

## 그림모델과 파티클필터를 이용한 인간 정면 상반신 포즈 인식

### Pictorial Model of Upper Body based Pose Recognition and Particle Filter Tracking

오치민, Chimin Oh\*, Md. Zahidul Islam\*\*, 김민욱, Minwook Kim\*\*\*, 이철우, Chilwoo Lee\*\*\*\*

**요약** 본 논문은 비전을 이용한 인간 정면 상반신 포즈를 인식 방법에 대해서 기술한다. 일반적으로 HCI(Human Computer Interaction)와 HRI(Human Robot Interaction)에서는 인간이 정면을 바라볼 때 얼굴, 손짓으로 의사소통 하는 경우가 많기 때문에 본 논문에서는 인식의 범위를 인간의 정면 그리고 상반신에 대해서만 한정한다. 인간 포즈인식의 주요 두 가지 어려움은 첫째 인간은 다양한 관절로 이루어진 객체이기 때문에 포즈의 자유도가 높은 문제점 때문에 모델링이 어렵다는 것이다. 둘째는 모델링된 정보와 영상과의 매칭이 어려운 것이다. 이를 해결하기 위해 본 논문에서는 모델링이 쉬운 그림모델(Pictorial Model)을 이용해 인체를 다수 사각형 파트로 모델링 하였고 이를 이용해 주요한 상반신 포즈를 DB화 해 인식한다. DB 포즈로 표현되지 못하는 세부포즈는 인식된 주요 포즈 파라미터로부터 파티클필터를 이용해 예측한 다수 파티클로부터 가장 높은 사후분포를 갖는 파티클을 찾아 주요 포즈를 업데이트하여 결정한다. 따라서 주요한 포즈 인식과 이를 기반으로 한 세부 포즈를 추적하는 두 단계를 통해 인체 정면 상반신 포즈를 정확하게 인식 할 수 있다.

**Abstract** In this paper, we represent the recognition method for human frontal upper body pose. In HCI(Human Computer Interaction) and HRI(Human Robot Interaction) when a interaction is established the human has usually frontal direction to the robot or computer and use hand gestures then we decide to focus on human frontal upper-body pose. The two main difficulties are firstly human pose is consist of many parts which cause high DOF(Degree Of Freedom) then the modeling of human pose is difficult. Secondly the matching between image features and modeling information is difficult. Then using Pictorial Model we model the human main poses which are mainly took the space of frontal upper-body poses and we recognize the main poses by making main pose database. using determined main pose we used the model parameters for particle filter which predicts the posterior distribution for pose parameters and can determine more specific pose by updating model parameters from the particle having the maximum likelihood. Therefore based on recognizing main poses and tracking the specific pose we recognize the human frontal upper body poses.

**핵심어:** *Human Pose Recognition, Upper Body, Pictorial Model, Particle Filter, Chamfer Distance, Diffusion Distance, Edge Model.*

본 논문은 정통부 및 정보통신연구진흥원의 정보통신선도기반기술개발사업의 연구비 지원에 의해 수행되었습니다.

\*주저자 : 전남대학교 전자컴퓨터공학과 석사과정 e-mail: sapeyes@image.chonnam.ac.kr

\*\*공동저자 : 전남대학교 전자컴퓨터공학과 석사과정 e-mail: zahid@image.chonnam.ac.kr

\*\*\*공동저자 : 전남대학교 전자컴퓨터공학과 석사과정; e-mail: purelover7@image.chonnam.ac.kr

\*\*\*\*공동저자 : 전남대학교 전자컴퓨터공학과 교수; e-mail: leecw@chonnam.ac.kr

## 1. 서론

인체 포즈는 HCI, HRI 등 다양한 곳에서 활용된다. HCI의 예로는 가상 터치스크린에서 가상터치를 인식하기 위해 다수 사용자를 구분하고 각 사용자의 손을 인체로부터 구분하는 방법을 사용한다. HRI의 예로는 로봇이 사람에게 물건을 주기 위해 사용자의 손을 세그멘테이션 한 후 손의 모양이 물건을 받을 수 있는 모양일 경우 물건을 전달하는 방법이 있다. 인체 포즈 인식 후 결정된 신체 정보는 이용해 이러한 의사소통에 활용될 수 있다. 따라서 이를 활용한 분야는 무궁무진하다.

