

# 실감만남 공간에서의 비전 센서 기반의 사람-로봇간 운동 정보 전달에 관한 연구

## Vision-based Human-Robot Motion Transfer in Tangible Meeting Space

최 유 경<sup>1</sup> · 나 성 권<sup>2</sup> · 김 수 환<sup>3</sup> · 김 창 환<sup>4</sup> · 박 성 기<sup>5</sup>

YuKyung Choi<sup>1</sup> · SyunKwon Ra<sup>2</sup> · Soowhan Kim<sup>3</sup> · ChangHwan Kim<sup>4</sup> · Sung-Kee Park<sup>5</sup>

**Abstract** This paper deals with a tangible interface system that introduces robot as remote avatar. It is focused on a new method which makes a robot imitate human arm motions captured from a remote space. Our method is functionally divided into two parts: capturing human motion and adapting it to robot. In the capturing part, we especially propose a modified potential function of metaballs for the real-time performance and high accuracy. In the adapting part, we suggest a geometric scaling method for solving the structural difference between a human and a robot. With our method, we have implemented a tangible interface and showed its speed and accuracy test.

**Keywords:** Tangible Space, Marker-less Human Motion Capture, Motion Imitation, Stereo Vision

### 1. 서론

최근 지능 로봇에 대한 활발한 연구는 로봇의 활용 범위를 넓혀주고 있다. 일반적으로 사람과 로봇의 직접 대면에 의한 서비스 로봇의 응용뿐만 아니라, 로봇이 가상과 실제 환경이 혼합된 환경 (Mixed Reality) 에도 응용될 수 있다. 본 연구에서는 실감 공간 기술 (Tangible Space Technology)<sup>[1]</sup> 이라는 혼합 환경 속에서 로봇이 서로 다른 공간의 사람에 대한 운동 정보 전달 매체로서 기능하는 연구를 다루고자 한다.

실감 공간 기술의 전형적인 적용의 예는 그림 1과 같은 원격 미팅 분야이다. 기존의 원격 미팅 시스템은 얼굴영상과 음성만을 전달하여 기본적인 통신이 되도록 하고 있지만, 인간의 대화법은 언어의 교환에만 의존한 것이 아니라 대화를 나누고 있는 당사자를 둘러싼 상황 정보 (implicit situational information 혹은 context) 를 적극

적으로 활용하기 때문에 화자들이 이와 같은 상황 정보를 공유할 때 보다 효과적인 대화를 나눌 수 있다. 이러한 원격 미팅에서 로봇은 실감 존재의 역할을 할 수 있으며, 로봇은 원격지 대화 상대의 동작에 대한 모사를 필요로 한다고 할 수 있다.

본 연구에서는 이러한 원격 로봇이 실제 사람의 동작 특징들을 표현 할 수 있도록 하기 위해 스테레오 비전 기반의 3D 모션 캡처 (motion capture) 기술을 구현하고자 한다.

기존의 이러한 휴먼 모션 캡처 기술은 움직임을 추출하고자 하는 사람에 마커 (marker) 를 붙여 데이터를 추출하는 마커 기반 휴먼 모션 캡처 (marker human motion capture) 기술과 별도의 마커를 부착하지 않고 비전 센서만을 활용하여 데이터를 추출하는 마커리스 휴먼모션 캡처 (mark-less human motion capture) 기술로 나눌 수 있는데, 본 연구는 마커 부착의 번거로움이 필요 없는 빠른 사용자 변환과 고가 장비의 사용 없이 로봇-사람 간의 정보전달을 위해 마커리스 휴먼 모션 캡처에 기반을 둔 연구를 수행하였다.

모델기반의 마커리스 휴먼 모션 캡처는 지금까지 많은 연구 결과들이 있었다<sup>[2, 3]</sup> 그 중에서 본 연구에서와 같이 스테레오 비전 센서 기반의 대표적인 연구들은 몇

※ 본 연구는 KIST 의 ‘실감공간 기술개발’ 과제의 지원을 받아 수행되었습니다.

<sup>1</sup> KIST/연세대학교 학연협동 석사과정(unizard@kist.re.kr)

<sup>2</sup> 한국과학기술연구원 인지로봇연구단 연구원(sykra@kist.re.kr)

<sup>3</sup> 한국과학기술연구원 인지로봇연구단 연구원(kimsoohwa@kist.re.kr)

<sup>4</sup> 한국과학기술연구원 인지로봇연구단 연구원(ckim@kist.re.kr)

<sup>5</sup> 한국과학기술연구원 인지로봇연구단 연구원(skee@kist.re.kr)

