

다중 카메라를 이용한 3차원 움직임 복원 및 분석에 관한 연구

윤 상 민*

1. 서 론

카메라를 이용한 물체의 3차원 복원(reconstruction) 및 움직임 분석과 관련된 연구는 컴퓨터 비전 및 컴퓨터 그래픽스 분야에서 매우 중요한 역할을 담당해왔다. 특히, 사람의 움직임 분석 및 복원에 관한 연구는 영화, 게임, 스포츠, 의료 분야 등의 다양한 분야에서 응용되고 있다.

카메라를 이용한 사람의 3차원 움직임 분석 및 복원과 관련된 연구는 크게 사람의 몸에 마커(Marker)를 부착하는 방법과 마커를 사용하지 않는 방법으로 나눌 수 있다. 마커를 이용하여 사람의 움직임을 분석하는 방법은 매우 정확하고 빠르게 사람의 움직임을 분석할 수 있다는 장점이 있지만, 비싼 장비 및 환경이 필요하다는 단점을 가지고 있다. 특히, 사람의 몸에 마커를 부착하고, 부착된 마커를 정확하게 추적하는 실험 과정은 자연스러운 사용자의 움직임 분석에 방해된다. 이러한 과정은 인간과 컴퓨터 상호작용(Human Computer Interaction) 측면에서, 사용자에게 매우 불편할 뿐만 아니라 정확한 사람의 움직임을 분석하기 어렵게 만든다. 이에 반하여, 마커를 부

착하지 않고 사람의 움직임을 분석하는 방법은 사용자의 움직임이 마커를 이용하는 방법에 비해 더 자유스럽고 특별한 장치도 필요로 하지 않는다는 장점이 있지만 자동으로 사람의 특징점 추출 및 분석등의 기술을 필요로 한다.

본 논문에서는 마커를 이용하지 않고 실시간으로 사람의 움직임을 복원 및 분석하기 위한 시스템을 제안한다. 3차원으로 복원된 사용자의 모습을 분석하기 위해서, 제안하는 시스템은 크게 두 부분으로 나눌 수가 있다. 첫번째로, 다중의 카메라를 통해 입력된 영상을 통해 사용자의 모습을 3차원으로 복원하는 방법과, 두번째로, 사람의 움직임을 분석하여 인터랙티브(interactive)하게 사용자가 쉽게 볼 수 있도록 표현하는 방법이다.

그림 1은 본고에서 설명하는 다중 카메라를 이용한 사람의 3차원 복원 및 움직임 분석의 순서도이다. 각각의 입력된 영상으로부터 얻은 사람의 실루엣 정보를 이용하여 3차원 복원 및 움직임을 분석한다.

본고는 다음과 같이 구성된다. 2장에서는 3차원 복원 및 움직임 분석과 관련된 배경 지식에 대하여 설명하고, 3장에서는 사람의 움직임 복원 방법과 복원된 물체 분석을 위한 구체적인 방법에 대해 설명한다. 4장에서는 제안하는 기법의 실험 결과를 설명한다. 마지막으로 5장에서는 결론을 맺는다.

* 교신저자(Corresponding Author): 윤상민, 주소: Fraunhoferstrasse 5, Darmstadt, 64283, Germany, 전화: +49-6151-155-618, FAX: +49-6151-155-669, E-mail: sangmin.yoon@gris.informatik.tu-darmstadt.de

