

논문 2008-45SP-6-6

모션 캡처를 위한 다시점 영상의 효율적인 분석법

(An Efficient Analysis Method of Multiple View Images for Motion Capture)

서용호*, 박유신*, 구떠올라*, 두경수**, 최종수***

(Yung-Ho Seo, You-Shin Park, Ddeo-Ol-Ra Koo, Kyoung-Soo Doo, and Jong-Soo Choi)

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

컴퓨터 비전 및 컴퓨터 그래픽스 기술의 발달함에 따라 사람의 자연스러운 움직임 및 사실적인 얼굴 표정의 변화가 3차원 가상 캐릭터에 그대로 재현되고 있다. 이러한 모션 캡처 기술은 영화의 특수효과, HCI(human communication interface) 시스템 구축을 통한 게임의

인터페이스, 정형외과 환자의 자동 진단 시스템, 운동선수들의 성취도 향상을 위한 시스템 및 공공장소에서의 군중들의 행동 분석을 위한 감시 시스템 등 다양한 분야에서 활용되고 있으며, 또한 많은 연구가 진행되고 있다^[1~2].

사람의 움직임을 3차원 디지털 데이터로 복원하는 모션 캡처 기술은 인체의 각 관절 부위에 기계장치 부착을 통해 취득하는 기계식, 자기장을 발생하는 센서를 부착하고 그 변화를 측정하는 자기식, 인체의 특정 부위에 마커를 부착하여 취득된 다시점 영상으로부터 3차원 좌표를 추정하는 광학식 방법 등이 있다^[3]. 이러한 시스템은 고가의 하드웨어 장비, 사용자 움직임의 제약 및 사용자의 복잡도 등의 여러 가지 단점들을 가지고

* 학생회원, ** 정회원, *** 평생회원, 중앙대학교
첨단 영상대학원 영상학과
(Dept. of Image Engineering, GSAIM, Chung-Ang
University)

※ 본 연구는 서울시산학협력사업(과제번호 10570),
서울시기술이전사업(과제번호 TR080601), 2단계
BK21 지원과제임.

접수일자: 2008년8월2일, 수정완료일: 2008년10월9일

있다. 따라서, 하드웨어 장치로부터 독립되고, 자유로운 사용자의 모션 캡처를 위해 마커를 사용하지 않은 영상 기반 움직임 해석에 대한 연구가 진행되고 있다.

영상 기반 모션 캡처에 대한 연구는 단일 카메라로부터 훈련된 데이터(training data)를 이용^[4], 동영상으로부터 시간적 및 공간적 제약을 통하여 특징 공간을 제한^[5], 3차원 인체 모델을 영상에 투영^[6], 인체의 각 부위를 검출하는 매커니즘^[7] 등 여러 가지 방법으로 진행되고 있다. 하지만, 실시간으로 사람의 자세를 추정하기에는 불가능하거나 계산량이 많다. 또한, 동작자의 움직임 제약조건으로 인해 다양한 동작에 대해서 자세 추정에 대한 어려움이 존재한다.

따라서, 다시점 영상을 사용하여 다각도에서 취득한 영상으로부터 동작자의 자세 추정에 대한 연구가 활발히 진행되고 있다. 인체의 각 부위에 마커(illuminated landmark)를 부착^[8]하거나 인체의 각 부위를 자동 검출^[9]하여 3차원 위치를 추정하는 방법, 실루엣(silhouette) 영상으로부터 3차원 체적소 상에서 인체의 각 부위를 분할(segmentation)하는 방법^[11], 모션 캡처 데이터로부터 훈련된 모션 DB에 의해 자세를 추정하는 방법^[12], 모션 DB로부터 검색(retrieval)을 통한 자세 추정^[24] 등 다시점 영상의 특징으로부터 동작자의 자세를 추정하는 방법들이 연구되고 있다.

일반적으로 다시점 영상에서 정확한 실루엣을 취득하기 어려우며, 잡음에 상당히 민감하다. 또한 훈련된 DB를 사용하는 방법은 정확한 동작의 추정이 어려우며, 동작자의 움직임에 제한적이고, 많은 DB를 사용하게 될 경우 자세 추정 시 계산량이 많아진다. 또한, 다수의 카메라를 사용하면, 카메라 영상에 대해 각각 처리를 수행해야 하므로 많은 하드웨어가 필요하며, 카메라 영상의 동기화를 맞춰야하는 문제가 발생한다.

마커를 사용하지 않은 영상기반 모션캡처 시스템에서는 객체의 실루엣^[4] 또는 객체의 외곽선^[17], 화소 기반의 특징점^[18], 인체의 얼굴, 손 및 발을 검출^[3]하는 등의 특징들을 추출한다. 영상으로부터 객체의 실루엣 또는 외곽선을 추출하는 방법은 많은 연구가 진행되고 있지만, 실제로 정확한 형태를 추정하기 위해서는 환경적인 제약 및 최적화를 위한 연산 수행 시간이 많이 소요된다.

또한 화소 기반의 많은 특징점을 사용하는 경우, 대응점 정합과 같은 어려운 문제점을 안고 있다. 역시, 정확한 대응점을 찾기 위해서는 많은 계산이 필요하다. 따라서, 영상으로부터 동작자의 움직임을 추정하기 위

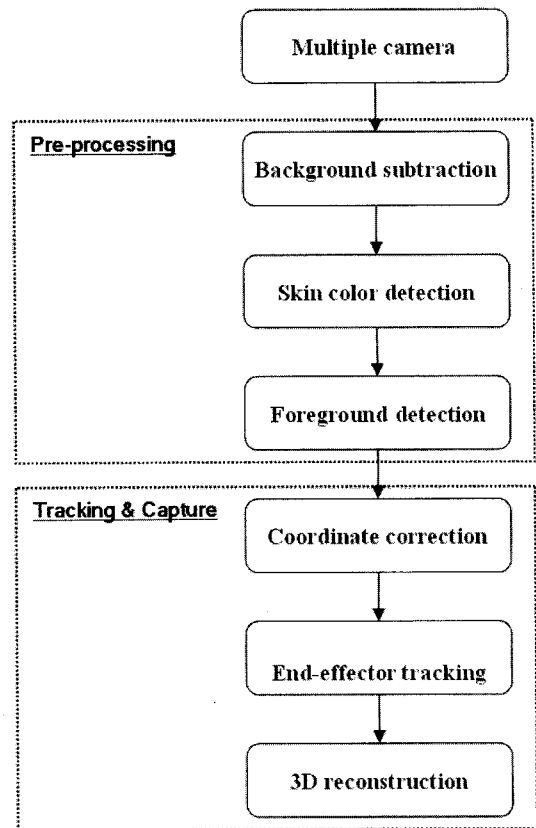


그림 1. 전체 시스템 흐름도
Fig. 1. Flowchart of proposed system.

해 보다 단순하며, 쉽게 추정 가능한 방법이 필요하다. 기존 연구 논문^[15]에서는 얼굴은 템플릿 정합을 이용하여 추출하였고, 손은 피부색을 이용하여 추출하였다. 다수의 영상 및 넓은 영역에 대한 강건한 템플릿 정합은 많은 계산량을 요구한다. 따라서 단순하며 빠르게 찾을 수 있는 인체의 공통된 특징인 피부색과 빠른 모션 캡처를 위한 3차원 복원 알고리즘을 사용한다.