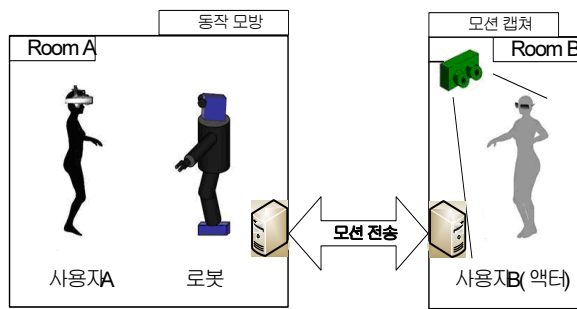


그림 1. 실감 만남 시스템의 구성도

몇 연구자들에 의해서 선행되어왔다<sup>[4-7]</sup>. Demirdjian<sup>[4]</sup>, Plankers/Fua<sup>[5]</sup>는 캡처된 점군데이터와 이를 형상화한 모델 사이의 차이를 최소화 (error minimization) 함으로써 인간의 동작을 추정하는 방법을 각각 제안하였다. 특히, Plankers/Fua는 인체 표면 전체를 하나의 함수로 정의하고 실루엣 추출 기술을 통해 샘플링되어 들어오는 3차원 정보와 모델간 거리의 차를 최소화 하는 방법을 제안하였다. 그러나 Plankers/Fua 연구는 여러 계층으로 구성된 세밀한 인체모델링으로 실시간 모션 캡처가 이루어 질 수 없다는 단점이 있다. 이에 반하여, Demirdjian<sup>[4]</sup>의 연구는 원통이라는 단순한 모델의 사용으로 실시간 모션 캡처는 가능하나 모션캡처의 정확성이 떨어지는 문제점이 존재한다. Steffen Knoop<sup>[6]</sup> 역시 마찬가지로 점군데이터와 모델 사이의 차이를 최소화 하는 방법을 제안하였으며, 특히 3차원 정보만을 사용했을 때 생기는 모션예측의 오차를 개선하고자 2차원 색상정보의 트래킹 방법을 추가적으로 제안하였다. 그러나 Steffen Knoop은 3차원 점군 데이터의 불확실성을 보완하기 위해 별도의 T.O.F (time of flight) 장비를 사용하였다. 마지막으로 L. Molina-Tanco<sup>[7]</sup>는 앞서 설명한 연구들과 같은 오차의 최소화 (error minimization) 방법을 사용하지 않고, 피부색상정보와 템플릿 매칭 (template matching) 을 통해 각 부분 (상박, 하박, 손)를 트래킹 한 후, 그 위치의 깊이 정보로부터 모델의 자세를 추정하는 방법을 제안하였다. 하지만 색상정보와 템플릿 매칭은 조명에 민감하고, 특히 템플릿 매칭은 회전운동과 병진운동에 영향을 받으며 몸의 일부가 가려지는 경우 (occlusion) 매칭에 실패한다는 문제점을 가지고 있어 환경에 많이 제약적이다.

이렇게 기존의 연구들은 데이터의 실시간 처리, 민감한 환경적 제약 그리고 원격 로봇 운동의 모사를 위한 정확도 확보에 어려움이 있어 본 연구와 같은 로봇 응

용에 적용하기에 부적절해 보인다.

따라서 본 연구는 복잡하지 않지만 모든 인체의 동작을 자연스럽게 표현할 수 있고, 모델과 데이터간 오차함수를 최적화할 때 지역적 극소에 빠지지 않는 메타볼의 포텐셜 함수를 새롭게 제안하여 모션 캡처의 정확성을 확보하는 새로운 방법의 스테레오 비전 기반 모션 캡처방법을 제안한다. 또한 실시간성의 확보를 위해 상이도 영상의 격자화, 모델의 기구학적 제약을 적용한 순차적 보간 방법을 추가적으로 제안 하고자 한다. 한편, 사람의 움직임 정보를 로봇에 적용할 때 발생하는 사람과 로봇 사이의 관절의 수, 관절의 길이와 같은 구조적인 문제점을 해결하기 위해, 사람과 로봇 사이의 자세 오차를 최소화 할 수 있는 로봇의 인간 동작 모방 알고리즘을 제안한다.

본 논문의 구성은 다음과 같다. 2장에서는 메타볼과 새롭게 제안된 메타볼의 포텐셜 함수를 설명한다. 3장, 4장에서는 휴먼 팔 모션을 캡처하고 그것을 로봇에 적용하는 모션 모방 시스템의 두 부분을 세부적으로 설명 한다. 그리고 5장에서는 제안된 방법들의 성능을 실험을 통해서 증명하고 6장에서 결론을 맺는다.

## 2. 팔의 수학적 모델링

이번 장에서는 메타볼<sup>[8]</sup>을 사용하여 구현한 3차원 팔 모델을 기술한다. 특히 트래킹 (tracking) 성능을 개선하고, 지역적 극소 (local minima) 에 빠지는 것을 피하기 위해 시스템에 적합한 포텐셜 함수를 새롭게 제안하고자 한다.

### 2.1 팔 모델의 구조

모델링에 사용된 팔 모델의 구조는 그림 2(a)와 같으며 팔 모델은 상박, 하박, 손으로 구성된다. 각 링크는 그림 2(b)처럼 회전운동과 관련하여 각각 3 자유도, 1 자유도, 3 자유도를 갖으며, 7개의 회전각 ( $\theta = [\theta_1, \dots, \theta_7]$ )은 우리가 최종적으로 추정하고자 변수가 된다. 3차원상의 팔 모델을 구현하기 위해 여러개의 메타볼을 사용하며, 하나의 표면으로 매끄럽게 연결되어 있는 팔 모델(isosurface) 은 상박 4개, 하박 4개, 손 1개의 메타볼로 구성한다. 그림 2(b)는 9개의 메타볼로 렌더링(rendering) 된 팔을 보여주고 있다.

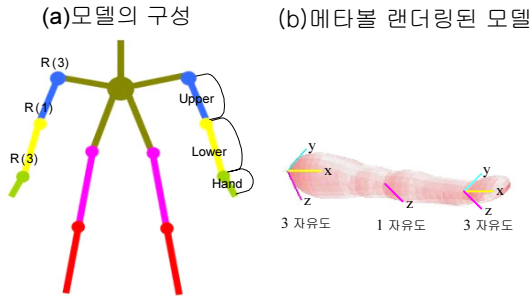


그림 2. 모션 캡처에서 사용된 팔 모델

2.2 메타볼

인체와 같이 가변형 물체는 메타볼을 통해 쉽게 모델링 할 수 있다<sup>[9]</sup>. 특히 메타볼은 적은 양의 데이터로 물체를 잘 표현 할 수 있는 장점을 가지고 있어 복잡한 곡면을 모델링 하는데 있어 널리 사용된다. 메타볼의 표면은 동일한 포텐셜 필드 값을 갖는 점들의 집합으로 정의된다. 따라서 여러 개의 메타볼로 표현된 하나의 팔 모델 S는 다음과 같이 표현된다.

$$S = \{[x, y, z] \in R^3 | F(x, y, z) = C\} \tag{1}$$

$$F(x, y, z) = \sum_{i=1}^n f_i(-2d_i(x, y, z)) \tag{2}$$

여기서,  $f_i$ 는  $i$ 번째 메타볼의 포텐셜 함수 값,  $d_i$ 는 유클리디언 거리, 그리고  $C$ 는 포텐셜 상수 값이다<sup>[9]</sup>.

