

FisheyeNet: 딥러닝을 활용한 어안렌즈 왜곡 보정

이흥재, 원재성, 이다은, 이성배, *김규헌
경희대학교

jimmy9704@khu.ac.kr, 2016104008@khu.ac.kr, daeun197@khu.ac.kr,
rhee@khu.ac.kr, *kyuheonkim@khu.ac.kr

FisheyeNet: Fisheye Image Distortion Correction through Deep Learning

Hongjae Lee, Jaeseong Won, Daeun Lee, Seongbae Rhee, *Kyuheon Kim
Kyunghee University

요약

Fisheye 카메라로 촬영된 영상은 일반 영상보다 넓은 시야각을 갖는 장점으로 여러 분야에서 활용되고 있다. 그러나 fisheye 카메라로 촬영된 영상은 어안렌즈의 곡률로 인하여 영상의 중앙 부분은 팽창되고 외곽 부분은 축소되는 방식 왜곡이 발생하기 때문에 영상을 활용함에 있어서 어려움이 있다. 이러한 방식 왜곡을 보정하기 위하여 기존 영상처리 분야에서는 렌즈의 곡률을 수학적으로 계산하여 보정하기도 하지만 이는 각각의 렌즈마다 왜곡 파라미터를 추정해야 하기 때문에, 개별적인 GT (Ground Truth) 영상이 필요하다는 제한 사항이 있다. 이에 본 논문에서는 렌즈의 종류마다 GT 영상을 필요로 하는 기존 기술의 제한 사항을 극복하기 위하여, fisheye 영상만을 입력으로 하여 왜곡계수를 계산하는 딥러닝 네트워크를 제안하고자 한다. 또한, 단일 왜곡계수를 왜곡모델로 활용함으로써 layer 수를 크게 줄일 수 있는 경량화 네트워크를 제안한다.

1. 서론

Fisheye 카메라로 촬영된 영상은 일반 영상보다 넓은 시야각의 영상이라는 장점이 있으며, 현재 VR을 이용한 영상 통화, 게임 및 360 image slam, 아라운드 뷰 등에서 활용되고 있다. 그러나 fisheye 카메라로 획득한 영상은 영상의 중앙 부분이 팽창되고 외곽 부분은 축소되어 보이는 방식 왜곡이 나타나기 때문에, 이를 보정 하지 않고 활용하기에는 어려움이 있다.

상기와 같은 방식 왜곡을 보정하기 위하여 기존 영상처리 분야에서는 렌즈의 곡률을 수학적으로 계산하거나 영상 속 사물의 형태 변화도를 통해서 영상의 왜곡을 보정하기도 하였다. 그러나 fisheye 카메라에 부착된 렌즈는 종류마다 서로 다른 곡률을 가지고 있기 때문에 렌즈가 바뀔 때마다 렌즈의 곡률을 수학적으로 계산하거나, fisheye 영상의 변화도 추정을 위해서 많은 양의 GT (Ground Truth) 영상이 필요하다는 제한 사항이 있다.

기존 기술의 제한 사항을 극복하기 위하여 fisheye 영상만을 입력 영상으로 활용하여 왜곡 파라미터를 추정할 수 있는 딥러닝 네트워크를 제안하고자 하며, 딥러닝 네트워크를 통해서 왜곡 파라미터를 추정하는 선행 연구의 제한 사항들을 분석함으로써, 선행 기술의 제한 사항을 극복할 수 있는 방법을 제안하고자 한다.

이에 본 논문의 2장에서는 fisheye 영상의 왜곡을 보정 하는 선행 연구를 분석하고자 하며, 3장에서는 본 논문에서 제안하는 딥러닝 네트워크인 FisheyeNet에 대하여 자세히 설명한다. 또한, 4장에서는 테스트 영상을 통해서 실험함으로써 본 논문에서 제안하는 기술의 효용성을 검증하고자 한다. 이어지는 5장에서는 결론을 지으며 마무리한다.