HCI, HRI와 같은 상호작용에서는 기본적으로 대상의 구분과 대상의 인체 영역을 세부적으로 세그멘테이션 정보를 이용해야 한다. 따라서 비전기반 연구에서는 이러한 상호작용에 이용될 수 있도록 인체 포즈를 정확히 인식하고 각 인체 영역을 세그멘테이션 할 수 있는 방법을 개발하여야 한다. 대부분 손을 이용한 정보교환이 이루어지며 정확한 손 정보는 정확한 상반신 포즈를 인식할 수 있는 방법이 필요하다. 인체 포즈는 주요 정보가 상반신에 포함되어 있고 HCI와 HRI에서는 주로 손을 이용하여 터치 또는 손짓 제스처를 전달하기 때문이다. 따라서 영상에서 상반신을 정확히 모델링 하는 방법이 필요하다.

Haritaoglu[1]은 사람 포즈를 구성요소인 손, 얼굴, 몸통을 인체 외곽선의 최외곽거리 지점을 모델링 하였고 2차원 영상특징에서 인체의 자유도와 연결하기 위해 세그멘테이션된 인체 실루엣 영상의 컨벡스(Convex), 콘케이브(Concave) 점을 이용하여 인체포즈를 구하였다. 하지만 영상에 사람의 전신이 나타나야 한다는 단점이 있다. Urtasun[2]은 인체를 3차원 모델로 정의하였고 모션 벡터 영상특징과 3차원 모델을 매핑하는 방법을 사용하고 있다. 이 방법은 다양한 환경에서 강건한 결과를 나타내지만 특정한 모션에 대해 PCA(Principle Component Analysis)를 얻어 매핑 하였으므로 제약된 모션만을 인식하는 단점이 있고 고유공간에서 모든 모션을 표현할 만큼 강건한 특징을 얻어야 한다는 단점이 있다. Felzenszwalb[3]은 객체를 다수의 파트로 모델링하는 방법을 제시하였다. 얼굴과 인체를 모델링 하여 Condensation을 통해 객체를 추적하는 방법을 나타내었다. 하지만 2차원 모델을 사용하고 있으므로 복잡한 객체의 움직임을 추적하는 데는 무리가 있으므로 간단한 객체 모델링에 적합하고 이를 향상시켜 사용해야 한다. 본 논문에는 Felzenszwalb의 그림구조 모델을 상반신을 모델링 하는데 사용하였다.

포즈인식은 미리 정의한 포즈 모델 정보의 실제 값을 영상의 특징과 매칭하는 것이다. 하지만 인체 상반신 포즈는 다양한 파트로 구성되어 있어 큰 사이즈의 포즈공간에 하나로 포함된다. 이러한 포즈공간에서 효과적인 포즈를 추적하

기 위해 제시된 한 방법으로 노명철[4]는 선택된 주요한 상반신 포즈들을 근간으로 하여 넓은 포즈공간에서 하나의 포즈를 결정하는 방법을 제시하였다. 주요 포즈만을 DB화 하여 이를 인식하는 방법이다. 모든 포즈공간에 해당하는 포즈를 인식하기 힘들기 때문이다.

주요 포즈 DB와 비교하여 영상과 DB의 가까운 포즈를 결정한 후 가까운 DB 포즈로부터 세부 포즈를 추적할 수 있다. 파티클필터[5,6]는 결정된 주요포즈 정보로부터 세부포즈를 예측하고 검증, 업데이트 하는 과정을 통해 세부포즈를 추적할 수 있다. 예측은 마르코프 랜덤분포를 사용하며 검증은 챔퍼매칭(Chamfer Matching)[7] 또는 적분영상[9]을 이용한 빠른 컬러매칭을 사용할 수 있다. 업데이트는 검증과정에서 확률이 높게 나타난 파티클을 중심으로 다시 예측하는 과정이다.

본 논문에서 인체 상반신 포즈를 모델링하기 위해 그림모델[3]을 이용하였다. 포즈 인식 공간을 줄이기 위해 12개의 주요포즈를 DB화 하였다. 주요포즈[4]는 챔퍼매칭을 통해 특징벡터를 만든 후 발산거리(Diffusion Distance)[10]을 통해 가까운 주요포즈를 선택하여 결정하였다. 주요포즈로부터 세부포즈를 추적하기 위해 파티클필터를 사용하였다. 그림모델은 파라미터로 구성되어 있기 때문에 이 파라미터로부터 일정 분포로 떨어진 다수의 파티클을 생성하였다. 가장 유사한 파티클을 선택하기 위해 챔퍼매칭을 사용하여 세부포즈를 추적하였다.

본 논문은 2장에서 상반신 포즈를 인식하는 과정을 소개한다. 상반신 포즈를 그림구조로 모델링 한 뒤 주요포즈 DB를 작성하고 이를 기반으로 주요포즈를 인식하는 과정을 서술한다. 3장에서는 파티클필터를 통해 세부포즈를 추적하는 과정을 소개한다. 먼저 파티클 필터 소개와 함께 어떻게 주요포즈를 통해 파티클필터를 초기화하는지 설명한다. 또한 예측과정과 검증 및 업데이트과정을 통해 세부포즈가 어떻게 결정되는지 설명한다. 4장에서는 실험 결과와 5장에서는 결론으로 논문을 맺는다.