본 논문은 동작자의 모션 캡처를 위한 정확하고, 효율적인 영상 분석 방법을 제안한다. 또한 마커를 사용하지 않으며, 다시점 영상에 대해 실시간 처리가 요구되어야 한다. 먼저, 취득된 다시점 영상을 축소(down sampling)하고, 배경 차분 및 피부색 추출을 통하여 동작자의 각 인체 영역을 검출한다. 다음으로, 카메라 보정 매개변수를 이용하여 검출된 2차원 영상 좌표를 보다 정확한 좌표로 보정한다. 보정된 좌표에 대한 칼만 필터(kalman filter)^[13] 및 각 부위별 특징을 이용하여 추적한다. 마지막으로, 카메라 보정에 의해 얻어진 투영 행렬과 추적된 좌표로부터 동작자의 각 인체 영역에 대한 모션 캡처를 얻는다.

다시점 영상 분석을 통하여 동작자의 부위 검출 시 계산량을 줄이며, 모션 캡처 시 정확성을 높이고, 각 부

위를 추적함으로써 동작자의 움직임을 안정적으로 추정한다. 또한, 제안된 방법을 통하여 시스템의 자원(resource)을 줄일 수 있으며, 실시간 모션 캡처가 가능하다.

본 논문의 구성은 II장에서 카메라 보정 및 에피플라 기하학 정보 추출에 대해 기술하며, III장에서는 인체의 각 부위를 검출 및 추적하는 방법에 대해 기술하며, IV장에서는 제안된 방법에 의한 실험 결과를 보여주며, 마지막으로 V장에서는 결론을 맺는다.

II. 카메라 보정 및 에피플라 기하학

이 장에서는 두 대의 카메라에 대해 카메라 보정^[14]과 에피플라 기하학 정보를 추정하는 방법에 대하여 기술한다.

1. 내부/외부 매개변수 추정

본 논문에서는 2차원 평면 격자 영상을 이용한 Z. Zhang의 강건한 카메라 보정 방법^[14]을 사용하여 두 카메라의 투영행렬을 추정한다. 기존 연구 논문^[15]에 자세히 기술되어 있다.

투영 행렬은 3차원 공간 상의 한 점이 2차원 영상 평면에 투영되는 관계를 표현한다. 수식 (1)은 투영 행렬을 내부 및 외부 매개변수로 분해해서 보여주고 있다.

$$P = K[R|T] \quad (1)$$

실세계의 2차원 평면을 $Z = 0$ 으로 가정하고, 영상 내의 좌표와 2차원 평면 좌표계 사이의 호모그래피 H를 계산한다.

$$s \begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix} = K \begin{bmatrix} r_1 & r_2 & t \end{bmatrix} \begin{bmatrix} X_i \\ y_i \\ 1 \end{bmatrix} \quad (2)$$

$$s\tilde{\mathbf{x}} = H\tilde{\mathbf{X}} \text{ with } H = K \begin{bmatrix} r_1 & r_2 & t \end{bmatrix} = \begin{bmatrix} h_1 & h_2 & h_3 \end{bmatrix}$$

호모그래피의 H 성분 중 회전행렬(rotation matrix)의 성질과 내부 매개변수의 특성을 이용하여 유일한 매개변수를 추정한다.

2. 렌즈 왜곡에 의한 매개변수 보정

실험에는 초점거리 6mm, 4mm인 렌즈를 사용하므로, 영상의 외곡으로 갈수록 방사왜곡(radial distortion)이 심해진다. 방사왜곡을 고려하지 않는다면, 영상의 주점(principal point)으로부터 멀어질수록, 직선이 곡선의

형태로 보이게 된다^[15~16].

이러한 방사왜곡을 교정하기 위해서는 격자 영상을 이용하여, 수식 (3)에 의해서 왜곡 계수 k_1, k_2 를 고려하여 좌표계를 수정한다^[16]. 방사 왜곡을 보정하기 위해서는 계수를 2개만 써도 충분하다^[23].

$$\begin{bmatrix} (u-u_0)(x^2+y^2) & (u-u_0)(x^2+y^2)^2 \\ (v-v_0)(x^2+y^2) & (v-v_0)(x^2+y^2)^2 \end{bmatrix} \begin{bmatrix} k_1 \\ k_2 \end{bmatrix} = \begin{bmatrix} u_{rd} \\ v_{rd} \end{bmatrix} \quad (3)$$

(u, v) 는 왜곡이 없는 이상적인 영상의 좌표, (x, y) 는 왜곡이 없는 정규화된(normalized) 영상의 좌표를 의미한다.

또한, 렌즈 표면의 곡률의 중심이 항상 동일 평면에 있는 것은 아니다. 이것은 접선 왜곡(tangential distortion)이라고 하며, 수식 (4)와 같이 정의된다^[23]. p_1, p_2 는 접선 왜곡 계수를 의미한다.

$$\begin{bmatrix} 2(u-u_0)(v-v_0) & ((x^2+y^2)+2(u-u_0)^2) \\ ((x^2+y^2)+2(v-v_0)^2) & 2(u-u_0)(v-v_0) \end{bmatrix} \begin{bmatrix} p_1 \\ p_2 \end{bmatrix} = \begin{bmatrix} u_{td} \\ v_{td} \end{bmatrix} \quad (4)$$

위에서 기술한 두 가지 왜곡에 대해 고려하면 수식 (5)와 같다. (\tilde{u}, \tilde{v}) 는 취득된 영상의 좌표이다.

$$\begin{bmatrix} u_{rd} \\ v_{rd} \end{bmatrix} + \begin{bmatrix} u_{td} \\ v_{td} \end{bmatrix} = \begin{bmatrix} \tilde{u} - u \\ \tilde{v} - v \end{bmatrix} \quad (5)$$

3. 최대우도추정(maximum likelihood estimation)

렌즈의 방사왜곡으로 인해 영상의 좌표계 또한 왜곡되었기 때문에, 추정된 내부 및 외부 매개변수 또한 교정이 필요하다. 따라서 최대우도 추정법을 적용하여, 내부 및 외부 매개변수, 방사왜곡 계수, 영상의 교정된 좌표들을 최적화해야한다. 영상에서 추출한 격자의 각 좌표와 공간상의 격자를 영상에 투영한 좌표 사이의 차이를 최소화함으로써, 최적의 매개변수 및 좌표들을 추정한다^[16].

$$\sum_{i=1}^n \sum_{j=1}^m \left\| \mathbf{m}_{ij} - \tilde{\mathbf{m}}(\mathbf{K}, k_1, k_2, p_1, p_2, \mathbf{R}_i, \mathbf{t}_i, \mathbf{M}_j) \right\|^2 \quad (6)$$

\mathbf{m}_{ij} 는 i 번째 영상에서 j 번째 격자 영상의 좌표를 의미하고, \mathbf{R}_i 와 \mathbf{t}_i 는 i 번째 영상의 외부 매개변수인 회전행렬과 이동행렬을 의미하며, \mathbf{M}_j 는 j 번째 격자 영상의 3차원 공간 상의 좌표를 의미하고, $\tilde{\mathbf{m}}(\dots)$ 는 격자의 각 좌표들이 추정된 투영행렬에 의해 영상에 투영되는

좌표를 의미한다.

