

# 신경망을 이용한 카메라 보정에 관한 연구

정경필\*, 우동민\*\*, 박동철\*\*

\* 명지대 전기공학과, \*\* 명지대 제어계측공학과

## A Study on Camera Calibration Using Artificial Neural Network

KyongPil Jeong\*, DongMin Woo\*\*, DongChul Park\*\*

\* Electrical Eng. Myong-ji Univ., \*\* Control & Instrumentation Eng. Myong-ji Univ.

### Abstract

The objective of camera calibration is to obtain the correlation between camera image coordinate and 3-D real world coordinate. Most calibration methods are based on the camera model which consists of physical parameters of the camera like position, orientation, focal length, etc and in this case camera calibration means the process of computing those parameters. In this research, we suggest a new approach which must be very efficient because the artificial neural network(ANN) model implicitly contains all the physical parameters, some of which are very difficult to be estimated by the existing calibration methods. Implicit camera calibration which means the process of calibrating a camera without explicitly computing its physical parameters can be used for both 3-D measurement and generation of image coordinates. As training each calibration points having different height, we can find the perspective projection point. The point can be used for reconstructing 3-D real world coordinate having arbitrary height and computing image coordinate of arbitrary 3-D real world coordinate. Experimental comparison of our method with well-known Tsai's 2 stage method is made to verify the effectiveness of the proposed method.

### 1. 서론

카메라 보정(camera calibration)의 목적은 컴퓨터 영상의 점과 실세계(real world)상의 점간의 상관관계를 결정하는 과정으로서 정확한 보정이 이루어져야 컴퓨터 영상 좌표와 3차원 실세계 좌표간의 변환이 가능하므로 컴퓨터 영상 해석 분야에서는 매우 중요한 선행 과정이다. 일반적으로 카메라 보정은 카메라의 위치, 회전각, 초점거리, 왜곡변수등과 같은 카메라의 물리적인 변수들로 카메라 모델을 설정하여 이 변수들을 추정하는 방식[1]-[5]으로 연구되어 왔다. 그 중 Tsai의 2단계 방법(two-stage method)[5]은 대부분의 변수들을 선형 방정식에 의해 효율적으로 구하면서도 렌즈의 왜곡을 고려한 정확한 보정 방법이므로 널리 사용하고 있다. 그러나 이러한 방법들의 가장 큰 문제점은 렌즈의 왜곡에 대해 완전한 모델을 가정할 수 없으며, 영상의 중심 및 uncertainty factor와 같은 변수들을 구하는 데 어려움이 있다는 것이다.

본 연구에서는 기존의 방식에서 설정한 물리적인 카메라 모델 대신에 신경망 모델을 적용하는 새로운 방식을 채용하였다. 이러한 신경망을 이용한 방식은 카메라의 물리적인 변수들이 직접적으로 나타나지는 않으나, 간접적으로 신경망내에 내재되어 있는 것으로 볼 수 있다. 따라서 본 연구에서 실행한 카메라 보정의 결과로써 카메라의 물리적인 변수들은 구할 수는 없으나, 컴퓨터 영상

해석 응용 분야에서 카메라 보정의 목적은 이러한 실세계 좌표와 컴퓨터 영상 좌표상의 좌표 변환이므로 정확한 좌표 변환만 계산될 수 있으면 매우 효율적인 카메라 보정 방법이 될 수 있다. 또한 실 공간에서 서로 다른 높이의 평면을 각각 학습함으로써 영상 좌표의 한 점에 대해 자기 다른 높이의 두 실 공간 좌표를 추측하였다. 이러한 방법을 이용하여 영상의 중심을 구하였고, 완전한 3차원 재구성이 가능하게 되었을 뿐 아니라, 역변환을 통해 임의의 3차원 공간상의 좌표에 대한 2차원 영상 좌표의 예측이 가능하도록 시도하였다.

### II. 카메라 보정을 위한 신경망

#### 카메라 보정을 위한 신경망의 구조

카메라 보정을 위하여 서로 다른 높이의 균일한 간격을 갖는 점들에 대한 영상을 취득한다. 이렇게 취득된 영상으로부터 학습과 테스트를 위한 보정점들을 얻게 되면, 이 점들의 영상 좌표(U,V)가 그림 1의 다층구조 퍼셉트론 형태의 신경망의 입력이 된다. 또한 보정점들은 실제 공간상의 한 점을 기준으로 일정한 간격을 두고 존재하는 점들이므로, 각 보정점들의 실제 위치 (X,Y)가 목표값이 되어 신경망을 구성하게 되고, 역전파 학습 알고리즘에 의해 신경망을 학습하게 된다.