* TU Darmstadt Informtaik

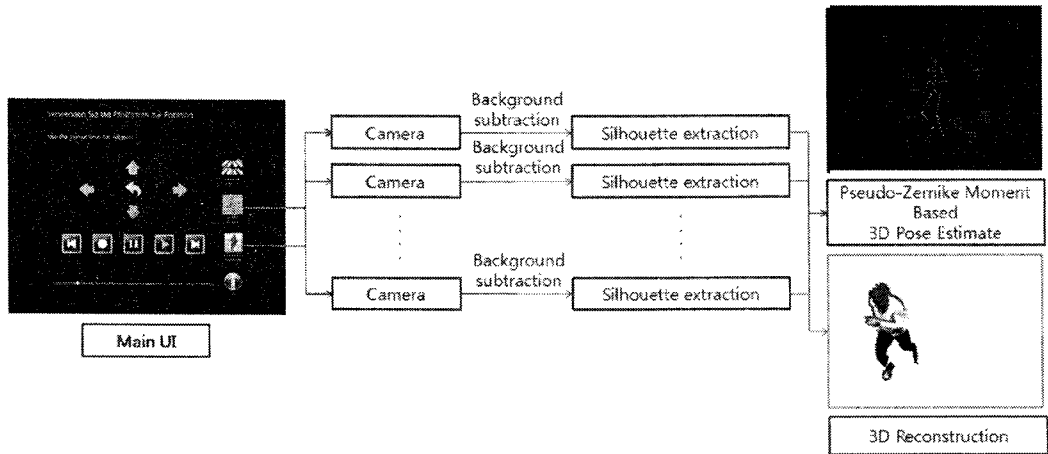


그림 1. 사람의 움직임 복원 및 분석을 위한 순서도

2. 배경 지식

고대 시대부터 이어져 온 사람의 움직임 분석과 관련된 연구는 대부분 의사, 예술가 등의 주도 하에 이루어 졌다. 이후, 카메라를 이용한 사물의 움직임 분석에 대한 연구는 1887년 Edward Muybridge의 “Animal Locomotion”에서부터 시작되었다. 그는 다수의 카메라를 통해 얻어진 사진을 통해 다양한 물체의 움직임을 분석하였다. 최근에는 카메라 및 GPU (Graphic Processing Unit) 기술의 발달에 힘입어 컴퓨터 비전 및 컴퓨터 그래픽스 연구자들의 주도하여 연구가 이루어지고 있다. 본 장에서는 컴퓨터 비전 및 컴퓨터 그래픽 분야에서 다중 카메라를 이용한 3차원 복원 방법과 모델을 기반으로 한 3차원 움직임 분석과 관련된 기존의 연구 결과를 고찰하도록 한다.

2.1 다중 카메라를 이용한 3차원 복원

3차원 물체 복원에 대한 연구는 스테레오 카메라를 이용한 3차원 복원에서부터 시작되었다[1]. 하지만 두 대의 카메라를 이용한 3차원 복원은 부분 폐색현상 (partial occlusion) 등의 궁극적인

문제를 해결할 수 없기 때문에 여러 대의 카메라를 이용한 3차원 복원 방법이 제안되었다[2]. 각각의 카메라에서 얻어진 깊이 정보를 정합하여 3차원 복원하는 방법이 제안되었다[3]. 최근에는, voxel coloring 및 Image Based Visual Hull [IBVH] 방법이 제안되었다[4,5]. GPU 기반으로 한 voxel coloring 및 IBVH 방법을 이용하여 실시간 3차원 복원이 가능하게 되었다.

2.2 3차원 포즈 복원

여러 대의 카메라를 통해 얻어진 영상을 이용하여 사람의 3차원 포즈를 예측을 위해 많은 방법들이 제안되었다. 특히 학습을 통한 3차원 복원과 관련된 연구는 크게 사람의 전체 실루엣 정보를 이용하는 방법과 머리, 몸통, 팔, 다리 등 각 부분으로 나누어진 사람의 신체적 특성을 이용하는 방법으로 나눌 수 있다.

본고에서 사용된 Pseudo-Zernike Moment (PZM) 방법을 이용한 사람의 움직임 분석 방법은 기존에 필기체 인식 및 얼굴 표정 연구 등에 응용되었다[6,7]. Teh 및 Chin 은 복잡한 자유도 (degree of freedom)을 갖는 물체를 분석하기 위

하여, Regular moment, Legendre moment, Zernike moment, Pseudo-Zernike moment 등을 비교 분석하였다[8]. Pseudo-Zernike moment를 이용한 움직임 분석 방법이 다른 방법에 비해 성능이 우수하다는 것을 증명하였다[9-11].