4. 에피폴라 제약조건(epipolar constraints)

두 대의 카메라 사이의 에피폴라 기하학은 각 카메라 중심을 잇는 기준선(baseline)이 포함된 수많은 평면들과 각 영상 평면이 교차하게 된다는 것을 의미한다^[14]. 이러한 특징은 특징점 정합(matching) 또는 정합을 위한 제약조건(constraints)로 사용된다.

정확한 대응점이 추출되었다면, 각각의 대응점은 각각의 에피폴라선 위에 존재하게 된다. 즉, 왼쪽 영상의 한 점 x 에 대응점은 오른쪽 영상에 대응되는 에피폴라선(l') 위에 존재하게 된다. 수식 (7)은 이러한 관계를 설명해주는 기본행렬(F : fundamental matrix)를 보여주고 있다^[14].

$$l' = Fx \tag{7}$$

위와 같은 성질에 의해 대응점 정합의 오차는 수식 (8)에 의해 얻어진다. N 개의 대응점에 대해서, 각 대응점과 대응되는 에피폴라선 사이의 거리의 평균을 평균 에피폴라 거리 오차로 정의한다^[14].

$$\frac{1}{N} \sum_i^N d(x'_i, Fx_i)^2 + d(x_i, Fx'_i)^2 \tag{8}$$

III. 인체 영역 추출 및 추적

이 장에서는 각 카메라 영상으로부터 머리, 손, 발에 해당되는 인체의 각 영역을 추출하고, 추적하는 방법에 대해 기술한다.

1. 인체 후보 영역 추출

실시간으로 입력되는 카메라 영상으로부터 동작자의 각 인체 영역을 추출하기 위해 인체 영역의 후보들을 추출해야 한다. RGB 입력 영상의 배경 차분을 통하여 동작자로 인식되는 후보 영역을 추출한다. 영상의 배경 차분 과정을 수행했다 하더라도, 동작자의 영역 외에 조명에 따라 변화된 주변 환경, 그림자 영역 및 잡음 등이 포함되어 있다. HSI 색 공간에서 색상(hue)과 채도(saturation)를 이용하여 그림자의 영역 및 조명에 의한 주변 환경을 제거한다.

위와 같은 영상처리 과정을 수행하면, 동작자의 영역 외에 조명의 영향으로 인한 잡음 등이 포함되어 있다. 본 논문에서는 동작자의 영역 중에서 피부색으로 표현

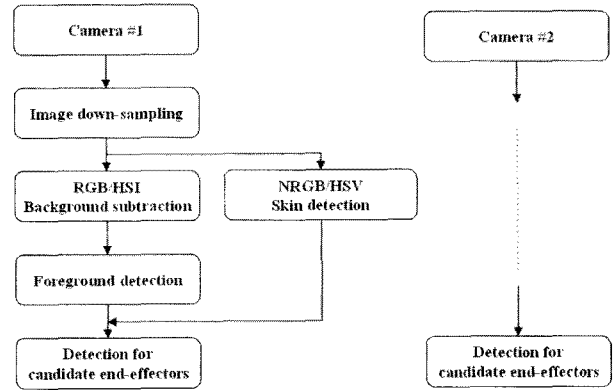


그림 2. 인체 후보 영역 검출
Fig. 2. Detection for candidate end-effector of human.

되는 머리, 손, 발의 위치를 추출하고, 실시간으로 동작자의 모션 캡처를 복원하는데 목적이 있다.

따라서 카메라로부터 입력 받은 영상에 대해 1/2 축소하고, RGB와 HSI 색 공간을 이용하여 추출된 영상에 대해 사람의 피부색 특성을 추출한다. 영역의 크기가 작은 잡음의 경우, 모폴로지를 이용하여 쉽게 제거가 가능하다. 하지만, 조명 변화 및 객체의 움직임 변화에 따라 배경 영역에서도 피부색으로 판단되는 영역이 존재한다. 따라서 일정 크기 이상의 영역은 모두 후보영역으로 지정한다.

피부색은 RGB, HSI, YCbCr 등의 색 공간을 바꿔서 경험적인 방법에 의해서 추출한다^[19,20]. 또한, 정확한 피부영역을 추정하기 위해서는 위의 방법을 조합하거나 적응적 알고리즘 등을 이용한다^[20,21]. 본 논문에서는 조명 변화에 둔감하고 빠른 수행 속도를 위해 정규화된 RGB 색 공간을 사용하여 추출한다^[19].

2. 인체 영역 추출 및 추적

다시점 영상을 사용하여 다중 객체를 추적하는 경우, 각 카메라 영상에서 취득되는 객체들을 정합하는 과정이 필수적이다. 본 논문에서는 동작자의 초기 자세로부터 후보 영역 중 인체 영역 추출 및 정합하는 방법을 사용한다.

그림 3은 검출된 5개의 인체 영역을 보여주고 있다. 그림 3(a)와 그림 3(b)는 두 대의 카메라로부터 취득된 각각의 인체 후보 영역을 보여주고 있으며, 그림 3(c)와 그림 3(d)는 찾아진 후보 영역에 대하여 동작자의 초기 자세 정합에 의해 얻어진 5개의 인체 영역을 검출하고, 원래의 입력 영상에 합성하여 보여준 결과이다. 그림 3(a)와 그림 3(b)에서와 같이 후보 영역은 실제 인체의 영역과 그 외 잡음 등도 동시에 취득된다. 일단, 동작자

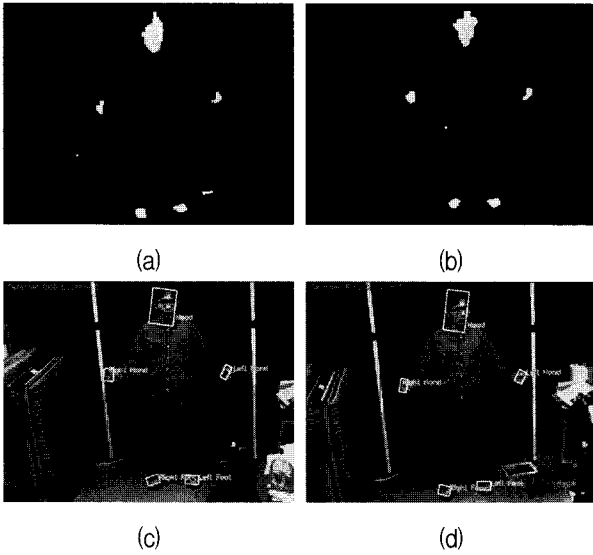


그림 3. 인체 영역 검출

(a), (b) 검출된 인체 후보 영역
(c), (d) 검출된 영역을 원래 입력 영상에 합성

Fig. 3. End-effectors detection of human.

(a), (b) Detected candidate end-effectors
(c), (d) Overlapped image for detected end-effectors

는 카메라를 향해 T 자세를 취한다. 다음으로, 인체의 후보 영역을 추출하고, 영역의 크기에 따라서 정렬한다. 일반적으로, 잡음에 해당되는 부분은 영역의 크기가 후보 영역보다 작다. 마지막으로, 각 후보 영역의 위치 정보를 이용하여 머리, 양손, 양발의 위치를 설정한다.