2.3 제안된 포텐셜 함수

그림 3(b)는 기존에 연구된 여러 개의 포텐셜 함수이고, 그림 3(a)는 함수에 대응되는 그래프이다<sup>[9]</sup>. Fua/Planker<sup>[5]</sup>는 알고리즘의 처리 성능보다는 완벽한 인체 모델링이 연구 목적이기 때문에 Blinn이 제안한 지수 타입의 포텐셜 함수를 사용하였다. 하지만 본 연구는 시스템의 특성상 빠른 처리속도가 필요하기 때문에 연산 양이 많고 복잡하여 처리속도가 떨어지는 지수 타입의 포텐셜 함수 대신 다항식 타입의 포텐셜 함수를 사용하였다.

기존에 제안된 포텐셜 함수들은 메타볼간의 자연스러운 연결 형태를 만들어 가변형 물체를 어색하지 않고 자연스럽게 렌더링 하는 것이 목적이었다. 그러나 본 연구는 컴퓨터 그래픽스 분야의 연구목표인 매끄러운 렌더링보다 팔 모델을 점군 데이터에 좀더 정확하게 보간 하는 것이 연구 목적이므로, 이에 맞는 적합한 형태의 포텐셜 함수를 다음과 같이 제안한다.

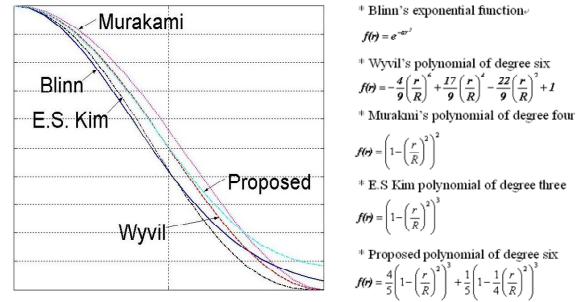


그림 3. (a) 포텐셜 함수의 그래프 (b) 포텐셜 함수의 방정식

$$f(r) = \frac{4}{5} \times (1 - r^2)^3 + \frac{1}{5} \times \left(1 - \frac{1}{4} \times r^2\right)^3 \tag{3}$$

여기서  $r = \sqrt{\frac{(x-x_c)^2}{R_x^2} + \frac{(y-y_c)^2}{R_y^2} + \frac{(z-z_c)^2}{R_z^2}}$  이고,  $x_c, y_c, z_c$ 는 메타볼의 중심 좌표,  $R_x, R_y, R_z$ 는 메타볼의 각축에 대한 반지름의 길이이다.

그림 4(a)와 (b)는 각각 김은석<sup>[9]</sup>의 포텐셜 함수로 만들어진 잔차 곡면 (residual surface) 과 앞서 제안한 포텐셜 함수로 만들어진 잔차 곡면의 한 예이다.

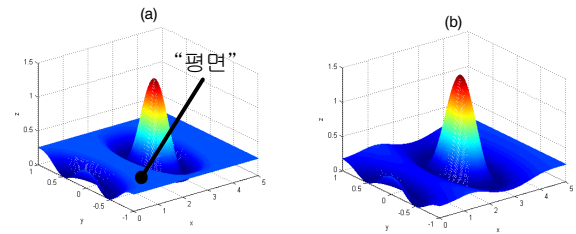


그림 4. (a) 기존 연구의 메타볼로 구성된 점차 곡면 (b) 본 연구에서 제안한 메타볼로 구성된 점차 곡면

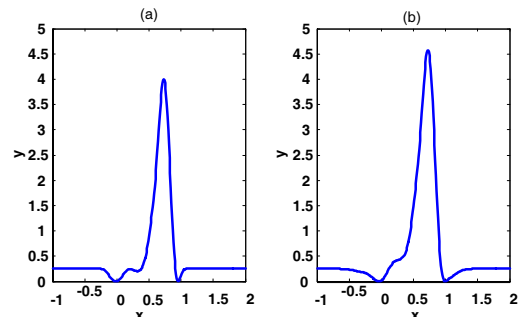


그림 5. (a) 기존 연구의 메타볼로 구성된 점차 함수 (b) 본 연구에서 제안한 메타볼로 구성된 점차 함수

본 논문에서는 잔차가 가장 작은 값이 되는 자세를 찾아 사람의 운동 정보를 추정한다. 김은석의 포텐셜 함수로 만들어진 그림 4(a)의 잔차곡면의 경우, 그림 상에 평평한 구간이 존재하며, 이는 오차의 최소화 문제 (minimization problem) 해결 시에 지역적인 극소 (local minima) 에 빠지는 원인이 된다. 그러나 제안된 포텐셜 함수는 모델 초기화 과정에서 발생할 수 있는 평평한 구간 제거되기 때문에, 지역적 극소를 피하여 쉽게 전역적 극소를 찾을 수 있다.

또한 초기화 과정에서 메타볼이 한쪽으로 편향 모델링 된 경우, 그림 5(a)의  $x = 0.3$  근처에는 지역적 극소 값이 존재하였지만, (b)에서는 지역적 극소 값이 제거될 수 있음을 보여주고 있다.

### 3. 액터의 팔 모션 캡처

모션 모방 시스템은 사람의 모션을 캡처하는 부분과 캡처한 데이터를 원격지의 로봇에 적용하는 모션 모방 부분으로 나누어 볼 수 있다. 이번 장에서는 비전 센서인 스테레오 카메라로부터 사람의 동작 정보를 획득하는 방법을 순차적으로 기술한다.