2. 관련 연구

기존 왜곡 보정 방법에는 카메라 내부 파라미터를 활용하는 방법과, 영상 내 특징점(Keypoint)을 활용하는 방법이 있다[1,2,3]. 먼저, 카메라 내부 파라미터를 활용하는 방법은 렌즈의 곡률과 영상이 투사되는 평면적을 통해서 왜곡모델을 계산하고 좌표 변환하여 왜곡을 보정하는 방법이다. 따라서 카메라 내부 파라미터를 활용하는 방법으로 영상의 왜곡을 보정하기 위해서는 촬영하는 카메라의 구체적인 제원을 알아야 한다는 제한 사항이 존재한다.

다음으로 영상 내 특징점을 활용하는 방법은 이동, 회전, 확대 및 축소 등의 변환 과정에도 동일한 지점으로 나타나는 강인한 지점을 활용하는 방법으로, 체크보드와 같이 특수한 무늬에서 직접 특징점을 지정하는 방법[4]과 특징점 검출 알고리즘을 통해서 추정하는 방법으로 구분되지만, 공통적으로 왜곡 영상과 GT 영상에서 동일한 지점의 특징점으로 변환 관계식을 추정하여 왜곡을 보정 한다[5,6,7]. 이는 앞선 카메라 내부 파라미터를 활용하는 방법과 달리 카메라의 구체적인 제원이 필요 없다는 장점이 있지만, 변환식은 매칭되는 점들의 평균적인 변환 관계이기 때문에 위치에 따라 왜곡 정도가 크게 다른 fisheye 영상을 보정하는 것에는 효과적이지 못하다는 제한 사항이 있다.

최근에는 상기와 같은 제한 사항을 극복하기 위하여 딥러닝을 활용한 영상의 왜곡 보정 연구가 진행되고 있다[8,9]. 그러나 해당 연구에서는 입력 영상의 가장자리 부분을 잘라내어 CNN(Convolutional Neural Network)으로 학습시키기 때문에 영상의 가장자리로 갈수록 강해지는 왜곡을 왜곡계수로 표현하는 것이 제한되며, 깊은 layer를 필요로 하고 왜곡계수의 추정 정확도 또한 낮게 나타난다.