따라서 동작자의 초기 T자세로부터 사람의 인체 구조적 특징을 사용하여 각 카메라 영상의 인체 영역을 결정하며 정합한다.

앞의 과정을 통하여, 초기 자세로부터 인체의 5개 특징 영역을 추출하고, 정합을 수행하였다. 하지만, 동작자의 움직임에 따라 5개의 특징 영역을 추출 및 정합하는 것은 쉽지 않다. 따라서 일반적으로 추적 알고리즘에서 많이 사용되는 칼만 필터를 이용하여 추적한다.

칼만 필터는 이산 신호 처리에서 이전 상태를 기반으로 이후 상태를 추정하기 위해 일반적으로 사용된다. 수식(9)와 수식(10)은 각각 시스템 방정식과 관측 방정식을 보여주고 있다. x 는 해당 시간의 알고자 하는 변수들의 상태 벡터를 의미하며, z 는 해당 시간의 변수들의 관측 벡터를 의미하고, A 는 이전 상태와 현재 상태의 관계를 정의해주는 상태 천이행렬을 의미하며, H 는 이전 상태와 현재 관측치와의 관계를 정의해 주는 관측 행렬이고, w 와 v 는 각각 시스템 오차, 관측 오차를 의미한다. w 와 v 는 평균값이 0이며, 분산이 각각 Q , R 을

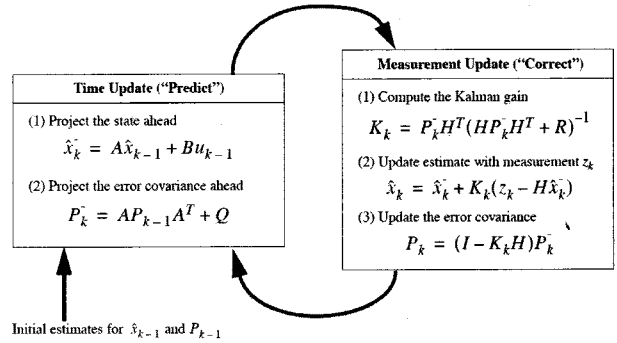


그림 4. 칼만 필터 수행 흐름도

Fig. 4. Diagram of the operation of the Kalman filter.

갖는 정규확률분포를 따른다.

수식 (11)은 2차원 영상에서 칼만 필터를 적용하기 위한 상태 천이행렬 A 와 관측 행렬 H 를 표현한 것이다.

$$x_{k+1} = Ax_k + w_k \quad p(w) = N(0, Q) \tag{9}$$

$$z_{k+1} = Hx_k + v_k \quad p(v) = N(0, R) \tag{10}$$

$$A = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad H = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \tag{11}$$

그림 4는 칼만 필터가 실제 수행되는 흐름도를 보여 주고 있다. 예측 부분에서 Bu_{k-1} 은 선택적 추가 입력으로서, 본 논문에서는 이 값을 “0”으로 설정하였다. P_k 와 P_k 는 각각 이전 상태 벡터와 현재 상태벡터에 대한 연역적(a priori) 추정 오차 분산과 귀납적(a posteriori) 추정 오차 분산을 의미한다.

칼만 필터는 시스템의 상태 벡터를 추정하는 예측 과정과 관측치로부터 칼만 필터를 갱신하는 교정과정으로 구성된다. 본 논문에서는 두 대의 카메라를 사용하고, 각 5개의 인체 영역을 추출하므로, 독립된 10개의 칼만 필터를 구성한다.

카메라에서 취득된 영상으로부터 각 인체 영역들이 추출되거나 짧은 시간동안 일부 영역들이 사라지거나 겹쳐지는 경우에는, 칼만 필터를 통하여 추적이 가능하다. 그림 5는 추출된 객체가 연결되어 보이거나 자기폐쇄(self-occlusion)가 발생할 경우, 칼만 필터를 통한 추적 예이다.

인체 영역이 실제로 겹쳐지거나 영상처리 이후 겹쳐지게 보이는 현상이 생기더라도, 짧은 시간 동안 겹쳐지기 때문에 칼만 필터를 이용하여 쉽게 추적이 가능하

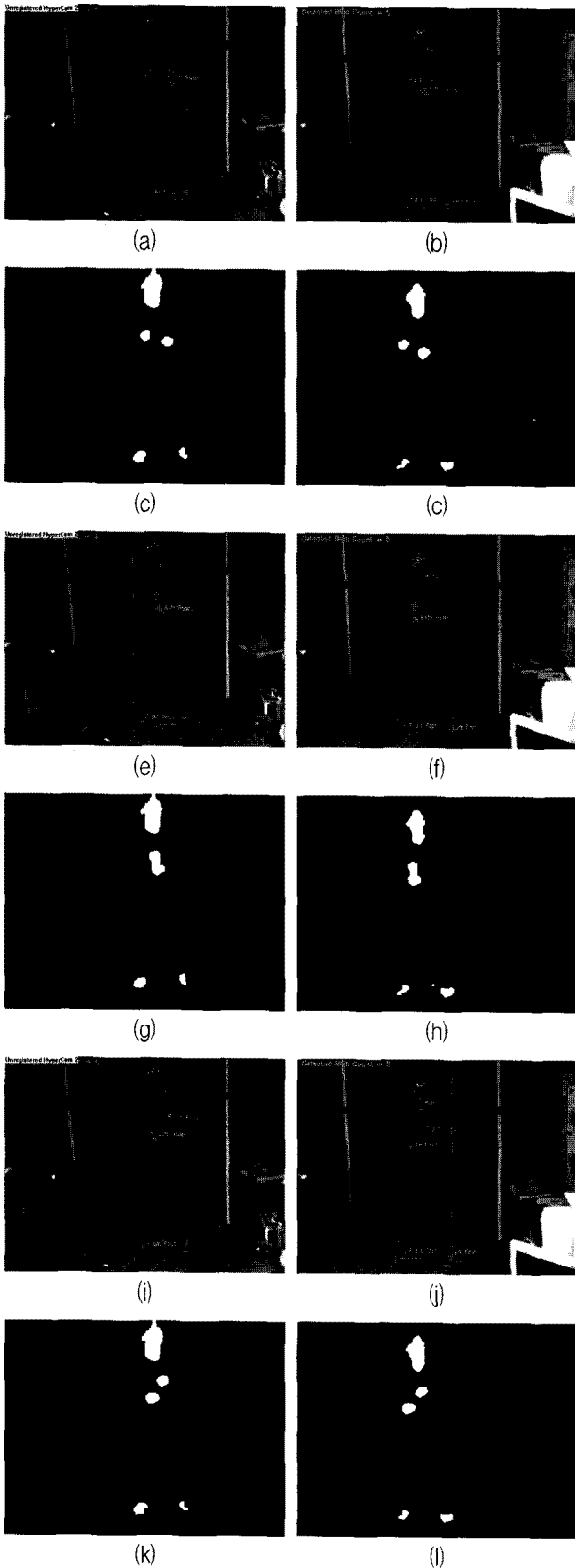


그림 5. 인체 영역 추적

(a), (b), (c), (d) 34번 프레임 (e), (f), (g), (h) 36번 프레임 (i), (j), (k), (l) 38번 프레임

Fig. 5. End-effectors tracking of huma7n.

