

카메라 캘리브레이션의 원리와 기술 동향

최준혁, 최성록(서울과학기술대학교 컴퓨터공학과)

I. 서론

카메라는 3차원 공간을 2차원 평면에 투영(projection)하여 이미지를 생성한다. 카메라의 3차원-2차원 투영 과정, 즉 카메라 투영 모델은 카메라의 초점거리나 렌즈의 왜곡계수와 같은 카메라 고유의 속성을 통해 표현된다. 카메라 캘리브레이션(camera calibration)은 주어진 카메라 투영 모델과 사용되는 이러한 속성값을 찾는 과정이다.

카메라 캘리브레이션은 3차원 공간과 2차원 이미지 사이의 기하학적 관계를 설명하는 카메라 투영 모델을 완성하기 때문에, 카메라 영상을 이용한 3차원 또는 기하학적 해석이 필요한 작업에 반드시 선행되는 작업이다. 이동로봇의 측위(localization)에 사용되는 visual odometry나 visual SLAM에 앞서 카메라의 캘리브레이션이 선행되어야 하고, 로봇팔의 조작(manipulation)이나 잡기(grasping) 작업에 사용되는 visual servoing에도 마찬가지로 카메라 캘리브레이션이 미리 진행되어야 한다. 또한 VR/AR 단말기의 카메라 자세추정과 다양한 계측과 자동화를 위한 머신비전(machine vision) 기술에도 카메라 캘리브레이션은 꼭 필요한 사전 작업이다.

본 기고에서는 카메라 캘리브레이션의 원리와 최근 기술 동향에 대해 설명한다. 카메라 캘리브레이션의 원리에 앞서 다양한 응용들에 사용되는 카메라 투영 모델들을 간략히 소개한다. 그리고 다양한 카메라 캘리브레이션 기법들을 통찰할 수 있는 전체적인 알고리즘 구성과 그 특징을 소개한다. 또 OpenCV에 포함된 현재 가장 널리 사용되는 Zhang's method에 대해 알아보고, Zhang's method를 통해 카메라 캘리브레이션을 성공적으로 마치기 위한 가이드라인을 제시한다. 카메라 캘리브레이션은 그 필요성과 중요성 때문에 관련 기술들이 여전히 많이 연구되

고 있다. 본 기고에서는 이러한 최신 기술들을 각각의 목적에 따라 중요한 요점들을 리뷰하고자 한다. 본 기고를 통해 독자들이 카메라 캘리브레이션의 원리와 기술 흐름을 이해하고, 실제 카메라 캘리브레이션을 사용할 때 큰 도움이 될 수 있는 통찰과 가이드라인을 얻길 기대한다.

II. 카메라 캘리브레이션의 원리

1. 카메라 투영 모델

카메라 투영 모델(camera projection model)은 3차원 공간 상의 점(point)을 2차원 카메라 이미지로 투영시키는 과정을 나타내는 수학적 모델이다. 여러가지 상황에 적합한 다양한 카메라 투영 모델들이 존재하고, 일반적으로 [그림 1]과 같이 매우 작은 구멍을 통해 입사되는 빛을 평면에 투영하는 핀홀 카메라 모델(pinhole camera model)을 사용한다. 핀홀 카메라에서의 작은 구멍은 투영된 모든 빛이 지나가는 초점(focal point)이다. 수학적으로 핀홀 카메라를 표현할 때는 [그림 1]과 같이 이미지 평면이 초점 앞에 있어 맺힌 상이 뒤집히지 않는 형태로 표현한다. 핀홀 카메라 모델은 렌즈 왜곡이 없는 이상적인 상황에서 수학적으로는

$$\mathbf{x} = \mathbf{K}(\mathbf{R}\mathbf{X} + \mathbf{t}) = \mathbf{K}\hat{\mathbf{x}}$$

와 같은 수식으로 표현된다. \mathbf{X} 는 월드좌표계(world coordinate)에서 표현된 3차원 점 $[X, Y, Z]^T$ 이고, \mathbf{x} 는 첫번째 픽셀을 영점으로 하는 평면(image plane)에 투영된 2차원 점이다. \mathbf{x} 는 동차좌표(homogeneous coordinate)로 $[x, y, w]^T$ 와 같이 표

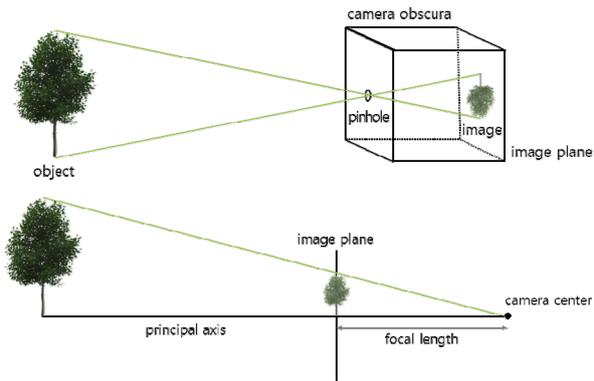


그림 1. 핀홀 카메라 모델 (실제 핀홀 카메라와 일반적으로 사용하는 단순화 예).

