일반논문 (Regular Paper)
방송공학회논문지 제28권 제5호, 2023년 9월 (JBE Vol.28, No.5, September 2023)
https://doi.org/10.5909/JBE.2023.28.5.603
ISSN 2287-9137 (Online) ISSN 1226-7953 (Print)

개선된 분류 기법 기반 단일 이미지 카메라 캘리브레이션 알고리즘

원 종 수[a)], 한 종 기[a)‡]

Modified Classification Algorithm for Single Image Camera Calibration

JongSu Won[a)] and JongKi Han[a)‡]

요 약

카메라 캘리브레이션은 가상 현실, 3D 복원, 왜곡 보정 등 컴퓨터 비전의 다양한 분야들의 기초이다. 대규모 3D 복원을 수행하기 위해 많은 개수의 이미지들이 필요한데, 정밀한 3D 복원을 가능하게 하려면 정확한 카메라 내부 파라미터가 필요하다. 이 과정에서 고가의 전문 장비를 활용하면 제작 비용이 높아지는 문제가 있다. 이에 대한 대안으로 휴대폰을 사용하여 영상 정보를 획득할 수 있는데, 이 경우 각 촬영 영상에 대한 카메라 내부 파라미터를 정확하게 추정하는 방법에 어려움이 있다. 보통의 카메라 캘리브레이션은 체커보드와 같은 교정의 기준이 되는 물체가 사용되고, 이 과정은 매우 번잡하여, 많은 개수의 영상 정보에 적용하기에는 어려움이 많다. 따라서 체커보드 등 추가의 과정이 사용되지 않고, 촬영된 영상 정보만을 가지고, 카메라 내부 파라미터를 추정할 수 있는 딥러닝 네트워크를 이용하여 초점 거리와 왜곡 계수를 구하는 과정은 매우 중요한 연구 주제이다. 지금까지 이 분야의 전문가들에 의해 제안된 방법들에서는, 초점 거리와 왜곡 계수 모두를 딥러닝 네트워크로 계산하는 것에 복잡도와 정답률 관점에서 만족할 만한 성능을 보이지 못하고 있다. 본 논문에서는 딥러닝 네트워크의 학습에 사용되는 코스트 함수값을 새로 제안하여, 실제 값과 예측 값의 차이를 반영하도록 설계하였다. 이러한 방법을 기반으로 컴퓨터 시뮬레이션을 수행한 결과, 향상된 정확도를 얻을 수 있었다.

Abstract

Camera calibration is the foundation of various fields in computer vision, such as virtual reality, 3D reconstruction, and distortion correction. To perform large-scale 3D reconstruction, a large number of images are required, and accurate camera intrinsics are necessary for precise 3D reconstruction. However, using expensive specialized equipment for camera calibration increases production costs. As an alternative, smartphones can be used to acquire video information, but there are difficulties in accurately estimating the camera intrinsics for each captured image. Conventional camera calibration uses a calibration object such as a checkerboard, and this process is very cumbersome, making it difficult to apply to a large amount of image data without additional steps such as using a checkerboard. Therefore, it is a very important research topic to estimate the focal length and distortion coefficients using only captured video information without additional steps, using a deep learning network. So far, the methods proposed by experts in this field have not shown satisfactory performance in terms of complexity and accuracy in calculating both the focal length and distortion coefficients using a deep learning network. In this paper, we propose a new cost function value for use in deep learning network training, designed to reflect the difference between predicted values and actual values. Based on this method, computer simulations resulted in improved accuracy.

Keyword : Camera Calibration, Deep learning, Focal length, Distortion coefficient

Ⅰ. 서 론

카메라 캘리브레이션은 왜곡 보정, 3D 복원, 가상현실(VR) 등 컴퓨터 비전의 많은 분야에서 활용된다[1]. 예를 들어, 대규모 3D 복원을 수행하기 위해 많은 개수의 이미지들이 필요하다. 이 촬영된 영상들에 대해 카메라 캘리브레이션 과정을 거치는데, 이때 초점거리와 같은 카메라 내부 파라미터가 필요하다. 이를 위해 전문 장비를 사용하여 정확한 카메라 내부 파라미터를 얻는다면 비용이 너무 많이 든다. 비용을 줄이기 위해서 휴대폰과 같이 비교적 접근이 쉬운 장비를 이용하여 찍은 이미지를 사용할 수 있다. 하지만 휴대폰을 이용하여 찍은 이미지를 이용할 경우 다음과 같은 문제가 발생한다. 휴대폰으로 사진을 찍어도 내부 파라미터를 알 수 있지만 초점을 맞추기 위해 자동 초점을 사용하게 된다. 이는 부정확한 카메라 내부 파라미터를 얻게 하며 추가로 카메라 렌즈에 의한 왜곡 또한 보정해야 한다.

카메라 캘리브레이션에 대한 많은 연구들이 진행되었지만 대부분 교정의 기준이 되는 체커보드를 이용하여 카메라 내부 파라미터를 구한다[2][3][4][5]. 이 경우 상당히 정확한 카메라 내부 파라미터를 추정할 수 있지만 3D 복원을 하기 위해 사용되는 사진들에 일일이 보정의 기준이 되는 체커보드를 집어넣을 수는 없다. 따라서 본 논문에서는 체커보드가 필요 없는 단일 이미지만을 이용하여 카메라 내부 파라미터를 추정하는 방법이 필요하고 이를 딥러닝을 이용하여 해결한다[6][7][8][9][10].