## 2. 상반신 주요 포즈 인식

### 2.1 그림구조(Pictorial Structure)

그림구조모델[3]은 변형가능한 객체(Deformable Object)를 다수의 파트를 가진 객체로 모델링한다. 그림모델은 그래프  $G = (V, E)$ 로 표현할 수 있다. 정점  $V = v_1, \dots, v_n$ 은 각 파트를 가리키며  $(v_i, v_j) \in E$ 는 파트간의 연결성을 가리킨다. 이것은 객체는 객체 파트의 수만큼 정점을 정의할 수 있고 일부 파트사이에는 연결성을 추가 속성으로 가질 수 있는 것이다.

그림구조의 각 파트는 두 가지 형태의 특징을 가지고 있다. 하나는 자신의 형상에 대한 영상특징이고 다른 하나는 자신이 위치가 이웃하는 파트와의 관계에서 적절한지 나타내는 특징이다. 이 특징은 향후 후보 그림구조 모델을 검증할 때 사용한다.  $L_t = (l_1, \dots, l_n)$ 은  $t$ 시간의 모든 파트의 부모 파트로부터의 상대적인 위치, 회전, 크기정보를 갖고 있다.  $L_0$ 은 객체의 초기 특징을 정확히 표현하고 있으며 이를 바탕으로 예측된 모델  $L_t^{i=1:n}$  중에서 어떤 후보모델이 가장 정확한 두 가지의 특징정보를 만족하는  $L_t^*$  인지 식 (1)을 통해 검증할 수 있다.  $m_k(l_k)$ 는 각 파트 위치에 존재하는 영상특징이  $L_0$ 의 그것과의 유사도를 나타낸다.  $d(l_j, l_k)$ 는 이웃하는 두 파트 관계는 현재 예측된 위치에서 존재가 가능한지 척도를 나타낸다.

$$L_t^* = \operatorname{argmin} \left( \sum_{k=1}^n m_k(l_k) + \sum_{(v_j, v_k)} d(l_j, l_k) \right) \quad (1)$$

## 2.2 상반신 그림구조 디자인

상반신 포즈는 신장 170cm를 갖는 사람의 상반신을 그림구조 모델로 정의하였다. 상반신 그림구조는 얼굴, 왼손위, 왼손아래, 오른손위, 오른손아래, 몸통 총 6개 파트를 가진 객체로 나타낸다. 각 파트의 부모는 얼굴, 몸통, 양손순서이다. 최초의 상반신 포즈를 얻기 위해 그림 1처럼 실제 사람의 신체 크기를 실측한 모델을 작성하였다. 그림구조 모델은 원점을 좌 상단으로 하는 영상좌표에서 정의된다. 그림 1의 각 파트는 회전과 이동의 원점을 가지고 있으며 파트 자신의 영상좌표에서의 위치와 가로 세로크기를 가지고 있다. 크기단위는 실측정보인 cm단위이며 이를 그대로 이용하기 위해 영상에서 픽셀의 간격도 1cm로 간주한다.



그림 1. 그림구조 모델과 파트 위치 및 크기정보(x,y,w,h),단위cm

표 1. 상반신 그림구조모델 파트 자유도 (얼굴 제외 파트에서는 x축,y축 이동이 얼굴에 의존적임)

파트명	x축이동	y축이동	크기변화	회전
얼굴	○	○	○	○
몸통			○	○
왼손위			○	○
왼손아래			○	○
오른손위			○	○
오른손아래			○	○

상반신 그림구조 모델은 각 파트에서 x축, y축, 크기변화, z축 회전에 대한 자유도를 갖는다(표 1). x, y축 이동은 영상범위 내에서 이동을 가리킨다. 크기변화는 파트의 가로, 세로크기에 대한 스케일 정보이다. 회전은 xy평면의 2차원 모델이므로 z축에 대한 회전량이다. 표 1에서는 일부 파트는 x,y축 이동 자유도를 갖지 않고 있는데 이는 얼굴 파트가 상반신모델의 모든 파트의 부모이므로 x, y축 이동이 일어날 경우 모든 파트가 함께 이동하기 때문이다.

## 2.3 상반신 주요 포즈 DB 만들기

상반신 주요포즈 DB는 3가지 종류이다. 첫째는 인체 정면 상반신 포즈가 가질 수 있는 포즈 중 대표할 수 있는 포즈영상 12개 이다. 둘째는 각 포즈 영상대해 작성한 그림구조 모델 12개이다. 셋째는 챔퍼매칭(Chamfer Matching)을 이용해 한 포즈영상에 대해 12개의 그림구조모델의 거리 값을 하나의 특징벡터로 하는 특징벡터이며 모든 주요포즈영상에 대해 계산하므로 12개의 특징벡터이다.

