

논문 2010-47SP-2-3

인체의 구조적 특성과 역운동학을 이용한 모션 캡처

(Motion Capture using both Human Structural Characteristic and Inverse Kinematics)

서용호*, 두경수*, 최종수***, 이철우****

(Yung-Ho Seo, Kyoung-Soo Doo, Jong-Soo Choi, and Chil-Woo Lee)

요약

기존 모션 캡처의 경우, 고가의 장비나 사용의 복잡도, 동작자의 움직임 제한 등 모션 캡처의 어려움이 있었다. 최근 실시간으로 모션 캡처가 가능한 컴퓨터 비전 기반 시스템에 대한 연구가 활발히 진행되고 있다. 본 논문에서는 다시점 영상으로부터 쉽고, 빠르게 추출할 수 있는 피부색과 정확한 3차원 복원을 위한 2차원 영상 좌표 보정을 사용하여 효율적인 다시점 영상 분석 알고리즘을 제안한다. 동작자의 피부색을 검출하고, 카메라 보정 및 에피폴라 기하학 정보를 이용하여 보다 정확한 영상 분석, 그리고 칼만 필터(Kalman filter)를 사용한 추적 등을 통해 보다 안정적인 모션 캡처가 가능하게 된다. 실험결과를 통하여, 제안된 방법은 보다 정확한 위치 추정 및 실시간 모션 캡처를 위한 알고리즘임을 보여주고 있다.

Abstract

Previous hardware devices to capture human motion have many limitations; expensive equipment, complexity of manipulation or constraints of human motion. In order to overcome these problems, real-time motion capture algorithms based on computer vision have been actively proposed. This paper presents an efficient analysis method of multiple view images for real-time motion capture. First, we detect the skin color regions of human being, and then correct the image coordinates of the regions by using camera calibration and epipolar geometry. Finally, we track the human body part and capture human motion using kalman filter. Experimental results show that the proposed algorithm can estimate a precise position of the human body.

Keywords: Motion capture, Epipolar geometry, Camera calibration, 3D reconstruction, Kalman Filter

I. 서론

최근 컴퓨터 비전 기술의 발달에 따라 영상 기반으로

* 정회원, **** 평생회원, 전남대학교 전자컴퓨터공학부 (School of Electronics and Computer Engineering, Chonnam National University)

** 정회원, *** 평생회원, 중앙대학교 첨단영상대학원 영상학과 (Dept. of Image Engineering, Graduate School of Advanced Image Science, Multimedia, and film, Chung-Ang University)

※ 본 연구는 지식경제부 및 정보통신산업진흥원의 대학 IT연구센터 지원사업(NIPA-2010-C1090-1011-0008)과 서울시기술이전사업(TR080601)의 연구결과로 수행되었음.

접수일자: 2009년8월14일, 수정완료일: 2010년2월10일

동작자의 움직임을 인식하여 사용자 인터페이스로 사용하는 멀티미디어 시스템이 등장하고 있다^[1]. 동작자의 모션을 캡처하는 기술은 부가적 하드웨어 장비나 마커로부터 자유로워질 수 있어서 다양한 응용시스템에 적용이 가능하다. 영화 또는 게임 속 캐릭터의 사실적인 움직임 표현, 리모컨, 키보드, 마우스와 같은 기존 컨트롤러로부터 보다 자유롭고 친근감 있는 사용자 인터페이스, 병원 및 노인보호시설에서 사람들의 갑작스러운 이상 상태 감시, 지하철 등 공공장소에서의 비정상적인 행동 분석을 위한 감시 등 다양한 분야에서 활용 가능하다. 이러한 모션 캡처에 대한 연구는 높은 잠재력을 갖고 있어 많은 연구가 진행되고 있다^[2-3].

모션 캡처란 3D 공간상에서 대상체의 움직임에 대한

위치와 방위를 측정하고, 컴퓨터가 사용할 수 있는 형태의 정보로 기록하여 분석 및 응용하는 기술을 의미한다^[4-5]. 3D Studio MAX나 MAYA와 같은 기존의 키프레임 애니메이션 기법은 캐릭터의 움직임을 직접적으로 자세히 제어할 수 있다는 장점이 있지만, 수작업의 양이 많고 부자연스럽거나 비현실적인 움직임을 만들 수 있다는 단점이 있다. 따라서 이를 해결하기 위해 다양한 방식의 모션 캡처 시스템들이 등장하고 있다. 데이터를 추출하는 방식에 따라 광학식, 기계식, 자기식으로 구분된다^[6]. 그 중에서 현재 활용되고 있는 영상 기반 광학식 모션 캡처 시스템은 동작자의 특정 부위에 마커를 부착하고, 다수의 카메라 영상을 취득한 후 3차원 데이터를 추정하는 방법이다. 이러한 마커의 부착은 사용자의 불편함이 존재하게 된다. 따라서 본 논문은 마커로부터 자유로운 영상에 기반한 광학식 모션 캡처에 대한 연구이다.

일반적으로, 영상 기반 모션 캡처 시스템은 단일 카메라 또는 다시점 카메라를 이용하는 두 가지 방법으로 분류된다. A. Agarwal^[7]은 훈련된 데이터를 이용하여 동작자의 3차원 자세를 추정하였다. C. Sminichisescu^[8]는 동영상 시퀀스에서 실루엣 영상을 이용하여 시간적 및 공간적 제약을 돕으로써 움직임 특징 공간을 제한하여 자세를 추정하였다. Y. Kameda^[9]와 M. Tong^[10]은 3차원 공간 상의 가상 모델을 영상에 투영하여 3차원 인체 모델의 관절 매개 변수를 추정하였다. M. W. Lee^[11]와 P. Correa^[12]는 인체 부위의 특징 및 위치 정보에 의한 검출 확률적 접근을 통하여 모션 캡처를 얻었다. 이와 같이 단일 카메라 영상을 이용할 경우, 실시간으로 동작자의 자세를 추정하기에는 불가능하거나 계산량이 많아진다. 또한 동작자의 움직임 제약조건으로 인하여 다양한 동작에 대한 자세 추정에 어려움이 존재한다^[13].

따라서, 다시점 영상을 사용하여 다각도에서 취득한 영상으로부터 동작자의 자세 추정에 대한 연구가 활발히 진행되고 있다. T. Chaichana^[14]는 인체의 각 부위에 마커(illuminated marker)를 부착하여 마커를 추적함으로써 각 부위의 3차원 데이터를 얻었다. S. E. Kim^[15]은 인체의 각 부위를 초기 자세로부터 자동 검출하여, 해당 부위를 추적 및 3차원 재구성을 이용하여 동작자의 자세를 추정하였다. F. Guo^[16]와 A. Sundaresan^[17]는 다수의 카메라를 이용하여 실루엣 영상으로부터 3차원 복셀 상에서 복원한 후 인체의 각 부위를 분할

(segmentation)하는 방법으로 동작자의 자세를 얻어냈다. H. Ohbuch^[18]는 각 움직임에 대한 모션 DB를 저장해 두고, 이로부터 영상 검색(image retrieval) 알고리즘을 통한 유사한 동작을 추정하였다.

일반적으로, 다시점 영상에서 정확한 실루엣을 취득하기 어려우며, 잡음에 상당히 민감하다. 훈련된 DB를 사용하는 방법은 정확한 동작의 추정이 어렵고 동작자의 움직임이 제한적이며, 많은 DB를 사용하게 될 경우 자세 추정 시 계산량이 많아진다. 또한 다수의 카메라를 사용할 경우 카메라 영상에 대해 각각의 영상처리를 수행해야 하므로 부가적인 하드웨어가 필요하며, 카메라 영상의 동기를 맞춰야 하는 문제가 발생한다^[13]. 따라서, 영상으로부터 동작자의 움직임을 추정하기 위해 보다 단순하며, 쉽게 추정 가능한 방법이 필요하다.

본 논문은 정확한 영상 분석을 통하여 동작자와 유사한 모션을 생성하는 효율적인 모션 캡처 시스템을 제안한다. 또한 마커를 사용하지 않으며, 다시점 영상에 대해 실시간 처리가 요구되어야 한다. 먼저, 취득된 다시점 영상에 대하여 배경 차분 및 피부색 추출을 통하여 동작자의 각 인체 영역을 검출한다. 다음으로 카메라 보정 매개변수를 이용하여 검출된 2D영상 좌표를 보다 정확한 좌표로 보정한다. 보정된 좌표에 대한 칼만 필터(Kalman filter)^[19]를 적용하여 인체 영역을 안정적으로 추적한다. 추적된 각 인체 영역에 대하여, 카메라 보정에 의해 얻어진 투영 행렬과 추적된 좌표로부터 동작자의 모션 캡처를 얻는다. 마지막으로, 역운동학 해법을 이용하여 중간 관절들의 움직임을 추정한다.