지금까지 전이학습을 이용하여 카메라 내부 파라미터를 추정하는 딥러닝 연구들이 진행되어왔다. [6][11][12][13]에서는 이미지 분류 문제를 해결하기 위해 사용한 GoogLeNet[11], DenseNet[12][13]과 같은 모델을 이용하여 초점 거리, 왜곡 계수 등 카메라 내부 파라미터를 추정하는 전이 학습을 진행한다. [6]에서는 AlexNet 모델을 이용하여 초점거리를 추정하고, [11]에서는 GoogLeNet(Inception v3) 모델을 이용하여 초점 거리를 왜곡 계수를 추정한다. [12]에서는 DenseNet-161 모델을 이용하였고 초점 거리 대신 시야각(Field of View) 추정하며 변환식을 통해 초점 거리를 구한다.

기존의 딥러닝 모델을 이용한 방법들은 초점 거리 및 왜곡 계수를 추정하는 회귀 문제를 분류 문제로 변형하여 목표 값들을 추정하였다. 회귀 문제의 목표는 실제 값과 예측 값의 차이를 줄이는 것이고 분류 문제의 목표는 실제 값과 예측 값이 같아지는 것이다. 이렇게 할 경우 정답을 예측하도록 학습을 수행하지만, 회귀 문제의 목표인 예측 값과 실제 값의 오차를 줄이도록 학습하지 못하는 한계가 있다. 본 논문에서는 분류 문제의 목표인 정답을 예측하도록 학습하면서, 회귀 문제의 목표인 예측 값과 실제 값의 오차를 줄이도록 학습하는 손실 함수를 제안한다.

본 논문의 구성은 다음과 같다. Ⅱ장에서는 기존의 딥러닝을 이용한 카메라 캘리브레이션 연구를 소개하고 기존 연구의 문제점을 파악한다. Ⅲ장에서는 기존 연구에서 발생한 문제점을 보안하는 손실 함수를 제안한다. Ⅳ장에서는 데이터 생성 과정과 실험 환경을 설명하고, 기존의 방법들을 이용하여 학습한 결과와 제안한 손실 함수를 사용하여 학습한 결과를 비교 및 분석한다. Ⅴ장에서는 결론을 내린다.

Ⅱ. 기존 연구

1. 기존 연구

[6]에서는 최초로 딥러닝을 활용하여 카메라 파라미터를 추정하였다. 이때 사용된 딥러닝 모델은 AlexNet으로, 전이학습 이용하여 초점 거리를 추정하였다. 전이학습이란 이미지넷에서 이미지를 분류하기 위해 개발된 모델을 다른 분야의 문제를 해결할 때 사용하는 방법을 말한다. 전이학습을 이용하여 카메라 파라미터를 추정하는 방법은 아래와 같다.

[11]에서는 Inception v3를 전이학습을 이용하여 파라미

a) 세종대학교(Sejong University)
‡ Corresponding Author : 한종기(Jongki Han)
E-mail: hjk@sejong.edu
Tel: +82-2-3408-3739
ORCID: https://orcid.org/0000-0002-5036-7199
※This work was supported by the National Research Foundation of Korea (NRF) under Grant 2022R1F1A1071513 funded by the Korea government through the Ministry of Science and ICT (MSIT).
· Manuscript May 19, 2023; Revised August 22, 2023; Accepted August 22, 2023.

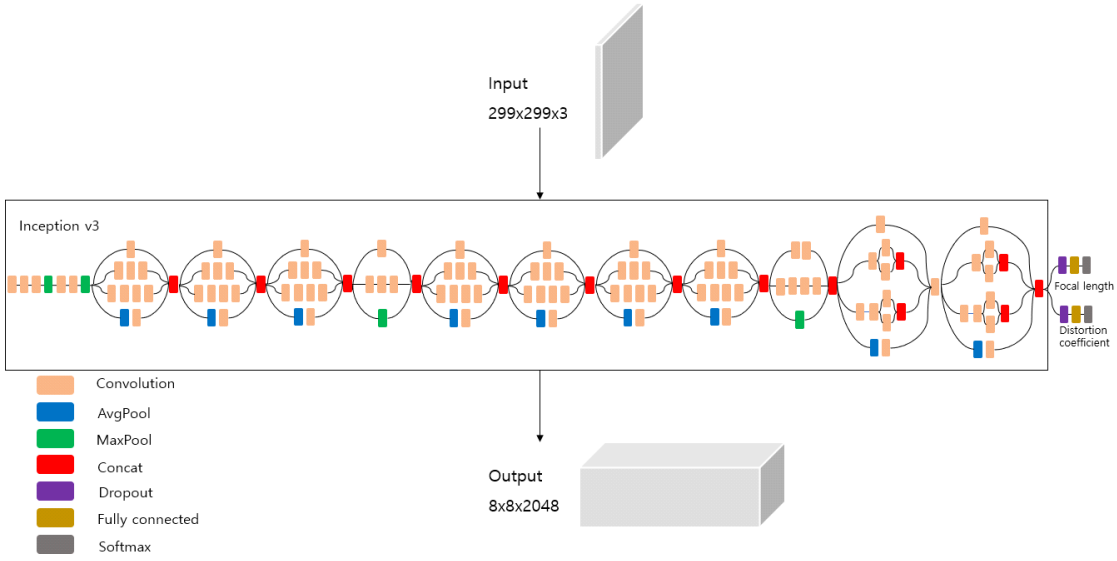


그림 1. 인셉션 모델 구조[14]
Fig. 1. Architecture of the Inception v3 model

터를 추정한다. 전이학습 모델의 초기 파라미터는 이미지넷 데이터셋을 이용해 미리 학습된 파라미터를 사용한다. Inception v3는 그림 1의 구조를 가지고 있고 출력으로 8x8x2048 크기를 가지는 특징맵이 나온다. 이 특징맵을 분류 또는 회귀 알고리즘에 적용하여 초점거리와 왜곡 계수를 예측한다. 그림 1의 특징맵을 출력층과 fully connected layer로 연결하여 학습한다.

