

모션캡처 시스템에서 3차원 데이터 획득의 정확도 개선을 위한 연구

한창호^o, 서평원, 김순철, 유영기, 오춘석

선문대학교 정보통신공학부

liberman@paran.com, {scvictor, ds3hoc}@naver.com, {ryu, csoh}@sunmoon.ac.kr

The study of improving the accuracy in the 3D data acquisition of Motion Capture System

Changho Han^o, Soonchul Kim, Pyoungwon Seo, Youngkee Ryu, and Choonsuk Oh

Department of Information and Communications, SunMoon University

요 약

기존 논문에서 CCD 카메라를 이용해 3차원 데이터를 획득하는 시스템을 소개했었다. 하지만, 정확도에서는 기존에 개발한 다른 PSD카메라를 이용한 시스템과 큰 차이를 보이지 않았다. 본 논문에서는 정확도를 개선하기 위한 여러 가지 실험을 했으며, 결과를 통해 상당히 향상되었음을 알 수 있다. 제안된 방법은 왜곡제거, z축 보정 등이며, 각 방법의 비교실험을 통해 얼마나 개선되었는지 살펴본다.

1. 서 론

3 차원 데이터를 획득하는 시스템은 모션캡처 시스템에 주로 이용되고 있으며, 현재 모션캡처의 기술은 날로 발전하고 있다. 가장 유명한 VICON 사의 모션캡처 카메라의 성능을 살펴보면 400 만 화소의 mx4 는 데이터 캡처속도가 2000fps 이며, 136,000 개의 마커를 인식하며, Static Calibration 의 정확도가 0.1mm 로 그 오차가 아주 작으며 6~7m 거리에서 2mm 마커를 캡처할 수 있을 정도의 해상력을 갖고 있다. 이러한 고성능 캡처 시스템으로 많은 영화에 사용되고 있다. 단점은 역대가 넘는다라는 것이다. 이에 따라 저렴한 모션캡처 장비를 위해 다양한 연구가 진행되고 있는 실정이다.

본 논문에서는 이전에 소개한 일반 CCD 카메라와 마커를 사용한 모션캡처 시스템을 살펴보고, 정밀도를 높이기 위한 주제로 왜곡과 z 축 보정에 관한 연구를 하였다. 그리고 기존의 시스템과 비교하여 실험결과를 보여주고, 새로 개선된 방법이 효과적임을 소개한다.

2. 기존 모션 캡처 시스템

기존 모션캡처 시스템을 살펴보면 먼저, 두 대의 CCD 카메라를 이용하여 실시간으로 움직이는 영상을 캡처하고, 전처리 과정을 거쳐 2차원 데이터를 추출하고, 그 후, 스테레오 기법을 이용하여 3차원 데이터를 구성하게 된다. 이렇게 얻어진 3차원 데이터는 후에 모션인식에서 사용된다.

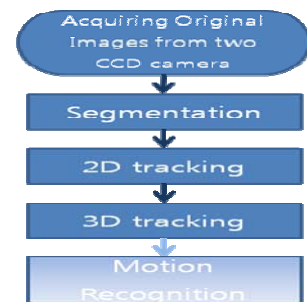


그림 1 CCD 카메라 모션캡처 시스템 흐름도

2.1 전처리 과정

입력 받은 두 개의 영상으로부터 마커의 신호를 분류해 내기 위해서는 전처리 과정이 필요하다. 현재 사용된 카메라는 적외선을 이용하지 않고 가시광선을 이용하므로 주변광의 영향이 크다고 할 수 있다. 따라서 입력 받은 영상에 대해 주의할 수 밖에 없다. 다음 그림에서 보듯이 초기 입력 영상은 약간 어둡게 해서 획득하게 된다.



그림 2 두 개의 원 입력 영상

이렇게 획득된 영상은 다음 그림 3과 같은 전처리 과정을 거쳐 각각 두 개의 2차원 데이터를 구하게 된다. 전처리 과정에는 간단하게 설명하면, 이진화 방법을 사용했으며, 구해진 마커의 점들은 각각 크기로 구별된다. 현재 권투 모션인식 시스템에 연결하여 작업을 수행하도록 되어 있기 때문에 네 개의 마커 신호를 사용하고 있으며, 전체 몸의 애니메이션 데이터를 획득 하기 위해서 더 많은 수의 마커를 필요로 한다면, 서로 대응하는 마커의 위치를 구하는 알고리즘을 추가해야 할 것이다.