대부분의 휴먼 모션 캡처 방법은 크게 자세 초기화 (initialization), 트래킹 (tracking), 추정 (estimation), 및 인식 (recognition) 과정으로 나누어 볼 수 있다<sup>[2]</sup>. 물론 기존의 모든 방법론들이 4가지 과정을 모두 포함 하는 것은 아니지만 대체로 위와 같은 범주 안에서 표현 된다. 이와 같은 기능적 분류 중에서 인식 과정은 본 논문의 연구 범위에 해당되지 않는다. 따라서 본 시스템 알고리즘은 인식과정을 제외한 자세 초기화, 트래킹, 추정 과정만으로 구성하였다.

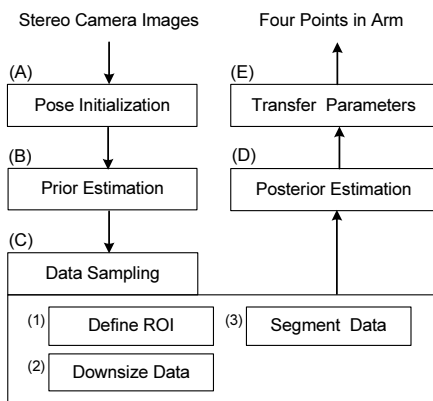


그림 6. 모션 캡처 시스템의 알고리즘 흐름도

그림 6는 본 연구의 알고리즘 흐름도이며, (A)는 자세 초기화, (B)는 트래킹, (C)와 (D)는 추정 과정에 해당된다.

#### 3.1 자세 초기화

자세 초기화 단계는 알고리즘이 시작하면서 마우스를 통해 사용자 입력을 받으며, 이때 액터는 자신의 오른 팔을 몸 앞으로 뻗는 부동 자세를 취해야 한다. 사용자 입력 값은 3차원 위치 정보이며 각각의 벡터는 어깨, 팔꿈치, 손목, 손 끝을 나타낸다. 앞으로 4개의 관절의 위치는  $p_1, p_2, p_3, p_4$  라고 칭하자. 그리고  $p_1$ 을 로컬 좌표계의 원점으로 하고, 좌표축은 팔의 방향을 x축, 팔의 위쪽방향을 y축으로 정한다. 여기서의 모든 과정이 완료되면, 최종적으로 역기구학 문제를 풀어 얻은 7개의 관절 각을 통해 팔 모델은 초기 자세를 취하게 된다.

#### 3.2 사전 추정

사전 추정 과정은 3차원 모델을 상이도 영상 (disparity map) 의 점군 데이터에 보간 하기 전에, 팔의 모션을 예측하여 해당 위치로 모델을 미리 이동시키는 방법이다. 본 연구는 현 과정에서 광학류 (optical flow) 방법과 시간적 이동 (temporal tracking) 방법을 결합하여 사용하였으며, 특히 광학류 방법의 경우에는 시간적 이동방법과 같이 3차원상의 이동경로 예측이 불가능하므로, x, y축에 대한 예측 결과만을 이용하여 모델에 적용하였다.

사전 추정과정은 위에서 언급한 두 방법을 통해 각각 구한  $p_1, p_2, p_3, p_4$  결과의 평균 값을 모델에 적용하게 된다. 그러나 광학류 방법이 영상 특징 (visual feature) 의 부족으로 올바른 결과를 얻을 수 없는 경우에는 시간적 이동 방법만으로 팔 모델의 위치를 사전 추정하게 된다. 비전 센서의 2차원 영상으로부터 현저한 영상 특징을 추출하고<sup>[10]</sup>, Lukas-kanade 광학류 알고리즘<sup>[11]</sup>을 이용하여 광학류를 구한다. 그리고 각 광학류 움직임의 평균 속도 패턴을 이용하여 이동경로를 산출한다. 추출된 영상 특징은 사람의 움직임에 가려질 수 있으므로, 입력 되는 영상마다 새로운 영상 특징을 뽑아 사용하며, 이를 통해 영상 특징의 소실을 방지하였고, 영상 특징 위치의 오차누적도 예방하였다. 광학류 방법에 의해서 시간 t에 추정된 위치 벡터 (positional vector) 를  $\tilde{p}_{OF,t}$  고 하자. 그리고 시간 t-1 에 팔 관절의 상태 벡터 (state vector)  $S_{t-1}$  는 식(4)와 같다.

$$S_{t-1} = [\hat{o}_{t-1}^T, \hat{w}_{t-1}^T]^T \tag{4}$$

여기서  $\hat{\theta}_{t-1}^T$  와  $\hat{w}_{t-1}^T$  는 전 스텝의 사후 추정 (posterior estimation) 을 통해 얻은 회전각과 각속도 값을 나타낸다.

시간적 이동 방법에 의하면, 현재의 회전각  $\hat{\theta}_t^T$  을 추측할 수 있으며, 방법은 식(5)와 같다.

$$\hat{\theta}_t^T = \hat{\theta}_{t-1} + \hat{w}_{t-1} \Delta t \quad (5)$$

여기서는  $\Delta t$  전체 알고리즘을 수행하는데 걸린 시간 간격을 나타낸다.

따라서, 순기구학 (forward kinematics) 을 통해 시간적 이동방법에 의한 팔 모델의 위치를 추정할 수 있고 이것을  $\hat{p}_{TT,t}$  라고 하자. 결론적으로 t 시간의 팔 관절의 위치 추정 값  $\hat{p}_t$  은 다음과 같다.

$$\hat{p}_t = \frac{1}{2} (\hat{p}_{OF,t} + \hat{p}_{TT,t}) \quad (6)$$

### 3.3 데이터 샘플링

본 절에서는 휴먼 팔 모션 캡처의 빠른 연산을 위해 제안하는 데이터 샘플링 방법을 소개한다. 일반적으로 스테레오 카메라를 통해 얻은 상이도 영상은 많은 데이터를 갖는다. 따라서 알고리즘 처리 속도의 향상을 위해서는 점군 데이터의 표본화와 분할 과정이 필요하다. 그러나 사후 과정이 샘플 데이터에 많은 영향을 받으므로, 점군 데이터는 본 과정에서 정확하고 조심스럽게 표본화되어야 한다. 데이터 샘플링 과정은 1)관심 영역의 정의 2)데이터 크게 줄이기 및 3)데이터 추출하기로 구성된다.