(a), (b), (c), (d) 34th frame (e), (f), (g), (h) 36th frame (i), (j), (k), (l) 38th frame

다. 인체 영역의 안정적인 추적을 위해, 후보 영역 중에서 이전 프레임에서 찾아진 영역의 면적 및 평균 색상에 대한 정보와 칼만 필터를 이용한다.

3. 모션 캡처 및 오차 보정

각 카메라 영상으로부터 동작자의 인체 영역을 추출하고, 추적 알고리즘을 사용하여 동작자의 움직임에 따른 각 인체 영역을 추적하였다. 현재 얻어진 2차원 영상 좌표계에서 렌즈의 왜곡이 보정되지 않았기 때문에, 3차원 상에서 좌표계로 변환하면 오차가 커지게 된다. 따라서, 렌즈의 왜곡에 대한 좌표 보정이 필요하다. 보정된 좌표(corrected coordinate)와 보정된 카메라 매개변수(calibrated camera parameters)를 사용하여 수식 (8)에 의해 3차원 공간 상에서 실 좌표를 얻는다. 이것은 선형 삼각화 방법으로서 2개 이상의 영상으로부터 3차원 복원을 위해 사용되는 방법이다^[14].

따라서 각 영상에서 취득된 5개의 인체 영역에 대하여 각각 좌표 보정 이후, 수식 (12)를 사용하여 동작자의 모션 정보를 얻어낸다. (x,y) 는 왼쪽 영상의 좌표를 의미하며, (x', y') 은 오른쪽 영상의 좌표를 의미한다. p^{1T}, p^{2T}, p^{3T} 는 왼쪽 영상의 투영행렬에서의 각각 행벡터를 의미하고, $p'^{1T}, p'^{2T}, p'^{3T}$ 는 오른쪽 영상의 투영행렬에서의 각각 행벡터를 의미한다. (X, Y, Z) 는 얻고자 하는 3차원 공간상의 점을 의미하고, 행렬 A, X 는 수식을 단순화하기 위해 정의한 행렬이다.

$$AX = \begin{bmatrix} xp^{3T} - p^{1T} \\ yp^{3T} - p^{2T} \\ x'p'^{3T} - p'^{1T} \\ x'p'^{3T} - p'^{2T} \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} = 0 \quad (12)$$

IV. 실험결과

본 논문은 실시간 모션 캡처를 위해 2대의 카메라를 사용하여 보정된 카메라 매개변수(calibrated camera parameters)를 사용하여 동작자의 얼굴, 손, 발의 추출 및 추적의 정확도를 향상시키는 방법을 제안한다. 카메라 보정 후 렌즈의 왜곡에 대한 오차를 보여 주고, 동작자의 각 인체 영역의 추정의 오차를 실험한다.

1. 실험 환경

본 논문에서는 실시간으로 모션 캡처 및 캡처 시스템의 계산량 및 단순화를 통해 시스템의 자원을 줄이는

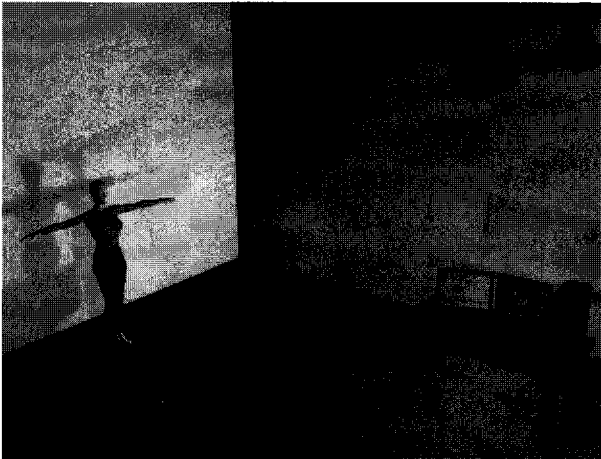


그림 6. 실험 환경

Fig. 6. The experimental environment.

표 1. 알고리즘 수행 시간 결과

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

알고리즘	수행시간
배경 차분	7.43 ms
피부색 추출	2.57 ms
레이블링	1 ms
좌표 보정	0 ms
영역 추적	5 ms
모션 캡처	0 ms

방법을 제안했다. 그림 6과 같이 하나의 Intel(R) Core(TM)2 X6800에 두 대의 카메라를 연결하였으며, 4mm/6mm렌즈를 장착한 두대의 1394B 타입 Dragonfly Express 카메라를 연결하였다.

표 1은 제안된 시스템의 실시간 수행 가능에 대한 실험 결과이다. 수행시간은 1,500프레임을 대상으로, 각각의 단계를 수행한 후, 프레임 당 평균 시간을 측정하였다. 각 알고리즘 수행 시간 결과는 본 시스템은 두 대의 카메라를 동시에 실시간으로 처리가 가능함을 보여준다. 좌표 보정 및 모션 캡처 과정은 5개의 대응점에 대해서 선형 방정식을 푸는 문제이므로, 매우 짧은 시간에 처리가 가능하다. 따라서 수행시간이 0ms에 가깝게 얻어졌다.

피부색 추출 알고리즘의 수행시간은 YCbCr, 정규화된 RGB, Peer^[24]에 의한 방법을 수행한 결과 각각, 3.0ms, 2.57ms, 2.79ms가 소요되며, YCbCr과 정규화된 RGB를 통해 추출한 피부색은 비슷한 결과를 얻었다. Peer는 RGB 색 공간에서 R, G, B 사이의 상관 관계를 이용하여 피부색을 정의한다. 또한 레이블링 알고리즘의 경우, 런길이 부호화 방법은 1ms가 소요되지만,

GrassFire방법은 1.89ms가 소요되었다. 10개의 인체 영역의 추적하기 위한 갈만 필터의 응답 및 계산 시간은 5ms가 소요되었다.

2. 카메라 보정 및 왜곡

실험에 사용된 두 대의 카메라에 대해, 미리 카메라 보정 알고리즘을 적용하여, 각 카메라의 내부 및 외부 매개변수와 렌즈 왜곡 계수들을 추정하였다. 카메라 보정에 사용되는 격자 영상은 간격 10cm이고, 크기 0.5m×1.5m의 격자 무늬를 사용하였다.

그림 7은 3차원 실세계 격자의 각 교차점을 영상에 재투영한 것이다. 그림 7(a)와 그림 7(b)는 각 카메라에 대해 격자를 재투영한 후, 재투영된 점과 원래 영상에서 취득된 점 사이의 유클리디안 거리이다. BLD와 ALD는 렌즈의 왜곡을 보정하기 전과 렌즈의 왜곡을 보정한 후를 의미한다. 렌즈의 왜곡을 보정하기 전에는 각각 평균 오차가 0.73화소, 0.95화소이었지만, 렌즈의 왜곡을 보정한 후에는 각각 0.2화소, 0.24화소로 오차는 약 74%로 줄어들었다.