3. 제안하는 방법

본 장에서는 여러 대의 카메라를 이용한 사람의 3차원 복원 및 움직임 분석을 위해서 사람의 위치 추적 (tracking)을 통한 3차원 복원 및 학습을 통한 3차원 포즈 복원 방법에 대하여 제안 한다.

3.1 다중 카메라를 이용한 3차원 복원 방법

본 장에서는 2장에서 언급한 voxel coloring 및 image based visual hull (IBVH)의 결점을 보완하기 위해서 물체를 추적을 통한 3차원 복원 방법을 제안한다. 물체의 3차원 복원을 위해서 먼저 카메라 캘리브레이션 (calibration) 정보 및 사람의 실루엣 (silhouette) 검출이 필요하다. 본 논문에서는 실루엣 정보를 얻기 위해 다중 커널을 이용한 배경 추출 (background subtraction) 방법을 이용하였다[13].

기존의 voxel coloring 및 IBVH 방법은 카메라의 수 및 환경에 따라 성능 및 처리 시간이 결정된다. 특히, 넓은 환경에서 적은 수의 카메라를 이용하여 3차원으로 복원하는 경우에는 제안하는 방법이 기존의 방법에 비해 처리 시간 및 성능에서 우월하다. 왜냐하면 물체의 추적을 통해 얻어진 물체의 3차원 후보지역 내에서 3차원 복원을 하기 때문에 전 3차원 공간에 대하여 체크할 필요가 없다.

제안하는 3차원 복원 방법은 그림 2와 같다. N 개의 카메라를 이용하여 물체를 3차원 복원한

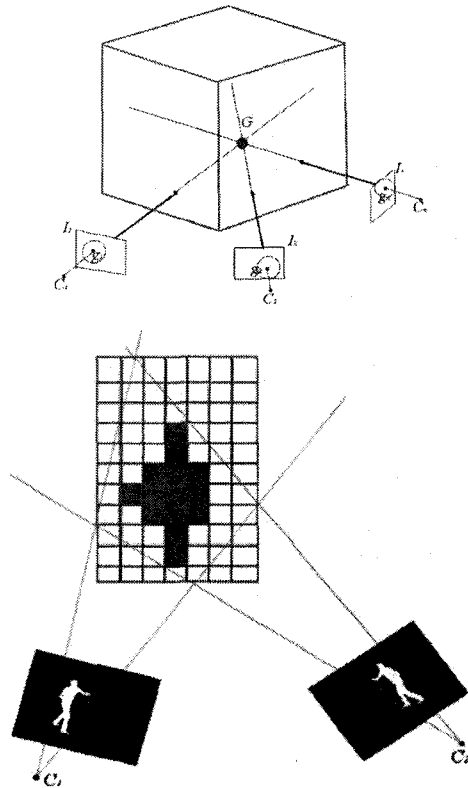
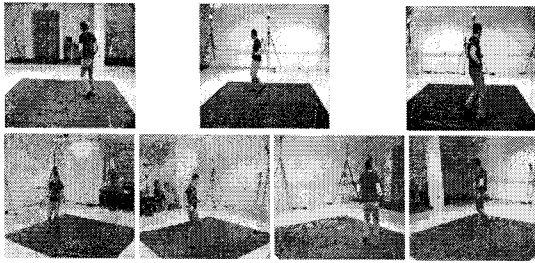


그림 2. 실시간 물체 추적을 통한 사람의 움직임 분석 개념도

다고 가정했을 때 각 카메라를 통해 얻어진 사람의 무게 중심은 g_1, g_2, \dots, g_N 으로 표현된다. 무게 중심과 inverse perspective projection matrix를 이용하여 3차원 공간에서의 G 값을 구한다. 추적된 G 값과 각 영상에서 얻어진 사람의 convex hull 을 이용하여 3차원 공간에서 사람이 위치할 후보지역을 설정한다. 3차원 후보지역 안의 어떤 복셀이 복원하고자 하는 물체와 관련되어 있는지 아닌지는 복셀들이 각 영상에 투영된 칼라 값의 편차를 이용한다. 각 영상에 투영된 칼라 값이 일정 문턱값 (threshold)을 넘는 경우에는 후보 지역 안에 있는 복셀들을 깎아낸다.

