Group Attention U-Net을 이용한 단일 영상 연무제거

홍찬의[1], 박진효[2], 최현덕*

전남대학교

217643@jnu.ac.kr, 206070@jnu.ac.kr, *ducky.choi@jnu.ac.kr

Single Image Dehazing based on Group Attention U-Net

Hong Chan Eui[1], Park Jin Hyo[2], Choi Hyun Duck*

Cheonnam National University

요 약

본 논문에서는 CA(Channel Attention), SA(Spartial Attention), ResNet[3]로 구성된 BottleNeck을 U-Net[1]구조로 접목시켜서 활용한 영상 연무 제거 모델을 제안한다. 영상에서 연무는 균형적으로 형성되지 않고, 특정 부분에 더 많이 일어나고 있음을 알 수 있다. 기존의 연무제거 모델로 주로 사용하는 모델구조인 ResNet[3]은 CNN(Convolution Neural Network)과 skip connection 방식을 사용하여 진행하였다. 이 방식에서는 지역적인 부분에만 작동된다는 점에서 큰 약점이 있다. CA과 SA을 추가하여 글로벌한 정보에서도 손실이 없도록 개선하였다. 제안하는 방법은 이전에 연구된 image dehazing 방법에 비해 PSNR(Peak Signal-to-Noise Ratio)과 SSIM(Structural Similarity Index)에서 우수한 결과를 나타냈다.

Ⅰ. 서 론

최근 인공지능 기술들이 발전하면서 영상을 통해 정보를 추출하고 이를 활용하는 컴퓨터 비전 연구들이 활발하게 진행되고 있다. 대기 상황이 안 좋은 경우 여러 입자들의 영향을 받아 색상 왜곡, 낮은 콘트라스트 등으로 영상이 손상되어 컴퓨터 비전을 토대로 진행하는 방법의 성능을 저하시킨다. 특히 객체 감지가 필요한 자율주행 자동차, 방범용 카메라 등에는 인명피해가 발생하는 치명적인 문제로 작용할 수 있다. 이런 문제를 해결하기 위해 연무로 인해 손상된 영상에서 연무가 제거된 영상을 복구하는 영상 연무 제거 기술을 목표로 한다.

Ⅱ. 본론

단일 영상 연무 제거 기술은 가시성이 낮아진 영상으로부터 연무가 제거된 영상을 복원하는 것을 목표로 한다. 대기 산란 모델은 연무의 간단한 근사치를 제공하며 다음과 같이 공식화된다.

$$I(z) = J(z)t(z) + A(1 - t(z)) \quad (1)$$

I(z)는 관측된 연무 영상, A는 지구 산란광, t(z)는 대기투과율, J(x)는 연무가 제거된 영상이다. 대기 산란 모델은 영상의 dehazing이 A,t(z)에 대한 지식 없이 결정되지 않는 문제임을 보여준다. 다음과 같이 공식화된다.

$$J(z) = \frac{(I(z) - A)}{t(z)} + A \quad (2)$$

공식 1과 2를 통해 연무가 낀 영상에 대해 지구 산란광과 대기 투과율을 적절하게 추정하면 연무가 제거된 영상을 복원할 수 있음을 알 수 있다.

초기 대기 산란 모델은 시각 정보를 추출하여 단일 영상의 연무를 제거하는 연구가 많이 진행되고 있다. DCP(Dark Channel Prior)[5]는 성능이 뛰어난 사전 지식 기반 방법들 중 하나이며 연무가 없는 영상은 R,G,B 채널 중에서 적어도 하나의 채널은 낮은 명도 값을 갖는다라는 가정을 기반으로 제안한다. 그러나 사전 지식 방법은 특정 사례에서는 잘 작동하지 않는다. 최근 딥러닝 기법이 떠오르면서 AOD-Net[6], DehazeNet[7]과 같이 신경망을 사용하는 연무제거 기술이 많이 발달하고 있다. 기존 방법과 비교하여 딥러닝 방법은 end-to-end 방식을 사용하여 처음부터 끝까지 기계가 알고리즘을 찾는 방법으로 사람이 개입함으로써 생기는 오류를 줄인다. 이를 빅데이터에 적용하면 우수한 품질의 영상을 얻을 수 있다.

본 논문은 CA, SA, ResNet[3]을 bottleneck구조로 구성한 GAB(Group Attention Block)을 기반으로 U-Net[1] 구조를 활용한 새로운 end-to-end 방식의 신경망을 제안한다. 그림 1은 제안하는 신경망 구조이다.

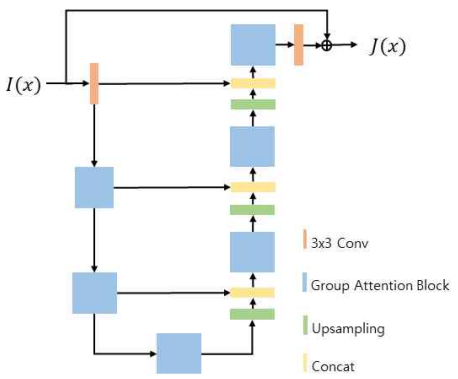

그림 1. 제안하는 GAB 기반 연무제거 모델

U-Net[1]은 영상의 크기를 줄이면서 이미지 내의 중요한 정보를 직접 전달(skip-connection)하여 디코더에서 선명한 이미지를 얻게 해 정확하게 연무 제거를 하는 데에 효과적인 구조를 가진다. 신경망의 구조에서 GAB 구조는 아래의 그림 2처럼 구성되어 있다.