현되어, 실제 2차원 이미지 상의 점은 $(x/w, y/w)$ 와 같이 정규화하여 구해진다.

카메라 행렬(camera matrix) \mathbf{K} 는 초점거리(focal length)나 주점(principal point)과 같이 카메라의 내부 파라미터(intrinsic parameters)를 포함하는 3x3 행렬이다. 가장 일반화된 형태로 카메라 행렬은 X/Y방향 초점거리 f_x 와 f_y , 주점의 X/Y 좌표 c_x 와 c_y , 기울임 변수(skewness factor) s 와 같은 5개의 파라미터를 이용하여

$$\mathbf{K} = \begin{bmatrix} f_x & s & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix}$$

와 같이 표현된다. 카메라 행렬은 상황에 따라 몇 가지 가정을 넣어 3개의 파라미터 또는 1개의 파라미터를 이용하여

$$\mathbf{K}_1 = \begin{bmatrix} f & 0 & c_x \\ 0 & f & c_y \\ 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{K}_2 = \begin{bmatrix} f & 0 & w/2 \\ 0 & f & h/2 \\ 0 & 0 & 1 \end{bmatrix}$$

와 같이 간소하게 표현되기도 한다. 파라미터 1개를 이용하는 가장 간소한 모델인 \mathbf{K}_2 은 주점이 $w \times h$ 크기의 이미지의 중심에 있다는 가정을 바탕으로 한다. 이러한 간소한 모델은 인터넷에서 무작위로 얻은 영상과 같이 초점거리 f 를 모르는 많은 수의 이미지를 이용한 3차원 복원(structure-from-motion) 문제에서 미지수를 효과적으로 줄일 수 있는 모델로 사용된다.

카메라의 자세(camera pose), 즉 카메라의 시점과 위치는 각각 3차원 회전행렬 \mathbf{R} 과 \mathbf{t} 로 간접적으로 표현된다. 실제 \mathbf{R} 과 \mathbf{t} 는 월드좌표계에서 표현된 3차원 점 \mathbf{X} 를 카메라좌표계에 대한 점 $\hat{\mathbf{x}}$ 으로 바꿔주는 변환(transformation)이고, 카메라의 방향각과 위치는 해당 변환의 역(inverse)인 \mathbf{R}^T 와 $-\mathbf{R}^T \mathbf{t}$ 로

표현된다. 카메라 투영에 사용되는 변환 \mathbf{R} 과 \mathbf{t} 는 카메라의 위치와 시점에 따라 달라지기 때문에 외부 파라미터(extrinsic parameter)로 불려진다.

실제 카메라는 핀홀이 아닌 렌즈를 통해 빛을 한 점, 즉 초점으로 모은다. 또 실제 렌즈는 이상적인 핀홀과 달리 기하학적 왜곡(을 포함하고 있다. 따라서 카메라 투영 모델에서 이러한 렌즈에 의한 기하학적 왜곡을 반영되어야 한다.

렌즈 왜곡계수(distortion coefficient)는 렌즈의 기하학적인 변형을 표현하는 모델에 정의된 파라미터이다. 초점거리나 주점과 같이 카메라의 시점과 위치 변화에도 바뀌지 않는 값이기 때문에 카메라의 내부 파라미터(intrinsic parameter)로 구분된다. 렌즈가 만들어내는 기하학적 왜곡은 렌즈의 형상과 굴절률과 같은 재료의 광학적 특성에 영향을 받는데, 복잡한 물리적 현상을 효과적으로 표현하는 모델은 존재하지 않는다. 대신 렌즈의 기하학적 왜곡을 간단한 수식으로 근사(approximate)한 모델을 사용한다. 가장 널리 사용되는 것은 렌즈의 기하학적 왜곡을 다항식으로 묘사하는 다항식 왜곡 모델(polynomial distortion model)이다. 다항식 왜곡 모델은 각 픽셀의 주점으로부터의 거리 r 에 대해

$$\begin{bmatrix} \hat{x}_d \\ \hat{y}_d \end{bmatrix} = (1 + k_1 r^2 + k_2 r^4 + \dots) \begin{bmatrix} \hat{x} \\ \hat{y} \end{bmatrix} + (1 + p_3 r^2 + \dots) \begin{bmatrix} 2p_1 \hat{x} \hat{y} + p_2 (r^2 + 2\hat{x}^2) \\ 2p_2 \hat{x} \hat{y} + p_1 (r^2 + 2\hat{y}^2) \end{bmatrix}$$