이러한 과정을 거쳐 얻어진 결과는 그림 3과 같다. 그림 3-(a)은 미국 Brown 대학에서 제공하



(a) 입력 영상



(b) 복원된 후보지역 및 최종 복원된 모습

그림 3. 제안하는 사람 추적을 통한 3차원 복원 방법

는 HumanEva 데이터로 3개의 영상은 칼라 영상으로 4개의 영상은 흑백 영상을 제공한다. 입력 영상으로부터 사람의 실루엣을 추출한 후에 제안하는 3차원 복원 방법을 적용하여 추출한 후보지역과 최종적으로 얻어진 3차원 사람 모델은 그림 3-(b)와 같다.

3.2 모델 기반 3차원 포즈 분석 방법

본 장에서는 여러 대의 카메라로부터 추출된 사람의 실루엣 정보를 3D 모델에 매칭하여 포즈를 분석하는 방법을 사용하였다. 3D 모델을 기반으로 한 여러 가지 포즈 분석 방법 중에 Pseudo-Zernike Moment (PZM) 방법을 이용한 3차원 복원 방법을 사용하였다[12].

최적의 3차원 포즈를 찾기 위하여 미리 계산된 카메라 캘리브레이션 정보와 미리 저장된 다수의 실루엣 정보를 이용한다. 여러 대의 카메라를 통

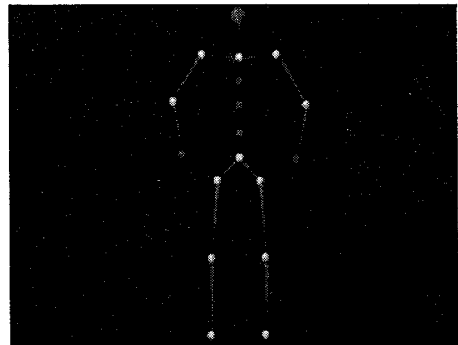


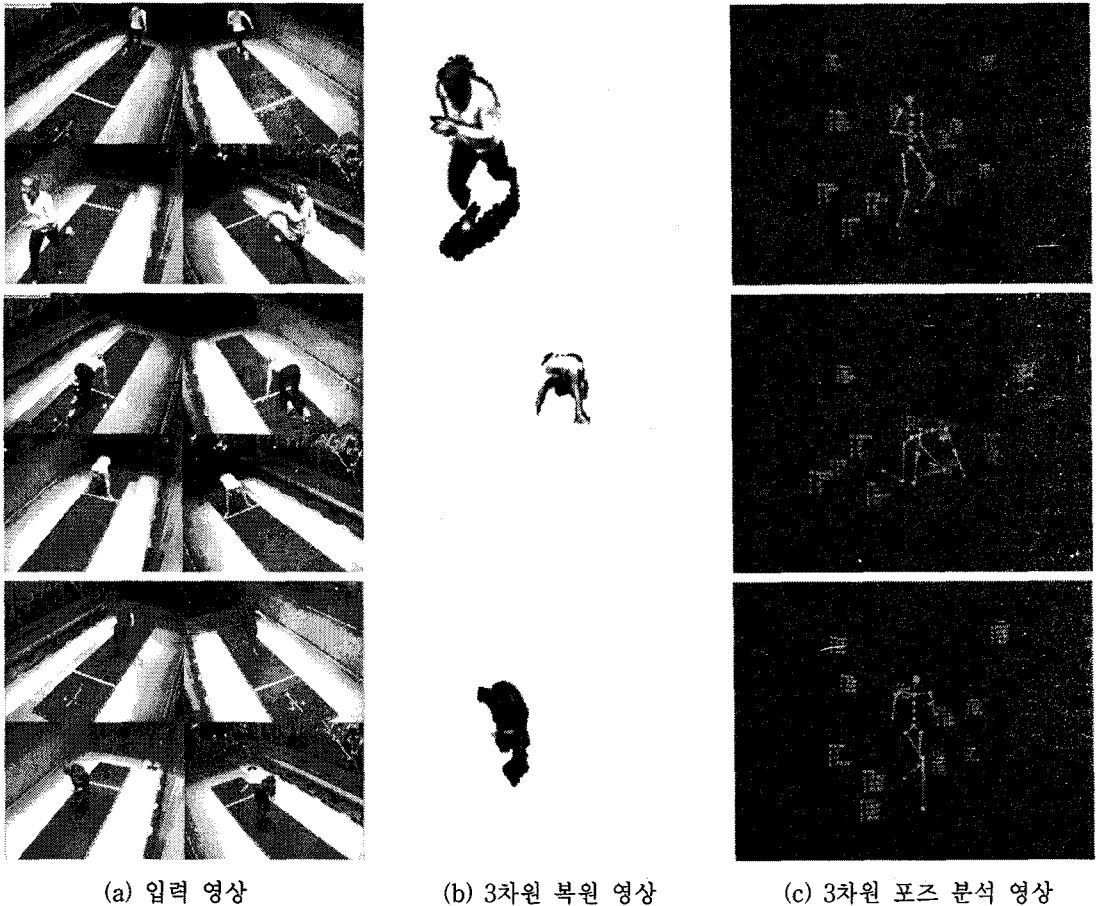
그림 4. 3차원 포즈 분석을 위한 아바타의 예

해 얻어진 사람의 실루엣 정보와 미리 저장된 사람의 실루엣과의 오차가 가장 적은 포즈를 통해서 3차원 포즈를 예측하도록 한다.