미리 추출된 카메라 교정, 에피폴라 기하학 정보 및 칼만 필터에 의해 각 영역의 좌표를 교정함으로써, 오차를 줄이며 추적의 안정성을 높였다. 또한, 인체의 부분적인 정보만을 취득하더라도 역운동학 해법을 이용하여 동작자의 유사한 모션 캡처를 얻을 수 있었으며, 다양한 관절 비율을 갖는 캐릭터에게도 동일한 움직임을 부여할 수 있다. 또한 혼합된 역운동학 해법과 의도적인 움직임 제약을 통해 관절의 위치 추정을 위한 반복 수행 횟수를 줄일 수 있다. 이러한 과정을 통하여 마커를 사용하지 않은 실시간 모션 캡처가 가능하다.

본 논문의 구성은 II장에서는 모션 캡처를 위한 2차원 영상분석에 대한 기존 연구 수행 내용을 기술하며, III장에서는 인체 영역 추정을 위한 영상 처리에 대하여 기술하며, IV장에서는 중간 관절 복원을 위한 역운동학에 대하여 기술하고, V장에서는 제안된 방법에 의한

실험결과를 통해 분석하고, 마지막으로 VI장에서 결론을 맺는다.

II. 기존 연구 수행 내용

이 장에서는 기존 연구 논문^[13]에서 다루었던 내용을 간략히 기술한다.

1. 카메라 보정 및 에피폴라 기하학

영상으로부터 3차원 데이터를 추출하기 위해서는 각 영상에 대한 카메라 보정(camera calibration)이 필수적이다. 2차원 평면 격자 영상을 이용한 Z. Zhang의 카메라 보정 방법을 이용하였으며, 렌즈의 왜곡 성분인 방사 왜곡(radial distortion)과 접선 왜곡(tangential distortion)에 대해서도 교정이 필수적이다. 자세한 기술은 기존 연구 논문에 기술되어 있다. 또한, 두 영상 사이의 대응점 및 대응 영역을 추정하기 위해 에피폴라 제약조건이 효율적이다. 두 대의 카메라 사이의 에피폴라 기하학은 각 카메라 중심을 잇는 기준선(baseline)이 포함된 수많은 평면들과 각 영상 평면이 교차하게 된다는 것을 의미한다^[13]. 에피폴라 제약조건을 사용하여 대응되는 영역의 후보를 줄일 수 있다.

인체 영역의 3차원 정보는 영상으로부터 추출된 2차원 영상좌표와 카메라 보정 파라미터에 의해 추정이 가능하다. 이것은 선형 삼각화 기법(linear triangular method)을 이용하면 빠르게 3차원 정보를 얻어낼 수 있다^[19].

2. 인체 영역 추출 및 추적

실시간으로 입력되는 각 카메라 영상으로부터 동작자의 머리, 손, 발에 해당되는 인체 영역을 추출해야 한다. 실시간 모션 캡처를 위해서는 각 영상에 대한 영상 처리는 빠른 시간 내에 수행되어야 한다. 따라서, 본 논문에서는 인체의 공통된 특징인 피부색을 이용하여 각 인체 영역을 추출했다. RGB와 HSI 색공간을 이용하여 배경 차분(background subtraction)을 수행하였다. 차분된 영상에 대하여 정규화된(normalized) RGB 색공간을 사용하여 피부색을 추출하였다. 정규화된 RGB는 조명 변화에 둔감하고, 빠른 수행 속도를 얻을 수 있다.

동작자의 초기 T-자세로부터 각 카메라 영상에서 추출된 인체 후보 영역은 인체의 구조적 정보를 이용하여 쉽게 정합이 가능하다. 각 인체 영역은 초기 정합 정보

와 칼만 필터를 이용하여 추적한다. 취득된 영상으로부터 인체 영역이 짧은 검출이 되지 않거나 자기 폐색이 발생할 경우에도 정상적인 추적이 가능하다.

III. 인체 영역 추정을 위한 영상처리

인체 영역 추정을 위해 필요한 영상처리는 기존 연구 내용에서 기술하였지만, 본 논문에서는 피부색 영역 추출의 안정성을 향상시킨다. 먼저, 입력되는 카메라 영상에서 RGB 영상차분을 수행하고, 차분된 RGB 영상에 대하여 HSI 영상차분을 수행한다. 이 과정은 그림자 및

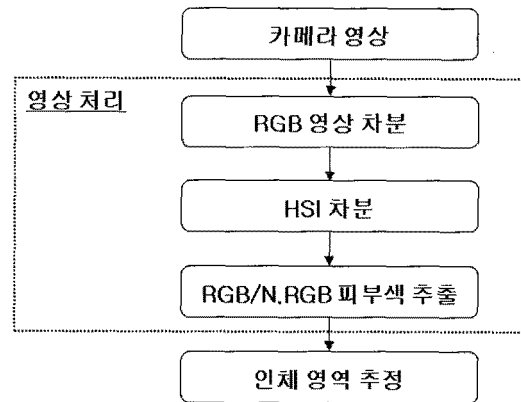


그림 1. 인체 후보 영역 검출

Fig. 1. Detection for candidate body parts of human.

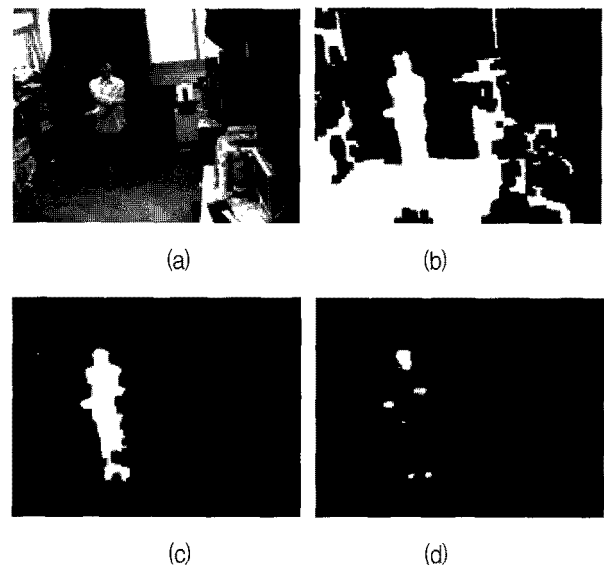


그림 2. 피부색 검출

(a) 입력영상 (b) RGB 차분 영상
(c) HSI 차분 영상 (d) 피부색 추출 영상

Fig. 2. Skin Detection.

(a) Input image (b) RGB subtraction
(c) HSI subtraction (d) Skin image

조명의 변화로 인해 추출된 영역을 제거하는데 효율적이다. 다음으로, RGB 색공간과 정규화된 RGB 색공간을 동시에 적용하여 인체의 후보 영역을 추정한다. 그림 1은 각 카메라 영상에 대한 2차원 영상 처리 과정이다. 그림 2는 자연광이 존재하는 환경에 대하여 피부색 추출 결과를 보여주고 있다.

그림 2(a)는 실시간으로 입력되는 현재 영상이며, 그림 2(b)는 RGB 영상 차분의 결과 영상이다. 그림 2(c)는 HSI 영상 차분의 결과 영상을 보여주고 있으며, 그림 2(d)는 RGB 색공간과 정규화된 RGB 색공간을 이용한 피부색 추출 영역을 보여주고 있다. 본 논문에서 RGB 색공간과 정규화된 RGB 색공간을 같이 사용하는 것은 색공간 변환하는 시간이 절약되며, 피부색 추출의 안정성을 향상시킨다.

IV. 중간 관절 복원

1. 가상 모델과 복원 모델사이의 좌표계 변환

기존 연구 수행 내용에서 각 인체 영역의 3차원 정보

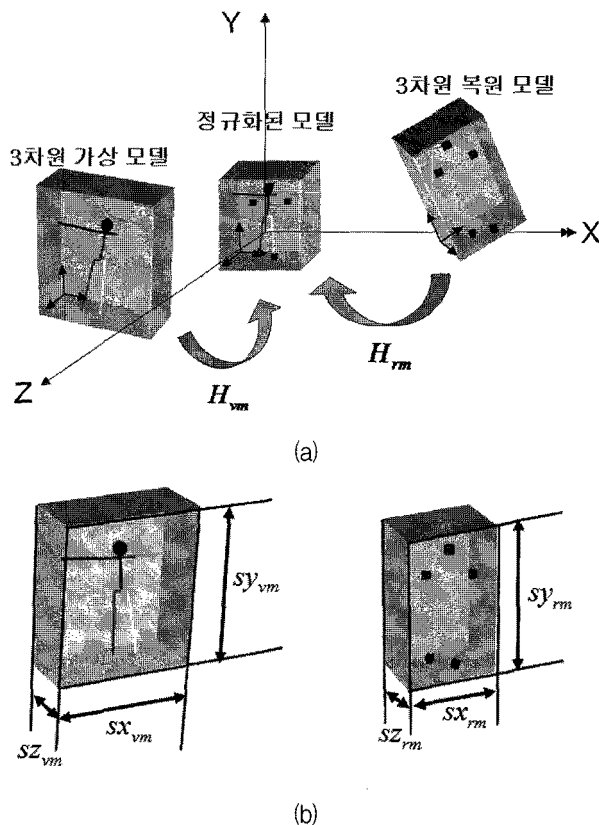


그림 3. 가상 모델과 복원 모델 사이의 좌표 변환
Fig. 3. Coordinate transformation between 3D virtual model and 3D reconstructed model.