$$(r^2 = \hat{x}^2 + \hat{y}^2)$$

와 같이 왜곡된 픽셀 위치 (\hat{x}_d, \hat{y}_d) 를 고차의 다항식으로 스케일링하여 표현한다. 다항식 왜곡 모델의 첫 번째 항과 파라미터 k_i 은 방사왜곡(radial distortion)을 표현하고, 두 번째 항과 파라미터 p_i 는 접선왜곡(tangential distortion)을 표현한다. 최근 제조 기술의 발달로 경우 카메라 이미지에 접선왜곡이 거의 나타나지 않기 때문에, 기하학적 왜곡으로 방사왜곡만 고려하는 경우가 많다. FOV 왜곡 모델(field-of-view distortion model)은 다항식으로 모델링하기 어려운 어안렌즈(fisheye lens)의 급격한 방사왜곡을 탄젠트(tangent) 함수로 근사하는 또 다른 왜곡 모델이다.

2. 카메라 캘리브레이션 문제 정의

카메라 캘리브레이션(camera calibration)은 카메라의 내외부 파라미터를 찾는 과정이다. 즉 카메라 캘리브레이션은 주어진

1) 실제 렌즈는 기하학적(geometric) 왜곡 외에 광도계적(photometric) 왜곡도 발생한다.

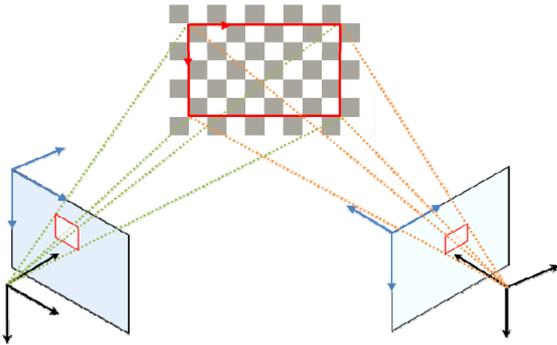


그림 2. 체스보드를 이용한 카메라 캘리브레이션의 예 (월드좌표계는 체스보드의 첫 번째 교차점에 위치).

카메라로 촬영한 이미지들을 이용하여 카메라 투영 모델을 설명하는 카메라 행렬, 렌즈 왜곡계수, 그리고 카메라의 자세를 계산하는 과정이다. 카메라 캘리브레이션에는 체스보드와 같이 사전에 정해진 마커(marker)를 이용하는데, 이를 통해 3차원 공간상의 점 \mathbf{X}_i 의 위치를 사전에 특정할 수 있다. 예를 들어 [그림 2]와 체스보드의 첫 번째 교차점을 월드 좌표계로 정의하면 나머지 다른 교차점들의 3차원 위치는 간단한 수열로 월드 좌표계에서 쉽게 계산된다. 마커를 촬영한 카메라 이미지에서 주어진 모든 3차원 점이 모든 이미지에 투영되어 관찰되는 점을 모두 찾으면 카메라 캘리브레이션을 위한 준비를 마치게 된다. i 번째 3차원 점 \mathbf{X}_i 가 j 번째 이미지에 투영된 점 \mathbf{x}_i^j 라고 표시할 때, 카메라 캘리브레이션은 아래와 같은 방정식의 해를 구하는 과정으로 볼 수 있다.

- 미지수: 카메라의 내외부 파라미터

$$(f, c_x, c_y, k_1, k_2, \dots, \mathbf{R}_j, \mathbf{t}_j)$$

- 주어진 정보: 3차원 점들 $\mathbf{X}_i (i = 1, \dots, n)$ 와 각 이미지에 투영된 점들 $\mathbf{x}_i^j (i = 1, \dots, n), (j = 1, \dots, m)$
- 주어진 조건: 카메라 투영 모델

$$\mathbf{x}_i^j = \text{proj}(\mathbf{X}_i; f, c_x, c_y, \dots, \mathbf{R}_j, \mathbf{t}_j)$$

카메라 캘리브레이션은 기술적으로 크게 두 부분으로 구성된다. 전단계(front-end)는 주어진 3차원 점들이 주어진 이미지들에 투영된 점들을 검출하는 단계이다. 예를 들어 체스보드를 마커로 사용하는 경우 체스보드를 촬영한 이미지들에서 체스보드의 교차점을 검출하는 단계이다. 후단계(back-end)는 전술한 조건들을 이용하여 카메라의 내외부 파라미터를 계산하는 단계이다. 보통 두 번째 단계도 2개의 세부 과정으로 나뉜다. 1) 첫 번째 세부 과정은 렌즈 왜곡 등을 고려하지 않거나 또

는 단순한 렌즈 왜곡만 고려하여 대략적인 (그러나 전역적으로 최적(globally optimal)에 가까운) 내외부 파라미터를 계산하는 과정이다. 2) 두 번째 세부 과정은 앞의 과정에서 얻어진 대략적인 내외부 파라미터를 초기값으로 하고 실제 타겟하는 복잡한 투영 모델에 해당하는 내외부 파라미터를 지역적 최적화(local optimization)를 통해 얻는 과정이다.