예측하고자 하는 두 가지 파라미터(초점거리, 왜곡계수)들은 연속적인 값을 갖기 때문에 회귀 문제로 해결해야 한다. 이때 출력층 노드의 개수는 예측하고자 하는 파라미터의 개수만큼 설정한다. 회귀 방법에서는 초점거리와 왜곡계수를 예측하기 때문에, 출력 노드의 개수는 2개이다. 출력층의 노드 각각은 하나의 파라미터를 예측하며, 손실 함수로 MSE, MAE, Huber Loss 등을 이용하여 학습한다.

[11]에서는 초점 거리 및 왜곡 계수를 정확히 추정하는 딥러닝 모델을 찾기 위해, 여러 가지 모델을 생성하고 비교하였다. 초점 거리와 왜곡 계수를 하나의 특징맵에서 모두 추정하는 모델, 초점 거리와 왜곡 계수를 독립된 두 개의 특징맵을 이용하여 추정하는 모델, 초점 거리를 먼저 구하고 이를 추가 입력으로 하여 왜곡 계수를 추정하는 모델을 비교하였다. 이중 하나의 특징맵에서 초점 거리 및 왜곡 계수를 추정하는 딥러닝 모델이 가장 높은 성능을 보였다. 추가로 회귀 방법과 분류 방법 각각을 적용하여 실험한 결과, 분류 방법을 이용했을 때 더 높은 성능을 보이는 것을 확인했다.

회귀 문제를 분류 문제로 변형하기 위해서는 클래스를 정의할 필요가 있다. 예측하고자 하는 파라미터의 값을 클래스로 표현할 수 있다면 분류 문제로 해결하는 것이 가능하다. 이를 위해 파라미터의 범위가 한정되어 있어야한다. 그림 2와 같이 일정한 간격을 가지는 파라미터의 값들을 클래스로 만들고 분류 문제로 해결한다. 이때 출력층의 개수는 예측하고자 하는 파라미터의 개수와 같고, 출력층 각 노드의 개수는 클래스의 개수와 같다. 손실 함수로는 교차 엔트로피를 사용한다.

[11]에서는 그림 2와 같이 초점 거리는 50~500의 범위 내에 존재하도록 설정했는데, 이는 다양한 카메라 모델을 수용할 수 있기 때문이다. 초점 거리는 10 단위로 샘플링하여 총 46개의 클래스를 만들었다. 왜곡 계수 또한 다양한 카메라 모델을 수용하는 범위인 0~1.2으로 설정했다. 왜곡 계수는 0.02 단위로 샘플링하여 총 61개의 클래스를 만들었다.

Inception v3의 출력은 그림 1과 같이 8x8x2048의 크기를 가지는 특징맵이다. 이 특징맵을 flatten 한 히든 레이어

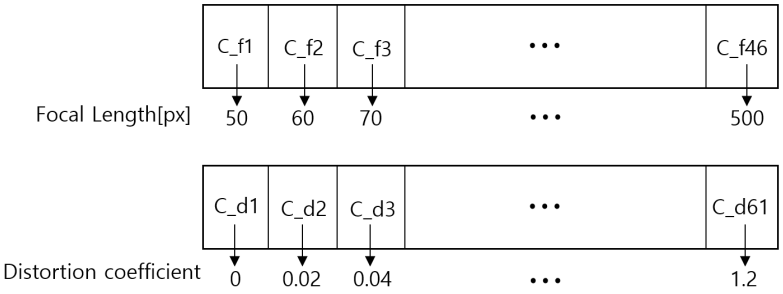


그림 2. 초점 거리 및 왜곡 계수 클래스 할당
Fig. 2. Assignment of focal length and distortion coefficient classes

와 출력층을 fully connected layer로 연결하여 학습을 진행한다. 초점 거리와 왜곡 계수를 구해야하므로 두 개의 fully connected layer가 존재하며 초점 거리를 예측하는 출력층 노드의 개수는 46개이며 왜곡 계수의 경우 61개이다. 활성화 함수로는 소프트맥스를 사용하는데, 활성화 함수를 거치고 나면 각 클래스를 예측할 확률을 나타내게 된다. N개의 클래스를 가지는 출력층의 각 노드가 나타내는 값을 L_i라고 할 때($i=1,2,...,N$), 소프트맥스 활성화 함수를 통과한 값 S_i는 다음과 같다.

$$S_i = \frac{e^{L_i}}{\sum_{i=1}^{N} e^{L_i}} \quad (1)$$

S_i는 모두 더했을 때 값이 1이 되기 때문에 확률 값으로 사용된다. 소프트 맥스 활성화 함수를 거쳐 확률로 표시된 값들을 손실 함수로 교차 엔트로피를 사용하여 실험한다. 교차 엔트로피는 실제 분포를 알지 못하는 상태에서, 모델링을 통해 예측 확률 분포를 구하여 실제 확률 분포를 예측하는 것이다. 딥러닝이 최종으로 도달하고자 하는 실제 확률 분포를 C_i라고 하자($i=1,2,...,N$). 예측 확률 분포는 소프트맥스 활성화 함수를 거친 확률 값인 S_i이다. 식으로는 다음과 같다.