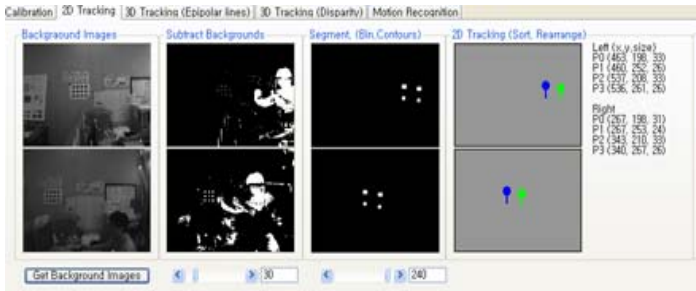


그림 3 전처리 과정 및 2차원 데이터 획득 프로그램

본 시스템에서 사용된 마커는 다음 그림과 같다. 가시광선을 사용하기 위해 일반 LED를 사용했으며, 전원 장치를 갖고 있다. 빛의 눈부심을 없애기 위해 확산판을 사용했으며, 대체로 인식 가능 각도는 약 125도 이내이다.



그림 4 마커의 앞, 뒤 모습

2.2 캘리브레이션

두 대의 CCD 카메라는 거리를 정확히 계산하여 일정 거리를 유지 했으며, 카메라의 평행을 유지하도록 시스템을 꾸렸다. 또한 사용된 캘리브레이션 패턴은 그림과 같은 8x6체스 보드[8]를 이용했다. 캘리브레이션 입력 및 계산을 위해 OpenCV 라이브러리를 이용해 개발했다.

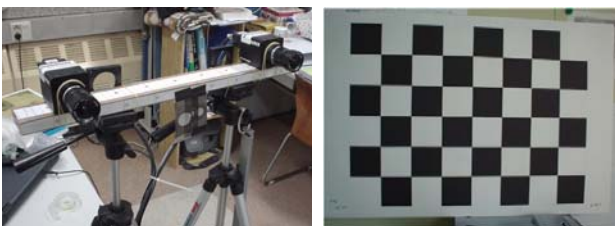


그림 5. 카메라 시스템과 캘리브레이션 패턴

두 대의 카메라의 평행을 설정하기 위해 최단 거리에서 체스보드 사이의 간격을 확인하고, 먼 거리에서 다시 체스보드 사이의 거리를 측정해서 일치하는가로 평행함을 증명하였다. 물론 캘리브레이션 실험 자체의 오차를 고려해야 할 것이다. 그림에서 보듯이 왼쪽영상의 체스보드의 중심 위치와 오른쪽 영상의 체스보드 중심 위치를 찾아내어, 실제 체스보드의 간격과 카메라의 간격을 비교하면 평행한지 알 수 있다. 측정 거리가 멀면 멀수록 더욱 정밀한 측정이 가능하다.

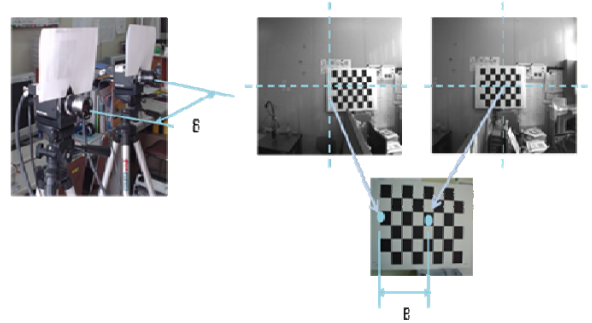


그림 6 카메라 평행 유지 방법에 대한 그림

2.3 3차원 데이터 획득

간단히 카메라 캘리브레이션을 마치고 나서 두 영상으로부터 얻은 데이터는 다시 전처리 과정을 통해 2차원 데이터를 얻게 된다. 얻어진 2차원 데이터는 시각차 (disparity) 알고리즘[1-7]을 통해 3차원 데이터로 재구성되는데, 이때 사용된 2차원 정보는 체스보드를 통해 얻게 된다. 이때는 캘리브레이션 테스트를 위해 전처리 과정을 하지 않고, 직접 체스 보드로부터 꼭지점 데이터를 얻게 된다. 체스 보드로부터 얻는 데이터는 총 7x5, 35개의 데이터가 영상으로부터 얻어지며, 다음과 같은 알고리즘을 통해 3차원 데이터로 바뀌며[9,10], 이 데이터는 실제좌표인 월드좌표와 비교하여 오차를 계산하게 된다.