를 추정하는 방법이 기술되어 있다. 본 논문에서는 5개의 인체 영역 정보를 이용하여 18개의 관절을 갖는 인체의 모델링을 하고자 한다. 먼저, 3차원 공간 상에서 인체의 가상 모델을 초기화해야 한다. 추정된 5개의 인체 영역에 대한 3차원 좌표값, 즉 복원 모델을 인체 가상 모델에 정합한다.

그림 3은 3차원 가상 모델과 5개의 인체 영역에 대한 복원 모델에 각각 PCA를 수행하여 좌표계를 변환하는 과정과 각 모델의 비율 성분들을 보여주고 있다. 그림 3(a)에서, H_{vm} 는 3차원 가상 모델에 대한 PCA 변환행렬을 의미하고, H_{rm} 는 3차원 복원 모델에 대한 PCA 변환행렬을 의미한다. 그림 3(b)에서, 첨자 vm 은 가상 모델(virtual model)을 의미하며, rm 은 복원 모델(reconstructed model)을 의미한다. 수식 (1)과 수식 (2)는 각각 가상 모델과 복원 모델의 비율 행렬을 보여주고 있다.

$$S_{vm} = \begin{bmatrix} sx_{vm} & 0 & 0 \\ 0 & sy_{vm} & 0 \\ 0 & 0 & sz_{vm} \end{bmatrix} \quad (1)$$

$$S_{rm} = \begin{bmatrix} sx_{rm} & 0 & 0 \\ 0 & sy_{rm} & 0 \\ 0 & 0 & sz_{rm} \end{bmatrix} \quad (2)$$

$$X_{vm} = S_{vm} S_{rm}^{-1} H_{vm}^{-1} H_{rm} X_{rm} \quad (3)$$

PCA를 수행하게 되면, 각 모델에서의 축 정보를 얻을 수 있으며 좌표계를 정규화(normalization)할 수 있다. 하지만, PCA를 수행하게 되면 좌우 또는 상하가 뒤집어지는 경우가 발생한다. 우리는 초기 자세로부터 가상 모델과 복원 모델의 머리, 왼손, 오른손, 왼발, 오른발의 위치를 이미 알고 있다. 따라서, PCA를 수행한 후 3차원 공간 상의 위치 정보는 자동으로 교정이 가능하다.

초기 가상 모델과 복원 모델로부터 얻어진 변환 행렬과 비율 행렬을 이용하여 수식(3)으로부터 연속되는 동작자의 모션 캡처가 가능하다. X_{vm} 은 3차원 가상 모델의 좌표이고, X_{rm} 은 3차원 복원 모델의 좌표이고, S_{vm} 과 S_{rm} 은 각 모델의 비율 행렬을 의미한다.

2. 역운동학(inverse kinematics)

일반적으로 역운동학은 관절을 갖는 로봇의 제어 및 캐릭터 애니메이션에서 많이 사용된다. 본 논문에서는 5개의 추출된 인체 영역을 이용하여 모션 캡처를 목적으로 하고 있다. 따라서 추출되지 않은 나머지 관절 위치에 대해서는 역운동학을 이용하여 추정한다. 역운동학 알고리즘은 크게 해석적인 방법(analytical method)와 수치적인 방법(numerical method)으로 구분된다. 수치적인 방법은 초기값에 기반을 두고 반복적인 계산을 통해 수렴값을 찾아나간다. 일반적으로 수치적인 방법보다 해석적인 방법이 모든 가능한 해를 얻을 수 있고, 빠르며 신뢰도가 더 높기 때문에 선호한다^[20].

하지만, 본 논문에서는 가상 모델과 복원 모델 사이의 초기 정합을 통해서, 정확한 초기값을 알고 있다. 또한, 동작자의 연속되는 움직임을 추정하므로, 즉 갑작스러운 변화가 존재하지 않는다. 따라서, 해 공간이 이전 값의 근처에서 존재하므로, 해석적인 방법을 사용하여 빠른 수렴이 가능하다. 본 논문에서는 해석적인 방법과

수치적인 방법을 조합하여 역운동학의 해를 구한다.

일반적인 인체의 모든 자유도를 복원하는 것은 매우 어려운 문제이다. 그림 4(a)는 왼쪽 팔에 대한 역운동학 구조를 보여주고 있다. 팔의 경우, 어깨에 대한 2개의 회전 성분, 팔꿈치에 대한 1개의 회전 성분을 갖는 3개의 자유도를 갖는다. 또한 다리의 경우에도 기준 좌표계만 달라질 뿐 팔과 같은 자유도를 가지며, 몸 전체의 기저 이동(base translation)에 대한 3개의 자유도를 갖는다. 그림 4(b)는 해석적인 방법으로부터 중간 관절값을 추정하는 과정을 보여주고 있다.

$$\|F(\theta_1, \theta_2, \theta_3) - G\| = 0 \tag{4}$$

먼저, 우리는 역운동학을 이용하여 수식 (4)를 만족하는 새로운 관절값을 찾는다. 여기에서, F는 현재의 관절값으로부터 얻어진 손의 위치를 의미하며, G는 $(x_g, y_g, z_g)^T$ 으로서, 캡처된 손의 위치를 의미한다. 왼쪽 팔의 경우, 현재의 관절값으로부터 손의 위치를 추정하는 방법은 수식 (5)와 같다. 이것은 2개의 링크(upper arm, lower arm)로 구성되어 있으며, 3차원 상에서 위치를 추정하는 것이다^[21~22].

$$F(\theta_1, \theta_2, \theta_3) = \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} l_{ua} \sin(\theta_1 + \theta_3) + l_{la} \sin \theta_1 \sin \theta_2 \\ l_{ua} \cos(\theta_1 + \theta_3) + l_{la} \cos \theta_1 \\ l_{ua} \sin(\theta_1 + \theta_3) + l_{la} \sin \theta_1 \cos \theta_2 \end{bmatrix} \tag{5}$$

다음으로, 극소증감벡터 Δ는 수식 (5)의 자코비안 행렬(Jacobian matrix)을 이용하여 찾을 수 있다. 자코비안 행렬에서 특이치(singularity)가 발생할 경우, 매우 작은 단위행렬을 더함으로서 항상 해를 얻을 수 있다. 이 단위행렬은 항상 해의 추정이 가능하도록 만들며, 매우 작은 값을 사용하므로 구하고자 하는 해에 근접하게 된다^[23].

$$\Delta = (J + \lambda I)^{-1} dF = \begin{bmatrix} \frac{\partial x}{\partial \theta_1} + \lambda & \frac{\partial x}{\partial \theta_2} & \frac{\partial x}{\partial \theta_3} \\ \frac{\partial y}{\partial \theta_1} & \frac{\partial y}{\partial \theta_2} + \lambda & \frac{\partial y}{\partial \theta_3} \\ \frac{\partial z}{\partial \theta_1} & \frac{\partial z}{\partial \theta_2} & \frac{\partial z}{\partial \theta_3} + \lambda \end{bmatrix}^{-1} \begin{bmatrix} dx \\ dy \\ dz \end{bmatrix} \tag{6}$$

추정된 극소증감벡터에 따라 다시 F값을 추정하여 오차를 계산하며, 얻어진 오차값이 정해진 오차값 이하가 되면 수렴된 것으로 판단하고 반복수행은 멈춘다. 하지만, 본 논문에서는 제한된 자유도를 갖는 인체 모델을 사용하므로, 중간 관절의 추정값은 해석적인 방법