3. 가장 널리 사용되는 Zhang's Method

Zhang's method[1]은 OpenCV 라이브러리[2]에 포함되어 있어 가장 널리 사용되는 카메라 캘리브레이션 방법이다. Zhang's method는 평면 위의 3차원 점들을 이용하는 방법으로 평면 패턴으로는 체스보드가 일반적으로 많이 사용된다. Zhang's method는 후단계(back-end)에 해당하는 방법론이고, 전술한 것과 같이 2개의 세부 과정으로 이뤄져 있다.

첫 번째 세부 과정으로 렌즈 왜곡을 고려하지 않는 핀홀 카메라 모델에 대해 closed-form 형태로 카메라 내외부 파라미터를 계산하는 단계이다. 해당 단계에서는 모든 3차원 점들이 동일한 평면에 있다는 조건($z = 0$)을 이용하여 카메라 투영 모델을 3x3 행렬 형태의 호모그래피(homography)로 표현하는 것이 핵심이다. 그리고 호모그래피 행렬의 각 열이 직교(orthonormal)한다는 조건을 이용하여 카메라의 내외부 파라미터가 closed-form 형태로 계산된다.

두 번째 세부 과정에서는 렌즈 왜곡을 고려한 복잡한 모델과 앞 단계에서 구한 해를 초기값으로 이용하여 Levenberg-Marquardt algorithm과 같은 비선형 (지역적) 최적화 기법을 통해 최적의 파라미터를 얻는 단계이다. 최적화의 비용함수로 재투영 에러(reprojection error)가 사용된다. 재투영 에러는 이미지에서 검출된 교차점 \mathbf{x}_i^j 과 주어진 3차원의 점 \mathbf{X}_i 를 현재까지 얻은 카메라 내외부 파라미터를 사용하여 이미지에 투영시킨 점 사이의 직선 거리이다. 재투영 오차는 픽셀(pixel) 단위의 직선거리이기 때문에 이를 통해 검출된 교차점과 캘리브레이션을 통해 얻은 내외부 파라미터가 얼마나 일관성(consistency)있는지를 판단할 수 있다. Zhang's method에 사용되는 최적화의 비용함수는 재투영 에러를 고려하여

$$\sum_{i=1}^n \sum_{j=1}^m \| \mathbf{x}_i^j - \text{proj}(\mathbf{X}_i; f, c_x, c_y, \dots, \mathbf{R}_j, \mathbf{t}_j) \|^2$$

와 같이 표현할 수 있다. 이러한 비용함수는 visual SLAM이나 SFM(structure from motion)에 많이 사용되는 bundle adjustment와 유사한데 3차원 점의 위치 \mathbf{X}_i 가 고정되어 있는 것이 일반적인 bundle adjustment와 차이점이다.

체스보드와 Zhang's method를 이용하여 카메라 캘리브레이션을 수행하며 갖게 되는 개인적인 경험과 가이드라인을 Q&A 형태로 아래와 같이 정리한다.

Q) 카메라 캘리브레이션에 몇 장의 이미지가 적절한가?

Zhang's method의 논문에는 최소 2장 이상의 이미지를 요구하고 있다. 그러나 간단한 카메라 투영 모델을 사용하는 경우, 1장의 영상으로도 캘리브레이션이 가능하다. 당연히 주어진 이미지의 개수가 많을 수록 정확하고 안정적인 캘리브레이션 결과가 나올 가능성이 높다. 그렇다고 무턱대고 너무 많은 이미지를 획득하는 것은 많은 시간과 노력이 필요한 힘든 일이다. 이때 사용할 수 있는 지표가 캘리브레이션의 비선형 최적화 단계에서 결과로 얻은 제곱평균제곱근(root mean square; 이하 RMS) 형태의 평균적인 재투영 에러 값이다. 즉 카메라 이미지의 개수를 늘려도 RMS 에러가 크게 줄지 않는다면 더 이상 많은 이미지로 캘리브레이션에 할 필요는 없다. OpenCV의 카메라 캘리브레이션 함수는 RMS 에러 값을 기본 출력으로 반환하므로 이를 참고하면 된다. 또는 직접 3차원 점을 투영시켜 검출된 점과의 거리를 확인해 볼 수 있다.