먼저 DB를 생성하기 위해서는 포즈 영상을 선택과 그림구조 모델링과정이 필요하다. 12개 그림구조 모델들은 각 포즈영상에서 얼굴인식 후 모델링 툴로 모델을 세팅하여 얻는다. 그림 2를 통해 주요포즈 12개의 영상과 그림구조모델링 툴을 볼 수 있다. 그림 3은 12개 그림구조 모델결과이다.

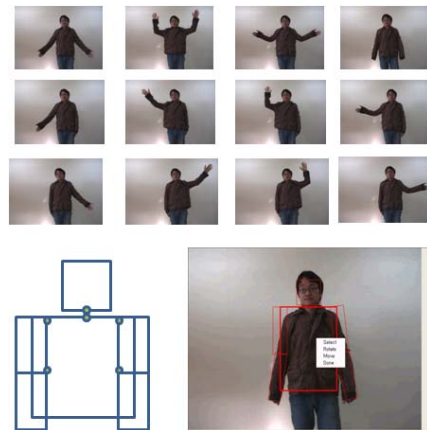


그림 2. 주요포즈(상단 12개 영상)와 그림모델로 모델링툴(하단 오른쪽영상)



그림 3. 주요포즈로부터 모델링한 12개 그림구조모델

주요포즈 DB의 세 번째 자료는 주요포즈에 대한 12개의 특징벡터이다. 특징벡터는 포즈 영상과 그림모델간의 챔퍼매칭을 통해 얻는다. 챔퍼매칭[7]은 객체가 영상에 존재 여부를 얻기 위해 객체 에지모델(Edge Model)과 에지영상과의 유사도를 챔퍼거리(Chamfer Distance)[7]를 통해 구하는 방법이다. 챔퍼거리는 에지영상에서 거리변환을 구하여 에지모델에서 추출한 특징점들의 위치의 픽셀 값들을 식(2)와 같은 RMS(Root Mean Square)값을 구한 것이다. 만일 에지모델이 똑같은 모양을 가진 에지영상에 위치하고 있을 경우 챔퍼거리는 0을 갖게 된다.

$$\frac{1}{3} \sqrt{\frac{1}{n} \sum_{i=1}^n v_i^2} \quad (2)$$

따라서 챔퍼매칭을 위해 사용될 두 가지 특징인 포즈영상의 에지 거리변환영상과 에지모델이 필요하다(그림 4). 에지 변환영상은 포즈영상의 캐니(Canny)연산자를 이용한 에지영상을 얻은 후 에지로부터 모든 픽셀의 거리 값을 나타낸 거리변환을 통해 얻는다. 에지모델은 그림구조모델 모든 파트에서 에지점 추출을 통해 얻는다. 그림1의 그림구조모델은 각 파트가 사각형 모양으로 외관을 가진다. 이러한 사각형에서 에지 특징점을 추출하기 위해 식 3을 사용한다. 각 파트의 사각형은 네 꼭짓점으로 구성된다. 특징점을  $n$ 개 추출할 때 각 사각형 에지의 에지점 수는  $g$ 이다.  $p$ 는 꼭짓점의 위치이다. 두 꼭짓점 사이의 에지 특징점들은 두 꼭짓점 사이를  $g$ 등분하는 점들이다.  $dx$ ,  $dy$ 는 두 꼭짓점 사이에 존재하는 에지 특징점들의 간격을 가리킨다.

$$g = n \frac{1}{4} \quad (3)$$

$$dx = \frac{(p_{i+1,x} - p_{i,x})}{g}, dy = \frac{(p_{i+1,y} - p_{i,y})}{g}$$

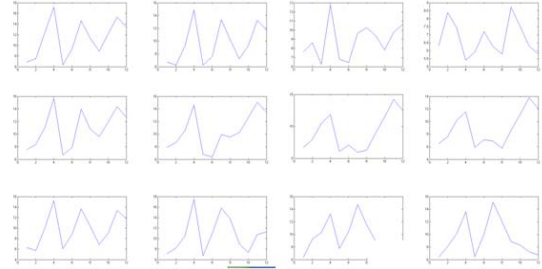
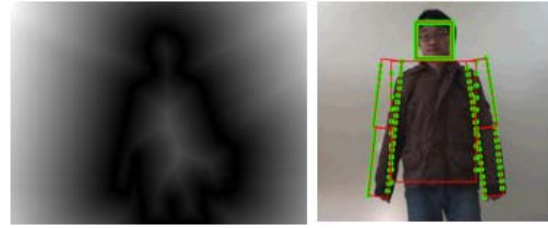


