

직접 감쇠 모델을 사용한 단안 내시경 비디오에서의 자가지도 깊이 예측 방법

이민호, 박민규, 김제우, 윤주홍

한국전자기술연구원

{amino7, mpark, jwkim, jhyoon}@keti.re.kr,

Self-Supervised Depth Prediction from Endoscopic Monocular Video
Using Direct Attenuation Model

Min Ho Lee, Min-Gyu Park, Je Woo Kim, Ju Hong Yoon,

요 약

내시경 검사는 내장기관의 이상 유무를 점검할 수 있는 효과적인 의료 기술이다. 해당 논문에서는 자가지도 방식의 직접 감쇠 모델(DAM, Direct Attenuation Model)[3]을 사용한 내시경 비디오 기반 깊이 예측을 제안한다. 단안 카메라의 비디오 영상에서 DAM 을 이용한 빛의 밝기에 따른 깊이 변화 정보와 Normal 정보를 사용하여 깊이와 자세 예측 네트워크 모델 학습을 효과적으로 수행한다. 실험을 통해 제안하는 방법은 기존의 깊이 추정 네트워크 대비 다양한 내시경 비디오 영상에서 더 정확하게 깊이를 추정함을 확인하였다.

1. 서론

Global Cancer Statistics 2020 에 따르면 암 사망자 약 100,000 명 중 상당수가 소화기관 암으로 사망했다[1]. 이러한 질병을 내시경 검사로 치료나 예방할 수 있으나, 소장 내부의 방대한 비디오영상의 모든 프레임을 눈으로 확인하는 것은 체력적 소모가 매우 크다. 따라서 컴퓨터 비전의 인식 기술을 이용하여 중요한 키프레임을 검출하고 키프레임에 대해서만 전문의가 집중적으로 검사하는 방안이 최근에 많이 시도 되고 있다 [2].

비디오의 키프레임을 검출하기 위해서는 영상 인식 기술을 포함하여 깊이 추정 기술도 많이 활용되고 있다. 최근에는 딥 러닝 기술 발전으로 단안카메라를 이용한 내시경 이미지에 대한 깊이 추정 기술이 제안되었다[2,3,4]. [4]에서 제안한 기술은 깊이와 Ego-motion 을 동시에 추정하여 깊이 추정에 좋은 성능을 보였다. 그러나 단안 영상만으로 학습을 할 경우 일부 영역의 예측된 깊이가 부정확함을 발견할 수 있다.

앞서 제기된 문제를 해결하기 위해 본 논문에서는 직접 감쇠 모델(DAM, Direct Attenuation Model)을 사용한 단안 내시경 비디오에서 자가지도 깊이 추정 방법을 제안한다. DAM 은 외부의 빛이 제한된 공간에서 광원과 멀어질수록 빛이 어두워지는

원리를 사용한 방법이다[3]. 즉, 빛의 감쇠는 거리의 증가가 된다. 내시경의 촬영 환경에서는 광원이 카메라에 부착되어 있어 내시경 영상의 깊이를 추정하기 위한 딥 러닝 네트워크 학습의 사전정보로 활용 가능하다.

깊이와 자세를 예측하는 네트워크에 DAM 을 사용하여 깊이 추정과 이미지 재건의 정확도를 올렸다. 우리가 제안하는 방법은 기존에 제안된 방법에 대한 성능 향상을 보였고, 더 다양한 내시경 영상에서 좋은 깊이 추정 결과를 보였다.

2. 제안하는 방법

DAM 은 Image Dehazing[6]과 비어-람베르트 법칙이 기반인 모델이다. 외부의 빛이 차단된 내시경에서 빛의 파장길이는 깊이와 같다고 가정했을 때 영상에서 빛의 밝기와 표면반사광에 대한 관계, 빛의 흡수율과 파장의 길이에 대한 관계를 정리하여, 빛의 밝기에 따른 깊이를 구했다. 여기서 구한 깊이와 Normal 정보는 그림 1 의 DepthNet 에서 Decoder 의 각 Layer 초기값을 가이드 해주고, 이미지 재건 손실함수(Image reconstruction loss function)의 입력 값으로 쓰인다.

