

신경망을 이용한 렌즈의 왜곡모델 구성 및 카메라 보정

김민석, 남창우, 우동민
 명지대학교 정보제어공학과

Camera Calibration And Lens of Distortion Model Constitution for Using Artificial Neural Networks

Min-Suk Kim, Chang-Woo Nam, Dong-Min Woo
 Dept. of Information & Control Eng. Myong-Ji UNIV

Abstract - The objective of camera calibration is to determine the internal optical characteristics of camera and 3D position and orientation of camera with respect to the real world. Calibration procedure applicable to general purpose cameras and lenses. The general method to revise the accuracy rate of calibration is using mathematical distortion of lens. The effective og calibration show big difference in proportion to distortion of camera lens. In this paper, we propose the method which calibration distortion model by using neural network. The neural network model implicity contains all the distortion model. We can predict the high accuracy of calibration method proposed in this paper. Neural network can set properly the distortion model which has difficulty to estimate exactly in general method. The performance of the proposed neural network approach is compared with the well-known Tsai's two stage method in terms of calibration errors. The results show that the proposed approach gives much more stable and acceptabke calibration error over Tsai's two stage method regardless of camera resolution and camera angle.

1. 서 론

카메라 보정은 컴퓨터 영상의 점과 실세계상의 점의 상관관계를 결정하는 과정이다. 일반적으로 카메라 보정시 고려하여야 할 요소들은 카메라의 위치, 회전각등과 같은 외부변수들과 초점거리, 왜곡 변수등과 같은 카메라의 고유 변수등이 있다. 이러한 물리적인 고려 요소로서 보정을 하는 방식을 외재적 보정이라 한다. 외재적 보정과 다른 개념으로 내재적 보정이란 방법이 있다. 내재적 보정이란 컴퓨터 영상 해석에 있어서의 보정이란 영상 좌표로부터 실제 좌표를 구하거나 주어진 실세계 좌표로 올바른 영상 좌표를 찾는 것이므로 외재적 보정이 아닌 가상적인 중개변수를 가진 모델을 설정하고 그러한 모델을 이용하여 좌표변환을 하는 방법을 말한다. 본 연구에서는 위의 외재적 보정방법에서의 렌즈의 왜곡을 무시한 보정방법을 사용하였다. 그리고 렌즈의 왜곡계수의 모델을 구성하기 위하여 신경망에 렌즈의 왜곡계수 모델을 내재 시키는 방법을 병용하였다. 렌즈의 왜곡을 무시한 보정 방법으로 z축으로의 위치 이동을 제외한 카메라의 외부 변수를 선형 방정식으로 구하고 여기서 구해진 수치들을 다시 신경망 모델에 적용시켜 렌즈의 왜곡을 적용시킨 결과를 얻을 수 있다. 그러므로 기존의 방법들에서 얻을 수 없는 물리적인 변수들을 구할 수 있으므로 기존에 제안된 방법들에 비해 정확도가 높을 것으로 기대된다. 본 논문에서는 왜곡계수가 큰 카메라

와 왜곡계수가 작은 카메라에 대해 카메라 보정 실험을 수행하였다. 본 논문의 2.1장에서는 왜곡이 무시된 카메라 모델의 보정 방법을 설명하고 2.2장에서는 적용된 신경망 알고리즘과 이에 근거한 카메라 보정방법을 설명한다. 2.3장에서는 영상 좌표로부터 실세계 좌표를 구하는 카메라 보정에 대해 실험하고 3장에서는 본 논문의 결론이 도출된다.