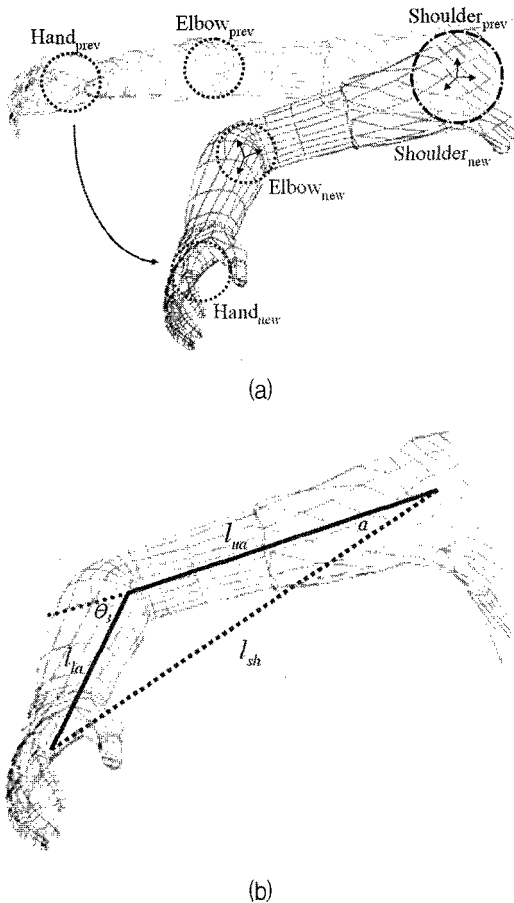


그림 4. 왼쪽 팔의 역운동학
Fig. 4. Inverse kinematics for left arm

을 통하여 미리 추정이 가능하다. 수식 (7)은 캡처된 손의 위치로부터 중간 관절의 추정값을 계산하는 수식이다.

$$\alpha = \cos^{-1} \left(\frac{l_{ua}^2 + l_{sh}^2 - l_{la}^2}{2l_{ua}l_{la}} \right) \tag{7}$$

$$\theta_2 = \sin^{-1} \left(\frac{l_{sh}}{l_{la}} \sin \alpha \right)$$

V. 실험 결과

1. 인체 영역 검출 및 안정적 추적

본 논문에서는 인체의 공통된 피부색을 검출하기 위해 다양한 색공간의 조합으로 실험하였다. 그림 5, 그림 6과 그림 7은 취득된 영상으로부터 피부색 추출 과정에 대한 결과 영상들을 보여주고 있다. 그림 5는 형광등 조명 아래에서의 환경, 그림 6은 일부 자연광을 포함한 환경, 그림 7은 많은 자연광이 포함된 환경 하에서 동작자가 피부색 옷을 입은 상태에서 인체 영역 검출 결과를 보여주고 있다. 그림 5(a)는 1번 프레임,

그림 5(b)는 51번 프레임, 그림 5(c)는 101번 프레임, 그림 5(d)는 151번 프레임, 그림 5(e)는 201번 프레임 보여주고 있다. 각 그림에서 첫 번째 행은 입력영상을 보여주고 있으며, 두 번째 행은 RGB 배경 차분에 의해서 얻어진 결과 영상을 나타내며, 세 번째 행은 HSI 차분에 의해서 얻어진 결과영상, 네 번째 행은 최종적으로 검출된 피부색 영상을 보여준다. 그림 6과 그림 7 역시 동일하다.

RGB 배경 차분에 의해서 얻어진 영상을 보면, 제한된 환경 내에서 동작자가 나타남으로 인하여 동작자의 그림자 외에 동작자로 인하여 주변 물체 및 환경의 상당한 변화가 생기는 것을 알 수 있다. 하지만 피부색 영역은 정상적으로 검출되고 있음을 알 수 있다.

하지만, 그림 7의 결과로부터 우리는 제안된 모션 캡처 시스템에서 정상적인 인체 영역 검출의 어려움을 알 수 있다. 이것은 제안된 시스템이 피부색에 의존하므로, 피부색과 유사한 색상을 가지는 옷을 입은 경우, 제안된 시스템에서는 올바른 모션 캡처가 어렵다. 따라서, 본 논문에서는 피부색 범위에 포함되는 옷에 대해서는 제한을 두었다.

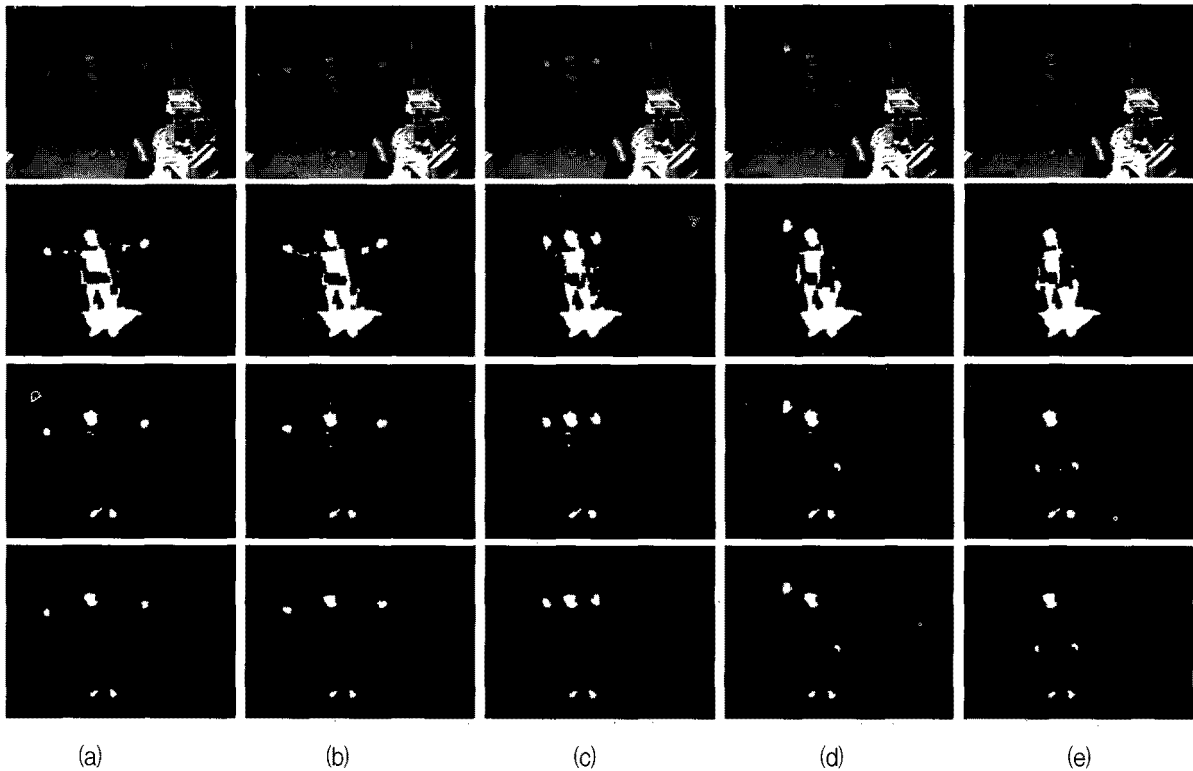


그림 5. 인체 영역 검출 실험 (환경-1)

Fig. 5. The experimental result of detection for human body part (Environment-1).

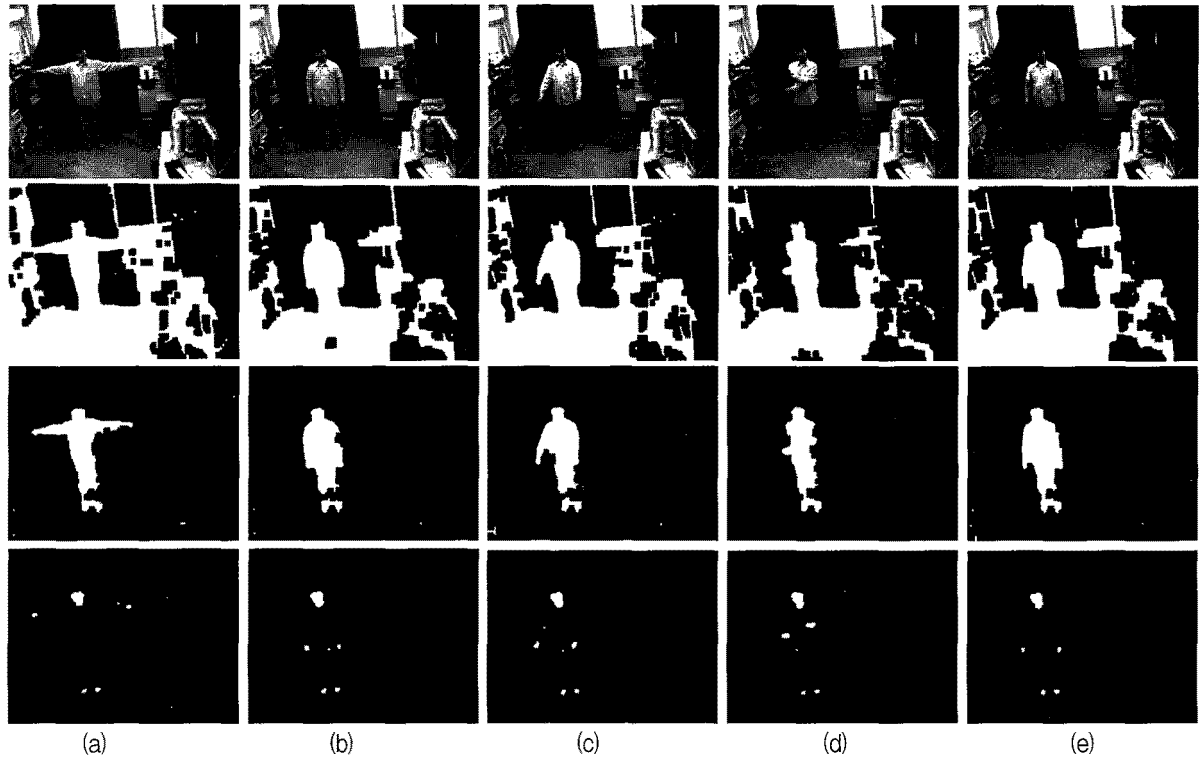


그림 6. 인체 영역 검출 실험 (환경-II)