1) 관심 영역의 정의: 상이도 영상 위에 존재하는 팔 모델의 기준으로 관심 영역을 찾는다. 우선 깊이 히스토그램 (depth histogram) 을 이용하여 배경 분리를 수행하고, 다음으로 라벨링과 그룹핑을 수행하여 샘플의 수가 가장 많은 영역을 관심 영역으로 지정한다. 하지만 비록 샘플의 수가 작은 영역이더라도, 모델 데이터와 연관성이 높다면 그 작은 영역들을 관심 영역 안에 포함 시킨다.

2) 데이터 크기 줄이기: 실시간 처리라는 시스템 조건을 만족하기 위해, 사후 추정 전에 데이터 크기를 줄이는 과정은 반드시 필요하다. 구체적으로 설명하자면, 3.4 절에서 소개될 ICP (Iterative Closet Point) 알고리즘과 LM (Levenberg-Marquardt) 알고리즘의 경우 반복적으로 모든 데이터와의 연산이 필요하기 때문에, 데이터의 양이 증

가할수록 연산량은 기하 급수적으로 증가하게 된다.

본 과정의 핵심 제안은 먼저 여과 (filtering) 를 통해 잡음으로 판단되는 분산이 큰 데이터를 제거하고, 상이도 위에 영상 5\*5의 평균 마스크를 적용하여 상이도 영상을 격자화하는 것이다. 이제 일정 간격으로 데이터 크기를 줄인 점군 데이터는 격자화된 상이도 영상 (gridded disparity map) 이라고 불리게 된다.

3) 데이터 추출하기: 여기서는 격자화된 상이도 영상의 관심영역에서 팔의 각 부분에 해당되는 점군 데이터를 추출한다. 3.2절에서 팔의 모션을 예측할 때는 부분적으로 2차원 상에서 액터의 팔을 추정하였다. 이는 3차원 팔 모델이 마치 2차원 상에서만 움직이는 것으로 임의 축소해석 한 것이므로 변환된 팔 모델이 정확한 위치에 놓여 있다고 판단 할 수 없다. 따라서 팔 모델이 놓인 위치로부터 각 부분에 해당되는 점군 데이터를 추출할 때는 사전 추정의 오차를 고려하여 메타볼 영역을 약간 확장한 메타볼 인근의 점군 데이터까지 포함한다.

### 3.4 사후 추정

이제 3차원 팔 모델을 샘플링 된 점군 데이터에 보간할 준비가 되었다. 데이터 추출과정을 통해서 m개의 샘플 데이터가 주어졌다고 가정하면,

$$(x_1, y_1, z_1), (x_2, y_2, z_2), \dots, (x_m, y_m, z_m),$$

식(2)의 isosurface 식에 대한 잔차 함수 (residual function) 식은 다음과 같이 표현된다.

$$r_i(\mathbf{x}_i, \boldsymbol{\theta}) = F(\mathbf{x}_i, \boldsymbol{\theta}) - C, (i = 1, \dots, m) \quad (7)$$

여기서  $\mathbf{x}_i = [x_i, y_i, z_i]$  는 샘플데이터의 3차원 좌표,  $\boldsymbol{\theta} = [\theta_1, \dots, \theta_7]$  는 7개의 관절 회전 각도,  $F(\mathbf{x}_i, \boldsymbol{\theta})$  는 팔 모델이  $\boldsymbol{\theta}$  의 값을 갖는 3차원 위치에 존재할 때, 팔 모델과 샘플데이터 사이의 포텐셜 값을 의미한다. 또한, 잔차 벡터를 식(8)로 정의한다.

$$\mathbf{r}(\mathbf{x}, \boldsymbol{\theta}) = (r_1(\mathbf{x}_1, \boldsymbol{\theta}), \dots, r_m(\mathbf{x}_m, \boldsymbol{\theta})), (r : R_n \rightarrow R_m) \quad (8)$$

실제로 모델의 팔 움직임을 데이터에 잘 보간하기 위해서는 식(8)의 잔차 벡터의 길이가 최소가 되는  $\boldsymbol{\theta}$  를 구해야 한다. 즉, 샘플데이터들의 포텐셜 필드 값과  $\boldsymbol{\theta}$  로 결정되는 팔 모델의 움직임 사이의 최적해를 구하여 메타볼로 구성된 팔을 점군데이터에 가까이 접근 시키는 것이다.

$$\text{minimize}=\|\mathbf{r}(\mathbf{x},\boldsymbol{\theta})\|^2 \quad (9)$$

정확한 보간의 해결책인 잔차 벡터의 전역 극소 값 (global minima) 을 구하기 위해 다음과 같은 방법을 적용한다. 먼저 일차적으로 ICP 알고리즘<sup>[12-13]</sup>을 사용하여 모델과 데이터간 최소거리를 추정 한다. 그리고 다음 스텝으로 LM 알고리즘<sup>[14]</sup>을 적용하여 일정한 포텐셜 필드 값으로 정의한 팔 모델과 각 샘플 데이터들의 포텐셜 필드 값의 차이가 최소가 되도록 한다. 즉, ICP 알고리즘을 사용하여 전역 극소 값이 존재할 만한 영역으로 모델을 보간하고, LM 알고리즘으로 더욱 정밀 하게 검색하여 지역 극소 값에 빠지지 않는 잔차 벡터 최소 값을 찾음으로써 올바른 동작 정보를 추정하는 것이다.