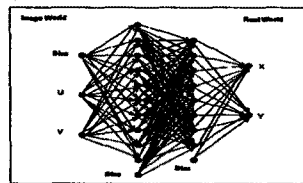


그림 1 사용된 다층구조 퍼셉트론 형태의 신경망  
Fig. 1 Example of multilayer perceptron type neural network

#### 영상의 중심

그림 2에서와 같이 임의의 영상 좌표 (u, v)에 대해 서로 다른 높이를 갖는 3차원 실 공간상에는 2개의 서로 다른 위치가 대응하게 된다. Z = 0인 보정면의 두 점의 위치를 각각 (x<sub>1</sub>, y<sub>1</sub>, 0), (x<sub>2</sub>, y<sub>2</sub>, 0)라 하고, 각 점들에 대해 Z=40인 높이의 보정면에 대응하는 점의 위치를 (x<sub>1</sub>, y<sub>1</sub>, 40), (x<sub>2</sub>, y<sub>2</sub>, 40)라 할 때 각각의 2점을 지나는 직선의 방정식은 다음의 식으로 표현될 수 있다.

$$\bar{P} = (x_1, y_1, z_1) + t(x'_1 - x_1, y'_1 - y_1, 40) \quad (1)$$

$$\bar{Q} = (x_2, y_2, z_2) + t(x'_2 - x_2, y'_2 - y_2, 40) \quad (2)$$

$$\bar{P} = \bar{Q} \quad (3)$$

식 (1), (2), (3)에 의해 구해진 직선의 방정식은 한 점에서 일치하게

되는데 이 점을 영상의 중점으로 볼 수 있다.

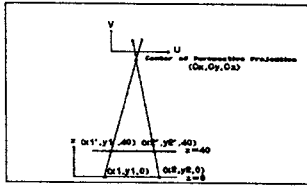


그림 2 영상의 중점

Fig. 2 The center of perspective projection

### 3차원 공간 재구성

그림 3에서와 같이 공간상에서 임의의 Z값에 대한  $(x,y,z)$ 가 있고 그 점에 대응하는 영상 좌표  $(u,v)$ 가 있다고 가정하자. 영상 좌표  $(u,v)$ 는 실 공간상의  $(x,y,z)$ 뿐만 아니라  $(x1,y1,0)$ 와  $(x2,y2,40)$ 과도 서로 대응하게 된다. 따라서 이들 세 점은 모두 일직선상에 존재하게 되므로  $(x1,y1,0)$ 와  $(x2,y2,40)$ , 그리고 영상의 중점, 세 점을 이용하여 직선의 방정식을 구하게 되면, 대응하는 임의의 Z값에서 좌표값을 알게 된다. 즉, 주어진 Z값에서 임의의 영상 좌표에 대응하는 실 공간상의 위치 예측이 가능하게 된다.

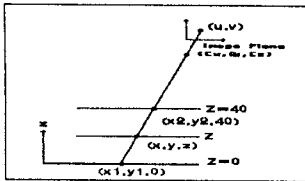


그림 3 3차원 공간 좌표 재구성

Fig. 3 3-D real world coordinate reconstruction

### 영상 좌표 예측

$Z = 0$ 과  $Z = 40$ 인 실 공간상의 두 보정면으로부터 얻은 정보를 이용하여 임의의 높이의 좌표에 대해 대응하는 영상 좌표를 예측할 수 있다. 그림 4에서와 같이 Z값은 다르지만 X와 Y 좌표값은 같은 공간상의 3개 점은 영상 좌표가 모두 다르게 나타난다. 따라서 임의의 높이의 점 p1은 영상의 중점과 직선의 방정식을 이룰 수 있는데, 이 직선의 방정식이 우리가 이미 알고 있는 공간상의 높이인  $Z = 0$  또는  $Z = 40$ 인 보정면과 교차하게 된다. 따라서, 그 직선의 방정식을 이용하여  $Z = 0$  또는  $Z = 40$ 인 보정면과의 교차점  $(x1,y1,0)$  또는  $(x1,y1,40)$ 을 구하게 되면 그 점에 대한 영상 좌표값  $(u1,v1)$  또는  $(u2,v2)$ 은 2장에서 구성한 다중 구조 퍼셉트론 타입의 신경망을 이용하여 예측할 수 있었으므로, 결과적으로 실 공간에서 임의의 높이를 갖는 좌표에 상응하는 영상 좌표 예측이 가능하다.