$$H_S(C) = -\sum_{i=1}^{N} C_i \log(S_i) \quad (2)$$

[12]에서는 기존의 이미지 분류 모델보다 파라미터 개수를 줄여 학습시간을 단축시킨 DenseNet-161을 전이학습을 이용하여 문제를 해결한다. 모델의 초기 파라미터 역시 이미지넷에서 미리 학습된 파라미터를 사용한다. [12]에서는 딥러닝을 이용하여 추정하기 어려운 파라미터들이 존재하며, 그 파라미터와 변환 및 역변환이 가능하면서 더욱 정확하게 추정할 수 있는 파라미터들을 제안하였다. 그중 초점 거리의 경우 딥러닝을 통해 단일 이미지에서 직접 구하기 어려운 파라미터이며, 초점 거리와 변환이 가능하며 단일 이미지에서 비교적 추정하기 쉬운 시야각(Field of View) 파라미터를 대신 구하여 초점 거리를 얻는 방법을 제안하였다. 시야각과 초점 거리는 식 (3)과 같은 관계를 가지기 때문에 시야각을 추정하면 초점 거리를 얻을 수 있다.

$$FOV = 2^{*}\tan^{-1}\left(\frac{height}{2f}\right) \quad (3)$$

2. 기존 연구의 문제점

[11]에서는 초점 거리와 왜곡 계수를 추정하는 회귀 문제를 초점 거리와 왜곡 계수의 범위를 설정하여 샘플링하고 클래스로 할당하여 분류 문제로 변환하였다. 분류 문제이기 때문에 손실 함수로 교차 엔트로피를 사용한다. 교차 엔트로피는 두 변수들의 확률들의 차이를 의미하는데, 확률들 중 하나는 최종적으로 딥러닝 모델이 추정해야하는 확률인 C_i이다. 정답 클래스일 경우 1을, 정답 클래스가 아닐 경우 0을 부여한다.

$$C_i = \begin{cases} 1, & i=k \\ 0, & otherwise \end{cases}, k = \text{ 답클래스} \quad (4)$$

다른 확률 하나는 딥러닝 모델이 예측하는 확률인 S_i 이다. 식 (2)와 같이 학습은 두 확률의 차이가 줄어드는 방향으로 진행된다. 즉 딥러닝 학습 목표는 정답 클래스를 예측할 때, S_i가 1이 되도록 학습하고, 다른 클래스를 예측할 때는, S_i가 0이 되도록 학습하는 것이다. 이러한 분류 기반 학습 방법에서는 예측 값의 클래스만을 정확하게 예측하도록 학습된다. 이 과정에서는 예측 값과 실제 값의 차이를 줄이는 과정이 생략된 문제점이 존재한다. 즉, 분류 기반 예측 방법에서는 정답을 예측하지 못했을 때 실제 값과 예측 값이 얼마나 차이 나는지는 알 수 없다. 정답을 예측할 경우 오차가 발생하지 않지만 정답을 예측하지 못했을 때 오차가 발생한다. 하지만 초점 거리와 왜곡 계수를 구하는 것은 회귀 문제이기 때문에 딥러닝 모델의 최종 목표는 실제 값과 예측 값의 차이를 줄이는 것이다. 정답을 예측하지 못한 경우 오차가 작아지도록 학습할 필요가 있다. 이를 해결하기 위한 방법을 다음 장에서 설명한다.

Ⅲ. 제안하는 방법

[12][13]에서 초점 거리 대신 시야각을 예측하는 학습을 진행하였을 때 더 높은 성능을 보여주는 것이 입증되었다. 따라서, 본 연구에서는 그림 3과 같이 초점 거리 대신 시야각을 이용하여 클래스를 만든다. 이때 식 (3)의 변환식을 통해 초점 거리와 동일한 범위를 가지게 설정했다. 시야각은 33~145.5의 범위를 가지며 2.5 단위로 샘플링하여 총 46개의 클래스를 만들어 학습을 진행하고, 왜곡 계수는 동일하게 0~1.2의 범위를 가지며 0.02 단위로 샘플링하여 총 61개의 클래스를 만들어 학습을 진행한다.

실제 값과 예측 값의 오차를 줄이기 위한 방법은 분류 문제의 목표인 정답 클래스를 맞추도록 학습하면서 회귀 문제의 목표인 오차까지 줄이는 것이다. 기존에는 식 (4)와 같이 정답 클래스는 S_i가 1과 가까워지도록 나머지 클래스들은 S_i가 0과 가까워지도록 학습하였지만, 본 논문에서는 이를 개선하여 정답을 예측하지 못하였을 때 오차를 줄이는 방법을 제안한다.

먼저 Fully Connected layer로 클래스를 분류를 하는 것의 물리적 의미를 이해할 필요가 있다. 그림 4는 그림 1에서 Inception v3의 출력인 히든 레이어와 출력 레이어를 Fully Connected layer로 연결시킨 것을 나타낸다. 히든 레

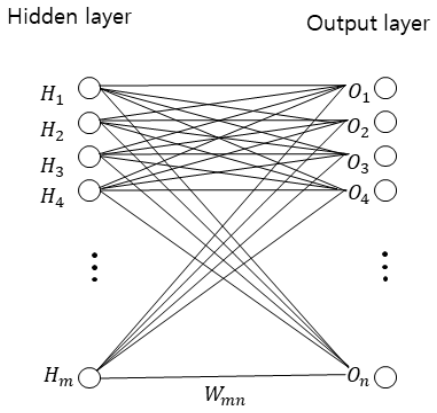


그림 4. 히든 레이어와 출력 레이어 사이의 가중치
Fig. 4. Weights between hidden layer and output layer