Fig. 6. The experimental result of detection for human body part (Environment-II)

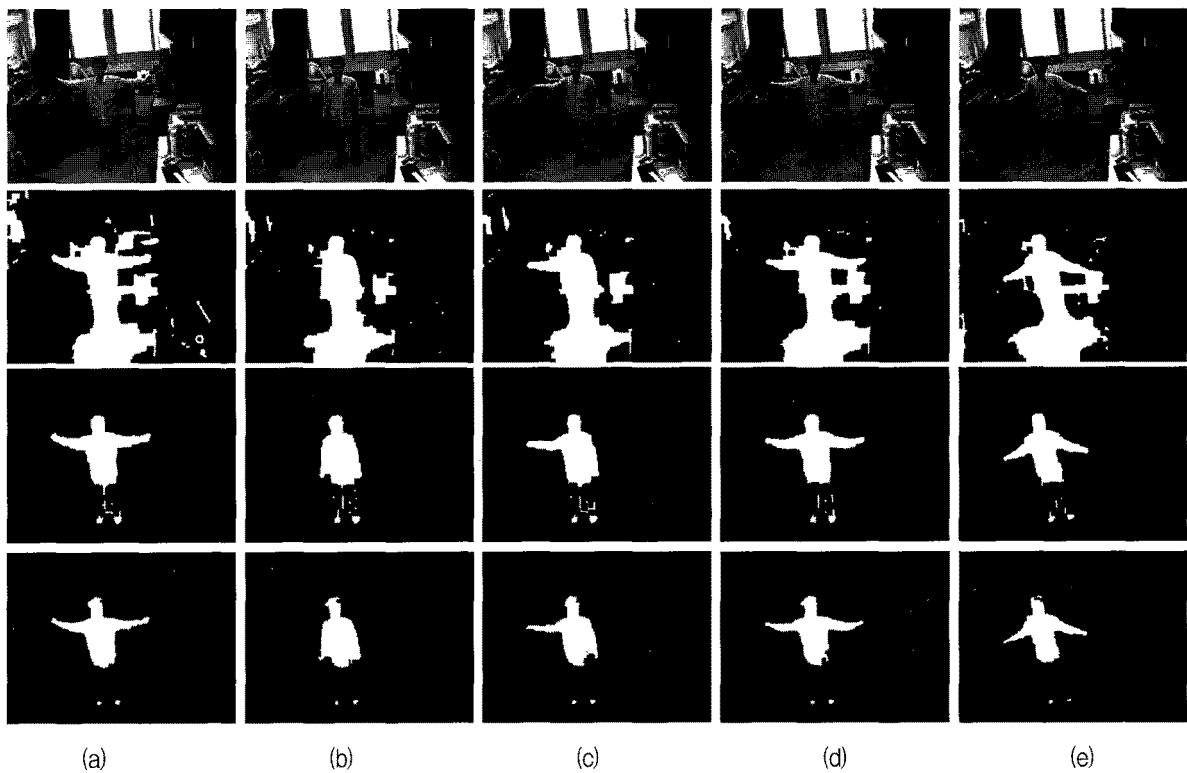
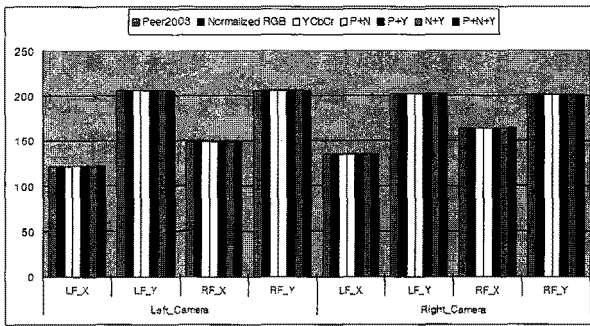
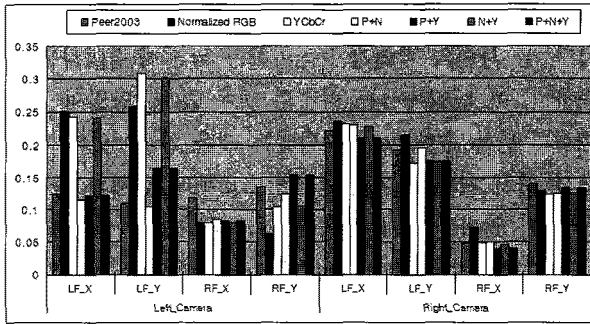


그림 7. 인체 영역 검출 실험 (환경-I)

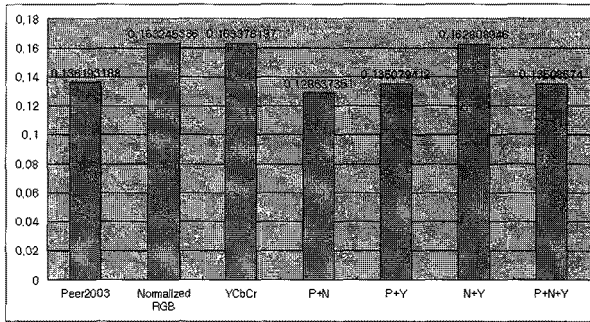
Fig. 7. The experimental result of detection for human body part (Environment-I)



(a)



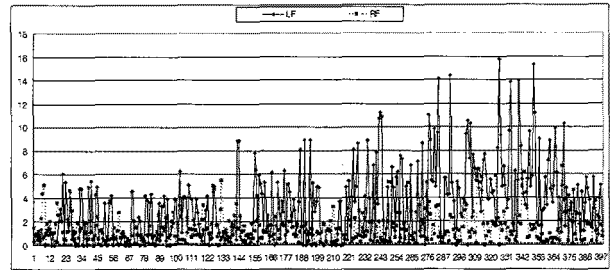
(b)



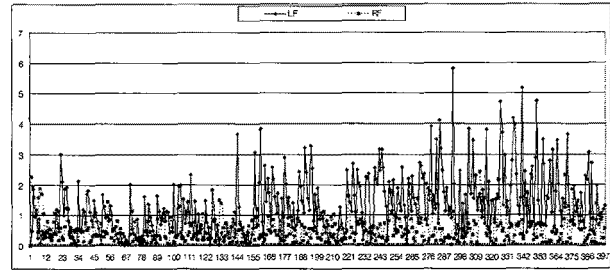
(c)

그림 8. 색공간에 따른 피부색 검출 실험
Fig. 8. The experimental result of skin detection.

그림 8은 안정적인 피부색 검출을 위해서 색공간을 조합한 실험을 보여주고 있다. 이 실험은 사용자의 의도적인 움직임이 없는 400프레임에 대한 결과를 보여준다. 그림 8(a), 8(b), 8(c)는 두 카메라에 취득된 발에 대한 평균, 표준편차와 표준편차의 평균을 보여주고 있다. 그림 8(a)에서 세로축은 영상의 평균 좌표를 의미하고, 그림 8(b)에서 세로축은 이전 프레임에서 취득된 값과 현재 프레임에서 취득된 값의 차이에 대한 표준편차를 의미한다. 범례에서 P는 Peer2003^[24]으로 RGB 색공간으로부터 취득한 방법, N은 정규화된 RGB 색공간으로부터 취득한 방법, Y는 YCbCr 색공간으로부터 취득 방법을 의미하며, “+”는 각각의 방법으로부터 추출된 결



(a)



(b)

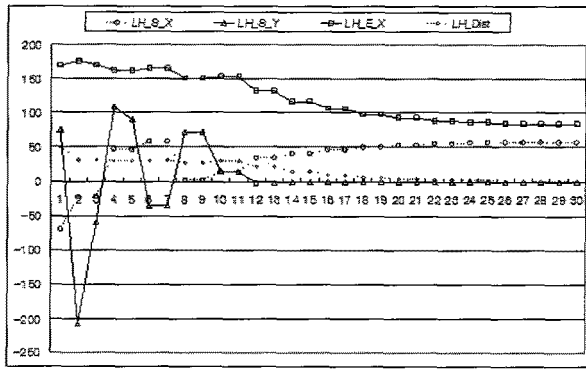
그림 9. 3차원 복원 데이터에 대한 안정성 실험
Fig. 9. The stability experiment of 3D reconstructed data.