그림 4. 에지 거리변환영상(상단 왼쪽), 그림구조 모델특징점(상단 오른쪽) 12개의 주요포즈 DB 특징벡터(하단)

특징벡터는 하나의 포즈영상으로부터 구한 거리변환영상에 대해 12개의 주요포즈 그림구조모델의 챔퍼거리를 가지고 있다. 이를 12개 포즈영상에 대해 반복적으로 시행하여 그림 4의 12개의 특징벡터를 얻는다(가로축-포즈번호, 세로축-챔퍼거리). 이를 통해 주요포즈인식에 온라인 특징벡터 추출에 사용될 12개의 그림구조모델과 온라인 특징벡터와 비교될 12개의 오프라인 특징벡터를 갖게 된다.

## 2.4 상반신 주요 포즈 인식

주요 포즈 인식은 입력된 포즈 영상에서 구한 특징벡터와 12개의 DB 특징벡터 중에서 유사도가 가장 가까운 포즈를 선택하여 결정한다. 유사도는 효과적으로 알려진 두 히스토그램 거리를 계산하는 확산거리(Diffusion Distance)[8]를 사용하였다. 즉 입력영상의 특징벡터가 어느 데이터베이스와 가까운지 계산하여 해당하는 주요포즈가 결정된다. 이러한 확산거리 중에서 Earth Mover의 거리법은 히스토그램의 분산 변화에도 효율적으로 계산을 하며 구현도 크게 어렵지 않은 장점이 있다. 계산의 복잡도는  $O(Nd \log \Delta)$ 이며  $N$ 은 특징벡터의 크기이고  $d$ 는 차원 수이다.

확산거리는 온도가 확산하는 원리를 히스토그램 차이를 반복적인 스무딩(Smoothing)을 하여 적은 차이도 구분할 수 있도록 하였다. 두 1차원 히스토그램 히스토그램  $h_1(x)$ ,  $h_2(x)$ 의 차이값은  $d(x) = h(x)_1 - h_2(x)$ 이다. 온도필드에서  $d(x) = T(x, 0)$ 이다. 시간에 대한 열의 확산 변화는 식(2)와 같다.

$$\frac{\partial T}{\partial t} = \frac{\partial T}{\partial x} \quad (4)$$

$T(x,t)$ 의 해는 초기  $T(x,0)$ 에 대해 가우시안 필터를 적용한 식 5이다. 가우시안 필터는 식 6과 같이 정의된다. 이를 통해 열의 확산 과정을 적용한 가우시안 필터가 적용된 두 히스토그램의 거리는 식 7을 통해 모든 히스토그램 원소 차이 값을 합산하여 얻을 수 있다.

$$T(x,t) = T(x,0) * \Phi(x,t) \quad (5)$$

$$\Phi(x,t) = \frac{1}{(2\pi)^{1/2}t} \exp\left(-\frac{x^2}{2t^2}\right) \quad (6)$$

$$\hat{K}(h1,h2) = \int_0^t k(|T(x,t)|)dt \quad (7)$$

그림 5에서는 주요포즈를 결정하기 위해 입력영상에서 그림모델의 비교과정과 결정된 주요포즈를 확인할 수 있다. 12개의 주요포즈 그림구조모델들은 얼굴 인식 결과를 통해 위치 및 스케일정보가 초기화된다. 초기화된 그림구조모델은 그림 5 상단 12개의 영상처럼 예지특징점을 추출할 수 있다. 모든 모델은 챔퍼매칭으로 입력영상과 12번의 비교를 통해 온라인 특징벡터를 만든다. 이 온라인 특징벡터는 미리 포즈 DB에 저장된 12개의 오프라인 특징벡터와 확산거리를 통해 가까운 포즈를 결정하게 된다. 하지만 주요포즈는 완벽하게 사용자의 포즈를 표현할 포즈공간을 갖고 있지 않다.