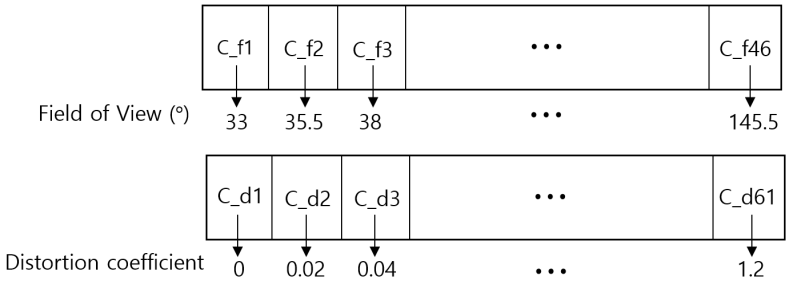


그림 3. 시야각 및 왜곡 계수 클래스 할당
Fig. 3. Assignment of field of veiw and distortion coefficient classes

이어의 노드 값을 $H_1, H_2, \cdots H_m$, 출력 레이어의 노드 값을 $L_1, L_2, \cdots L_n$ 이라고 하고, 가중치를 $W_{11}, W_{12}, \cdots W_{mn}$ 이라 하자.

Fully Connected layer로 분류를 하는 것은 H_m과 W_{mn}의 내적을 의미한다. 식으로 정리해보면 다음과 같다.

$$L_n = \sum_{i=1}^{m} H_i W_{in} \tag{5}$$

H_m과 W_{mn}의 내적 값이 n번째 클래스의 값, 즉 확률을 나타낸다. 이를 물리적으로 해석해보면 H_m과 W_{mn}이 얼마나 닮았는지를 의미한다.

보통 분류 문제에서는, 예를 들어 개와 고양이를 구별하거나 숫자 혹은 글자를 구별하는 등 클래스들은 서로 다르다. 각 클래스들이 서로 다르기 때문에 식 (4)와 같이 정답 클래스는 S_i가 1이 되도록 나머지 클래스들은 S_i가 0이 되도록 학습시킨다. 하지만 초점 거리와 왜곡 계수를 구하는 회귀 문제를 분류 문제로 변형해 해결하는 방법을 선택했기 때문에, 그림 3에서와 같이, 각 클래스들은 해당 목표값(예를 들면, 33, 35.5 또는 0, 0.02 등등)을 가지고 있으며, 경우에 따라 클래스들의 목표값들 간의 차이가 큰 경우와 작은 경우가 혼재되어 있다. 따라서, 실제 값과 예측 값의 오차를 줄이기 위해서는 실제 예측값과 목표값들 간의 오차가 큰 클래스보다 오차가 작은 클래스를 선택하도록 학습시켜야 한다.

그림 3에서 시야각은 33, 145.5까지 2.5단위로 총 46개의 클래스가 구성되어 있다. 기존의 분류 기반 알고리즘에서는 이웃된 클래스들이라도 서로 상관없는 클래스로 고려하여 학습하였으나, 본 연구에서는 이웃 클래스들간의 유사성을 고려하여 학습한다. 예를 들면, 기존의 분류 기반 알고리즘에서는, 교차 엔트로피 손실 함수를 이용할 때, 시야각 목표값이 63인 클래스, 145.5인 클래스, 60.5인 클래스들은 모두 서로간에 상관없고 다른 클래스로 고려되어 학습된다. 이에 반해, 본 연구에서는 목표값이 63인 클래스는 목표값이 60.5인 클래스와 유사한 클래스이고, 목표값이 145.5인 클래스와는 유사하지 않은 클래스임을 고려하여 학습한다. 즉, 회귀 문제를 분류 문제로 변형하여 해결할 때 실제 값과 예측 값의 오차가 작아지도록 학습하기 위해 식 (4)와 같이 정답 클래스에만 값을 주어 학습시키는 것이 아니라, 식 (6)과 같이 정답 클래스와 닮은 인접한 클래스에도 값을 부여하는 손실 함수를 제안한다. 첫 번째와 마지막 클래스의 경우 인접 클래스가 하나밖에 존재하지 않기 때문에 식을 조정한다.

제안한 방법으로 학습할 경우, 정답 클래스와 닮은 인접 클래스들을 예측하기 때문에 정답을 예측하지 못해도 인접한 클래스를 예측하여 실제 값과 예측한 값의 오차가 줄어들게 된다.

Ⅳ. 실험 결과 및 분석

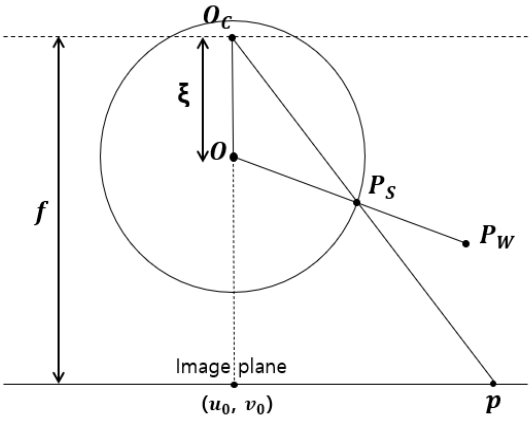
1. 데이터셋 생성

딥러닝 네트워크 모델의 학습을 위해 초점 거리와 왜곡 계수를 알고 있는 많은 개수의 훈련 데이터들이 필요하다. 하지만 이런 정답을 알고 있는 데이터들은 실제로 얻기 어렵다. 이를 위해서, 인터넷 상에서 얻을 수 있는 파노라마 이미지로부터 2D 훈련 영상들을 추출하여 사용한다. 아래 내용은 파노라마 영상으로부터 2D 영상을 추출하는 과정을 설명한다.