2. 본 론

2.1 렌즈의 왜곡을 무시한 보정

카메라 보정의 대부분의 방법은 먼저 카메라의 기하학으로부터 카메라 모델을 설정하고 그 모델의 변수들을 미리 주어진 보정점과 이에 대응하는 실세계 좌표점을 이용하여 구하는 방식으로 되었다. 이 과정의 핵심은 카메라 모델의 설정인데 근본적으로 왜곡계수를 완전히 반영하는 모델은 없고 단지 근사화일 뿐이며 왜곡계수를 고려하게 되면 전체모델이 비선형으로 이루어 지게 된다. 이러한 외재적 보정 방식중 가장 널리 쓰이는 Tsai의 2 단계 방법으로 외재적 보정 방법을 수행 하였다. 먼저 렌즈가 이상적이라는 가정하에 카메라 모델을 설정한다. 그림 1에서 나타난 카메라 모델에서 (X_w, Y_w, Z_w) 는 3차원 실세계 좌표상 물체 P이며 (x, y, z) 는 3차원 카메라 좌표 상 물체위치 P이다. 카메라 좌표계의 중심점은 광학적 중심 O이고 (X, Y) 는 O_0 에 중심을 가진 영상 좌표계이다. (X_u, Y_u) 는 이상적인 렌즈에 의한 영상 좌표점으로 렌즈의 왜곡계수가 전혀 고려 되지않은 좌표이다. (X_u, Y_u) 는 최종적으로 컴퓨터 영상 좌표계 상의 점 (X_f, Y_f) 로 변환된다. 이 모델을 가지고 좌표계간의 변화를 수행하면 (x_w, y_w, z_w) 는 물체의 점 P의 3차원 실세계 좌표점이고 (x, y, z) 는 렌즈의 광학적 중심 O에 중점을 두고 z축이 광학축과 같은 3차원 카메라 좌표에서의 물체의 점 P이다. 이때 두 좌표계간의 변환식은 다음과 같이 나타난다.

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} r_1 & r_2 & r_3 \\ r_4 & r_5 & r_6 \\ r_7 & r_8 & r_9 \end{bmatrix} \begin{bmatrix} x_w \\ y_w \\ z_w \end{bmatrix} + \begin{bmatrix} T_x \\ T_y \\ T_z \end{bmatrix}$$

실제로 이단계에서 $r_1, r_2, r_3, r_4, r_5, r_6, r_7, r_8, r_9, T_x, T_y, T_z$ 의 값이 보정 되어야 한다. 다음 단계로 3차원 카메라 좌표가 투시투영변환에 의해 이상적인 왜곡이 없는 영상 좌표점 (X_u, Y_u) 로 변환 된다.

$$\begin{aligned} X_u &= f(x/z) \\ Y_u &= f(y/z) \end{aligned}$$

이 단계에서는 초점거리 f 가 포함된다. 최종적으로 실제 영상 좌표가 영상 버퍼상의 컴퓨터 영상 좌표 (X_f, Y_f) 로 변환되는 과정은 다음의 식으로 나타난다.

$$\begin{aligned} X_d &= S_x(X_f - C_x) \\ Y_d &= S_y(Y_f - C_y) \end{aligned}$$

위의 식에서 $(X_d, Y_d) = (0, 0)$ 일 때 $(C_x, C_y) = (X_f, Y_f)$ 이므로 (C_x, C_y) 는 실제로 렌즈의 광학축과 영상 좌표

평면의 교점의 영상 비퍼 상의 컴퓨터 영상 좌표로서 영상의 중점이고 S_x 및 S_y 는 영상 비퍼의 A/D변환시 x축 및 y축으로의 확대 축소비를 계산하는 계수이다. 모든 식을 통합하여 보면 다음과 같은 식이 나온다.

$$S_x(X_f - C_x)(1 + G(X_d^2 + Y_d^2)) = f(r_1x_w + r_2y_w + r_3z_w + T_x) / (r_7x_w + r_8y_w + r_9z_w + T_z)$$

$$S_y(Y_f - C_y)(1 + G(X_d^2 + Y_d^2)) = f(r_4x_w + r_5y_w + r_6z_w + T_x) / (r_7x_w + r_8y_w + r_9z_w + T_z)$$

위의 식에서 실험 대상인 실세계 평면상의 좌표는 z_w 의 값은 0이 되고 왜곡을 고려하지 않은 보정이므로 왜곡 계수 $G = 0$ 이 되어 모든 변수의 값이 구해진다.