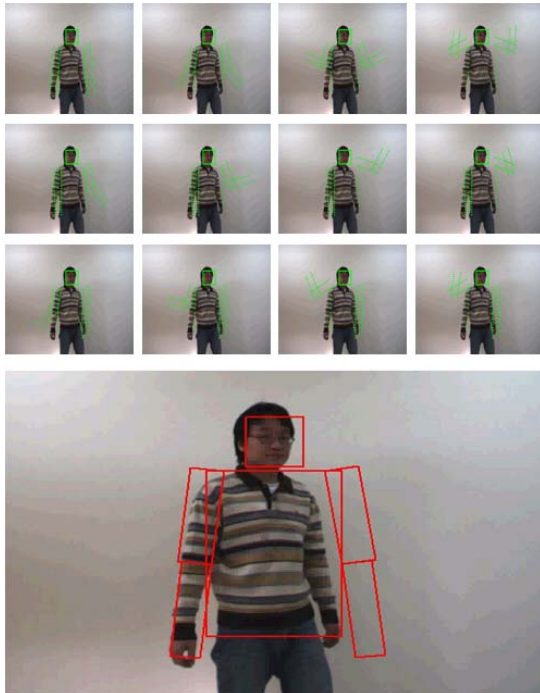


그림 5. 12개 그림모델의 특징점들과 인식된 주요포즈

### 3. 세부 포즈 추적

#### 3.1 파티클 필터(Particle Filter)

파티클 필터(Particle Filter)[5]는 베이시안 순차 중요성 샘플링 방법(Bayesian Sequential Importance Sampling) 방법으로 사후분포(Posterior Distribution)를 예측하고 검증과 업데이트 하는 과정으로 되어 있다. 미래의 객체 후보모델인 파티클은 예측된 파라미터 벡터  $x_t$ , 이전 파라미터벡터  $x_{t-1}$  그리고 초기 파라미터 정보  $x_0$ 를 가지고 있다. 위 파라미터 벡터는 상반신 객체가 갖는 이동, 크기변화 및 회전 자유도에 대한 파라미터를 갖고 있다. 식 8은 예측된 사후분포를 가리킨다.  $z_{1:t-1}$ 은  $t-1$ 시간까지의 영상이다. 예측된 사후분포는 예측된 파티클이 객체 모델에 맞게 예측되었는지 나타낸다.

$$p(x_t|z_{1:t-1}) = \int p(x_t|x_{t-1})p(x_{t-1}|z_{1:t-1}) \quad (8)$$

파티클의 예측은 마르코프 랜덤분포를 사용한다. 식 9에서 파티클의 새로운  $x_t$  예측은 이전 예측 값인  $x_{t-1}$  값으로 이동하며  $x_{t-1}, x_{t-2}, x_0$  값을 기반으로 예측한다. 즉 예측 벡터  $x_t$  초기 값으로부터 최근 값과 가우시안 랜덤  $g()$  값을 통해 어느 정도 분포 안에서 결정된다. 상수 값  $c_i$ 는 초기 값으로부터 랜덤한 예측 분포를 결정하며 실험에 의해 결정된다.

$$x_t = c_1(x_{t-1} - x_0) - c_2(x_{t-2} - x_0) + g(c_3) + x_0 \quad (9)$$

파티클 필터의 검증은 식 8에서 예측값을 갖는 파티클에 대한 사후분포를 현재 영상에서 예측된 값이 존재 하는지에 대한 사후분포로 변경한다. 즉 검증은 파티클필터의 예측된 파라미터로부터 해당하는 탐색영상과의 실제 유사도 계산을 통해 모든 파티클이 예측된 정보가 현재 영상에 존재하는지에 대한 확률을 갖는 사후분포로 변경한다. 이를 위해 식 10은 시간  $t$ 에서 탐색영상  $z_t$ 가 존재할 경우 Bayes' rule에 의해 변경된 사후분포를 표현한다.  $p(z_t|x_t)$ 는 예측된 정보가 탐색영상에서 존재하는 우도(Likelihood) 또는 탐색확률(Observation Probability)를 나타낸다.

$$p(x_t|z_{1:t}) = \frac{p(z_t|x_t)p(x_t|z_{1:t-1})}{p(z_t|z_{1:t-1})} \quad (10)$$

파티클 필터의 업데이트는 검증과정에서 결정된 각 파티클의 우도 중 높은 우도를 가진 파티클 위주로 리샘플링되는 과정을 말한다.  $n$ 개의 파티클이 있을 때 가장 높은 우도를 갖는 파티클을 위주로  $0.5n$ 개의 파티클이 리샘플링되거나 우도가  $T$ 값을 넘는 파티클들을 위주로 리샘플링 하는 방법이 있다. 이를 통해 파티클 필터는 객체를 추적할 때 높은 확률을 갖는 파티클을 위주로 다시 예측하여 더 강건한 추적결과를 낳는다.