그림 8은 기본행렬의 방사왜곡 교정을 통한 에피폴라 거리 오차를 보여주고 있다. 그림 8(a)는 RANSAC에 이용하여 추정한 기본행렬에 대한 결과를 보여주고 있으며, 그림 8(b)는 LMedS를 이용하여 추정한 기본행렬에 대한 결과를 보여주고 있다. 일반적으로, 정확한 기본행렬을 추정하는 방법은 반복에 의한 선형 최적화 방법, M-추정자, LMedS, RANSAC과 같은 강건한 방법들을 사용한다^[25]. 우리는 알고 있는 격자 영상을 사용하기 때문에, 모든 대응점은 정확한 정합이 된다. 따라서, 대응점 사이에 이상치(outlier)는 존재하지 않는다. 두 방법 모두 렌즈의 왜곡을 보정하기 전에는 각각 0.23화소, 0.21화소이었지만, 렌즈의 왜곡을 보정한 후에는 0.04화소, 0.07화소로 오차는 약 75%로 줄어들었다.

그림 9는 3차원 보정틀(3D frame)^[9]과 평면 격자 영상(planar pattern)을 이용한 카메라 교정 오차에 대한 실험결과이다. 3차원 보정틀을 사용하는 경우 단일 영상으로부터 카메라 교정이 가능하지만, 평면 격자 영상의 경우에는 다수의 방향에서 취득된 영상이 필요하다. 하지만, 보정틀의 경우 소수의 특징점 및 영상 전체의 특징을 알 수 없으므로, 영상의 외곽으로 갈수록 생기는 렌즈의 왜곡에 대해서 오차가 커진다. 영상의 가운데 부분에서는 두 방법 모두 오차가 작지만, 영상의 외곽 부분에서는 보정틀을 사용하는 경우 현저히 커진 것을 볼 수 있다.

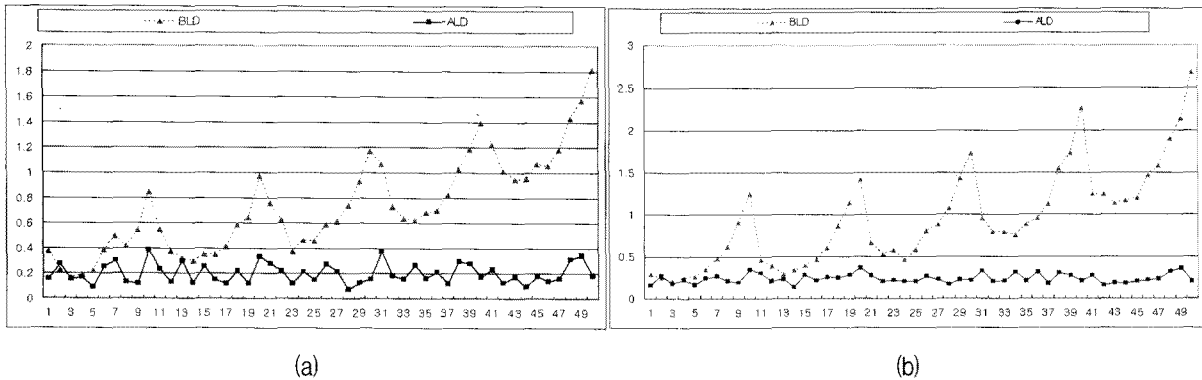


그림 7. 카메라 투영행렬의 방사왜곡 교정 실험

(a) 카메라 #1 (b) 카메라 #2

Fig. 7. The experimental result of lens distortion for camera projection matrix.

(a) camera #1 (b) camera #2

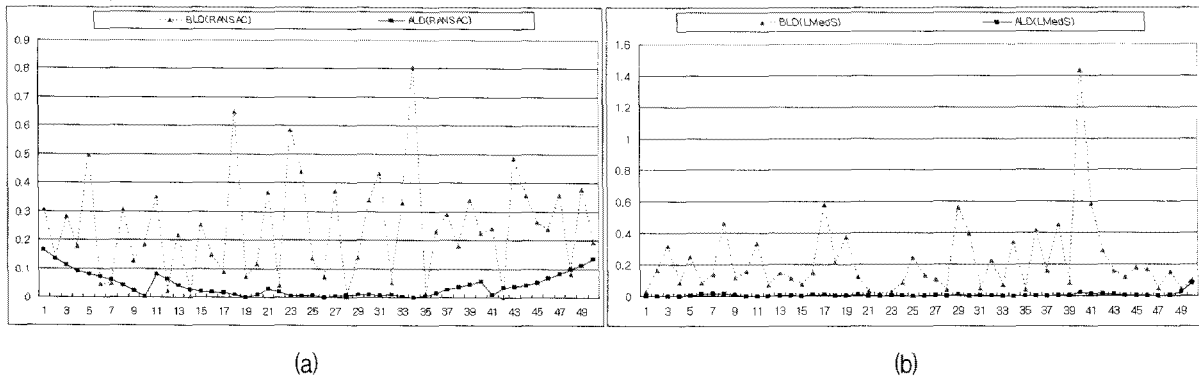


그림 8. 기본행렬의 방사왜곡 교정 실험

(a) RANSAC에 의한 방법 (b) LMedS에 의한 방법

Fig. 8. The experimental result of lens distortion for fundamental matrix.

(a) a method by RANSAC (b) a method by LMedS

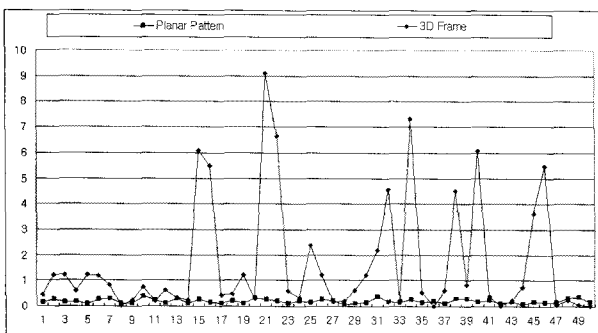


그림 9. 평면 패턴 및 보정틀에 대한 오차 실험

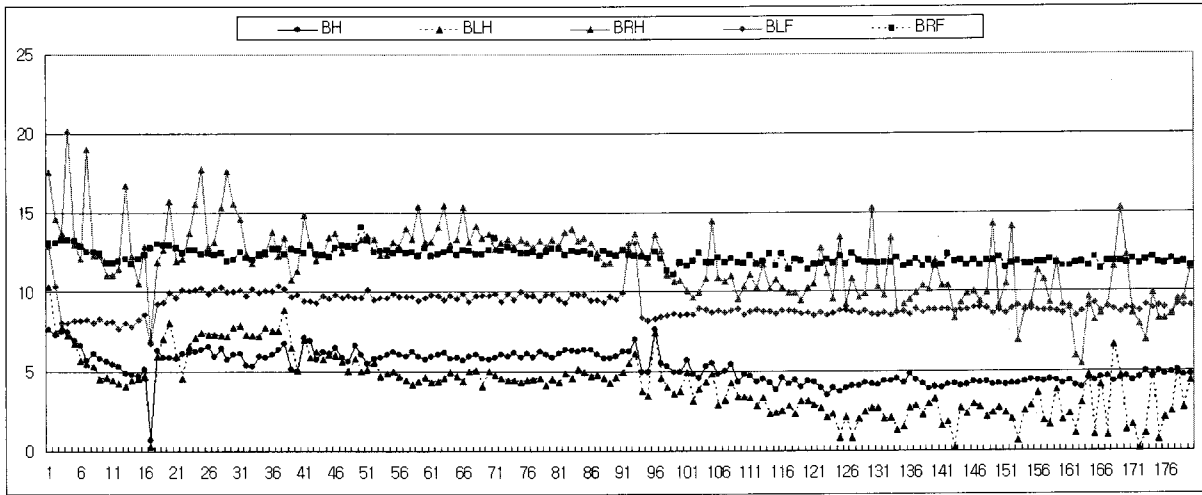
Fig. 9. The experimental result of calibration for both planar pattern and 3D frame.