2.2 렌즈의 왜곡계수 모델을 위하여 신경망을 사용한 방법

실제적으로 위의 2.1절에서 적용된 외재적 보정 방식을 사용하여 왜곡계수가 적용된 보정방식을 실행하면 여러 가지의 변수가 적용되면서 매우 복잡한 비선형 방정식이 도출되므로 정확하게 결과값을 얻어내기가 힘들어진다. 따라서 내재적 보정방법으로 이러한 왜곡계수를 신경망의 뉴런구조에 적용시켜 결과를 도출해내는 과정을 제안한다. 본 논문에서는 왜곡계수를 적용 시킬 알고리즘으로 역전파 학습 알고리즘을 제안한다. 이 구조는 특히 Mapping에 관한 부분에 매우 유용하다는 평가를 받고 있다. 적용된 역전파 알고리즘은 감독 학습 알고리즘이라서 주어진 입력 값에 대한 목표값이 필요하다. 이 알고리즘의 기본적인 개념은 주어진 목표값과 입력값 사이의 오차를 최대한으로 줄이는 것이다. 여기서의 오차는 다음과 같이 나타낸다.

$$E = \sum_{j=0}^N (t_j - O_j)^2$$

여기서 t_j 는 신경망의 총 N개의 출력뉴런중 j번째 목표값이고 O_j 는 출력 뉴런중 j번째 실제 출력값을 나타낸다. 위에서 보듯이 역전파 알고리즘은 주어진 입력에 대해 신경망을 통과한 출력이 최대한 목표값과 가까운 값을 가지기 위해 E값을 점차적으로 줄여나가는 Gradient decent 알고리즘을 이용한 것이다. 역전파 알고리즘의 신경망 내부의 뉴런간의 weight들은 다음의 식에 따라 오차를 줄이는 방향으로 갱신된다.

$$\Delta W_{ij} \propto - \frac{\partial E}{\partial W_{ij}}$$

$$\Delta W_{ij} = \eta \delta_j \cdot O_j$$

그리고 그 식은 위의 식과 같이 변할수 있다. 여기서 η 는 학습률을 나타낸다. 위에서 보듯이 역전파 알고리즘은 입력값과 목표값이 적절한 대응을 이루어야지만 올바른 결과를 얻어낼 수 있다. 이러한 적절한 쌍을 이룬 신경망을 가지고 학습을 한다면 학습에 사용되지 않은 임의의 데이터에 대해서도 믿음직한 출력값을 나타내주므로 렌즈의 왜곡계수를 내재시킬수 있다는 것이다.

2.3 카메라 보정 실험 및 결과

제안된 보정 알고리즘의 타당성을 입증하기 위하여 실제의 영상에 의한 실험을 수행하였다. 일단 이미 실제의 좌표를 알고있는 균일한 간격을 갖는 영상을 취득한다. 본 논문에서 사용된 카메라는 두 대의 CCD카메라로 하나는 고해상도인 PULNIX TM-7EX CCD 이고 또 하나는 저해상도의 일반 CCTV용 CCD 카메라이다. 이 두가지 카메라에 대한 사양은 표1에 주어져 있다. 이렇게 두 대의 카메라를 사용한 이유는 본 논문에서 제안된 방법과 일반적으로 사용되는 Tsai의 방법이 왜곡계수가 심한 경우 어떻게 동작하는가 예시하기 위하여 사용되었다. 보정을 위한 영상은 각기 다른 각도의 3가지 각도에서 취득하였고 본 논문에서 제안된 방법과 Tsai의 방법을 같이 적용하여 비교, 분석하였다. 본 논문에서 제안된 방법의 정확도를 판단하기 위하여 제안된 기준은 영상 좌표를 입력한 후 나오는 결과값인 실세계 좌표를 이미

알고있는 실세계 좌표와 비교 해보는 것으로 하였다. 이 평균 픽셀 단위 오차는 다음의 식과 같다.

$$AEIP = \frac{1}{N} \sum_{i=1}^N [(X'_i - X_i)^2 + (Y'_i - Y_i)^2]^{1/2}$$

여기서 (X'_i, Y'_i) 는 영상 좌표를 입력값으로 하여 나온 실세계 좌표의 추정치이다.