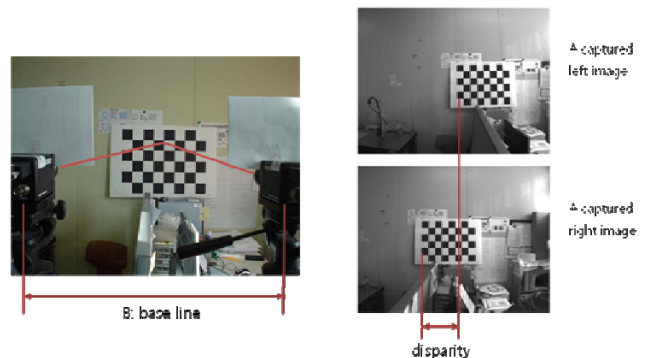


그림 7 시각차 측정 방법에 대한 그림

시각차는 스테레오 카메라의 영상으로부터 동일한 위치의 점들의 차이를 가지고 z값인 거리를 계산하는 알고리즘이다. 시각차가 클수록 거리는 가까워지고, 시각차가 줄수록 거리는 멀어진다고 할 수 있다.

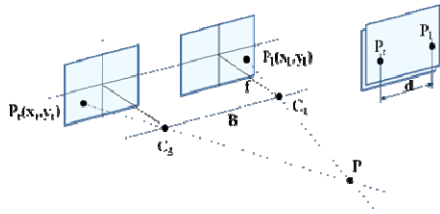


그림 8 시각차 측정 알고리즘

- b : baseline
- f : focal length
- x_l, x_r : x, y position on an CCD array
- X_l, X_r : x, y position on an image
- N_{fx} : number of pixels of CCD camera array
- L_x : length of CCD camera array
- Z: depth = $f*b / |x_l - x_r| = f*b*N_{fx} / (X_l - X_r) * L_x$

위 식에서 z는 거리를 의미하며, 카메라 사이의 베이스라인 b에 카메라 초점거리를 곱하고, 영상의 해당 위치의 시각차를 나눔으로 거리가 계산된다.

3. 정확도를 개선하기 위한 알고리즘

이번 논문에서는 거리측정의 오차를 줄이기 위한 방법으로 두 가지를 제안했다. 첫 번째가 카메라의 왜곡을 줄이는 방법이다. 그리고, 두 번째가 z축을 효율적으로 줄이는 방법을 제안했다.

3.1 왜곡제거

왜곡모델은 가장 간단한 기본 모델식[11]을 이용하여 구현했다. 왜곡 계수를 구하기 위해 OpenCV 를 사용하여 구했으며, 각 상수는 $k1=-0.24574789$, $k2=0.11125672$, $p1=0.00098713441$, $p2=-0.0000735978$ 이다

$$x_u = (1 + \sum_{i=1}^{\infty} K_i r^{2i})x_d + (2P_1 \bar{x}\bar{y} + P_2(r^2 + 2\bar{x}^2))(1 + \sum_{i=1}^{\infty} P_{i+2} r^{2i}) \quad (1)$$

$$y_u = (1 + \sum_{i=1}^{\infty} K_i r^{2i})y_d + (P_1(r^2 + 2\bar{y}^2) + 2P_2 \bar{x}\bar{y})(1 + \sum_{i=1}^{\infty} P_{i+2} r^{2i})$$

$$x_u = (1 + K_1 r^2 + k_2 r^4)x_d + (2P_1 \bar{x}\bar{y} + P_2(r^2 + 2\bar{x}^2)) \quad (2)$$

$$y_u = (1 + K_1 r^2 + k_2 r^4)y_d + (P_1(r^2 + 2\bar{y}^2) + 2P_2 \bar{x}\bar{y})$$

여기서 $\bar{x} = x_d - C_x$, $\bar{y} = y_d - C_y$, $r^2 = \bar{y}^2 + \bar{x}^2$. C_x 와 C_y 는 광학 중심이고, K_i 와 P_i 는 radial 과 tangential 왜곡 파라미터이다.

