

트랜스포머 기반 깊이 추정을 통한 안개 영상 화질 개선

김종호

순천대학교 인공지능공학부

e-mail: jhkim@scnu.ac.kr

Single Image Haze Removal using Transformer-based Depth Estimation

Jongho Kim

Dept. of Artificial Intelligence Engineering, Sunchon National University

요약

본 논문에서는 안개 성분을 효과적으로 제거하여 영상의 활용성을 증대시키기 위해 트랜스포머 기반 깊이 추정을 포함하는 방법을 제안한다. CNN 기반 모델 및 생성 모델을 통한 안개 제거 영상은 기존의 다양한 알고리즘 기반 방법에 의한 안개 제거 성능을 뛰어넘어 효과적으로 안개를 제거하지만, 합성곱 연산의 본질적 한계로 인해 장거리 종속성을 효과적으로 반영하지 못하고 이에 따라 성능 향상이 제한되었다. 이를 개선하기 위해 본 논문에서는 안개 영상의 깊이 정보를 효과적으로 예측하기 위해 CNN과 트랜스포머의 장점을 모두 활용하는 구조를 제안한다. 트랜스포머의 채택으로 인해 네트워크가 국부적인 디테일을 다루는 능력이 상실되지 않도록 게이트 기반 어텐션 메커니즘을 사용하는 디코더를 제안한다. 본 논문에서는 트랜스포머 구조에 의한 깊이 정보를 이용하여 거리에 따른 안개의 농도를 계산하고 이를 통해 안개 제거 영상을 복원한다. 폭넓은 실험을 통해 제안한 깊이 예측 모델은 세 가지 도전적인 데이터셋에서 가장 좋은(state-of-the-art) 예측 성능을 달성한다는 점을 보여준다.

1. 서론

실외 영상의 활용성을 개선하기 위하여 다양한 안개 영상 제거 방법이 제안되었지만, 최근 딥러닝 기술의 급격한 발전으로 기존의 알고리즘 기반의 방법을 대체하고 있다. 알고리즘 기반 방법은 저 복잡도 및 저 메모리 등과 같은 컴퓨팅 자원이 제한된 많은 응용에서 여전히 수요가 있지만, 안개 제거 성능을 향상시키기 위해서는 다양한 딥러닝 모델이 주목받고 있다. 획득한 영상에서 안개의 성분을 정확하게 예측하여 제거하는 것이 중요한데, 이를 위해서는 카메라와의 거리에 따른 안개 정도를 나타내는 전달량(transmission)을 정확하게 예측해야 한다. 이러한 전달량은 일반적으로 깊이 정보(depth information)와 매우 밀접한 관계가 있으므로 안개 제거 성능은 정확한 깊이 정보 예측에 따라 달라진다[1].

CNN(convolutional neural networks) 기반 모델은 다양한 컴퓨터 비전 작업에 큰 영향을 미쳤지만, 일반적으로 합성곱(convolution) 작업의 본질적인 지역성으로 인해 장거리 종속성을 명시적으로 모델링하지 못하는 한계를 보인다[2]. 초기에 자연어 처리를 위해 설계된 트랜스포머(transformer)는 장거리 종속성을 반영하기 위한 자기 주목(self-attention) 메커니즘을 갖춘 대체 구조로 등장했다[3]. 본 논문에서는 안개 영

상의 깊이 정보를 예측하기 위하여 CNN과 트랜스포머의 장점을 모두 활용하는 구조를 제안한다. 트랜스포머의 채택으로 인해 네트워크가 국부적인 세부 정보를 다루는 능력을 상실하는 것을 방지하기 위해 게이트 기반 주목(attention) 메커니즘을 사용하는 새로운 복호기(decoder)를 제안한다. 트랜스포머 기반의 깊이 예측 결과를 이용하여 효과적인 안개 제거 알고리즘을 구현하도록 한다. 폭넓은 실험을 통해 제안한 깊이 예측 모델은 세 가지 도전적인 데이터셋에서 가장 좋은(state-of-the-art) 성능을 달성한다는 점을 보여준다.

2. 제안하는 트랜스포머 기반 깊이 예측을 통한 안개 제거

영상의 안개 성분을 효과적으로 제거하기 위하여 본 논문에서는 기본적으로 식 (1)과 같은 안개 영상의 광학적 모델을 기반으로 한다.

$$I(\mathbf{x}) = J(\mathbf{x})t(\mathbf{x}) + A(\mathbf{x})(1 - t(\mathbf{x})) \quad (1)$$

이때, \mathbf{x} 는 픽셀의 위치, $I(\mathbf{x})$ 는 카메라를 통해 획득된 안개가 포함된 영상, $J(\mathbf{x})$ 는 복원해야 할 안개가 제거된 영상, $A(\mathbf{x})$ 는 영상의 안개값, $t(\mathbf{x})$ 는 전달량으로서 영상에 안개가 포함된 정도를 나타낸다. 일반적으로 안개값은 픽셀에 따라 다른 값을 가지지만, 보통 영상 전체에 같은 값을 갖는 것으로 모

텔링한다. 전달량 $t(\mathbf{x})$ 는 장면(scene)의 깊이에 밀접한 관련이 있는데, 제안하는 방법은 안개 제거의 성능을 좌우하는 $t(\mathbf{x})$ 를 예측하는데 트랜스포머 기반 모델을 사용한다.

감사의 글

본 논문은 2021년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업의 결과임 (NRF-2021R111A3056637)

참고문헌

- [1] K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 12, pp. 2341-2353, Dec. 2011
- [2] B. Cai, X. Xu, K. Jia, C. Qing, and D. Tao, "DehazeNet: an end-to-end system for single image haze removal," *IEEE Trans. Image Processing*, vol. 25, no. 11, pp. 5187-5198, Nov. 2016
- [3] Y. Song, Z. He, H. Qian, and X. Du, "Vision transformers for single image dehazing," *IEEE Trans. Image Processing*, vol. 32, pp. 1927-1941, 2023