그림 5의 (a)는 [15]에서 사용한 Unified Spherical Model을 나타내며 단일 왜곡 계수를 반영한 카메라 투영 모델이다. 이 모델을 사용할 경우 투영, 역투영이 자유롭기 때문에

$$C_i = \begin{cases} \begin{cases} 0.9 & ,i=k \\ 0.1 & ,i=k+1, k=1 \\ 0.1 & ,i=k-1, k=N, \\ 0 & ,otherwise \end{cases} & k=1, k=N \\ \qquad\qquad ,k= \text{ 답클래스} \\ \begin{cases} 0.8 & ,i=k \\ 0.1 & ,i=k\pm 1 \\ 0 & ,otherwise \end{cases} & otherwise \end{cases} \tag{6}$$

파노라마 영상을 이용하여 2D 영상을 생성할 수 있다. 또한 왜곡 계수의 범위가 넓기 때문에 휴대폰뿐만 아니라 다양한 종류의 카메라에서 찍은 이미지도 접근이 가능한 장점이 있다.



(a) Unified Spherical Model

Notation

3D world point P_w = (X, Y, Z)

Spherical point P_s = (X_s , Y_s , Z_s) = P_w / || P_w ||

2D image plane point = (x, y)

Sphere center O = (0, 0, 0)

Projection center O_c = (0, 0, ξ)

Principal point (u_0, v_0)

Distortion coefficient ξ

Focal length f

(b) notation

그림 5. 통합 구형 모델과 기호
Fig. 5. Unified Spherical Model and the notation

구에서 2D 이미지로 투영하기 위한 과정은 아래와 같다. 3차원 점 P_W를 단위 구로 위의 점 P_S로 정규화 시킨다. 그 후 투영 중심 O_C를 기준으로 단위 구 위의 점 P_S(X_S, Y_S, Z_S + ξ)를 Image plane의 점 p(x, y)으로 투영시킨다. p(x,y)는 식 (7)로 표현할 수 있다.

$$p=(x,y)=\left(\frac{Xf}{\xi\sqrt{X^2+Y^2+Z^2}+Z}+u_0,\frac{Yf}{\xi\sqrt{X^2+Y^2+Z^2}+Z}+v_0\right) \tag{7}$$

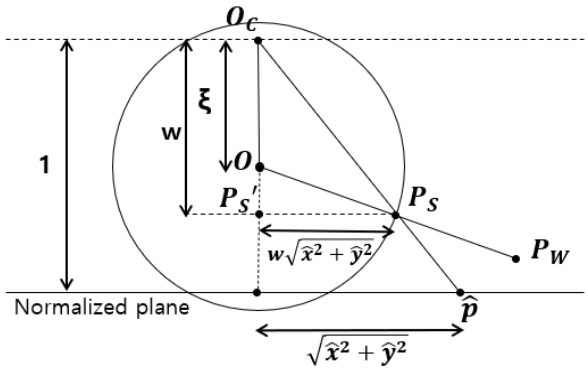


그림. 6. 정규화된 평면으로 표현된 통합 구형 모델
Fig. 6. Unified Spherical Model with normalized plane

2D 영상을 구로 역투영하기 위한 과정은 다음과 같다. 그림 6은 그림 5의 Image plane 위의 점 p(x, y)를 Normalized plane 위의 점 $\hat{p}=(\hat{x},\hat{y})$로 정규화 시킨 것을 나타낸다. 그림 6에서 투영 중심 O_C와 단위 구 위의 점 P_S의 z축 좌표 차이를 w라 할 때 역투영된 점 P_S는 다음과 같다.

$$P_S=(w\hat{x},w\hat{y},w-\xi) \tag{8}$$

Unified Spherical Model은 단위 구이므로 반지름이 1이다. 직각삼각형 OP_SP_S'의 세 변의 길이를 식으로 표현할 수 있기 때문에 피타고라스 정리를 이용하면 다음과 같이 w를 구할 수 있다.

$$w=\frac{\xi+\sqrt{1+(1-\xi^2)(\hat{x}^2+\hat{y}^2)}}{\hat{x}^2+\hat{y}^2+1} \tag{9}$$

따라서 파노라마 영상을 구로 역투영시키고, 초점 거리와 왜곡 계수를 아는 2D 영상으로 투영시키는 과정을 통해 이미지를 생성한다.

파노라마 이미지는 SUN360[16]과 3D60[17] 데이터셋에서 총 3만개를 얻었고 초점 거리와 왜곡 계수를 알고 있는 2D 이미지 20만개를 생성하였다.

2. 실험 조건

총 데이터는 20만개이며 훈련 데이터는 16만개, 검증 데이터와 시험 데이터는 각각 2만개로 설정하였다. 네트워크 구조는 그림 1과 같이 Inception v3의 출력인 특징맵과 출력층을 fully connected layer로 연결하였다. 초점 거리 대신 시야각을 추정하며, 시야각은 46개의 클래스를 가지고 왜곡 계수는 61개의 클래스를 가진다. 제안한 손실 함수의 성능을 비교하기 위해 기존의 방법들을 이용하여 실험한 결과와 제안하는 손실 함수를 이용하여 실험한 결과를 비교한다. 추가로 분류 알고리즘을 이용하여 실험한 결과와 회귀 알고리즘을 이용하여 실험한 결과를 비교한다.

미니 배치는 64, 아담 옵티마이저를 사용한다. 학습율은 10^{-4}에서 에포크를 두 번 돌 때마다 1/10로 줄이는 방법을 사용하며 과적합을 방지하기 위해 조기 멈춤 알고리즘을 사용한다.