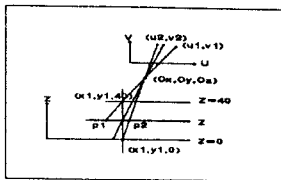


그림 4 영상 좌표 예측

Fig. 4 Image coordinate prediction

### III. 실험 결과

#### 시스템 구성 및 비교 방법

본 논문에서 제안한 카메라 보정 알고리즘의 타당성을 입증

하기 위한 실험 환경은 그림 5와 같다. UNIX운영 체제에서 동작하는 VME 버스 개발 시스템(VME 167)을 호스트로 하는 시스템에 영상 취득기(frame grabber)를 접속하고, LAN을 통해 실험과정의 모니터링과 사용자 인터페이스 및 전체 카메라 보정 알고리즘의 수행을 위한 워크스테이션(workstation)과 IBM PC를 연결하였다. PULNIX TM-7EX CCD카메라를 사용하여 카메라 보정을 위한 영상을 취득하였으며, 워크스테이션에서 보정점 취득 및 Tsai의 방법을 실험하였고, IBM PC에서 신경망의 학습과 테스트 및 3차원 재구성과 2차원 영상 좌표 예측에 관한 모든 실험을 수행하였다.

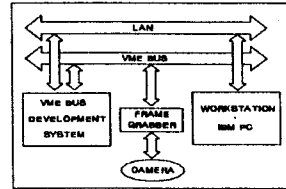


그림 5 시스템 구성도

Fig. 5 The schematic diagram of image processing system

보정을 위한 영상은 가로 25mm, 세로 20mm 간격으로 일정하게 배치되어 있고, 가로 11개, 세로 9개로 모두 99개의 보정점으로 이루어져 있으며, 각기 다른 세 위치( $z = 0, 20, 40$ )에서 취득되어졌고 신경망을 이용한 카메라 보정과 기존의 가장 보편화된 방식인 Tsai의 2단계 방법을 동시에 적용하여 비교, 분석하였다. 두 방식의 객관적 비교를 위해서는 보정의 정확도를 판별할 기준이 설정되어야 한다. 렌즈의 확대율과 무관하게 정확도를 규정하기 위하여 평균 픽셀 단위 오차 AEIP(Average Error In Pixel)[11]를 사용하였다.

#### 신경망의 학습 및 테스트

신경망의 학습을 위하여 각각 전체 보정점의 80%인 79개를 사용하였고, 테스트를 위하여 나머지 20개를 사용하였다. 제안된 방법의 객관적 평가를 위하여 Tsai의 2단계 방법과 비교 연구 하였다. Tsai의 방법은 카메라의 물리적 변수를 이미 알고 있는 3차원 실 공간상의 좌표와 영상 좌표와의 상관관계로써 구하게 되고 그 물리적 변수를 구하기 위하여 사용한 79개 보정점을 가지고 최종 오차를 구하는 방식이므로 신경망에서의 방법과는 차이가 있다. 따라서 Tsai의 방법에서는 79개 보정점을 이용하여 물리적 변수를 구하고, 그 변수를 이용하여 학습 오차를 구하였다. 그리고 앞에서 구한 물리적 변수를 이용하여 나머지 20개 보정점의 테스트 오차를 구하였다. 표 1은 두 방법을 실제 영상에서 취득한 데이터로 실험한 결과이고, 표 2는 입력값을 영상 좌표로 하고 목표값을 실 공간 좌표로 하여 학습과 테스트를 해 본 결과이다. 표 1의 결과를 보면 전체적으로 신경망을 이용한 방법이 더 우수함을 알 수 있다.