Q) 체스보드와 카메라 사이의 시점(각도와 거리)는 어떻게 선택하는 것이 좋을까?

중요한 점은 초점거리나 주점과 같은 카메라 파라미터는 투영된 체스보드에 의해 생기는 소실점(vanishing point)으로부터 도출된다는 점이다. 따라서 체스보드가 소실점없이 직사각형 형태로 보여지도록 촬영하는 것보다 소실점이 생기도록 어느정도 기울여 주는 것이 좋다. 그렇다고 체스보드를 너무 기울이는 경우 시점왜곡이 커져 검출된 체스보드의 교차점의 부정확해지기 때문에 주의하여야 한다. 카메라와 체스보드 사이의 거리는 체스보드가 모두 이미지에 보여지도록 잡는 것이 좋은데, 체스보드가 작게 보여지는 경우 마찬가지로 교차점 검출이 부정확해진다.

Q) 카메라가 오토포커스(auto-focus) 기능이 켜져있는지 확인할 수 있는가?

일반적으로 연구용 산업용으로 사용되는 카메라는 고정초점 카메라이다. 그러나 웹캠이나 스마트폰 또는 CCTV의 카메라와 같이 주변에서 쉽게 접하게 되는 카메라들은 대부분 초점이 움직이는 오토포커스이므로 주의하여야 한다. 오토포커스로 인해 카메라 캘리브레이션 당시의 초점거리와 실제 기하

학적인 분석을 수행할 때의 초점거리가 달라질 수 있고, 이는 분석 결과에 큰 오차를 만들어낸다. 기하학적 분석 알고리즘 자체가 초점거리가 오토포커스를 고려하여 만드는 것이 근본적인 해결책이지만, 이러한 방법은 알고리즘이나 문제의 복잡도가 크게 증가시킨다. 따라서 사용하고 있는 카메라가 오토포커스 기능이 켜져있는지 확인하고 이를 불능상태(disable)로 만드는 것이 간단한 방법이다. 카메라 설정에서 오토포커스를 확인할 수 있는 경우도 있지만, 그렇지 않은 경우에는 카메라 캘리브레이션 단계에서 이를 확인할 수 있다. 체스보드를 촬영한 이미지들을 촬영 거리에 따라 구분하고, 구분된 이미지들에 대해 각각 캘리브레이션을 수행한다. 만약 촬영 거리에 따라 얻어진 초점거리가 비례적으로 달라진다면 오토포커스 기능이 켜져 있는 상태일 수 있다.

III. 최근 카메라 캘리브레이션 기술 동향

카메라 캘리브레이션의 중요성과 필요성 때문에 2000년에 발표된 Zhang's method 이후에도 많은 연구가 이뤄지고 있다. 본 기고에서는 아래와 같이 세 가지 연구 방향에 따라 몇 가지 기술들을 간략히 리뷰하고자 한다.

1. Marker를 사용하지 않는 Self-Calibration

많은 카메라 캘리브레이션 기술이 3차원 참조점 X_i 를 쉽게 얻기 위해 체스보드와 같은 마커를 사용한다. 그러나 이미 높은 곳에 장착된 CCTV 카메라와 같이 체스보드 영상을 촬영하기 힘든 경우가 있고, 인터넷에서 다운받은 이미지와 같이 체스보드를 이용해 캘리브레이션 불가능한 경우도 있다. 따라서 체스보드와 같은 마커를 이용하지 않는 셀프-캘리브레이션(self-calibration) 또는 오토-캘리브레이션(auto-calibration) 기술들이 많이 연구개발되어 왔다.

간단한 예 중 하나는 규격이 정해진 스포츠 경기장의 바닥 패턴을 체스보드 대신 사용하는 것이다. Hu et al.[3]은 일반적인 농구 중계 영상에서 농구장의 바닥 패턴을 이용하여 Zhang's method와 유사한 평면기반의 카메라 캘리브레이션을 사용하였다. 스포츠 경기장의 경우 대부분 평면이므로 Zhang's method와 유사한 평면 방법이 잘 적용된다.