3. 실험 결과 및 분석

표 1. 시야각 및 왜곡 계수 학습 결과(정답률)
Table 1. Field of view and distortion coefficient learning outcome (accuracy)

Accuracy	Field of view	Distortion coefficient
DeepCalib[11]	57.56%	34.02%
Proposed method	58.48%	35.12%

표 1은 시야각(초점 거리)과 왜곡 계수의 정답률을 나타낸다. 실험 결과를 보면 제안한 손실 함수를 이용했을 때 정답률이 상승하였지만, 시야각의 정답률은 50%, 왜곡 계수의 정답률이 30%대로 낮은 것을 볼 수 있다. 시야각과 왜곡 계수를 예측하는 것이 복잡한 일이기 때문에 높은 정답률을 보이기 힘들 것이라 예상하였다. 이를 보안하기 위해 실제 값과 예측 값의 오차가 줄어들도록 정답 클래스뿐만 아니라 정답 클래스와 인접한 클래스까지도 예측하도록 학습하는 새로운 손실 함수를 제안하였다. 또한 왜곡된 사진을 복원했을 때 정답 클래스와 인접한 클래스를 예측한 파라미터를 이용하여 복원한 것과 실제 파라미터 값을 이용하여 복원한 결과 거의 차이가 없었다. 따라서 오차가 가장 작은 정답 클래스와 인접한 클래스를 예측한 것도 정답으로 간주하였을 때의 결과를 확인한다.

표 2. 정답 클래스와 인접한 클래스를 예측한 것까지 정답으로 간주하였을 때 시야각 및 왜곡 계수의 정답률
Table 2. Accuracy rate of field of view and distortion coefficient when considering both the correct class and adjacent classes predicted as correct.

Accuracy	Field of view	Distortion coefficient
DeepCalib[11]	95.59%	67.01%
Proposed method	95.91%	68.48%

정답 클래스와 인접한 클래스를 예측한 것까지 모두 정답으로 간주하면 표 2와 같이 높은 정답률을 보여준다. 시야각의 경우 기존 손실 함수로 학습할 경우 정답률이 95.59%이고 제안한 손실 함수로 학습했을 때 95.91%까지 상승하였다. 왜곡 계수의 경우 기존 손실 함수로 학습할 경우 정답률이 67.01%이고 제안한 손실 함수로 학습했을 때 68.48%까지 상승하였다.

왜곡 계수의 정답률이 시야각의 정답률 보다 낮게 나왔는데, 그 이유를 분석해보니 클래스의 수가 61개로 시야각의 클래스 수인 46개보다 많기 때문이다. 만약 왜곡 계수의 클래스 수를 시야각과 비슷하게 한다면 왜곡 계수 역시 더 높은 정답률을 보일 것으로 예상한다.

회귀 알고리즘을 사용할 경우 정답률을 측정하기 어렵기 때문에, 실제 값과 예측 값의 오차를 이용하여 성능을 비교한다. 시야각의 경우, 초점 거리를 구하는 것이 목표이기 때문에, 식 (3)을 이용하여 시야각을 초점 거리로 변환하여 오차를 구한다. 회귀 알고리즘을 사용한 결과로는 [11]논문에서 손실 함수를 MSE로 설정하여 실험을 하고, [12]논문에서 제안하는 방법으로 초점 거리와 단일 왜곡 계수만을 추정하는 모델을 만들어 추가 실험을 한다.

표 3은 초점 거리와 왜곡 계수의 실제 값과 예측 값의 평균 오차를 구한 것이다. 표3을 보면 회귀 알고리즘을 사용했을 때 평균 오차가 더 큰 것을 알 수 있다. 추가로 제안하는 방법으로 학습한 경우 평균 오차가 작은 것을 확인할 수 있다. 그 이유는 표 2에서 알 수 있듯이 교차 엔트로피를 사용하여 학습했을 때 분류 방법을 사용하여 학습을 진행

표 3. 초점 거리 및 왜곡 계수 평균 오차
Table 3. Mean error of focal length and distortion coefficient

	Mean error	Focal length	Distortion coefficient
Classification	DeepCalib[11]	7.22	0.0309
	Proposed method	6.95	0.0298
Regression	DeepCalib-MSE[11]	20.46	0.0689
	DeepSingle[12]	32.80	0.1157

했지만, 정답 클래스와 가까이 있는 클래스를 예측하도록 학습되었기 때문이다. 이는 회귀 알고리즘을 사용했을 때의 학습 목표와 비슷한데, 제안하는 손실 함수의 학습 목표와 방향이 같고, 실험 결과에서 알 수 있듯이 성능이 향상하는 것을 볼 수 있다.

V. 결 론

기존의 딥러닝을 활용한 카메라 캘리브레이션은 미리 학습된 Inception v3 및 DenseNet의 출력인 특징 맵을 이용하여 회귀 문제를 분류 문제로 변형하여 해결하였다. 초점 거리, 왜곡 계수가 연속적인 값을 가지기 때문에 보통 회귀 문제로 해결한다. 하지만 회귀 문제로 해결하는 것보다 분류 문제로 해결하는 것이 더 높은 성능을 보이며, 파라미터의 범위가 한정되어 있다면, 일정한 간격으로 샘플링하여 클래스를 만들어 분류 문제로 해결할 수 있다. 회귀 문제를 분류 문제로 변형하여 해결할 경우, 인접한 클래스끼리는 서로 무관하지 않고 닮은 부분이 존재하며, 값의 차이가 작은 특징을 가진다. 따라서 기존의 교차 엔트로피 손실 함수에서 정답 클래스에만 가중치를 부여하는 방법 대신 정답 클래스와 인접한 클래스에도 가중치를 부여하는 새로운 손실 함수를 제안한다. 제안한 손실 함수로 학습할 경우 정답률이 상승하였고 실제 값과 예측 값의 오차가 더 줄어드는 결과를 얻었다.