### 3.5 파라미터 전송

알고리즘 (A)에서 (D)까지의 과정을 통해 모션 정보를 얻은 후, 이 정보들은 기가비트 이더넷을 통해 원격의 로봇에 전송된다. 전송 전후로는 카메라 좌표계와 로봇 좌표계의 변환이 필요하다.

## 4. 로봇의 액터 동작 모방

유연한 관절을 가진 액터와 여러 개의 강체로 이루어진 로봇의 관절 사이에는 기하학적 그리고 구조적인 차이가 존재한다. 따라서, 모션캡처를 통하여 얻은 액터의 움직임 정보를 로봇에 그대로 적용할 수는 없다.

Riley<sup>[15]</sup> 는 액터의 전신 동작을 좀 더 간단한 역기구학 부분 문제로 나누어서 실시간으로 전환하는 방법을 제안하였다. 그러나 여기서는 액터와 인간형 로봇간의 기하학적 차이가 고려되지 않았다. C. Kim<sup>[16]</sup> 은 최적화의 과정을 거쳐 인간형 로봇이 인간의 동작을 모방하기 위한 방법을 개발하였다. 본 연구는 C. Kim와 유사하게 해를 도출하는 과정으로 최적화를 사용하였지만, 손 끝의 위치와 자세를 이용하지 않고, 어깨, 팔꿈치, 손목, 손의 위치만을 이용하였다는 점에서 상당한 차이가 있다.

### 4.1 알고리즘 흐름도

그림 7은 동작 모방 알고리즘의 데이터 흐름을 보여주고 있다.  $p_1, p_2, p_3, p_4$ 를 각각 액터의 어깨, 팔꿈치, 손목, 손의 3차원 위치벡터라고 하자. 이 벡터들은 액터의 동작을 표현하고 있으며 원격지의 로봇에 전달된다. 이 벡터정보가 기하학적 변환 과정을 거치면, 로봇에 적용

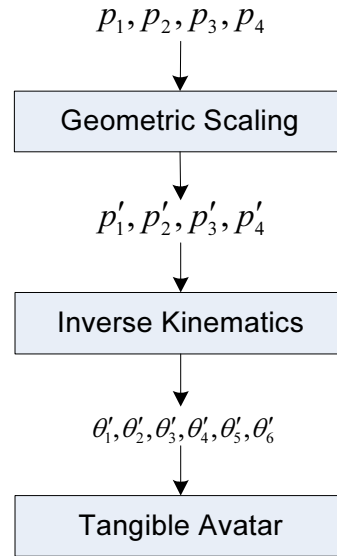


그림 7. 모션 적용 시스템의 알고리즘 흐름도

가능한 위치벡터  $p_1, p_2, p_3, p_4$ 로 바뀌게 된다. 이때  $p_1, p_2, p_3, p_4$ 는 각각 로봇의 어깨, 팔꿈치, 손목, 손의 위치 벡터이다. 우리의 최종 목표는 로봇을 움직이기 위한 로봇의 관절 각  $\theta_i$  ( $i = 1 \sim 6$ ) 를 얻는 것인데, 이것은 역기구학 과정을 통하여 계산된다.

### 4.2 기구학적 변환

액터의 팔 길이는  $l_i = \|p_{i+1} - p_i\|$  와 같이 계산 가능하다. 그런데, 이 길이는 사람들 간의 팔 길이의 상이성과 모션 캡처의 에러로 인하여 시간에 따라 변화하게 된다. 그러나 강체들로 이루어진 로봇의 팔 길이  $l'_i = \|p'_{i+1} - p'_i\|$  는 항상 일정하다. 이러한 차이를 보정하기 위해서 기하학적 길이변환 방법이 필요하다.

우리의 기하학적 길이변환 방법은, 아래와 같이 액터의 위치벡터를 로봇의 위치벡터로 구간구간 마다 선형 사상을 하는 과정을 통해 이루어진다.

$$p'_1 = p_1$$

$$p'_i = \frac{l'_i}{l_i} (p_1 - p_{i-1}) + p'_{i-1}, (i = 2, 3, 4) \quad (10)$$

이런 과정을 거치면, 액터 팔의 인접한 두 관절 사이의 거리가 로봇의 관절 길이에 적합하게 변환된다. 이 과정 전후에 인접한 위치벡터들이 이루는 각은 변하지 않고 유지된다.



### 4.3 역기구학

캡처된 사람 팔의 움직임은 어깨, 팔꿈치, 손목, 손의 3차원 좌표로 표현된다. 이 좌표정보를 이용해 로봇을 움직이기 위해서는 이 좌표 정보들을 관절 각 정보로 바꾸어야 한다. 결국 이 과정은 역기구학문제로 귀결된다. 로봇의 팔은 6개의 1자유도 관절로 이루어진 개연쇄 기구이다. 개연쇄 기구의 기구학은 다음과 같은 동차변환 행렬의 곱으로 표현할 수 있다. 기구의 특정 위치  $i$  에 대한 3차원 위치벡터를  $\bar{p}_i$  라 하면, 이는 아래와 같이 계산된다.

$$\bar{p}_i(\theta'_1, \dots, \theta'_n) = \exp(A_1, \theta'_1) \cdots \exp(A_i, \theta'_i) \cdot M_i \quad (11)$$

여기서,  $\theta_j$  는  $j$  번째 관절의 관절 각이고,  $A_1, \dots, A_n$   $M_i$  는 로봇의 기구학 파라미터 상수이다.  $1 \sim n$ 까지의 관절은 위치벡터  $\bar{p}_i$  의 변화에 영향을 미치는 관절들만 표현한 것이다.  $\exp(-)$  는 기구학 파라미터 상수  $A_i$  를 강제 위치와 자세를 나타내는 동차변환 행렬로 사상하는 함수이다.