실시간으로 3차원의 포즈를 얻기 위해서 다음과 같은 세 가지의 조건이 필요하다. 첫 번째로, 사람의 포즈를 분석하기 위한 사람의 중요 관절 부위를 최소화한다. 보통 사람의 몸은 200여개의 관절로 이루어져있지만 실시간으로 사람의 움직임을 분석하기 위해서는 모든 관절 부분을 복원할 필요가 없기 때문에 중요한 최소한의 관절 부분만을 복원하였다. 본 논문에서는 총 32도의 자유도 (degree of freedom)를 갖는 15개의 관절의 움직임을 복원하였다. 두 번째로 사람이 취할 수 있는 포즈를 미리 정의하여 실시간으로 입력되는 영상으로부터 빨리 3차원 포즈를 예측하도록 한다. 총 110개의 포즈를 여러 개의 뷰(view)에 따라 분류하여 복원하였다. 마지막으로 복원된 사람의 포즈를 효율적으로 표현하기 위해서 그림 4와 같은 아바타를 사용하였다. 아바타의 각 관절은 예측된 움직임 값에 의해서 표현된다.

4. 실험결과

본 장에서는 제안하는 3차원 복원 및 움직임 분석에 관하여 설명하였다. 본 논문에서 제안하는



(a) 입력 영상

(b) 3차원 복원 영상

(c) 3차원 포즈 분석 영상

그림 5. 사람의 움직임 복원 및 분석의 예

방법은 독일 파더본의 Heinz-Nixdorf 박물관에 설치되어 사용자가 직접 복원된 자신의 움직임을 분석할 수 있도록 전시되었다. 두 대의 모니터를 설치하여, 한 대는 사용자가 원하는 뷰를 자유 자재로 바꿀 수 있도록 하였으며, 다른 모니터는 조작된 뷰에 따라 복원된 모습 및 포즈를 볼 수 있게 하였다. 3차원 복원된 사람은 128×128×128의 복셀 크기를 통해 표현된다.