### 3.2 상반신 세부포즈 추적

상반신 세부포즈 추적은 주요포즈인식이 갖는 부정확한 포즈인식결과를 세밀하게 추적하여 더 나은 포즈를 갖도록 돕는다. 얼굴 인식을 통해 초기화된 모델과 최초 정의된 모델간의 변형관계를 나타내는 파라미터들을 예측하고 검증하는 파티클 필터를 통해 추적한다. 최초로 170cm의 신장을 갖는 사람을 모델링한 정보  $M$ 은 모든 파트의  $(x,y)$  위치 정보이다. 초기 모델의 파트의 회전량은 0이며, 가로, 세로 크기는 변화하지 않는 정보이기 때문에 모델정보에 포함되지 않는다. 이를 기반으로 작성된 특정 포즈에 대한 모델링 정보인 주요포즈 DB의 그림구조 모델은  $M$ 으로부터 변환된  $m$  모델이다. 식 9에 의해 예측된 파라미터 정보 또는 얼굴 인식 정보를 통해 얻어낸 정보인  $(s, R_z, dx, dy)$ 는 모델  $m$ 을 얻는데 식 11과 같이 사용한다. 이를 통해 모든 모델은 초기 모델  $M$ 에 근거하여 표현할 수 있고 파티클 필터와의 연동 또한  $(s, R_z, dx, dy)$ 을 예측하면 되기 때문에 쉽게 이뤄질 수 있다.

$$m = sR_zM + (dx, dy) \quad (11)$$

그림 6은 파티클 필터의 인식된 주요포즈로부터 세부포즈로 추적하는 과정을 나타낸다. 파티클 필터는 인식된 주요포즈로부터 세부포즈의 예측값을 갖는 10,000개의 파티클을 갖는다. 이는 인식된 주요 포즈를 초기값으로 하여 식 9에 의해 예측된 파라미터 값을 가지고 있으며 식 8과 같은 사후분포를 갖는다. 예측된 모든 파티클은 새로운 탐색영상  $z_t$ 이 들어오면 챔퍼매칭을 통해  $p(x_t|z_t)$  우도를 계산한다. 이를 통해 예측된 정보가 실제 영상에서 존재하는지에 대한 사후분포 식 10을 얻어낼 수 있다. 따라서 세부포즈의 결정은 가장 높은 우도를 갖는 파티클의 파라미터 정보를 식 11의 그림구조 모델  $m$ 을 얻어 결정할 수 있다. 파티클 필터 업데이트에서는 새로 인식된 얼굴위치로 초기화된 절반의 파티클과 이전 최대 우도를 갖는 파티클을 초기값으로 갖는 절반의 파티클로 구성한다.

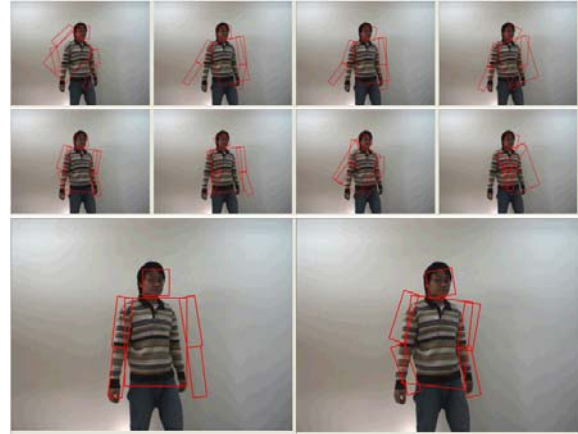


그림 6. 예측된 파티클(상단 8개 영상), 인식된 주요포즈(하단 왼쪽) 및 추적된 세부포즈(하단 오른쪽)

### 4. 실험결과

본 논문은 그림모델과 파티클 필터를 이용해 인체 정면 상반신 포즈를 주요포즈인식 후 세부포즈를 추적하였다. 시스템 환경은 Pentium D 3.40G, 3.0GB Ram이며 실험을 위해 사용된 영상은 640×480이다. 파티클 필터의 파티클 수는 10,000개로 하였으며 챔퍼매칭을 사용했기 때문에 실시간으로 강건하게 동작하였다. 그림 7과 8은 결과영상이다.