사람이 생활하는 실내의 공간은 많은 면과 선으로 이뤄져 있고 모두 직교(orthogonal)하는 특징이 있고, 이러한 공간을 촬영한 이미지에서 관찰되는 평행한 선분들이 만드는 소실점은 카메라 파라미터를 추정하는 단서가 된다. Kosecka와 Zhang[4]은 소실점을 이용하여 카메라의 내부 파라미터를 추정하는

기법을 제안하였다. 제안된 방법의 전단계(front-end)에서는 Hough transform으로 선분을 추출하고, 이를 세 방향의 평행한 선분 그룹으로 나누어 3개의 소실점을 얻는다. 후단계(back-end)에서는 3개의 소실점 ($\mathbf{v}_i, \mathbf{v}_j, \mathbf{v}_k$)의 기저벡터가 서로 직교하므로

$$\mathbf{v}_i \mathbf{K}^T \mathbf{K}^{-1} \mathbf{v}_j = \mathbf{v}_j \mathbf{K}^T \mathbf{K}^{-1} \mathbf{v}_i = \mathbf{v}_j \mathbf{K}^T \mathbf{K}^{-1} \mathbf{v}_k = 0$$

와 같은 3개의 제약조건을 이용하여 미지수에 해당하는 카메라 행렬 \mathbf{K} 을 얻을 수 있다. 유사하게 Nakano[5]는 길이가 동일하고 지면에 수직인 2개 이상의 선분들을 이용하여 카메라를 캘리브레이션하는 후단계(back-end) 기법을 제안하였다. 선분의 길이나 지면에 닿는 3차원 위치를 알지 못하는 경우에도 카메라 캘리브레이션이 가능하다. 실내외 공간에는 유사한 길이의 선이 반복적으로 나타나는데, Nakano의 방법은 매우 효과적으로 적용될 수 있다.

최근에는 딥러닝 기법을 이용하여 일반 영상에서 카메라 캘리브레이션을 수행하는 기술들도 많이 연구되고 있다. Lopez et al.[6]은 CNN 네트워크를 이용하여 초점거리와 방사왜곡계수 k_1, k_2 그리고 카메라의 상하(tilt)와 돌림(roll) 회전 각도를 회귀 분석(regression) 네트워크 문제로 접근하였다. 특징적인 부분은 정해진 크기로 변환(resize)된 CNN 입력 이미지에 적합하게 카메라 내부 파라미터를 새로 정의하고 사용하였다는 점이다. 또 해당 연구에서는 SFM 데이터셋으로부터 얻은 방사왜곡계수 k_1 과 k_2 사이의 상관관계가 있음을 보였다는 점도 흥미롭다.

카메라 이미지를 전혀 사용하지 않는 zero-shot 캘리브레이션 방법[7]도 있다. 핵심 아이디어는 카메라에 스펙으로 주어진 해상도와 시야각 정보를 이용하여 카메라의 영상을 보지 않고 바로 초점거리와 방사왜곡 파라미터를 계산하는 것이다. 카메라의 영상을 사용하지 않지만 렌즈 왜곡이 큰 어안렌즈 환경에서도 5% 이내의 꽤 정확한 결과를 만들어낸다.

2. Point를 사용하지 않는 Calibration

Zhang's method와 많은 후속 연구들은 3차원 점과 이들의 투영점을 이용한다. 그러나 점이 아닌 선이나 원과 같은 다른 기하학적 요소를 이용한 카메라 캘리브레이션 방법도 연구되었다.

Chung et al.[8]은 주선(principal line)이라는 개념을 이용하여 Zhang's method의 단점을 보완한 방법을 제안한다. 해당 연구는 Zhang's method의 두 가지 단점으로 1) 여러 체스보드 이미지 중 오히려 결과에 방해가 될 수 있는 나쁜 이미지를 판단하거나 제거하지 못한다는 점과 2) 고정된 초점거리를 가정하고 캘리브

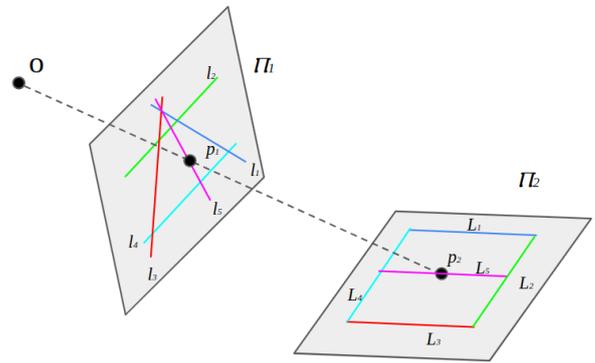


그림 3. Chung et al.[8]에서 제안하는 주선(principal line) l_5 (Π_1 은 이미지 평면, Π_2 는 마커의 평면, p_1 은 주점(principal point)를 나타냄).