본 실험을 위해 4개의 IEEE 1394 카메라를 펜티엄 4 1.2GHz 및 NVIDIA Geforce 8200이 내장된 컴퓨터를 사용하여 구현하였다. 실시간으로 3차원으로 물체를 복원하고 움직임 분석하기 위해

이와 관련된 배경 처리, 3차원 복원 및 포즈 분석 등의 모든 모듈(Module)은 GPU를 이용하여 병렬 처리되었다.

그림 5는 4대의 카메라를 통해 얻어진 칼라 영상, 3차원 물체 추적을 통한 3차원 복원, 및 움직임 분석에 대한 예를 보여준다. 사용자는 터치 스크린을 통해 자신이 원하는 뷰에서 3차원 복원된 자신의 모습과 포즈를 볼 수 있다.

5. 결 론

본고에서는 다중 카메라를 이용한 사람의 움직

입 복원 및 분석에 대해 설명하였다. 본고에서 제안하는 물체의 3차원 복원 및 움직임 분석 방법은 3차원 공간에서 물체를 계속적으로 추적함으로써 기존의 3차원 복원 방법에 비해 더 효율적으로 고해상도의 복원된 물체를 얻을 수 있다는 장점이 있다.

배경 추출 과정 중에 추출된 그림자 영역은 사람의 움직임을 분석하는 과정에서 많은 오차가 발생한다. 좀 더 정확하게 3차원 움직임을 분석하기 위해서는 환경에 영향을 강인한 배경 추출 및 사물의 실루엣 추출 방법이 필요하다. 사람 뿐만 아니라 다양한 물체의 움직임 분석을 위한 방법 또한 필요하다.

참 고 문 헌

[1] D. C. Marr, and T. A. Poggio, "A computation theory of human stereo vision," In proceeding of the Royal Society of London, 1979.

[2] M. Okutomi, and T. Kanade, "A multiple-based stereo," IEEE PAMI, pages, 353-363, 1993.

[3] S. B. Kang, R. Szeliski, and J. Chai, "Handling occlusion in dense multi-view stereo," Computer Vision and Pattern Recognition, 2001.

[4] W. Matusik, C. Buehler, S. J. Gortler, and L. McMillan, "Image-based visual hulls," ACM SIGGRAPH, 2000.

[5] S. M. Saitz and C. M. Dyer, "Photorealistic scene reconstruction by voxel carving," Computer Vision and Pattern Recognition, 1997.

[6] N.A. Zaidi, and N. M. Shiekh, "Character recognition using statistical parameter," SSIP, 2006. Shen, J. Mark, and A. Alfa, "Semi.

[7] J. Hoey, and J. J. Little, "Representation and recognition of complex human motion," Computer Vision and Pattern Recognition, 2000.

[8] C. H. Teh, and R. T. Chin, "On image analysis by the methods of moments," IEEE PAMI, 1998.

[9] A. B. Ahatia, and E. Wolf, "On the circle polynomials of zernike and related orthogonal sets," Philosophical Society of Cambridge, 1954.

[10] F. Zernike, "Beugungstheorie des schneidverfahrens und seiner verbesserten form," Physical 1, 1934.

[11] C. W. Chong, P. Raveendran, and R. Mukund, "An efficient algorithm for fast computation of pseudo-zernike moments," IEEE PAMI, 2003.

[12] T. W. Lin, and Y. F. Chou, "A comparative study of zernike moments," IEEE/WIC International Conference on Web Intelligence, 2003.

[13] B. Han, D. Comaniciu, and L. Davis, "Sequential kernel density approximation through mode propagation," BCCV, 2004.



윤 상 민

- 2000년 고려대학교 전자공학과 (공학사)
- 2002년 고려대학교 전자공학과 (공학석사)
- 2002년~2005년 삼성종합기술원 연구원
- 2007년~현재 다름슈타트 공대 컴퓨터공학과 박사과정
- 관심분야 : HCI, 컴퓨터 비전 등