GT는 참값(ground truth)를 의미한다. 렌즈의 왜곡을 보정하기 전에는 평균 13.8mm, 표준편차 15.4mm의 오차가 발생하지만, 렌즈의 왜곡을 보정한 후에는 평균 6.4mm, 표준편차 6.3mm의 오차가 발생하였다. 렌즈의 왜곡을 보정함으로써 오차는 절반 이하로 줄었다.

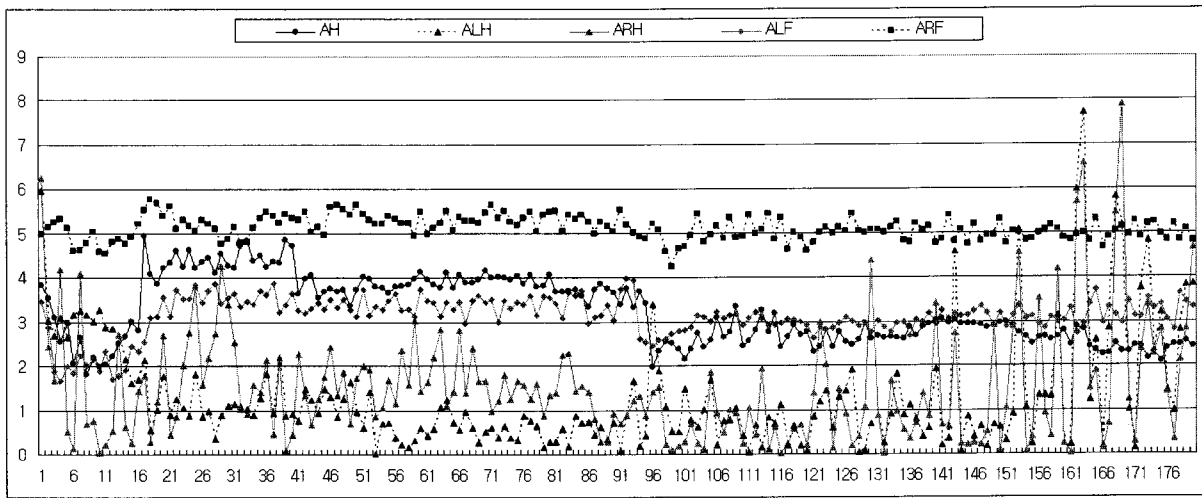
그림 10은 추정된 머리(H), 손(LH, RH), 발(LF, RF)에 대해 180프레임 동안 에피폴라 거리 오차를 구한 것이다. 그림 10(a)는 렌즈의 왜곡을 보정하기 전이며, 오차의 평균은 7.96화소, 표준편차는 1.12화소이다. 그림 10(b)는 렌즈의 왜곡을 보정한 후의 결과로, 오차의 평균은 2.83화소, 표준편차는 0.58화소이다. 따라서, 렌즈의 왜곡을 보정하게 되면, 오차는 65% 감소되었다. 영역에 기반한 객체 추적을 하기 때문에, 특정 부위에서는 에피폴라 거리 오차가 상대적으로 크게 발생했다.

3. 인체 영역 추출 및 추적

피부색을 통하여 인체의 각 부위를 추출하기 때문에, 칼만 필터를 이용하여 동작자의 움직임 변화에 따라 각 부위를 추적했다. 자기폐색(self-occlusion)의 경우에도 칼만 필터가 동작하면, 잘못된 예측값을 얻는다. 따라서, 카메라로부터 5프레임 이상 사라지면 더 이상 칼만



(a)



(b)

그림 10. 인체 영역에 대한 추적 실험 결과
(a) 렌즈 왜곡 교정 전 (b) 렌즈 왜곡 교정 후

Fig. 10. The experimental result of tracking for 5 end-effectors
(a) before we correct lens distortion (b) after we correct lens distortion.

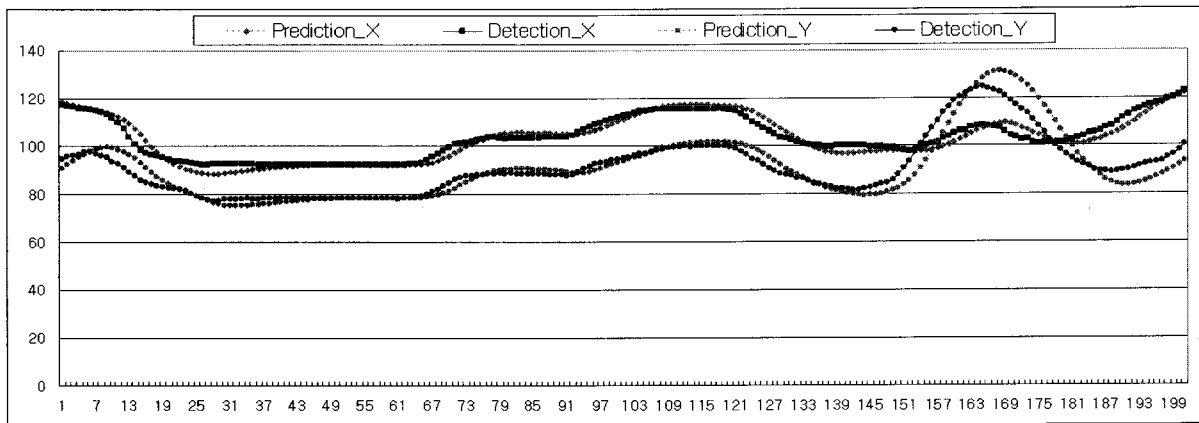


그림 11. 오른손에 대한 칼만 필터의 추적 실험
Fig. 11. The experimental result of Kalman tracking for right hand.



(a)



(b)



(c)



(d)



(e)

그림 12. 다양한 동작에 대한 모션 캡처
Fig. 12. Motion capture for various action.

필터가 동작하지 않도록 했다. 순간적으로 검출이 안되는 경우에도 예측이 가능하며, 또한 카메라에서 사라진 상태로 동작자의 특정 영역이 이동하는 경우 예측이 어려우므로 제약을 두었다.

그림 11은 움직임이 큰 손에 대해서 칼만 필터의 예측치와 검출값을 비교한 것이다. 빠른 움직임의 경우 3-4프레임 정도의 응답 시간이 있지만 정상적으로 인체 영역을 추적하였다. 그림 12는 동작자의 다양한 동작에 따른 인체의 각 부위 검출 및 추적과 모션 캡처 결과를 OpenGL 상에서 출력한 결과이다.