그리고 보정을 위한 영상은 가로 26mm, 세로 20mm 간격으로 일정하게 배치되어 있고 가로 11개, 세로 9개로 총 99개의 보정점을 가지고 있다. 이렇게 취득한 영상으로부터 영상 좌표를 얻게되면 이 영상 좌표가 제안된 보정 알고리즘의 입력값이 된다. 입력값을 일단 선형 계산 부분인 Tsai의 1단계 부분에 입력 시킨후 여기서 출력되는 왜곡계수를 적용시키지 않은 결과값을 다시 신경망의 입력값으로 입력 시킨다. 여기서 적용된 신경망은 역전파 알고리즘을 이용한 신경망이고 은닉층이 두

Frame Grabber	수평 해상도(X축)	512	
	수직 해상도(Y축)	512	
PULNIX TM-7EX CCD	셀의 치수	X축	8.4 μ m
		Y축	9.8 μ m
	셀의 개수	X축	768
		Y축	494
	초점거리	16 mm	
일반 CCTV용 CCD	셀의 치수	X축	9.6 μ m
		Y축	7.5 μ m
	셀의 개수	X축	243
		Y축	494
	초점거리	12 mm	

표 1 실험에 사용된 두 가지의 카메라 사양

개인 신경망을 사용하였다. 여러 가지 다른 구조의 신경망을 가지고 실험 한 결과 본 논문에서 요구되는 정확도를 가질수 있는 구조로는 은닉층이 충분한 숫자가 요구되었다. 실험 결과 각 은닉층당 약 10개내외의 뉴런이 필요하였다. 총 99개의 데이터 중 무작위적인 추출을 통해 79개의 데이터를 왜곡계수를 무시한 Tsai의 방법과 신경망의 학습에 이용하였고 나머지 20개의 보정점은 최종 결과를 테스트하는데 사용되었다. 마찬가지로 Tsai의 방법에서는 79개의 데이터로 카메라의 물리적인 변수들을 구하고 나머지 20개의 보정점 데이터로 테스트 오차를 구하였다. 각각의 보정에 사용된 영상이 3가지의 각도에서 존재하므로 각 데이터 Set이 3가지가 존재하게 된다. 위의 세가지 데이터 Set을 본 논문에서 제안된 방법과 Tsai의 방법, 두 가지에 각각 적용시켜 수행하였다. 실험의 결과는 아래의 그림과 같이 각각의 AEIP의 결과로서 보여지고 있다.

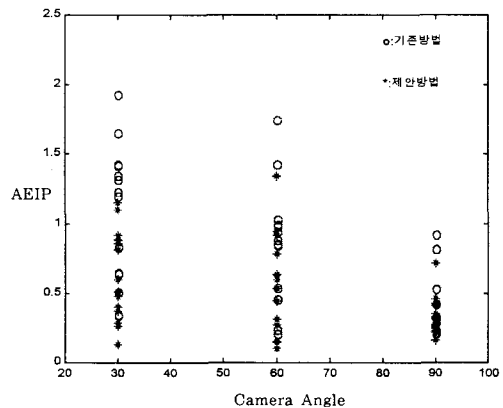


그림 2 저해상도 카메라

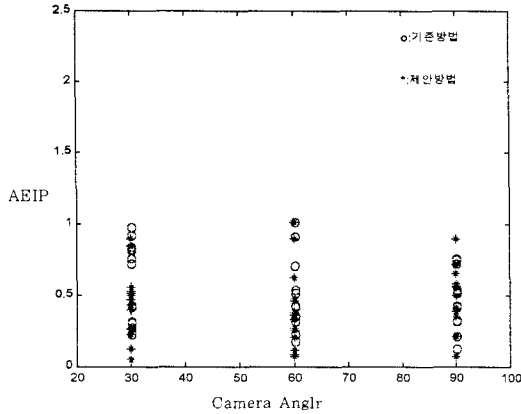


그림 3 고해상도 카메라

그림 1은 저해상도 카메라인 경우를 보여주고 있다. 그림에서 볼수 있듯이 카메라의 각도가 낮을때인 30°인 경우 Tsai의 방법은 본 논문에서 제안한 방법과 약 1.5배에서 2배까지의 차이가 나는 것을 볼수 있다. 즉 카메라의 각도에 따라 많은 차이를 나타내는 것이다. 그림 2는 고해상도 카메라의 경우를 보여주고 있는데 역시 마찬가지로 각도에 따라 본 논문에서 제안한 방법과 많은 차이를 보여주고 있다. 위의 결과를 다시 표2로 만들었다. 표2에서 보듯이 기존의 Tsai의 방법과 논문에서 제안한 방법간의 평균 값에서 약 1.5배의 차이를 보이고 있다.