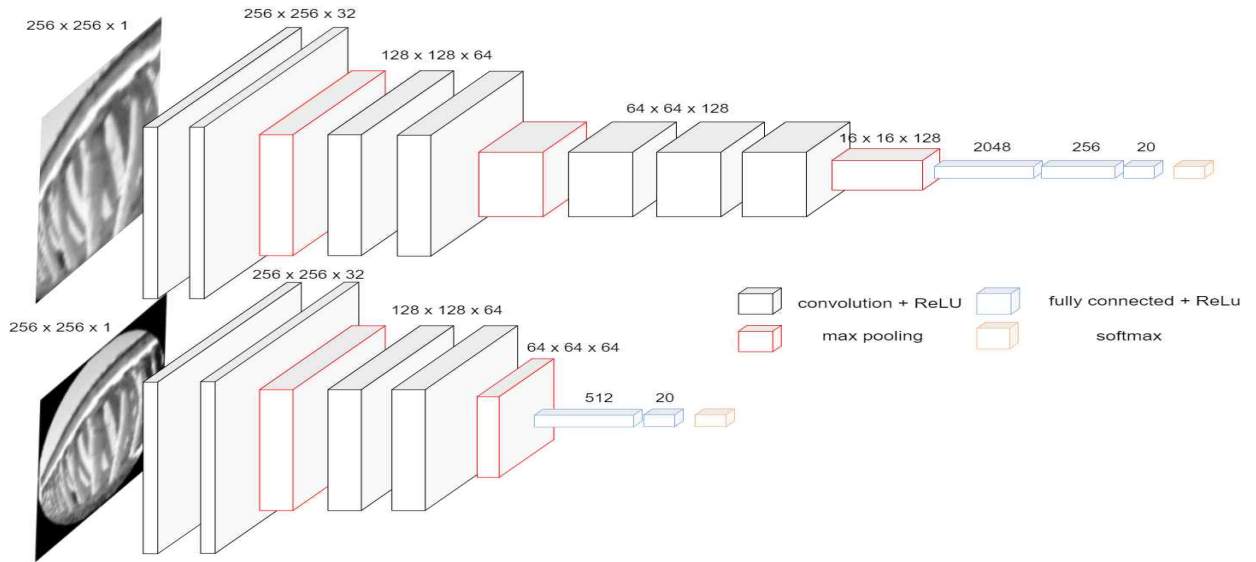


그림 1. 왜곡 보정 딥러닝 네트워크 비교, 위: 사각형 영상 학습 네트워크. 아래: 원형 영상 학습 네트워크

이에 본 논문에서는 가장자리 부분을 제거하지 않은 원본 fisheye 영상을 입력 영상으로 하여 높은 왜곡계수 추정 정확도를 획득할 수 있었다. 또한, 단일 왜곡모델을 활용하여 딥러닝 네트워크가 학습할 사항을 단순화함으로써 적은 수의 layer로 구성된 경량화 딥러닝 네트워크를 구성하였고 이를 이어지는 3장에서 자세히 설명하고자 한다.

3. 제안 기술 설명

3.1 왜곡모델

Fisheye 영상에서 나타나는 왜곡은 방사 왜곡과 접선 왜곡으로 구성된 왜곡모델로 표현할 수 있다. 방사 왜곡은 볼록렌즈의 굴절률로 인해 렌즈 중심과의 거리 차이에 따라 왜곡 정도가 결정되는 왜곡이며, 접선 왜곡은 카메라 제조과정에서 렌즈와 영상 센서가 수평을 이루지 못해 발생하는 왜곡이다. 위의 두 왜곡을 고려한 왜곡모델은 식 1과 같이 나타낼 수 있다. 식 1의 우변에서 더하기 연산자 앞에 있는 다항식은 방사 왜곡을 나타내며, 더하기 연산자 뒤에 있는 다항식은 접선 왜곡을 나타낸다.

$$\begin{bmatrix} x_{n,d} \\ y_{n,d} \end{bmatrix} = (1 + k_1 r_u^2 + k_2 r_u^4 + k_3 r_u^6) \begin{bmatrix} x_{n,u} \\ y_{n,u} \end{bmatrix} + \begin{bmatrix} 2p_1 x_{n,u} y_{n,u} + p_2 (r_u^2 + 2x_{n,u}^2) \\ p_1 (r_u^2 + 2y_{n,u}^2) + 2p_2 x_{n,u} y_{n,u} \end{bmatrix} \quad (1)$$

이때 fisheye 영상의 왜곡은 주로 방사 왜곡으로부터 발생되며, 방사 왜곡을 식 2에서 나타나는 바와 같이 이차항까지의 계산만으로 단일 매개 변수 왜곡모델을 구성함으로써[10], 딥러닝 네트워크의 학습 모델을 단순화하였다.

$$r_d = \frac{r_u}{1 + k r_d^2} \quad (2)$$

3.2 Dataset 구성

본 논문에서는 Place365 Dataset[11]을 이용하여 dataset을 구성하였고, dataset의 size를 최소화하기 위해 RGB 영상을 흑백 영상으로 변환하였다. 또한, 딥러닝 네트워크를 훈련함에 있어서 어려운 점은 라벨링 된 dataset를 구성하는 것이기에, 단일 방사 왜곡계수 k로 영상을 왜곡하여 200장의 valid set과 10,000장의 train set으로 dataset을 구성하였다. 이때, 왜곡계수는 1부터 20까지 총 20개로 설정했다.

3.3 네트워크 구조

최근 딥러닝 네트워크를 통해 영상의 왜곡을 보정하는 연구가 제안되고 있다. 하지만 대부분의 연구에서는 사각형 영상을 네트워크의 입력으로 사용하여 왜곡계수를 추정한다는 제한 사항이 있다[8,9]. 왜냐하면 방사 왜곡은 볼록렌즈의 굴절률에 의한 것으로, 왜곡의 정도가 중심에서 멀어질수록 강하게 나타나는데, 이를 네트워크에 학습시키기 위해 사각형 영상으로 크롭(crop) 하여 넣으면 강하게 왜곡된 영상의 외곽 부분이 손실되어 네트워크가 학습하는데 어려움이 발생하기 때문이다.