VI. 결 론

본 논문에서는 다시점 영상으로부터 취득된 영상을 이용하여 실시간 모션 캡처를 위한 정확한 영상 분석 방법을 제안했다. 영상 교정과 에피폴라 기하학 정보를 이용하여 검출된 피부 영역의 좌표를 교정함으로써 교정 전 오차보다 65%줄었으며, 각 영상에 대하여 전체 알고리즘을 수행하는데 소요되는 시간은 16ms 정도 소요되었다. 또한, 동작자의 다양한 움직임에도 칼만 추적을 사용하여 동작자의 각 영역을 추적한다. 따라서 제안된 알고리즘은 다시점 영상으로부터 모션 캡처를 위해 정확한 영역 추정 및 시스템의 자원을 줄여 실시간 시스템이 가능했다. 향후에는 제안된 캡처 시스템을 응용하여 인체 전체에 대한 모션 캡처가 가능한 시스템 구성할 계획이다.

참 고 문 헌

- [1] T. B. Moeslund, 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.
- [2] 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.
- [3] 이민기, 박성규, 박근표, 양선우, 이범렬, "모션캡처 기술 동향," *전자통신동향분석*, 제22권, 제4호, 2007.
- [4] A. Agarwal, B. Triggs, "Recovering 3D Human Pose from Monocular Images," *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 28, No. 1, pp. 44-58, Jan. 2006.
- [5] G. Rogez, C. Orrite-Uruñuela, and J. Martínez del Rincón, "A spatio-temporal 2D-models framework for human pose recovery in monocular sequences," *Pattern Recognition*, Vol. 41, pp. 2926-2944, 2008.
- [6] C. Sminichisescu, A. Telea, "Human pose Estimation From Silhouette: A Consistent Approach Using Distanace Level Sets," *Int. Conf. on Computer Graphics, Visualization and Computer Vision*, 2002.
- [7] M. W. Lee, R. Nevatia, "Integrating Component Cues for Human Pose Tracking," *Int Workshop on VS-PETS*, pp. 41-48, Oct. 2005.
- [8] T. Chaichana, M. Sangworasil, C. Pintavisrooj, S. Aootaphao, "Acclerate a Dlt Motion Capture System With Quad-Tree Searching Scheme," *Int. Sym. on Communications and Information Technologies*, pp. 1035-1038, Oct. 2006.
- [9] 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 Cybernatics and Intelligent Systems*, pp. 129-134, Dec. 2004.
- [10] N. Date, H. Yoshimoto, D. Arita, R. I. Taniquchi, "Real-time Human Motion Sensing based on Vision-based Inverse Kinematics for Interactive Applications," *Int. Conf. on Pattern Recognition*, Vol. 3, pp. 318-321, Aug. 2004.
- [11] A. Sundaresan, R. Chellappa, "Model driven segmentation of articulating humans in Laplacian Eigenspace," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 30, Issue 8, 2008.
- [12] F. Guo, G. Qian, "Human Pose Inference from Stereo Cameras," *IEEE Workshop on Applications of Computer Vision*, pp. 37-37, Feb. 2007.
- [13] G. Welch, G. Bishop, "An Introduction to the Kalman Filter," *UNC-Chapel Hill, TR 95-041*, July, 2006.
- [14] R. Hartley, A. Zisserman, "Multiple View Geometry," *Cambridge Univ. Press*, 2000.
- [15] 한윤상, 서용호, 두경수, 김진태, 최중수, "실세계 3차원 기하학 정보를 이용한 지시영역 추정 시스템의 구현 및 응용," *대한전자공학회논문지*, 제45권 SP편 제2호, 2008년 3월
- [16] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 22, No. 11, 1330-1334, 2000.
- [17] P. Azad, A. Ude, T. Asfour, G. Cheng, and R. Dillmann, "Stereo-based Markerless Human Motion Capture for Humanoid Robot Systems," *IEEE Int. Conf. on Robotics and Automation*, pp.

- 3951-3956, April, 2007
- [18] E. Aguiar, C. Theobalt, C. Stoll, H. P. Seidel, "Marker-less 3D Feature Tracking for Mesh-based Human Motion Capture," Workshop on Human Motion, LNCS 4814, pp. 1-15, 2007.
 - [19] L. Sigal, S. Sclaroff, V. Athitsos, "Estimation and Prediction of Evolving Color Distributions for Skin Segmentation Under Varying Illumination," IEEE Conf. on Computer Vision and Pattern Recognition, Vol. 2, pp. 152-159, June. 2000.
 - [20] S. K. Singh, D. S. Chauhan, M. Vatsa, R. Singh, "A Robust Skin Color Based Face Detection Algorithm," Tamkang J. of Science and Engineering, Vol. 6, No. 4, pp. 227-234, 2003.
 - [21] J. Yang, W. Lu, A. Waibel, "Skin-Color Modeling and Adaptation," Asian Conf. on Computer Vision, Vol. 2, pp. 687-694, Jan. 1998.
 - [22] Q. Zhu, C. T. Wu, K. T. Cheng, Y. L. Wu, "An Adaptive Skin Model and Its Application to Objectionable Image Filtering," ACM Int. Conf. on Multimedia, pp. 56-63, Oct. 2004.
 - [23] J. Heikkila, O. Silven, "A Four-step Camera Calibration Procedure with Implicit Image Correction," IEEE Conf. on Computer Vision and Pattern Recognition, pp. 1106-1112, Jun. 1997.
 - [24] H. Ohbuch, J. K. Tan, S. Ishikawa, "Reconverging Human Actions Using Eigenspace," Int. Joint Conf. SICE-ICASE, pp. 2698-2701, Oct. 2006.
 - [25] Y. H. Seo, S. H. Kim, K. S. Doo, J. S. Choi, "Optimal keyframe selection algorithm for three-dimensional reconstruction in uncalibrated multiple images," Optical Engineering, Vol. 47, Issue 5, May, 2008.

저자 소개



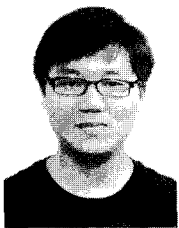
서 응 호(학생회원)
 2002년 중앙대학교 전자전기
 공학부 학사 졸업.
 2005년 중앙대학교 첨단영상대학
 원 영상공학과 석사 졸업.
 2005년~현재 중앙대학교 첨단
 영상대학원 영상학과
 박사과정.

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



박 유 신(학생회원)
 2002년 세명대학교 전기공학과
 학사 졸업.
 2004년 중앙대학교 첨단영상대학
 원 영상공학과 석사 졸업.
 2004년~현재 중앙대학교 첨단
 영상대학원 영상학과
 박사과정.

<주관심분야 : 영상처리, 3차원 영상검색, 컴퓨터
 비전>



구 떠 올 라(학생회원)
 2008년 남서울대학교 멀티미디어
 학과 학사 졸업.
 2008년~현재 중앙대학교 첨단
 영상대학원 영상학과
 석사과정.

<주관심분야 : 영상처리, 모션 캡
 처, 얼굴 인식>



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 종 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>