레이션이 수행된다는 점을 지적한다. 주선(principal line)은 [그림 3]의 l_5 와 같이 주점(principal point)와 평행한 두 선이 카메라 투영되어 만드는 소실점(vanishing point)을 지나는 선으로 정의된다. 제안하는 방법은 3차원 공간 내 직사각형 패턴과 이미지에 투영된 해당 사각형 사이의 호모그래피(homography)를 구하고, 얻어진 호모그래피 행렬의 원소들을 이용하여 주선(principal line)의 방정식을 도출한다. 이후 여러 영상들을 이용하여 2개 이상의 주선들이 만나는 교차점 구하여 주점(principal point)을 구한다. 여러 영상에서 투영된 직사각형의 호모그래피를 X축 방향의 회전과 Z축 방향의 병진이동이 있는 호모그래피로 모두 정규화하여 초점거리와 외부 파라미터를 구할 수 있다. 주선(principal line)을 통해 캘리브레이션 과정에서 사용하면 오히려 좋지 않은 이미지를 판단할 수 있고, 주점이 초점거리 등 다른 파라미터에 관계없이 도출되기 때문에 초점거리가 고정되지 않는 상황에도 적용하기 유리하다.

그 외에 Zhang et al.[9]은 체스보드의 교차점이 아닌 선분(line segment)을 이용해 캘리브레이션을 수행한다. 제안된 알고리즘 또한 평면 위의 선분을 이용하는데, 전단계(front-end)에서는 선분 위의 점들의 매칭과 법선벡터(normal vector)를 이용한 호모그래피(homography)의 추적(tracking)을 진행한다. 후단계(back-end)에서는 점과 선의 거리를 비용함수로 하는 선형 최적화를 통해 카메라 캘리브레이션을 수행한다.

3. 다양한 센서 데이터를 고려한 Multi-Sensor Calibration

카메라는 단독으로 사용되기도 하지만, 다양한 센서들과 함께 사용하기 한다. 예를 들어 스테레오 카메라나 CCTV 카메라 네트워크와 같이 두 개 이상의 카메라를 동시에 사용하기도 한다. 또 RGB 데이터 외에 깊이(depth) 데이터나 열화상

(thermal) 데이터를 획득하는 카메라를 함께 사용하기도 한다. 또 다양한 응용, 특히 로봇틱스와 자율주행 분야에서는 카메라와 함께 LiDAR, Radar, IMU, GPS 등의 센서를 함께 사용하는 것이 일반적이다. 다양한 센서 데이터를 융합하여 처리하기 위해서 가장 기본이 되는 것은 각 센서들이 설치된 상대적인 위치와 방향을 아는 것이다. 각 센서의 상대적 자세를 알고 있어야 센서 데이터를 융합하여 사용할 때 센서 데이터의 설치 위치를 고려하여 센서 데이터를 보정할 수 있기 때문이다. 다양한 센서들의 상대적인 설치 위치와 방향을 알아내는 과정을 멀티 센서 캘리브레이션(multi-sensor calibration)이라고 부른다. 또 센서가 설치된 상대적인 자세 정보는 카메라의 외부 파라미터에 해당하므로 이를 외부 캘리브레이션(extrinsic calibration)이라고 부르기도 한다.

우선 널리 사용되고 있는 OpenCV[2]의 경우 단일 카메라의 캘리브레이션과 스테레오 카메라의 캘리브레이션만을 지원하고 있다.

ETHZ에 만든 Kalibr[10]는 다양한 센서들을 캘리브레이션하는 가장 널리 알려진 툴이다. 단일 카메라 영상뿐만 아니라 시야각 겹침이 거의 없는 카메라 네트워크의 내외부 파라미터 캘리브레이션이 가능하다. Kalibr는 핀홀 카메라나 전방향 카메라(omni-directional camera)를 포함 총 4종의 카메라 투영 모델 지원하고, 다항식 왜곡모델과 FOV 왜곡모델을 포함하여 총 3종의 렌즈 왜곡 모델을 지원한다. 또 롤링-셔터(rolling shutter) 카메라를 위한 파라미터를 찾아내는 기능도 있다. 또 로봇 분야에서 개발된 만큼 로봇 분야에 많이 사용되는 IMU 센서에 대한 캘리브레이션이 가능한데, 카메라-IMU간 캘리브레이션과 다수의 IMU 캘리브레이션을 수행할 수 있다.

그 밖에 TU Delft에서 만든 멀티 센서 캘리브레이션 도구[11]는 다수의 카메라와 LiDAR, Radar 센서를 동시에 캘리브레이션 가능하다. 이 기술은 로봇 미들웨어인 ROS 기반으로 작성되어 센서에서 바로 들어오는 데이터나 rosbag 형태로 동기화되어 기록된 데이터를 이용하여 전단계(front-end)에 해당하는 각 센서 데이터에 해당하는 검출기(detector)와 후단계(back-end)에 해당하는 파라미터 최적화 모듈로 구성되어 있다.