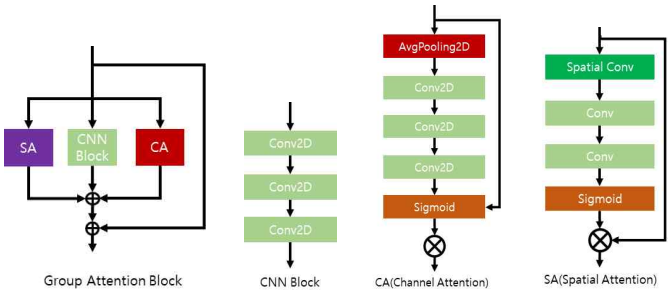

그림 2. GAB 구조

GAB은 CA, SA, CNN을 bottleneck 구조로 구성하고 skip connection을 통해 residual 구조를 사용한다.

CNN Block은 CNN을 통해 image feature을 생성한다. CA는 averagepooling을 이용해 $C\times W\times H$에서 $C\times 1\times 1$로 만들고 CNN을 통해 attention해서 channel feature를 추출한다. SA는 CNN을 이용해 $C\times W\times H$에서 $1\times W\times H$로 만들어 CNN를 통해 attention해서 spatial feature를 추출한다. CA, SA, CNN을 bottleneck 구조로 이용함으로써 채널단위, 픽셀단위, 영상단위로 구성된 다양한 관점으로 영상에서 연무의 특징을 분석하여 효과적으로 복원한다.

GAB의 bottleneck 구조는 ResNext[4]에서 사용되는 Group Convolution과 유사한 구조로 형성이 되어 있으며, Group Convolution에서는 CNN을 3개 이상 구성해야 좋은 성능이 나타난다. 기존에 CA, SA은 2개의 CNN으로 구성했지만, 각각 CNN이 3개씩 구성하여 보다 좋은 성능을 이끌어냈다.

제안하는 방법을 학습할 때 학습용 데이터셋으로 RESIDE (Realistic Single Image Dehazing)이라는 합성 및 실제 흐릿한 이미지로 구성된 대규모 벤치마크를 사용하였다. ITS (Indoor Training Set) 110,000여개, OTS (Outdoor Training Set) 313,959여개 등등으로 구성되어 있다. train data와 ground truth간의 MSE(Mean Square Error)를 비용함수(Loss Function)로 설정하고, 초기 learning rate를 0.0001로 기준으로 10epoch마다 0.2배씩 늘려가면서 image dehazing 실험을 진행하였다. 실험 결과 PNSR 지표는 30.96, SSIM 지표는 0.954로 보여주고 있다. 표 1에서 제안하는 방법이 기존의 연무제거 모델보다 연무제거에 효과적이며, 우수한 성능이 나타나는 것을 확인할 수 있다.

	SSIM	PSNR
DCP[5]	16.62	0.82
AOD-Net[6]	19.06	0.85
DehazeNET[7]	21.14	0.86
GFN[8]	22.30	0.880
Ours	31.07	0.954

표 1. 지표분석

Ⅲ. 결론

본 논문에서는 단일 영상의 연무를 효과적으로 제거하기 위해 CA, SA, ResNet[3]을 bottleneck방식으로 구성된 block으로 기반으로 U-Net[1]구조를 활용한 새로운 end-to-end 방식의 신경망을 제안했다. 제안하는 네트워크는 구조는 다양한 attention을 통해 다각도로 영상의 특징을 분석하여, 영상의 세부사항과 색상 복원에 강력한 이점을 갖고 있다. 제안하는 새로운 신경망은 RESIDE Dataset를 토대로 진행된 image dehazing 실험을 통해 효과적으로 검증되었으며, super-resolution, denosing, deraining과 같은 작업에도 탁월할 것으로 예상된다.

ACKNOWLEDGMENT

본 논문(저서)은 2020년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원(No. 2020R1G1A1101070)을 받아 수행된 연구임.

참 고 문 헌

[1] Ronneberger O., Fishcer P. and Brox T., “U-Net[1]: Convolutional Networks for Biomedical Image Segmentation,” Medicial Image Computing and Computer-Assisted Intervention-MICCAI 2015, pp 234-241, Nov. 2015.

[2] Park J., Woo S., Lee J.Y. and Kweon I.S., “BAM: Bottleneck Attention Module,” BMVC, Oct. 2018.

[3] He K., Zhang X., Ren S., and Sun J., “Deep Residual Learning for Image Recognition,” 2016 IEEE Conference on Computer Vision and Pattern Recognition, pp. 770-778, June. 2016.

[4] Xie S., Girshick R., Dollar P., Tu Z., and He K., “Aggregated Residual Transformations for Deep Neural Networks,” 2017 IEEE Conference on Computer Vision and Pattern Recognition, pp. 5987-5995, July. 2017.

[5] He K., Sun J. and Tang X., “Single Image Haze Removal Using Dark Channel Prior,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 33, no. 12, pp. 2341-2353, Dec. 2011.

[6] Li B., Peng X., Wang Z., Xu J. and Feng D., "AOD-Net: All-in-One Dehazing Network," 2017 IEEE International Conference on Computer Vision, pp. 4780-4788, Dec. 2017.

[7] Cai B., Xu X., Jia K., Qing C. and Tao D., “DehazeNet: An End-to-End System for Single Image Haze Removal,” IEEE Transactions on Image Processing, vol. 25, no. 11, pp. 5187-5198, Nov. 2016.

[8] Ren W., Ma L., Zhang J., Pan J., Cao X., Liu W. and Yang M., “Gated Fusion Network for Single Image Dehazing," 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 3253-3261, June. 2018.

[9] Li, B., Ren, W., Fu, D., Tao, D., Feng, D., Zeng, W., and Wang, Z., “Benchmarking single-image dehazing and beyond,” IEEE Transactions on Image Processing, vol. 28, no. 1, pp.492 - 505, Jan. 2018.