과에 대해 논리합(OR)을 의미한다. 그림 8(a)에 의하면, 각 7가지 방법에 의해 추출한 결과 추정된 값은 거의 비슷한 결과를 보여주고 있다. 하지만, 그림 8(c)에 의하면, RGB 색공간과 정규화된 RGB를 동시에 적용한 방법이 다른 방법에 비해 낮은 표준편차값을 얻었다. 이것은 상대적으로 안정적인 피부색 검출이 되고 있음을 보여주고 있다.

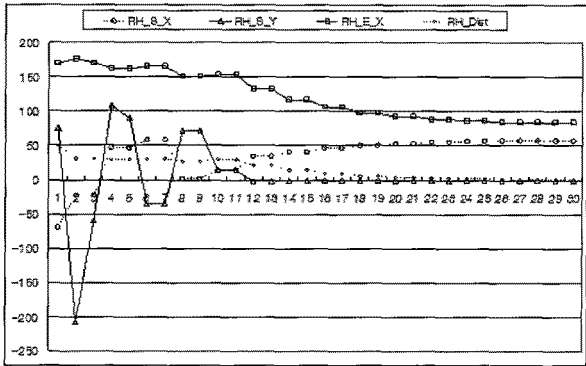
그림 9는 움직이 없는 양 발에 대한 복원된 3차원 좌표 추적의 안정성에 대한 실험을 보여주고 있다. 여기에서 오차는 사용자의 의도적인 움직임이 없는 영상(400프레임)에 대하여 현재 영상의 3차원 복원값과 이전 영상의 3차원 복원값 사이의 차이로 정의한다. 그림 9(a)는 칼만 필터를 적용하기 전의 오차값을 보여주며, 그림 9(b)는 칼만 필터를 적용한 후의 오차값을 보여준다. 칼만 필터를 적용하기 전의 경우에는, 왼쪽 발과 오른쪽 발의 오차값이 각각 3.15mm, 1.53mm였다. 하지만, 칼만 필터를 적용한 후의 경우에는, 왼쪽 발과 오른쪽 발의 오차값이 각각 1.26mm, 0.55mm로 40%이하로 줄었다. 이로부터, 본 논문에서는 1cm이하의 움직임이 있는 경우에는 의도적인 움직임이 없는 것으로 판단한다.

2. 역운동학 해법

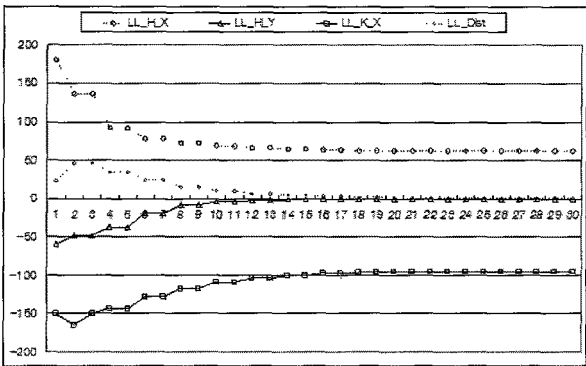
본 논문에서는 인체의 부분적인 정보를 추출하고, 이것으로부터 역운동학을 이용하여 동작자와 유사한 모션



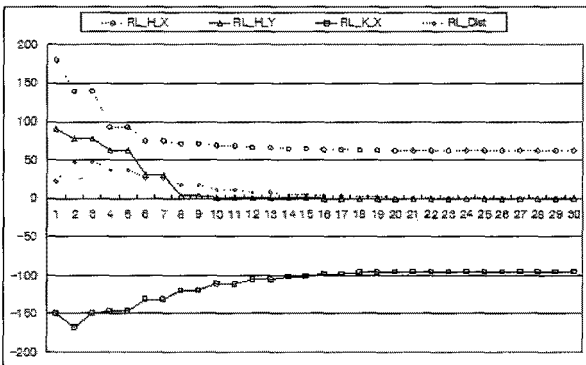
(a)



(b)



(c)

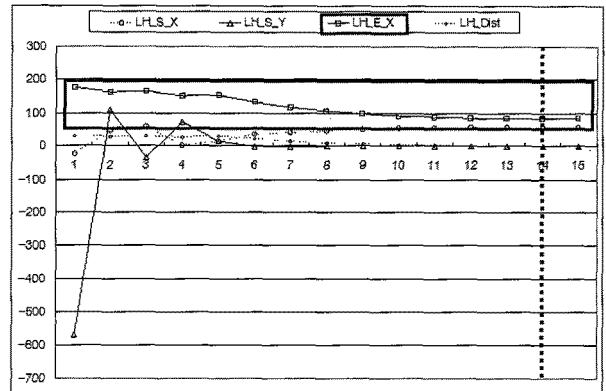


(d)

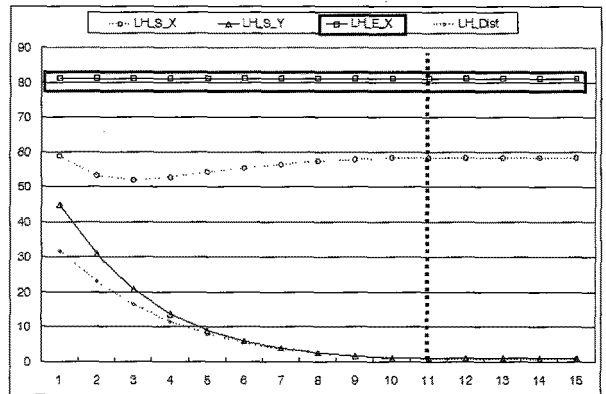
그림 10. 역운동학 모의실험
Fig. 10. Simulation result of inverse kinematics.

캡처가 가능한 시스템을 제안한다. 먼저, 우리는 5개의 인체 영역 정보를 이용한 수치적인 역운동학 해법에 대한 실험을 통해 모션 캡처의 가능성을 보여주려고 한다. 그림 10은 양손과 양발에 해당하는 4개의 목적점(goal points)을 설정하고, 역운동학 모의실험(simulation)을 통해 해를 구한 결과를 보여주고 있다. 그림 10(a)~(d)는 반복 수행(가로축)에 따른 왼손 관절, 오른손 관절, 왼발 관절, 오른발 관절의 회전각(세로축)에 대한 변화를 보여주고 있다. 왼손과 오른손의 목적점까지의 거리는 50cm, 왼발과 오른발의 거리는 22cm이다. 둘 사이의 거리가 1cm이하가 되면 수렴이라고 간주하며, 이 실험에서는 양손의 경우 28회, 양발의 경우 22회의 반복 수행 후에 수렴하는 것을 알 수 있다. 실제 모션 캡처 시에는 이전 움직임으로부터 현재의 움직임이 크지 않기 때문에 반복 수행의 횟수는 줄어든다.

본 논문에서는 수치적인 해법과 더불어 해석적인 방법을 동시에 적용하고자 한다. 이것은 해석적인 방법을 이용하여 미리 중간 관절의 회전각 정보를 계산하므로,



(a)



(b)

그림 11. 역운동학 모의실험
Fig. 11. Simulation result of inverse kinematics.