그림 7. 주요포즈에서 파티클 필터를 통해 추적된 세부포즈 1

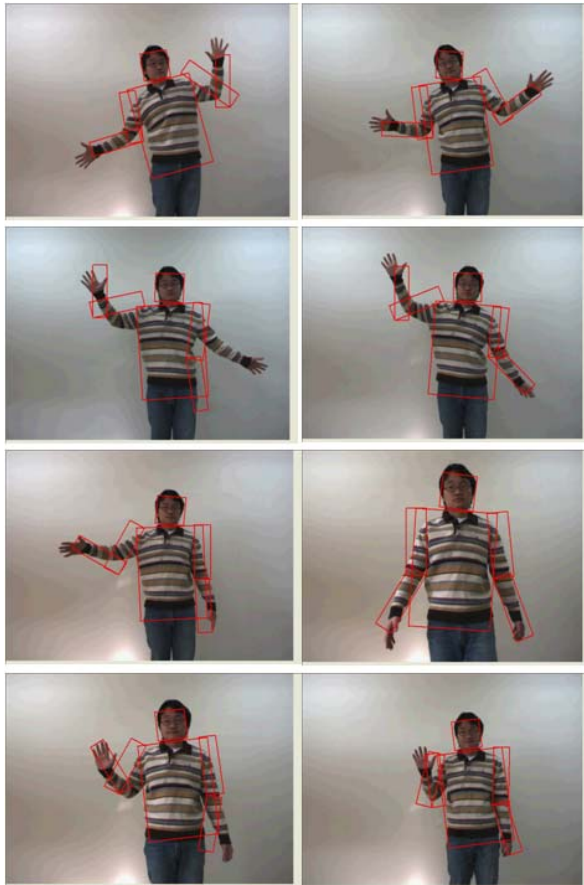


그림 7. 주요포즈에서 파티클 필터를 통해 추적된 세부포즈 2

실험결과는 어느 정도 대상자의 포즈와 유사한 결과를 나타내고 있다. 완벽하게 대상자의 신체에 맞는 그림구조모델 파트크기를 나타내지 못하고 있지만 이는 비교적 단순한 모델을 사용했기 때문이다. 주요포즈인식, 세부포즈 추적과정 모두 에지영상에 기반을 두어 작동하기 때문에 배경제거가 필요 없고 대상자의 옷, 텍스처, 키에 크게 구애받지 않는 장점이 있다. 또한 속도에서는 그림구조모델에서 얻은 특징점 위치로부터 단일 에지변환영상의 픽셀 값을 추적 유사도 계산에 사용하기 때문에 실시간으로 작동한다.

## 5. 결론

본 논문은 그림모델과 파티클 필터를 이용해 인체 정면 상반신 포즈를 인식하는 방법을 제시하였다. 인체는 큰 자유도를 갖고 있어서 모델링이 어려우며 모델이 복잡 할수록 영상에서 매칭이 어려운 단점이 있다. 본 연구를 통해 이러한 문제를 해결하고 쉽게 인식할 수 있는 방법을 제시하였다. 챔퍼매칭은 특징점과 매칭 방법에서 에지를 사용 하였으

므로 빛의 변화, 대상자의 옷 색, 옷 텍스처 변화에 강건하게 포즈인식이 작동하였다. 하지만 인체의 정면을 대상으로 했고 2차원 모델을 기반으로 했으므로 회전과 같은 복잡한 움직임에는 오류가 발생하는 단점이 있었다. 향후연구에서는 3차원 정보 추가로 그림구조 모델을 향상시키고 다양한 환경에서 챔퍼매칭을 효과적으로 구하는 방법을 연구할 것이다.

## 참고문헌

- [1] I. Haritaoglu, D. Harwood and L. Davis, "Ghost: A human body part labeling system using silhouettes Fourteenth International Conference on Pattern Recognition", Pattern Recognition, 1998. Proceedings, Fourteenth International Conference on ,vol1 , pp.77-82 1998.
- [2] R. Urtasun, P. Fua, "3D Human Body Tracking using Deterministic Temporal Motion Models", In European Conference on Computer Vision, pp. 92-196, 2004.
- [3] Felzenszwalb, P. F., Huttenlocher, D. P. "Pictorial Structures for Object Recognition", Int. J. Comput. Vision 61, 1 (2005), 55-79.
- [4] 노명철, 장혜민, 강승연, 이성환, "휴먼-로봇 상호작용을 위한 비전 기반3차원 손 제스처 인식", 2006년도 한국정보과학회 가을 학술발표 논문집, Vol.33, No.2, pp. 421-425.
- [5] C. Yang, R. Duraiswami and L. Davis, "Fast Multiple Object Tracking via a Hierarchical Particle Filter", ICCV05, pp 212-219, 2005.
- [6] Md. Z. Islam, C. M. Oh, J. S. Yang, and C. W. Lee, "DT Template based Moving Object Tracking with Shape Information by Particle Filter", IEEE Cybernetic Intelligent Systems, Vol. 1, pp. 127-132, 2008.
- [7] Borgefors, Gunilla "Hierarchical Chamfer Matching: A Parametric Edge Matching Algorithm", IEEE Trans. Pattern Anal. Mach. Intell., Vol. 6, pp. 849-865, Nov 1988.
- [8] H. Ling and K. Okada, "Diffusion Distance for Histogram Comparison", CVPR, Vol 1, pp 246-253, 2006.
- [9] P. Viola and M. J. Jones. "Robust Real-time Face Detection", Int'l Journal of Computer Vision, 52(2): 137-154, May 2004.
- [10] W. Yan, Q. Wang, Q. Liu, H. Lu, S. Ma, "Topology-Preserved Diffusion Distance for Histogram Comparison", BMCV 2007.