드라마 영상에 적용한 결과이다. 그림 (b)는 제안된 WOFT 모델의 평면 트래킹 결과이며, 물체가 놓일 평면을 나타낸다. 삽입하려는 객체 물체인 선인장의 3D 메쉬(c)가 드라마 장면 속에서 적절하게 삽입된 예시를 (d)에서 확인할 수 있다. 기존 기술인 Place anything[1]의 경우, 그림 (a)처럼 Harmonizer[6]와 그림자 생성이 이루어지지 않아 비교적 부자연스러운 장면이 연출됨을 확인할 수 있다.

Ⅲ. 결론

본 연구는 AI 기술을 활용하여 VPP 를 자동화하기 위한 새로운 방법을 제안한다. 제안된 파이프라인은 제품 배치 위치 탐색, 제품 렌더링, 영상 후처리의 세 단계로 구성되며, 이를 통해 적은 인력과 비용으로도 자연스러운 간접 광고 생성을 가능케 한다. 앞으로 이 기술은 광고 산업 뿐만 아니라 다양한 디지털 콘텐츠 제작 분야에서도 광범위하게 활용될 수 있을 것으로 기대된다.

ACKNOWLEDGMENT

이 논문은 2024 년도 SK 텔레콤의 재원으로 진행된 SKT AI Fellowship 과정에서 수행된 공동연구결과임 (C-12. VPP 를 위한 Object Insertion 기술 연구)

참 고 문 헌

[1] Ziling Liu, Jinyu Yang, Mingqi Gao, and Feng Zheng. Place anything into any video. *arXiv preprint arXiv:2402.14316*, 2024.

[2] Jona □s˘ S˘ery˘ ch and Ji˘r □ı Matas. Planar object tracking via weighted optical flow. In *Proceedings of the IEEE/CVF Win ter Conference on Applications of Computer Vision*, pages 1593–1602, 2023.

[3] Annika Hagemann, Moritz Knorr, and Christoph Stiller. Deep geometry-aware camera self-calibration from video. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 3438–3448, 2023.

[4] Lihe Yang, Bingyi Kang, Zilong Huang, Zhen Zhao, Xiao- gang Xu, Jiashi Feng, and Hengshuang Zhao. Depth anything v2. *arXiv preprint arXiv:2406.09414*, 2024.

[5] Nikhila Ravi, Jeremy Reizenstein, David Novotny, Taylor Gordon, Wan-Yen Lo, Justin Johnson, and Georgia Gkioxari. Accelerating 3d deep learning with pytorch3d. *arXiv preprint arXiv:2007.08501*, 2020.

[6] Zhanghan Ke, Chunyi Sun, Lei Zhu, Ke Xu, and Rynson WH Lau. Harmonizer: Learning to perform white-box image and video harmonization. In *European Conference on Computer Vision*, pages 690–706. Springer, 2022.