따라서 본 논문에서는 네트워크의 입력으로 크롭하지 않은 원형 fisheye 영상을 사용하는 FisheyeNet을 제안한다. 그림 1은 본 논문에서 제안하는 FisheyeNet(아래)과, VGGNet[12]을 기반으로 하여 사각형 영상을 입력으로 사용하는 본 논문의 비교 네트워크로 설계한 SquareNet(위)의 구조이다. FisheyeNet은 크롭하지 않은 원형 fisheye 영상을 입력으로 사용하여 크롭된 영상을 입력으로 하는 SquareNet에 비해 경량화된 네트워크 구조를 가질 수 있다.

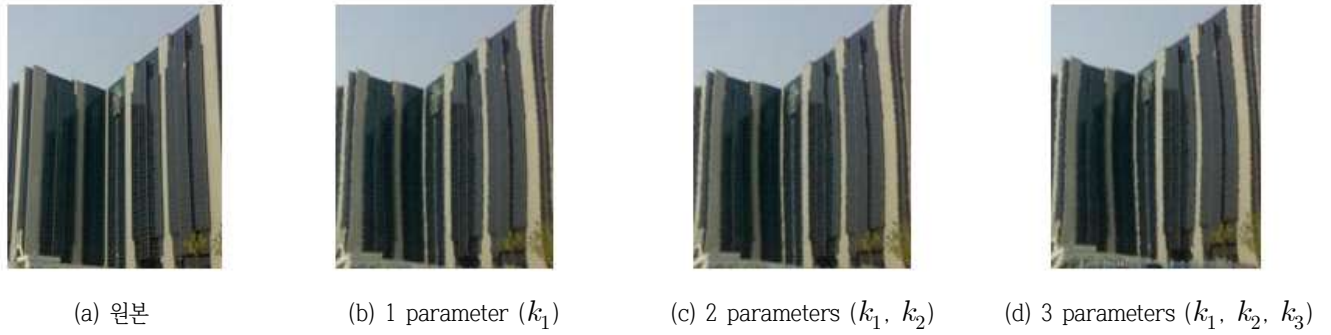


그림 2. 복원에 사용된 parameter의 수에 따라 복원된 fisheye 영상 (수식 (1) 계수)