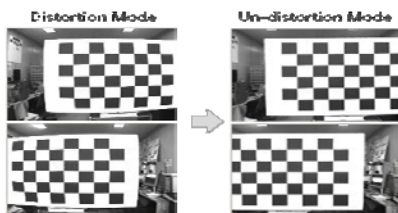


그림 9 왜곡 제거 영상

3.2 z축 보정

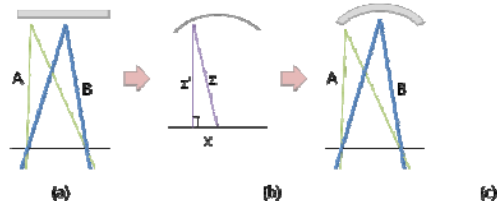


그림 10 z 축 보정 그림

$$z' = \sqrt{z^2 - x^2} \quad (3)$$

$$z' = \sqrt{z^2 - x^2 - y^2}$$

두 번째 방법으로 z축 보정을 적용했는데, 기존 시각차 이론에 거리에 대해 포물선 공식을 적용하여 거리를 줄였으며, 또한 z축 거리 자체의 스케일 조절을 하였다.

표 1 왜곡에 따른 z축 측정거리

World Z	Distortion		Un-Distortion	
	Z	Scale	Z	Scale
750	772.6346	0.970705	731.988	1.024607
1000	1003.642	0.996371	972.1869	1.028609
1250	1234.787	1.01232	1209.307	1.033649
1500	1467.691	1.022013	1444.041	1.038751
1750	1702.709	1.027774	1684.413	1.038938
2000	1952.764	1.024189	1934.106	1.03407
2250	2170.102	1.036817	2155.116	1.044027
avr		1.012884		1.034664

4. 실험

실험에서는 3가지로 나누어 했다. 첫 번째는 왜곡을 제거한 후 월드좌표와 비교하여 에러가 얼마나 발생하는지를 검사했으며, 두 번째는 z축 보정을 한 후, 세 번째는 왜곡과 z 축 보정을 동시에 한 경우에 대해 비교 실험을 했다.

표 2 각 조건별 모션 캡처 정확도 에러

Dist(mm)	Removing Distortion			Adjusting Distance			Removing and Adjusting		
	x	y	z	x	y	z	x	y	z
750	9.36	5.44	25.98	18.18	14.44	30.72	5.92	3.86	87
1000	4.05	5.61	40.35	8.21	10.71	26.24	4.5	4.56	64.42
1250	5.74	5	57.5	8.1	7.5	28.18	5.52	5.24	66.1
1500	5.18	4.69	64.77	8.7	6.33	38.61	4.16	8.14	59.76
1750	5.41	5.94	78.04	6.15	6.81	45.27	6.6	7.99	50.24
2000	9.71	4.17	91.99	10.83	5.67	49.77	5.89	7.11	72.28
2250	9.67	7	116.04	11	7.19	72.3	5.46	7.39	66
max	9.71	7	116.04	18.18	14.44	72.3	6.6	8.14	87

표 3 기존 시스템의 오차 결과 데이터와 비교

오차 거리(mm)	기존 시스템	왜곡제거	Z 축 보정	왜곡제거 +z 축보정
750	96	26	30	87
1000	79	40	26	64
1250	96	58	28	66
1500	123	65	38	60
1750	147	78	45	50
2000	164	92	49	72
2250	179	116	72	66

z축 보정한 영상의 결과 데이터와 기존 시스템과 비교하면 최대 3배 이상의 정확도가 나타남을 표를 통해 알 수 있다. 캘리브레이션의 오차를 제거한다면 에러를 좀 더 줄일 수 있을 것이다. 하지만, 750~1250mm의 일정 거리 내에서는 오차가 20~30mm 정도의 오차를 갖는 것을 알 수 있다. 이 정도의 오차를 감안하여도 체감형 게임, 또는 애니메이션 [12,13] 등의 모션 캡처 시스템으로 사용하기에는 불편함이 없을 것이다.