(단위 : 픽셀)

		Tsai의 2단계 방법	신경망을 이용한 방법
z=40인 영상	79개 학습 데이터	0.597062	0.413293
	20개 테스트 데이터	0.532232	0.503957
z=0인 영상	79개 학습 데이터	0.620149	0.479018
	20개 테스트 데이터	0.527716	0.521259

표 1 카메라 보정 오차의 비교

Table. 1 Comparison of The camera calibration error

(단위 : mm)

		신경망을 이용하여 계산한 결과
z=40인 영상	79개 학습 데이터	0.191225
	20개 테스트 데이터	0.283481
z=0인 영상	79개 학습 데이터	0.219389
	20개 테스트 데이터	0.409615

표 2 카메라 보정 오차

Table. 2 The error of camera calibration

3차원 실 공간 좌표 재구성 결과 및 오차

영상의 한 점을 선택하여 그 점에 대응하는  $Z = 0$ ,  $Z = 40$  인 실 공간상의 좌표를 구하고, 식 (1), (2), (3)을 이용하여 영상의 증점을 구하였다. 실제 실험하여 얻은 영상의 증점의 좌표는  $C_x = 556.123339$ ,  $C_y = 53.035429$ ,  $C_z = 634.266627$  이다. 임의의 영상 좌표 10개를 선택하여 그 영상 좌표에 대응하는  $Z = 0$ ,  $Z = 40$ 인 공간 좌표상의 대응점을 구하였다. 그리고 각각 10개의 직선의 방정식을 만들어  $Z = 20$ 인 면에서 3차원 공간 좌표를 예측해 보았다. 표 3은 10개의 3차원 공간 좌표에 대한 예측 결과이다.

(단위 mm)

Z = 20인 실 공간상의 좌표	신경망을 이용하여 계산한 결과	오차
(75,0)	(74.841134,1.62823)	1.172625
(175,0)	(173.893471,0.534052)	1.228665
(150,80)	(149.840561,62.177689)	2.183518
(100,80)	(100.637690,82.288989)	2.376070
(175,100)	(174.824252,101.696271)	1.897962
(225,100)	(224.902722,101.401265)	1.404638
(25,120)	(24.635060,122.020465)	2.053099
(150,140)	(149.911304,141.774392)	1.776906
(125,160)	(126.069663,158.623411)	1.743324
(175,160)	(175.395367,158.539569)	1.516487
평균 오차	1.715399	

표. 3 3차원 공간 좌표 재구성 오차

Table 3 3-D world coordinate reconstruction error

p

2차원 영상 좌표 예측 결과 및 오차

$Z = 20$ 인 면에서 10개의 임의의 점을 선택하여 그 점들과 앞에서 구한 영상의 증점을 있는 10개의 직선의 방정식을 구하여  $Z = 0$ 인 면과  $Z = 40$ 인 면과의 교차점을 구하였다. 그리고  $Z = 0$ 과  $Z = 40$ 인 면의 각 교차점에서 신경망을 이용하여 영상 좌표를 예측하였다. 예측이 끝나면 최종적으로 예측된 값과  $Z = 20$ 인 면을 취득한 영상에서 구한 실제 값과의 차이를 구하였다. 표 4는  $Z = 0$ 인 보정면과의 교차점으로부터 2차원 영상 좌표를 예측한 결과이고, 표 5는  $Z = 40$ 인 보정면과의 교차점으로부터 2차원 영상 좌표를 예측한 결과이다.

(단위 : 픽셀)

Z = 20인 보정면에 대한 영상 좌표	신경망에 의해 계산한 결과	오차
(204,83)	(204.769513,75.104715)	7.932697
(333,78)	(334.214268,71.244285)	6.863965
(126,134)	(126.158960,126.322497)	7.679145
(288,132)	(287.931383,125.650950)	6.349421
(163,232)	(163.347749,225.830374)	6.179419
(377,240)	(378.091826,236.667818)	3.506497
(286,289)	(285.316940,284.599820)	4.452882
(162,329)	(161.617908,325.186735)	3.833365
(329,345)	(329.653513,341.781785)	3.283899
(241,388)	(240.435836,384.827784)	3.221993
평균 오차(AEIP)	5.330328	

표. 4 2차원 영상 좌표 예측 오차( $Z = 0$ )

Table 4 2-D image world coordinate prediction error( $Z = 0$ )

(단위 : 픽셀)

Z = 20인 보정면에 대한 영상 좌표	신경망에 의해 계산한 결과	오차
(204,83)	(204.056000,75.358781)	7.641424
(333,78)	(334.644588,70.933576)	7.284497
(126,134)	(124.952924,127.837980)	6.447609
(288,132)	(287.896626,126.281966)	6.719949
(163,232)	(161.739498,228.103629)	4.096114
(377,240)	(377.509917,237.884484)	2.390534
(286,289)	(285.722879,287.759982)	1.270606
(162,329)	(160.282894,328.719465)	1.739892
(329,345)	(329.840383,346.242917)	1.500362
(241,388)	(241.305161,389.894324)	1.918746
평균 오차(AEIP)	4.000871	