액터의 팔 움직임을 로봇이 유사하게 따라 하기 위해서는 액터의 어깨, 팔꿈치, 손목, 손의 위치벡터와 로봇의 각 해당 부위 위치벡터들간의 차이를 최소화 할 필요가 있다. 이러한 역기구학 문제는 아래와 같은 최적화 문제로 그 해를 구할 수 있다.

$$\begin{aligned} \underset{\theta'}{\text{minimize}} H(\theta') &= \frac{1}{2} \sum_{i=1}^4 \|\bar{p}_i(\theta') - p'_i\|^2 \\ &= \frac{1}{2} \|\bar{p}(\theta') - p'\|^2 \end{aligned} \quad (12)$$

Subject to

$$\theta'_{upper} \leq \theta' \leq \theta'_{lower}$$

여기서,  $\theta' = [\theta'_1, \dots, \theta'_6]^T$ ,  $\bar{p}(\theta') = [\bar{p}_1(\theta'), \dots, \bar{p}_4(\theta')]^T$ ,  $p' = [p'_1, \dots, p'_4]^T$  이고, 그리고  $\theta'_{lower}$  와  $\theta'_{upper}$  는 각각 팔 기구부의 구조 등으로 인한 관절의 최소값과 최대값을 의미한다.

우리는 최적화 과정에서 해에 빨리 수렴하기 위해서는 목적함수에 대한 해석적 구배 함수를 사용하였다. 식(13) 에서 제시된 목적함수에 대한 구배 함수는 아래와 같이 계산 가능하다.

$$\nabla H(\theta') = \|\bar{p}(\theta') - p'\| J \quad (13)$$

여기서  $J$  는  $\bar{p}(\theta')$  에 대한 자코비안 행렬이다.

### 5. 실험 결과

본 논문은 비전센서를 기반한 사람과 로봇간의 운동 정보 전달 시스템을 직접 구현하였고, 이더넷으로 연결된 두 대의 PC와 비전센서를 통해 실험환경을 구성하였다. 한대의 컴퓨터는 스테레오 카메라를 사용하여 사람의 움직임 동작을 캡처하는 역할을 맡고, 다른 컴퓨터는 전송된 움직임 정보를 가지고 로봇을 제어하는 역할을 맡았다.

그림 8은 실험 결과의 일부를 보여준다. 그림 (a)와 (b)는 실험환경 속의 사람과 로봇이며, 그림 (c)는 스테레오 카메라를 통해 입력되는 2차원 영상을 나타내고, 그림 (d)는 모션캡처 알고리즘을 통해서 추정된 팔 모델의 자세를 보여준다. 모션 캡처 시스템은 대략 12~16hz로 동작하고, 로봇의 팔은 자연스럽게 움직인다.

특히 제안된 모션 캡처 시스템의 정확성을 검증하기 위해서 액터의 역할을 로봇이 대신 수행하는 동일한 실험도 수행하였다. 이는 로봇의 경우 움직인 자취를 기록할 수 있어 모션 캡처 시스템을 통해 얻은 추정 결과와 비교하여 정확성을 검증할 수 있기 때문이다.

그림 9는 캡처한 로봇 팔의 모션과 로봇 관절에 부착되어 있는 엔코더 기록을 통해 얻은 모션을 비교한 그림이며, 표 1은 실제 모션과 캡처된 모션 사이의 오차 결과를 각 관절에 따라 나타내었다. 이 실험 결과는 캡처된 움직임 정보가 액터의 움직임에 잘 추종하고 있음을 보여주고 있다.

Frame	(a)	(b)	(c)	(d)
#210				
#390				
#690				
#840				
#1110				
#1280				

그림 8. 실감만남 시스템의 실험 결과

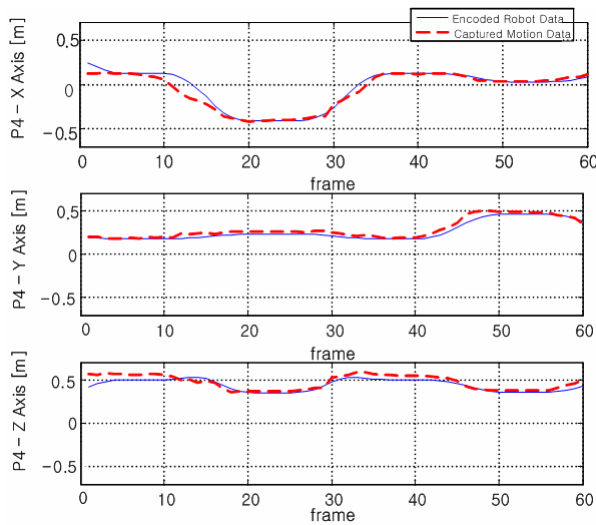


그림 9. 손 끝 위치에 해당되는 P<sub>4</sub>의 자취

표 1. 실제 모션과 캡처된 모션의 오차 결과 (단위:mm)

		P <sub>1</sub>	P <sub>2</sub>	P <sub>3</sub>	P <sub>4</sub>
평균	X	0.0	-12.2	-18.0	22.5
	Y	0.0	41.5	19.3	-27.1
	Z	0.0	23.4	35.8	-21.4
분산	X	0.0	3.7	2.5	2.0
	Y	0.0	0.7	0.3	0.8
	Z	0.0	0.8	0.5	1.4

## 6. 결론

본 논문에서는 기존의 모션 캡처 방법을 개선하여 실시간 처리가 가능한 휴먼 팔 모션 모방의 방법을 제안하였다. 모션 캡처 부분에서 컴퓨터의 연산량을 줄이고 점군 데이터의 사이즈를 줄여 상이도 영상을 격자화하였다. 또한 라벨링과 그룹핑을 통해 정의된 관심영역으로부터 3차원 점군데이터를 표본화 하였다. 그 결과 휴먼 팔 모션 모방 시스템은 12~16hz로 동작하였다.