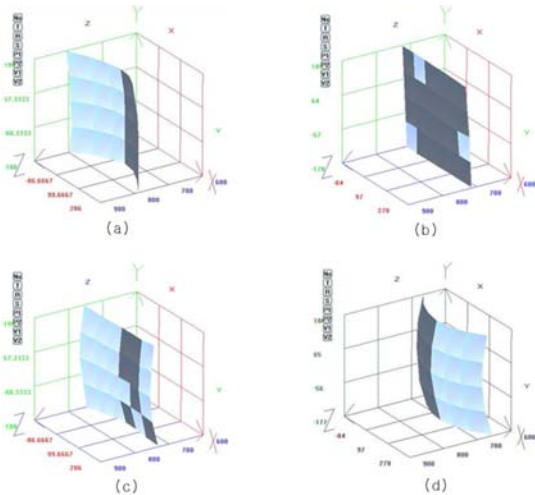


그림 11. 체스보드를 입력받은 3차원 영상; (a) 기존 시스템, (b) 왜곡제거를 한 3차원 영상, (c) z축 보정을 한 3차원 영상, (d) 왜곡제거와 z축 보정을 한 3차원 영상

5. 결론

본 논문에서는 기존 모션캡처 시스템의 정확도를 향상 시키고자 두 가지 방법을 이용하여 실험을 하였다. 왜곡 제거, z축 보정 방법을 이용하여 각각 정확도를 측정하여, 기존 시스템과의 비교를 해봤다. 위에 나온 결과 같이 3배 이상의 정확도가 향상 됨을 보였다. 하지만, 현재 사용한 CCD 카메라는 일반 가시광선으로 자연광의 영향을 많이 받게 되어 실험에 제한적이었다. 이러한 단점을 해결하고자, 다음 연구에서는 적외선 카메라를 이용하여 정밀 측정을 할 예정이다.

감사의 글

본 연구는 교육과학기술부와 한국산업기술재단의 지역혁신 인력양성사업으로 수행된 연구결과입니다.

References

[1] Trucco E., Verri A. 1998. Introductory Techniques for 3-D Computer Vision, Prentice-Hall (eds).

[2] Lucas B., Kanade T. "An Iterative Image Registration Technique with an Application to Stereo Vision", in *Proc. of 7th International Joint Conference on Artificial Intelligence (IJCAI)*, 1981, pp. 674-679.

[3] Gilbert S'ébastien, Laganière Robert, "Registration of a Moving Rigid Object Using a Stereoscopic Vision Setup", in *Proc. of the 3rd IEEE International Workshop on Haptic, Audio and Visual Environments and their Applications - HAVE 2004*, 2004, pp. 171-175.

[4] K.Kim, W.Woo, "3D Camera Tracking from Disparity Images," *VCIP*, pp. 1381-1388, 2005.

[5] Y.K.Yu, K.H.Wong, S.H.Or and M.M.Y.Chang, "Recursive recovery of position and orientation from stereo image sequences without three-dimensional structures", in *Proc. IEEE CVPR*, New York, Jun. 2006.

[6] Kanade, T. 1994. "Development of a video-rate stereo machine", In *Proc. of ARPA Image Understanding Workshop*. Monterey, CA, 549-558.

[7] C. L. Zitnick and T. Kanade, "A volumetric iterative approach to stereo matching and occlusion detection," *CMU Technical Report CMU-RI-98-30*, 1998.

[8] Chang Shu, Alan Brunton, and Mark Fiala. Automatic grid finding in calibration patterns using Delaunay triangulation. Technical Report NRC-46497/ERB-1104, National Research Council of Canada, Institute for Information Technology, 2003.

[9] Hartley, R. and Zisserman, A. 2000. Multiple view geometry in computer vision. Cambridge University Press, Cambridge, UK.

[10] DiFranco, D., Cham, T., and Rehg, J. 2001. Reconstruction of 3-D figure motion from 2-D correspondences, In *Proc. of Computer Vision and Pattern Recognition*, Kauai, HI, vol. 1: 307-314.

[11] D. Brown, "Decentering distortion of lenses," *Photogrammetric Eng.*, vol. 32, no. 3, pp. 444-462, 1966.

[12] W. Frey, M. Zyda, R. McGhee and W. Cockayne, "Off-The-Shelf, Real-Time, Human Body Motion Capture for Synthetic Environments," Technical Report NPSCS-96-003, Computer Science Dept., Naval Postgraduate School, Monterey, California 93943-5118, USA.

[13] Chai, J., Hodgins, and J. K., "Performance Animation from Low-dimensional Control Signals," *ACM Transactions on Graphics (SIGGRAPH 2005)*.