표. 5 2차원 영상 좌표 예측 오차( $Z = 40$ )

Table 5 2-D image world coordinate prediction error( $Z = 40$ )

제안된 보정방법은 기존의 방법과 같이 정확한 카메라 모델을 설정하고 모델의 변수들을 계산하는 것 대신에 신경망 모델을 구성하고, 그 신경망 모델을 학습시킴으로써 카메라 보정을 하는 방법으로, 카메라 보정분야에서는 최초의 시도이다. 기존의 방법들은 렌즈의 왜곡을 근사화하고 왜곡 계수를 구하는데 어려움이 있을 뿐아니라, 영상의 증점, uncertainty factor와 같은 변수는 추정이 힘든 문제점이 있으나, 제안된 방법은 학습된 신경망 모델이 이러한 모든 요소를 내포하고 있으므로 정확도가 훨씬 높다. 뿐만아니라 전체적인 과정이 자동적으로 수행되어질 수 있고, 오차도 Tsai의 2단계 방법과 비교했을 때, 보다 더 작음을 알 수 있다. 학습할때에는 비교적 많은 데이터와 느린 수로 비교적 많은 시간이 소요되지만, 일단 한 번 학습이 이루어 지고 나면 모든 것이 실시간 처리되어질 수 있으므로, 실시간 처리를 요구하는 많은 분야에 적용될 수 있다. 또한 신경망을 이용한 방법은 기존의 방법에 비해 카메라의 위치나 조망등에서 많은 제약을 받지 않으므로 보다 융통적이라 할 수 있고, 카메라의 왜곡에 영향을 거의 받지 않으므로 보다 일반적이라 할 수 있다. 신경망을 이용한 방법에서는 기존의 방법에서 예측치를 사용했거나, 근사치를 사용한 부분이 따로 정의되지 않고 영상의 좌표와 실 공간상의 좌표를 서로 매칭하여 학습하는 방법이므로 카메라의 성능에 관계없이 좋은 결과를 나타낼 것이다.

참고문헌

- [1] W. Faig, "Calibration of close-range photogrammetry systems: Mathematical formulation," Photogrammetric Eng. Remote Sensing, vol.41, pp.1479-1486, 1975
- [2] I. Sobel, "On calibrating computer controlled cameras for perceiving 3-D scenes," Artificial Intell., vol.5, pp.185-188, 1974.
- [3] J. Y. Itah, A. Miyauchi and S. Ozawa, "Direct measuring method using only simple vision constructed for moving robots," Proc. 7th Int. Conf. on Pattern Recognition, vol.1, pp.192, 1984.
- [4] Y. Yakimovsky and R. Cunningham, "A system for extracting three-dimensional measurements from a stereo pair of TV cameras," Computer Graphics Image Processing, vol.7, pp.195-210, 1978.
- [5] R. Tsai, "An efficient and accurate camera calibration technique for 3-D machine vision," Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition, pp.364-374, 1986.
- [6] H. A. Martins, J. R. Birk, and R. B. Kelley, "Camera models based on data from two calibration planes," Comput. Graphics and Automat., 1988, pp.1407-1413.
- [7] R. Mohr and L. Morin, "Relative positioning from geometric invariants," in Proc. IEEE Conf. on Comput. Vision Pattern Recognit., Hawaii, June 1991, pp.139-144.
- [8] G-Q. Wei, Z. He, and S. D. Ma, "Fast moment calculation: Discrete form, analytic form and parallel implementation," Proc. Int. Conf. Comput. Aided Technol. Hong Kong, 1988, pp. 331-340.
- [9] O. D. Faugeras and G. Toscani, "The calibration problem for stereo," in Proc. IEEE Conf. Comput. Vision and Pattern Recognit., 1986, pp. 15-20.
- [10] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning internal representations by error propagation," Parallel Distributed processing, vol.2, pp.318-362, MIT Press, Cambridge, 1987
- [11] 우동민, "렌즈의 왜곡모델을 이용한 카메라 보정에 관한 연구," 전자공학회의논문지, 31권 B편 2호, 1994.