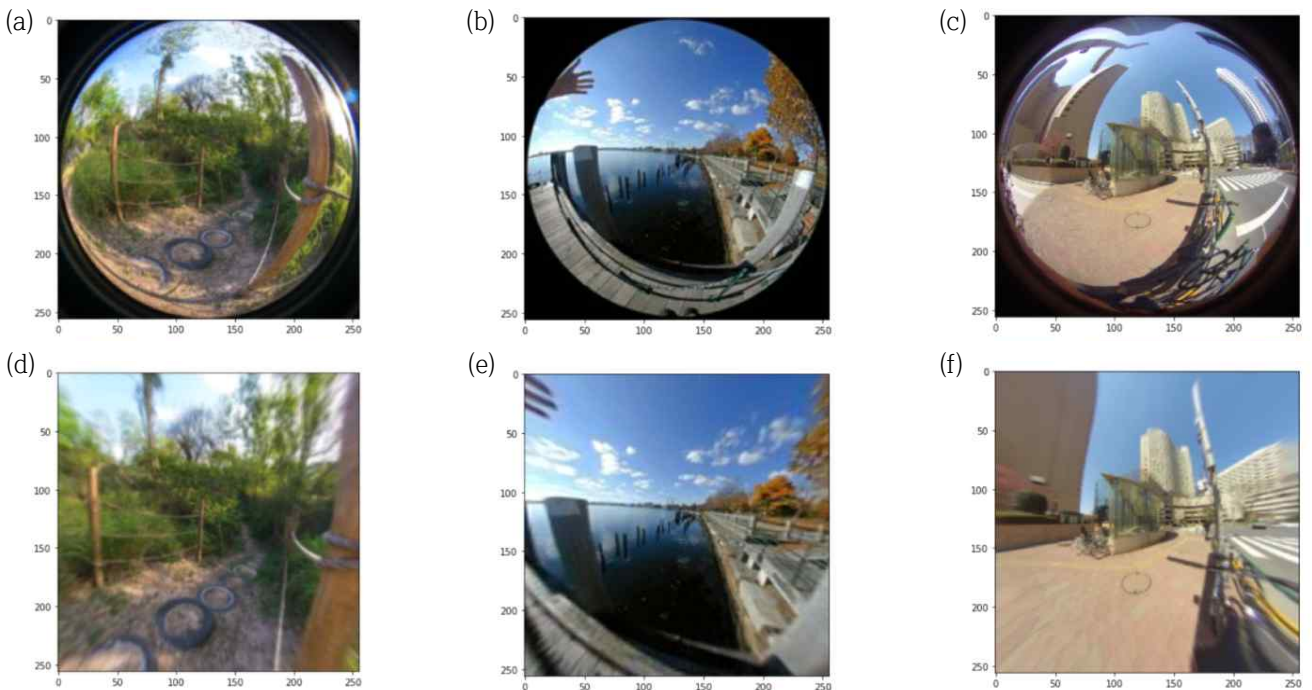


그림 3. 원본 fisheye 영상(a)(b)(c) FisheyeNet을 이용하여 왜곡 보정 한 fisheye 영상(d)(e)(f)

4. 실험 결과

4.1 왜곡모델에 의한 성능 비교

본 논문에서 제안하는 FisheyeNet은 네트워크가 추론할 계수를 최소화하기 위해 단일 왜곡 매개 변수를 이용한 왜곡모델을 사용하였다. 따라서 단일 왜곡계수 모델의 성능을 확인하기 위해 다항의 왜곡계수를 사용한 모델과 비교 실험을 진행한다. 이때 실험환경은 Google Colab 으로 실험하였다.

그림 2의 (a)에 나타나는 원본 영상을 k_1, k_2, k_3 개의 파라미터로 왜곡시킨 뒤, 각각 1, 2, 3개의 파라미터로 복원하였다. 각각의 결과 영상은 그림 2의 (b), (c) 그리고 (d)로 나타난다. 이때, 복원된 각각의 영상에 대해 원본 영상과의 PSNR, SSIM을 측정 한 결과는 표 1과 같이 나타났으며, 복원된 파라미터 수에 따른 PSNR과 SSIM의 차이가 크지 않기 때문에 단일 왜곡계수만으로도 fisheye 영상의 왜곡을 보정 할 수 있다

는 것을 확인할 수 있다.

	PSNR	SSIM
1 parameter	14.50843	0.305251
2 parameter	14.94706	0.284549
3 parameter	14.0249	0.256657

표 1. Parameter의 수에 따른 PSNR, SSIM

4.2 네트워크 성능 비교

본 논문에서 제안하는 FisheyeNet으로 fisheye 영상을 왜곡 보정한 결과는 그림 3과 같이 나타난다. 또한, VGGNet 기반의 SquareNet과 fisheye 영상에 특화된 FisheyeNet의 성능을 비교하기 위해 각 네트워크의 train time과 valid time, valid accuracy를 측정하여 표 2에서 비교해 보았다. 이때 두 네트워크 모두 256x256 흑백 영상 10,000장

을 이용하여 학습하였다. 원형 영상을 네트워크의 학습 데이터로 사용한 결과 FisheyeNet은 사각형 영상을 사용한 SquareNet에 비하여 layer를 줄일 수 있었다. layer의 수를 최소화한 결과로 FisheyeNet이 train time과 valid time에서 모두 SquareNet에 비해 좋은 성능을 보였으며, valid accuracy 또한 98%로 FisheyeNet이 높게 나타났다.

	Train time	Valid time	Valid accuracy
SquareNet	658.12(sec)	10.32(sec)	80%
FisheyeNet	224.555(sec)	4.91(sec)	98%

표 2. 네트워크의 train time, valid time, valid accuracy 비교

5. 결론

방사 왜곡 보정엔 앞에서 기술한 바와 같이 다양한 방법들이 있다. 하지만 기존에 왜곡을 보정하는 방법들은 결국 다양한 시점에서 얻은 영상을 통해서 캘리브레이션을 해야 하거나, 영상에 사용자가 사전 처리 작업으로 특징점을 일일이 설정해야 한다는 제한 사항이 있다.