자가지도(self-supervised) 학습 방식의 깊이 및 Ego-motion 추정은 자세 및 깊이 정보로부터 재건된 이미지를 통해

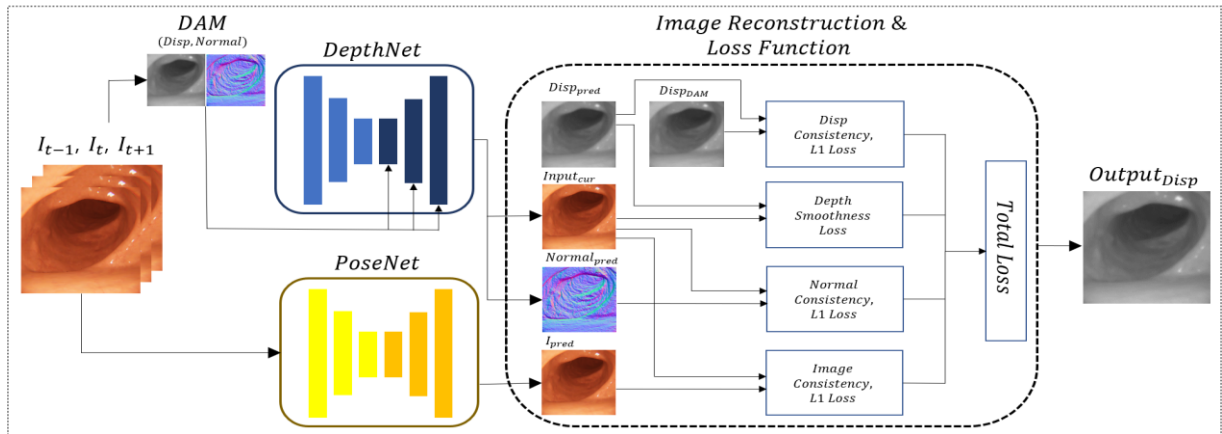


그림 1: 전체 네트워크 개략도

구하게 된다[4]. 전체 네트워크 개략도는 그림 1 과 같다. 깊이와 자세를 예측하는 네트워크가 동시에 학습이 되고, 네트워크는 이미지 재건 손실 함수와 Depth smoothness loss function 을 학습한다.

이미지 재건 손실함수는 Photometric Error(PE)의 최소값을 구한다[4]. PE 는 Target, Source 영상의 구조적 닮음에 대한 오류를 계산한다[7]. Depth smoothness loss 는 RGB 영상 경계선을 희미하게 해 줌으로 깊은 깊이에 대한 학습을 돕는다[2].

3. 실험 및 분석

학습에 사용된 데이터셋은 EndoSLAM 의 Unitycam Colon[5] 영상으로 11,432 개의 Triplet 입력 영상을 사용하였다.

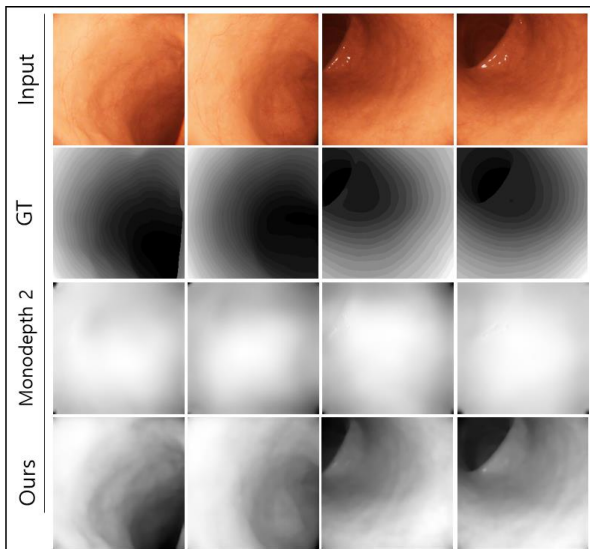


그림 2: 기존 깊이 예측 결과와 제안된 방법의 결과 비교.

그림 2 에서는 Monodepth2[4] 와 우리가 제안하는 방법의 Disparity 결과를 나타낸다. 단안 내시경 비디오 환경에서 Monodepth2 는 깊이 정보를 잘 예측하지 못하는 반면, 우리가 제안하는 방법은 비교적 깊이 정보를 잘 나타내고 있다.

본 논문에서 제안한 DAM 기반의 단안 내시경 비디오에서의

자가지도 깊이 예측 방법은 빛에 반사된 표면의 색 강도, 깊이, 자세, Normal 정보를 사용하여 기존의 깊이 예측 네트워크 학습에 도움을 주었다는 것을 실험적으로 증명하였다. 또한 양안 카메라 기반이 아닌, 단안 카메라의 비디오만을 이용하여 깊이 추정 네트워크 학습이 가능하여 [3]의 기술 대비 실용적임을 확인할 수 있다.

사사의 글

이 논문은 2022 년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (No.2020-0-00457, 중대형 공간용 초고해상도 비정형 플렌옵틱 동영상 저작/재생 플랫폼 기술 개발), (No.2020-0-00103, 가상공간구성을 위한 5G 기반 3D 공간 스캔 디바이스 기술 개발)

참고문헌

- [1] Sung H et al. Global Cancer Statistics 2020. in CA. (2021). 209-249, 71(3).
- [2] Hwang S et al. Unsupervised Monocular Depth Estimation for Colonoscopy System Using Feedback Network. in Sensors. (2021). 21, 2691.
- [3] Yoon J et al. Learning Depth from Endoscopic Images. in 3DV. (2019). pp.126-134.
- [4] Godard C et al. Digging Into Self-Supervised Monocular Depth Estimation. in ICCV. (2019). pp.3828-3838
- [5] Ozyoruk K et al. EndoSLAM Dataset and An Unsupervised Monocular Visual Odometry and Depth Estimation Approach for Endoscopic Videos: Endo-SfMLearner. in MIA. (2021). Volume 71, 102058
- [6] Raanan F. Single image dehazing. in ACM Transactions on Graphics. (2008). Volume 27, pp.1-9
- [7] Godard C, Mac Aodha O, and J.Brostow G. Unsupervised Monocular Depth Estimation with Left-Right Consistency. in CVPR. (2017). pp.270-279