IV. 맺음말

본 기고에서는 카메라 캘리브레이션의 원리와 기술 동향을 살펴보았다. 카메라 캘리브레이션의 배경 이론으로 카메라 투영 모델과 이에 포함된 다양한 의미를 갖는 카메라 파라미터들

을 살펴보았다. 카메라 캘리브레이션은 카메라 투영 모델의 파라미터를 추정하는 과정이다. Zhang's method는 카메라 캘리브레이션에 가장 널리 사용되는 방법으로 closed-form 형태로 초기 파라미터를 추정하고, 비선형 최적화를 통해 복잡한 렌즈 왜곡도 고려한 카메라 투영 모델을 알아낸다. 본 기고에서는 Zhang's method를 사용하면 흔히 갖게되는 이슈에 대한 가이드라인을 공유하였다. 카메라 캘리브레이션은 그 필요성과 중요성으로 인해 여전히 많은 연구가 이뤄지는 주제이고, 세 가지 관점에서 최근의 연구들을 살펴보았다. Zhang's method와 달리 마커를 사용하지 않는 기법들, 또 점을 사용하지 않는 기법들을 살펴보았다. 로봇 분야에서는 카메라를 다양한 센서들과 융합하여 사용하는 경우 많은데, 멀티 센서 캘리브레이션 기법은 여러 센서 데이터를 설치 위치를 보정하여 융합하는데 필요한 정보를 제공한다. 본 기고를 통해 독자들이 카메라 투영 모델과 카메라 캘리브레이션을 이해하고, 자신의 연구에 적합한 방법론과 절차를 선택하는데 도움이 되었으면 한다.

사사

본 기고는 과학기술정보통신부 및 한국연구재단의 'BRIDGE융합연구개발'사업의 지원으로 작성되었습니다. (과제명: AI기반 3차원 곡면에서의 위치 인식 및 이동 경로 생성 기술, 과제번호: 2021M3C1C3096810)

REFERENCES

- [1] Z. Zhang, "A Flexible New Technique for Camera Calibration," *IEEE Transactions on Pattern Recognition and Machine Intelligence (PAMI)*, vol. 22, no. 11, 2000 <https://doi.org/10.1109/34.888718>
- [2] OpenCV, <https://opencv.org/>
- [3] Min-Chun Hu et. al., "Robust Camera Calibration and Player Tracking in Broadcast Basketball Video," *IEEE Transactions on Multimedia*, vol. 13, no. 2, 2011 <https://doi.org/10.1109/TMM.2010.2100373>
- [4] Jana Kosecka, Wei Zhang, "Video Compass," *European Conference on Computer Vision (ECCV)*, 2002 https://doi.org/10.1007/3-540-47979-1_32
- [5] Gaku Nakano, "Camera Calibration using Parallel Line Segments," *International Conference on Pattern Recognition (ICPR)*, 2021 <https://doi.org/10.1109/ICPR48806.2021.941>

2183

- [6] Manuel Lopez et. al., "Deep Single Image Calibration with Radial Distortion," *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019 <https://doi.org/10.1109/CVPR.2019.01209>
- [7] Jae-Yeong Lee, "Zero-Shot Calibration of Fisheye Cameras," arXiv, 2020 <https://arxiv.org/abs/2011.14607>
- [8] Jen-Hui Chuang et. al., "Geometry-Based Camera Calibration Using Closed-Form Solution of Principal Line," *IEEE Transactions on Image Processing*, vol. 30, 2021 <https://doi.org/10.1109/TIP.2020.3048684>
- [9] Yueqiang Zhang et. al., "A Flexible Online Camera Calibration using Line Segments," *Journal of Sensors*, 2016, <https://doi.org/10.1155/2016/2802343>
- [10] Paul Furgale et. al., "Unified Temporal and Spatial Calibration for Multi-Sensor Systems," *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2013 <https://github.com/ethz-asl/kalibr>
- [11] Joris Domhof et. al., "An Extrinsic Calibration Tool for Lidar, Camera and Radar," *IEEE International Conference on Robotics and Automation*, 2019 https://github.com/tudelft-iv/multi_sensor_calibration

저자약력



최준혁

- 2015년~2022년 단국대학교 기계공학과 (공학사)
- 2022년~현 재 서울과학기술대학교 컴퓨터공학과 (석사과정)



최성록

- 2001년~2006년 서울대학교 기계항공공학부 (공학사)
- 2006년~2008년 KAIST 로봇공학학제전공 (공학석사)
- 2014년~2019년 KAIST 로봇공학학제전공 (공학박사)
- 2008년~2020년 ETRI 지능로봇스연구본부 (선임연구원)
- 2021년~현 재 서울과학기술대학교 컴퓨터공학과 (조교수)