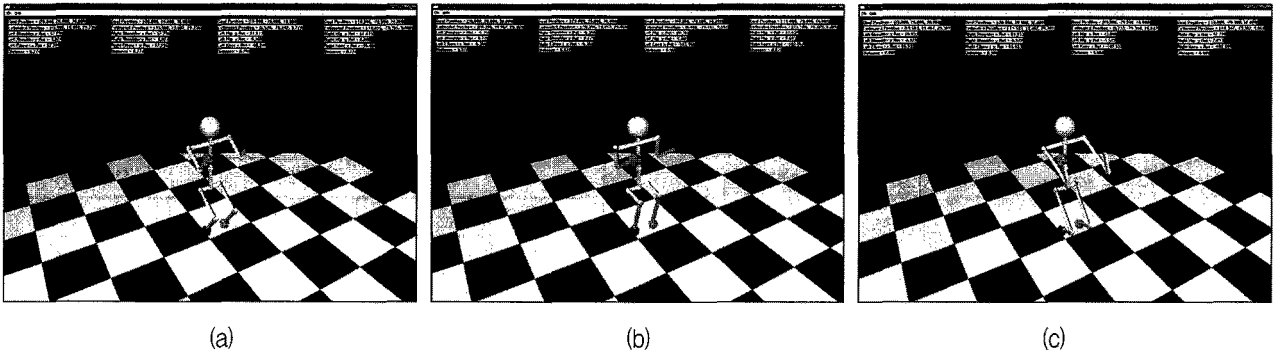
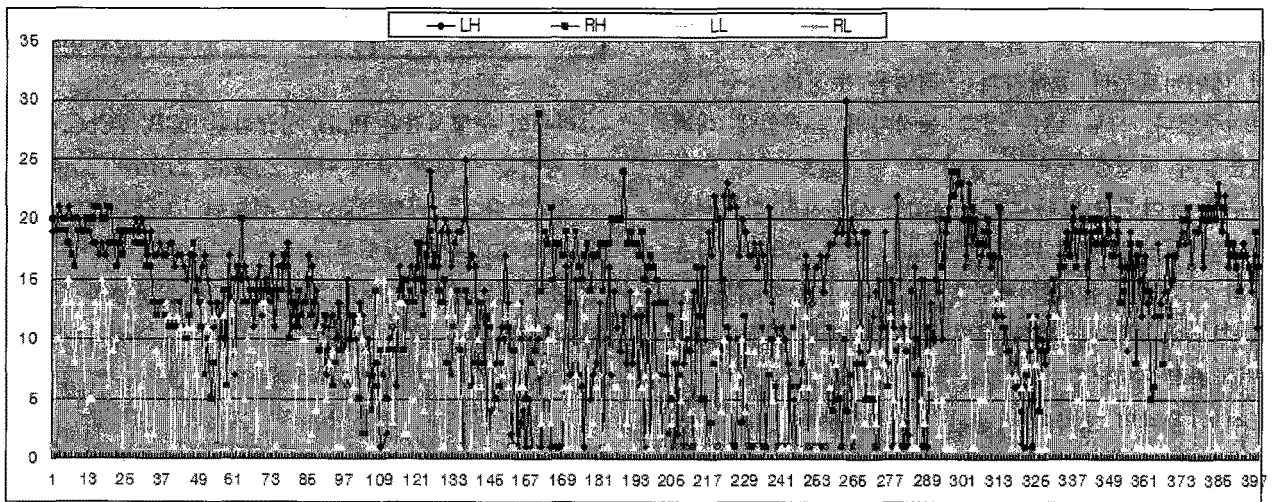
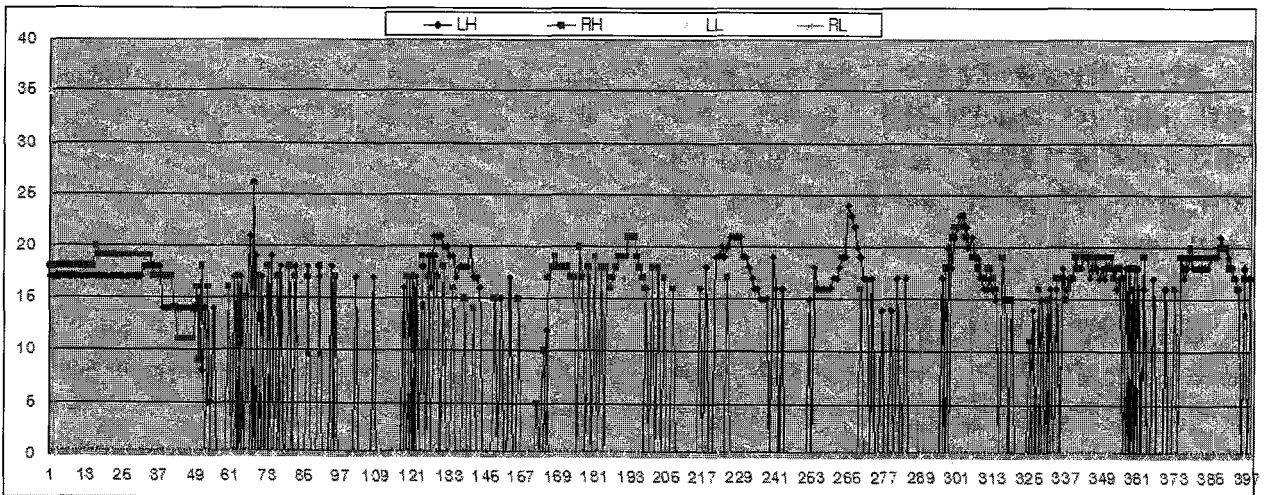


그림 12. 다양한 관절 비율에 대한 역운동학 실험
 Fig. 12. Simulation result of inverse kinematics for various length.



(a)



(b)

그림 13. 캡처된 움직임에 대한 역운동학 실험
 Fig. 13. Simulation result of inverse kinematics for captured motion.

반복 수행의 횟수를 줄일 수 있는 장점을 가지고 있다. 그림 11은 왼손에 대한 해석적인 방법을 적용하기 전과 적용한 후의 반복 수행 결과를 보여주고 있다. 그림

11(b)를 보면, 왼쪽 팔꿈치(LH_E_X)에 해당하는 값이 미리 계산이 되어 고정되어 있다. 해석적인 방법을 적용하기 전에는 14회의 반복 수행이 소요되었지만, 해석

적인 방법을 적용한 후에는 11회의 반복 수행 후 수렴했다. 따라서, 해석적인 방법을 동시에 적용한 결과 해에 더 빠르게 수렴하는 것을 알 수 있다.

그림 12는 취득된 모션 캡처 데이터를 이용하여 다양한 관절 비율을 갖는 모델에 적용한 실험을 보여주고 있다. 위쪽 팔(upper arm)과 아래쪽 팔(lower arm)의 관절 비율이 각각 1:1(그림 12(a)), 1:2(그림 12(b)), 2:1(그림 12(c))인 모델에 적용한 결과이다. 이것은 한번 취득된 모션 캡처 데이터가 다양한 관절 비율을 갖는 사람 또는 의인화된 동물 등 다양한 캐릭터에게 적용이 가능함을 의미한다.

모델에 대한 역운동학 실험을 통해, 취득된 모션 캡처 데이터로부터 동작자의 움직임을 추정하는 실험을 보여주었다. 그림 13은 실제 동작자의 움직임에 대해 역운동학 해법을 구하는 동안 각 부위별로 반복 수행 횟수를 보여주고 있다. 그림 13(a)는 사용자의 의도적인 움직임 제약을 두지 않은 경우이고, 그림 13(b)는 사용자의 의도적인 움직임 제약을 둔 경우이다. 그림 13(a)의 결과에 의하면, 왼쪽 발과 오른쪽 발의 역운동학을 위한 반복 수행 횟수는 평균 각각 1.5회, 4.2회이다. 의도적인 움직임 제약이 적용된 그림 13(b)의 경우, 왼쪽 발과 오른쪽 발의 반복 수행 횟수는 각각 0회, 0.08회이다. 또한, 양쪽 팔의 경우에 대해서도, 움직임 제약을 두지 않은 경우, 왼손과 오른손의 반복 수행 횟수는 12.4회, 11.6회이다. 반면에, 움직임 제약이 적용된 후에는 9.4회, 8.2회로 역운동학 해법을 위한 반복 수행 횟수가 상당히 감소되었다. 그림 13(b)에서 왼쪽 발과 오른쪽 발에 대한 반복 수행횟수는 거의 모든 프레임에서 0이 나왔다. 이것은 사용자의 의도적인 움직임이 거의 없었음을 보여주고 있다.

3. 실시간 모션 캡처 시스템

본 논문은 인체의 부분적인 정보로부터 역운동학 해법을 이용하여 실시간 마커프리 모션 캡처 시스템을 구현하고자 했다. 표 1은 제안된 시스템의 실시간 수행 가능에 대한 실험 결과이다. 수행시간은 1,500프레임을 대상으로 각각의 단계를 수행한 후, 프레임 당 평균 시간을 측정하는 것이다. 각 알고리즘 수행시간 결과는 두 대의 카메라를 동시에 실시간으로 처리가 가능함을 보여준다. 좌표 보정 및 모션 캡처를 위한 3차원 복원 과정은 5개의 대응점에 대해서 선형 방정식을 푸는 문제이므로, 매우 짧은 시간에 처리가 가능하다. 피부색 추출 알

표 1. 알고리즘 수행 시간 결과

Table 1. The experimental result for the elapsed time of each algorithm

알고리즘	수행시간	
	하나의 영상	두장의 영상
배경 차분	7.4 ms	14.8 ms
피부색 추출	5.4 ms	10.8 ms
레이블링	1 ms	2 ms
좌표 보정	0 ms	0 ms
영역 추적	2 ms	4 ms
3차원 복원	0 ms	0 ms
역운동학	29.7 ms	29.7 ms
전체	-	61.3 ms

고리즘의 수행시간은 RGB, 정규화된 RGB, YCbCr에 의한 방법을 수행한 결과 각각 2.79ms, 2.57ms, 2.79ms가 소요되었다. 두 카메라로부터 취득된 10개의 인체 영역을 추적하기 위한 칼만필터의 응답 및 추적 시간은 4ms가 소요되었다. 수치적인 방법에 의한 역운동학 해법을 한 번 계산하는데 걸리는 평균 시간은 1.7ms가 소요된다. 역운동학 해법을 위해 걸린 수행시간은 의도적인 움직임 제약이 없을 경우에는 한 프레임 당 평균 29회의 반복 수행으로 50ms가 소요되었으며, 의도적인 움직임 제약을 적용한 후에는 한 프레임 당 평균 17.7회로 29.7ms가 소요되었다. 하나의 PC를 이용하여 두 카메라의 영상 처리, 3차원 복원 및 역운동학 해법까지 걸리는 소요 시간은 61.3ms가 소요되었다. 이것은 16.3fps의 처리 속도가 나오며 실시간으로 모션 캡처가 가능함을 알 수 있다.