영상	Tsai의 방법		신경망을 이용한 렌즈의 왜곡모델 구성 방법	
	각도	평균 오차	각도	평균 오차
PULNIX	30°	0.653516	30°	0.435926
TM-7EX	60°	0.431502	60°	0.395995
CCD	90°	0.464781	90°	0.431470
일반CCTV-용 Camera	30°	0.992168	30°	0.644148
	60°	0.810672	60°	0.564963
	90°	0.388940	90°	0.325435

표 2 두 가지 방법에 의한 실험 결과(단위:AEIP)

본 논문에서 제안한 방법은 카메라의 각도나 해상도에 영향을 거의 받지 않고 매우 안정되어 있는 결과를 얻을 수 있음을 알수 있다. 이는 본 논문에서 제안한 방법이 카메라의 왜곡계수에 영향을 거의 받지 않고 항상 적절한 취득값을 얻을수 있음을 의미하는 것이다. 비교 대상인 Tsai의 방법은 카메라 내부의 왜곡계수의 영향을 많이 받는 영상인 경우, 그리고 상대적으로 저해상도의 카메라인 경우 그 성능이 상대적으로 매우 안 좋았음을 알수 있다.

3. 결 론

본 논문에서 제안된 방법인 Tsai의 방법과 신경망을 병합하여 사용하는 방법으로 기존의 Tsai의 방법과 같이 카메라 내부의 왜곡계수를 수식적으로 계산하여 모델의 변수들을 구성하는 대신 기존의 Tsai의 방법으로 수식적으로 계산할 때 적정도의 정확성을 가질수 있는 내용은 수식적으로 계산하고 수식적으로 적정의 정확도를

얻기 힘들어서 든 왜곡계수등의 비선형 계산 부분을 신경망에게 학습을 통하여 왜곡계수를 내재할수 있도록 하는 방법이다. 기존의 방법은 믿을수 있는 왜곡계수를 지니기 힘들고 기타의 비선형 부분의 변수들에서 정확한 추정이 힘든 단점이 있으나 본 논문에서 제안한 방법은 그러한 왜곡계수와 같은 비선형 부분의 변수들을 신경망 구조내내 내포하게 되므로 훨씬 안정적인 카메라 보정이 가능하다. 게다가 결과에서 보듯이 Tsai의 방법보다 상당한 성능개선 효과가 있다. 본 논문에서 제안한 방법은 신경망에서의 학습시 많은 데이터 양과 시간을 소요하지만 일단 학습이 되고 신경망의 Weight가 뉴런간에 내재되면 실시간적인 적용이 가능하다. 그러므로 실시간 처리를 요구하는 응용분야에서 적용이 가능하고 카메라의 각도에 대해 많은 영향을 받지 않으므로 보다 넓은 적용범위를 가질수 있다. 그리고 왜곡계수등의 영향을 많이 받지 않으므로 보다 일반적인 분야에도 적용이 가능하다고 할수 있다. 즉 카메라 보정의 적용 범위를 넓혀 현재까지 정확도의 문제나 카메라 보정 적용시 문제가 되었던 점들을 개선시킬수 있는 방법이라 할수 있다.

(참 고 문 헌)

- [1] Roger Y. Tsai, "A Versatile Camera Calibration Technique for high-Accuracy 3D Machine Vision Metrology Using Off-the-Shelf TV Cameras and Lenses", Proc IEEE Journal, Robotics and Automation, vol.RA-3, pp.323-343, 1987
- [2] Anuo Basu and Kavita Ravi, "Active Camera Calibration using Pan, Tilt and Roll", Proc IEEE Trans. Systems, Man, and Cybernetics, vol 27, pp. 559-566, 1997
- [3] Reimar K. Lenz and Roger Y. Tsai, "Calibrating a Cartesian Robot with Eye-to-Hand Relationship", Proc IEEE Trans. Pattern Analysis and Machine Intelligence, vol 11, pp.916-928, 1989
- [4] Mengxiang Li and Jean-Marc Lavest, "Some Aspects of Zoom Lens Camera Calibration", Proc IEEE Trans. Pattern Analysis and Machine Intelligence, vol 18, pp.1105-1110, 1996
- [5] 우 동 민, "렌즈의 왜곡모델을 이용한 카메라 보정에 관한 연구", 전자공학과 논문지, vol.31B, no.2, pp170-182, 1994