이와 같은 제한 사항으로 인해 최근엔 딥러닝을 활용한 왜곡 보정 연구가 활발하게 이뤄지고 있다. 하지만 해당 연구들은 네트워크의 입력을 원본 fisheye 영상이 아닌 크롭된 일부만을 사용한다. 이는 영상의 가장자리에 갈수록 강해지는 방사 왜곡의 특성을 온전히 반영하지 않게 되고 결국 방사 왜곡 추정 시 어려움이 생기게 된다.

이를 개선하기 위해서 본 논문에서 제시하는 FisheyeNet에서는 크롭하지 않은 원형 fisheye 영상을 네트워크의 입력으로 사용한다. 이를 통해 입력 영상의 가장자리를 최대한 살려 가장자리로 갈수록 지수적으로 커지는 방사 왜곡의 영향을 최대한 보존할 수 있다. 또한, 단일 왜곡 계수를 활용함으로써 layer를 줄이더라도 높은 valid accuracy를 획득하였으며, train, valid time을 65.88% 단축할 수 있었다.

본 논문을 바탕으로 아래와 같은 후속 연구를 제안할 수 있다. [13]에서는 두 개의 어안렌즈를 이용하여 얻은 영상을 스티칭하여 360 FOV 영상을 만들었다. 이때, 본 논문에서 제안하는 FisheyeNet을 이용한다면 ground truth가 없는 원형 fisheye 영상을 왜곡 보정한 후 스티칭 하여 왜곡이 없는 360 FOV 영상을 얻을 수 있을 것이다.

참고문헌

[1] imone Gasparini, Peter Sturm, and João P. Barreto. "Plane-Based Calibration of Central Catadioptric Cameras." In ICCV 2009.

[2] Christopher Mei and Patrick Rives. 2007. Single View Point Omnidirectional Camera Calibration from Planar Grids. In ICRA.

[3] Davide Scaramuzza, Agostino Martinelli, and Roland Siegwart. "A Toolbox for Easily Calibrating Omnidirectional Cameras." In IROS 2006. .

[4] Zhang, Z. "A flexible new technique for camera calibration." IEEE Transactions on Pattern Analysis and Machine Intelligence 2000

[5] Faisal Bukhari and Matthew N Dailey. "Automatic radial distortion estimation from a single image." Journal of mathematical imaging and vision, 45(1):31-45, 2013

[6] Ciaran Hughes, Patrick Denny, Martin Glavin, and Edward Jones. "Equidistant fisheye calibration and rectification by vanishing point extraction." IEEE Transactions on Pattern Analysis and Machine Intelligence, 32(12):2289-2296, 2010.

[7] Xianghua Ying and Zhanyi Hu. "Can we consider central catadioptric cameras and fisheye cameras within a unified imaging model." Computer Vision-ECCV 2004, pages 442-455, 2004.

[8] Jiangpeng Rong, Shiyao Huang, Zeyu Shang, and Xianghua Ying. "Radial lens distortion correction using convolutional neural networks trained with synthesized images." In Asian Conference on Computer Vision, pages 35-49. Springer, 2016.

[9] Xiaoyu Li, Bo Zhang, Pedro V Sander, and Jing Liao. 2019. "Blind Geometric Distortion Correction on Images Through Deep Learning." In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 4855-4864.

[10]. Claus, D., Fitzgibbon, A.W. "A rational function lens distortion model for general cameras." 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2005), vol. 1, pp. 213-219. IEEE (2005)

[11] B. Zhou, A. Khosla, A. Lapedriza, A. Torralba, and A. Oliva. "Places: An image database for deep scene understanding." arXiv preprint arXiv:1610.02055, 2016.

[12] Simonyan, K., Zisserman, A. "Very deep convolutional networks for large-scale image recognition" arXiv preprint arxiv:1409.1556 (2014)

[13] T. Ho and M. Budagavi, "Dual-fisheye lens stitching for 360-degree imaging," in Proc. of the 42nd IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'17), 2017