VI. 결 론

본 논문에서는 다시점 영상으로부터 추출된 인체 영역을 기반으로 동작자의 모션을 캡처하는 방법을 제안하였다.

인체 영역 검출을 위해 RGB와 정규화된 RGB를 사용함으로써, 처리 시간의 감소와 추적의 안정성을 향상시켰다. 또한, 칼만 필터와 움직임 제약조건은 각 인체 영역의 미세한 떨림(jittering)을 안정화하는데 기여하였다. 이것은 역운동학을 이용하여 해를 구하는 과정에서 반복 수행 횟수를 줄이는 효율적인 방법임을 보여주었다. 역운동학을 이용함으로써, 2차원 영상에서 추출하지 못한 인체의 중간 관절 정보들을 복원할 수 있었으며,

수치적인 방법과 해석적인 방법을 같이 사용하여 해를 구하기 위한 반복 수행의 횟수를 줄일 수 있었다. 또한, 취득된 동작자의 움직임은 각 관절의 비율이 다른 모델에 적용이 가능함을 보여주었다. 향후 사용자의 다양한 움직임을 추적하기 위해 중간관절의 정보를 보다 사실적으로 캡처할 수 있는 연구가 필요하다.

참고 문헌

- [1] <http://www.xbox.com/en-US/live/projectnatal/>
- [2] T. B. Moreslund, A. Hilton, V. Kruger, "A survey of advances in vision-based human motion capture and analysis," *Int. J. Computer Vision and Image Understanding*, Vol. 104, Issues 2-3, pp. 90-126, 2006.
- [3] R. Poppe, "Vision-based human motion analysis: An overview," *Int. J. Computer Vision and Image Understanding*, Vol. 108, Issues 1-2, pp. 4-18, 2007.
- [4] 이민기, 박성규, 박근표, 양선우, 이범렬, "모션캡처 기술 동향," *ETRI-전자통신동향분석*, 제22권, 제4호, 2007.
- [5] A. Menache, "Understanding Motion Capture for Computer Animation and Video Games," *Morgan Kaufmann Publishers*, 2000.
- [6] 이민기, "디지털 3D애니메이션지원을 위한 첨단영상제작장비 활용방안 연구," *세종대학교 영상대학원*, 2003.
- [7] A. Agarwal, B. Triggs, "Recovering 3D Human Pose from Monocular Images," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 28, No. 1, pp. 44-58, Jan. 2006.
- [8] C. Sminichescu, A. Telea, "Human Pose Estimation from Silhouette: A Consistent Approach Using Distance Level Sets," *Int. Conf. on Computer Graphics, Visualization and Computer Vision*, 2002.
- [9] Y. Kameda, "Human Pose and Motion Estimation with a Three Dimensional Articulated Object Model," *Ph. D. dissertation, Kyoto University*, Kyoto, 1999.
- [10] M. Tong, Y. Liu, T. S. Huang, "3D Human Model and Joint Parameter Estimation from Monocular Image," *Pattern Recognition Letters*, Vol. 28, No. 7, pp. 797-805, May, 2007.
- [11] M. W. Lee, R. Nevatia, "Integrating Component Cues for Human Pose Tracking," *Int. Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance*, pp. 41-48, Oct. 2005.
- [12] P. Correa, J. Czyz, F. Marques, T. Umeda, X. Marichal, B. Macq, "Bayesian Approach for Morphology-Based 2-D Human Motion Capture," *IEEE Trans. on Multimedia*, Vol. 9, No. 4, pp. 754-765, June, 2007.
- [13] 서용호, 박유신, 구떠올라, 두경수, 최종수, "모션 캡처를 위한 다시점 영상의 효율적인 분석법," *대한전자공학회논문지*, 제45권 SP편 제6호, pp. 44-56, 2008년 11월.
- [14] T. Chaichana, M. Sangworasil, C. Pintavisrooj, S. Aootaphao, "Accelerate a Dlt Motion Capture System with Quad-Tree Searching Scheme," *Int. Sym. on Communications and Information Technologies*, pp. 1035-1038, Oct. 2006.
- [15] S. E. Kim, C. J. Park, I. H. Lee, "A Tracking Method of End-effectors in a Vision-based Marker-free Motion Capture System," *Int. Conf. on Cybernetics and Intelligent Systems*, pp. 129-134, Dec. 2004.
- [16] F. Guo, G. Qian, "Human Pose Inference from Stereo Cameras," *IEEE Workshop on Applications of Computer Vision*, pp. 37-37, Feb. 2007.
- [17] A. Sundaresan, R. Chellappa, "Model Driven Segmentation of Articulating Humans in Laplacian Eigenspace," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 30, No. 10, pp. 1771-1785, Oct. 2008.
- [18] H. Ohbuch, J. K. Tan, S. Ishikawa, "Recovering Human Actions using Eigenspace," *Int. Joint Conf. on Society of Instrument and Control Engineers/Institute of Control, Automation and System Engineers*, pp. 2698-2701, Oct. 2006.
- [19] R. Hartley, A. Zisserman, "Multiple View Geometry," *Cambridge University Press*, 2000.
- [20] D. Tolani, A. Goswami, N. I. Badler, "Real-time Inverse Kinematics Techniques for anthropomorphic limbs," *Graphical Models and Image Processing*, Vol. 62, Issues 5, pp. 353-388, Sept. 2000.
- [21] J. U. Korein, N. I. Balder, "Techniques for Generating the Goal-Directed Motion of Articulated Structures," *IEEE Computer Graphics and Applications*, Vol. 2, Issues 9, pp. 71-81, Nov. 1982.
- [22] <http://www.nbb.cornell.edu/neurobio/land/OldStudentProjects/cs490-97to98/yen/>
- [23] W. H. Press, S. A. Teukolosky, W. T. Vetterling, B. P. Flannery, "Numerical Recipes in C++," *Cambridge University Press*, 2002.

[24] P. Peer, J. Kovac, F. Solina, "Human skin colour clustering for face detection," *IEEE Conf. on Computer as a Tool EUROCON2003*, Vol. 2, pp. 144-148, Sept. 2003.

저 자 소 개



서 용 호(정회원)
 2002년 중앙대학교 전자전기 공학부 학사 졸업.
 2005년 중앙대학교 첨단영상대학원 영상공학과 석사 졸업.
 2005년~현재 중앙대학교 첨단 영상대학원 영상학과 박사과정.

<주관심분야 : 영상처리, 컴퓨터 비전, 3차원 복원, 모션 캡처, HCI>



두 경 수(정회원)
 1989년 중앙대학교 전자공학과 학사 졸업.
 1991년 중앙대학교 대학원 전자공학과 석사 졸업.
 2001년 중앙대학교 대학원 전자공학과 박사 졸업.

2002년~2004년 일본 재)소프트피아 재팬 주임연구원

2005년~현재 중앙대학교 첨단영상대학원 영상학과 전임강사

<주관심분야 : 실시간 영상처리, 휴먼모션, 감시 시스템 개발>



최 중 수(평생회원)-교신저자
 1975년 인하대학교 전기공학과 학사 졸업.
 1977년 서울대학교 대학원 전자공학과 석사 졸업.
 1981년 Keio University(일본) 전기공학과 박사 졸업.

1981년 2월~1981년 9월 Aloka Co. Ltd 연구소 연구원.

1981년 9월~1999년 8월 중앙대학교 전자공학과 교수.

1999년 9월~현재 중앙대학교 첨단영상대학원 영상학과 교수.

<주관심분야 : 영상신호처리, 컴퓨터 비전, 증강 현실, 영상합성, 영상통신>



이 철 우(평생회원)
 1986년 중앙대학교 전자공학과 학사 졸업.
 1988년 중앙대학교 대학원 전자공학과 석사 졸업.
 1992년 동경대학교 대학원 전자공학과 박사 졸업.

1992년~1995년 이미지 정보과학 연구소 수석 연구원 오사카 대학 기초공학부 협력연구원

1995년 리즈메이칸 대학 특별초빙강사

1996년~현재 전남대학교 전자컴퓨터공학부 교수

<주관심분야 : 컴퓨터 비전, 멀티미디어 데이터베이스, 컴퓨터 그래픽스>