참 고 문 헌 (References)

[1] R. I. Hartley and A. Zisserman. Multiple View Geometry in Computer Vision. Cambridge University Press, 2003.

[2] Simone Gasparini, Peter Sturm, and João P. Barreto. 2009, “Plane-Based Calibration of Central Catadioptric Cameras,” proceedings of International Conference on Computer Vision, Kyoto, Japan, pp. 1195-1202, 2009.
doi: https://doi.org/10.1109/ICCV.2009.5459336

[3] Christopher Mei and Patrick Rives, “Single View Point Omnidirectional Camera Calibration from Planar Grids,” proceedings of International Conference on Robotics and Automation, Rome, Italy, pp. 3945-3950, 2007.
doi: https://doi.org/10.1109/ROBOT.2007.364084

[4] Davide Scaramuzza, Agostino Martinelli, and Roland Siegwart, “A Toolbox for Easily Calibrating Omnidirectional Cameras,” proceedings of International Conference on Intelligent Robots and Systems, Beijing, China, pp. 5695-5701, 2006.
doi: https://doi.org/10.1109/IROS.2006.282372

[5] B. Chen, Y. Liu and C. Xiong, “Automatic Checkerboard Detection for Robust Camera Calibration,” proceedings of IInternational Conference on Multimedia and Expo, Shenzhen, China, pp. 1-6, 2021.
doi: https://doi.org/10.1109/ICME51207.2021.9428389

[6] Workman, Scott, et al, "Deepfocal: A method for direct focal length estimation," proceedings of International Conference on Image Processing (ICIP), pp. 1369-1373, 2015.
doi: https://doi.org/10.1109/ICIP.2015.7351024

[7] J. Rong, S. Huang, Z. Shang and X. Ying, “Radial lens distortion correction using convolutional neural networks trained with synthesized images,” proceedings of Asian Conference on Computer Vision, pp. 35-49, 2016.
doi: https://doi.org/10.1007/978-3-319-54187-7_3

[8] X. Yin, X. Wang, J. Yu, M. Zhang, P. Fua, and D. Tao, “FishEyeRecNet: A multi-context collaborative deep network for fisheye image rectification,” proceedings of European Conference on Computer Vision, pp. 469-484, 2018.

[9] T. H. Butt and M. Taj, "Camera Calibration Through Camera Projection Loss," proceedings of International Conference on Acoustics, Speech and Signal Processing, Singapore, Singapore, pp. 2649-2653, 2022.
doi: https://doi.org/10.1109/ICASSP43922.2022.9746819

[10] Y. Hold-Geoffroy et al, “A Perceptual Measure for Deep Single Image Camera Calibration,” proceedings of Conference on Computer Vision and Pattern Recognition, pp. 2354-2363, 2018.

[11] O. Bogdan, V. Eckstein, F. Rameau, and J.-C. Bazin, “DeepCalib: A deep learning approach for automatic intrinsic calibration of wide fieldof-view cameras,” proceedings of European Conference on Visual Media Production, pp. 1 – 10, 2018.
doi: https://doi.org/10.1145/3278471.3278479
[12] M. Lopez, R. Mari, P. Gargallo, Y. Kuang, J. Gonzalez-Jimenez, and G. Haro, “Deep single image camera calibration with radial distortion,” proceedings of Conference on Computer Vision and Pattern Recognition, pp. 11817-11825, 2019.
doi: https://doi.org/10.1109/CVPR.2019.01209
[13] Nobuhiko Wakai, Takayoshi Yamashita, “Deep Single Fisheye Image Camera Calibration for Over 180-degree Projection of Field of View,” proceedings of International Conference on Computer Vision, pp. 1174-1183, 2021.
doi: https://doi.org/10.1109/ICCVW54120.2021.00137
[14] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, Zbigniew Wojna, “Rethinking the Inception Architecture for Computer Vision,” proceedings of Computer Vision and Pattern Recognition, pp. 2818-2826, 2016.
doi: https://doi.org/10.1109/CVPR.2016.308
[15] João P. Barreto, “A Unifying Geometric Representation for Central Projection Systems,” Computer Vision and Image Understanding, Vol.103, pp. 208-217, 2006.
doi: https://doi.org/10.1016/j.cviu.2006.06.003
[16] SUN360 panorama dataset, https://3dvision.princeton.edu/projects/2012/SUN360/
[17] 3D60 panorama dataset, https://vcl3d.github.io/3D60/

저 자 소 개

원 종 수

- 2018년 ~ 현재 : 세종대학교 전자정보통신공학과 학사과정
- ORCID : https://orcid.org/0009-0005-9398-9085
- 주관심분야 : 영상 신호처리, VR

한 종 기

- 1992년 : KAIST 전기및전자공학과 공학사
- 1994년 : KAIST 전기및전자공학과 공학석사
- 1999년 : KAIST 전기및전자공학과 공학박사
- 1999년 3월 ~ 2001년 8월 : 삼성전자 DM연구소 책임연구원
- 2001년 9월 ~ 현재 : 세종대학교 전자정보통신공학과 교수
- 2008년 9월 ~ 2009년 8월 : University California San Diego (UCSD) Visiting Scholar
- ORCID : https://orcid.org/0000-0002-5036-7199
- 주관심분야 : 비디오 코덱, 영상 신호처리, 정보 압축, 방송 시스템