모션의 정확도를 강화하기 위해, 메타볼의 개선된 포텐셜 함수를 제안하였다. 제안된 포텐셜 함수는 모션 캡처 시스템에 적합하여 최적화 문제 해결시 지역 극소값에 빠지는 위험을 줄여준다. 실험 결과는 모션 캡처 시스템이 원격지 사람과 대화하기에 충분히 정확한 모션을 생성함을 보여주고 있으며, 원격의 로봇은 자연스럽게 액터의 자세를 취할 수 있음을 확인하였다.

동작 모방 부분에서는 기구학적 스케일링 방법을 제안하였으며, 이 방법을 통해 사람과 로봇 사이의 구조

적 차이를 극복 할 수 있게 되었다.

본 연구의 장기적인 목표는 혼합공간에서의 사람 같은 로봇을 만들어 제공하는 것이며, 앞으로 현 연구의 트래킹 범위를 확장하여, 사람의 모든 신체 요소를 모션 캡처하는 연구를 진행 할 예정이다.

## 참고문헌

- [1] S. Park, "Tangible Space Initiative (TSI)," *Tangible Space Initiative Workshop in the 8th Int. Conf. on Virtual Systems and Multimedia (VSMM2002)*, Korea, 25-27, 2002.
- [2] Thomas B. Modeslund and Erik Granum, "A survey of computer vision-based human motion capture," *Computer Vision and Image Understanding*, vol.81, pp. 231-268, 2001.
- [3] L. Wang, W. Hu and T. Tan. "Recent Developments in Human Motion Analysis," *Pattern Recognition*, pp. 585-601, 2003.
- [4] D. Demirdjian, T. Ko and T. Darrell, "Constraining Human Body Tracking," *IEEE Int. Conf. on Computer Vision*, pp. 1071-1078, 2003.
- [5] Ralf Plankers and Pascal Fua, "Articulated Soft Objects for Multiview Shape and Motion Capture," *IEEE Transactions on Pattern Analysis And Machine Intelligence*, vol. 25, no. 9, pp. 1182-1187, 2003.
- [6] S. Knoop, S. Vacek and R. Dillmann, "Sensor fusion for 3D human body tracking with an articulated 3D body model," *IEEE Int. Conf. on Robotics and Automation*, pp. 1686-1691, 2006.
- [7] L. Molina-Tanco, J.P. Bandera, R. Marfil, F. Sandoval, "Real-time human motion analysis for human-robot interaction," *IEEE Int. Conf. on Intelligent Robots and Systems*, pp. 1402-1407, 2005.
- [8] J. F. Blinn, "A generalization of algebraic surface drawing," *ACM Transactions on Graphics*, vol.1, pp. 235-256, 1982.
- [9] J. J. Kim, E.-S. Kim and S.-K. Park, "An automatic description of volumetric objects using metaballs," *Computer Graphics International*, pp. 65-73, 1997.
- [10] Jianbo Shi and Carlo Tomasi, "Good Features to Track," *IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 593-600, 1994.
- [11] Carlo Tomasi and Takeo Kanade, "Detection and Tracking of Point Features," *Carnegie Mellon University Technical*



Report CMU-CS-91-132, 1991.

- [12] P.J. Besl and N.D. McKay, "A method for registration of 3-D Shapes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.14, pp. 239-256, 1992.
- [13] M. Covell, A. Rahimi, M. Harville and T. Darrell, "Articulated-Pose Estimation Using Brightness and Depth-Constancy Constraints," *Conf. Computer Vision and Pattern Recognition*, vol.2, pp. 438-445, 2000.
- [14] A. Ranganathan, "The Levenberg-Marquardt Algorithm," *Ananth Ranganathan 8<sup>th</sup>*, 2004.
- [15] M. Riley, A. Ude, K. Wade and C. G. Atkeson, "Enabling Real-time Full-Body Imitation: A natural Way of Transferring Human Movement to Humanoids," *IEEE Int. Conf. on Robotics and Automation*, pp. 2368-2374, 2003.
- [16] C.H. Kim, D. Kim, and Y. Oh, "Solving an inverse kinematics problem for a humanoid robots imitation of human motions using optimization," *Second Int. Conf. on Informatics in Control, Automation and Robotics*, pp. 85-92, 2005.



**최 유 경**

2006 숭실대학교 정보통신전  
자공학부(학사)  
2006~현재 연세대학교 전기전  
자공학과 석사과정

2006~현재 한국과학기술연구원 인지로봇연구단 학생  
연구원  
관심분야 : 컴퓨터 비전, 로봇 비전, 지능 로봇



**나 성 권**

2003 서울대학교 기계항공공  
학부(학사)  
2005 서울대학교 기계항공공  
학부(석사)

2005~현재 한국과학기술연구원 인지로봇연구단 연구원  
관심분야 : 다물체 동역학



**김 수 환**

2003 서울대학교 기계항공공  
학부(학사)  
2004 Univ. of Southern  
California, Computer  
Science (MS)

2006~현재 한국과학기술연구원 인지로봇연구단 연구원  
관심분야 : Computer Vision, Character Animation



**김 창 환**

1993 한양대학교 기계공학과  
(학사)  
1995 한양대학교 기계설계학  
과(석사)  
2002 아이오와주립대학교 기  
계공학과(박사)

2004~현재 한국과학기술연구원 인지로봇연구단 선임  
연구원  
관심분야 : 인간형 로봇의 동작생성 및 인간 동작  
모방, 인간동작 및 움직임 해석



**박 성 기**

1987 서울대학교 기계설계학  
과(학사)  
1989 서울대학교 기계설계학  
과(석사)  
2000 KAIST 자동화 및 설계  
공학과(박사)

2000~현재 한국과학기술연구원 인지로봇연구단 선임  
연구원  
관심분야 : 컴퓨터 비전, 로봇 비전, 지능 로봇