

뼈대-구조 능동형태모델을 이용한 사람의 자세 정합

(Human Pose Matching Using Skeleton-type Active Shape Models)

장 창 혁 ^{*}

(Changhyuk Jang)

요 약 본 논문은 뼈대-구조(skeleton) 형태의 Active Shape Models을 이용한 사람의 자세 정합에 대한 새로운 접근 방법을 제안한다. 제안된 방법은 모델 생성과 정합 과정에서의 빠른 수행 시간을 위해 기존 윤곽 형태(silhouette)의 모델이 아닌 뼈대-구조 형태의 모델을 적용하였다. 기존 Active Shape Models을 뼈대-구조 형태로 사람 자세 정합에 적용했을 경우 자세를 결정짓는 팔과 다리의 부정확한 정합은 사람 몸의 다양한 색상 정보와 전후(fore-rear direction)만을 고려한 특징점(landmark)의 방향정보로 인해 발생되며, 이러한 문제점은 입력 영상의 차영상 정보와 사람의 자세를 결정짓는 팔과 다리의 중요 특징점에 방향정보를 추가하여 해결하였다. 사람의 뼈대-구조 모델을 생성하기 위해 600개의 이미지를 사용하였으며, 생성된 형태 모델은 사람의 자세에 정합될 수 있는 17개의 특징점을 포함한다. 정합 과정에서 최대 30번 이하의 반복 과정을 수행 하며, 최대 수행 시간은 0.03초로 빠른 수행 시간의 결과를 얻었다.

키워드 : 능동형태모델, 뼈대-구조, 자세 정합

Abstract This paper proposes a novel approach for the model-based pose matching of a human body using Active Shape Models. To improve the processing time of model creation and registration, we use a skeleton-type model instead of the conventional silhouette-based models. The skeleton model defines feature information that is used to match the human pose. Images used to make the model are for 600 human bodies, and the model has 17 landmarks which indicate the body junction and key features of a human pose. When applying primary Active Shape Models to the skeleton-type model in the matching process, a problem may occur in the proximal joints of the arm and leg due to the color variations on a human body and the insufficient information for the fore-rear directions of profile normals. This problem is solved by using the background subtraction information of a body region in the input image and adding a 4-directions feature of the profile normal in the proximal parts of the arm and leg. In the matching process, the maximum iteration is less than 30 times. As a result, the execution time is quite fast, and was observed to be less than 0.03 sec in an experiment.

Key words : Active Shape Models, Skeleton-type, Pose matching

1. 서 론

최근 컴퓨터 비전 분야에서 사람의 자세를 인식할 때 마커(Marker)와 같은 물리적인 장비를 대신할 연구가 활발하게 진행되고 있으며, 카메라 감시 시스템과 HCI (Human-Computer Interaction) 등 영상 기술을 활용한 연구 분야에 다양한 해결책을 제시하고 있다. 고가의 모션 캡처 장비를 이용하지 않고 사람의 다양한 자세 (pose)을 분석하는 기술은 카메라를 이용한 여러 분야에 많은 연구 과제를 포함하고 있으며, 이를 해결하기 위한 다양한 컴퓨터 비전 관련 알고리즘이 연구되고 있다.

사람의 자세 분석에 이용되는 모델은 각 관절의 자유

^{*} 학생회원 : 숭실대학교 미디어학과
luee@ssu.ac.kr
논문접수 : 2009년 8월 19일
심사완료 : 2009년 9월 30일

Copyright©2009 한국정보과학회 : 개인 목적이나 교육 목적인 경우, 이 저작물의 전체 또는 일부에 대한 복사본 혹은 디지털 사본의 제작을 허가합니다. 이 때, 사본은 상업적 수단으로 사용할 수 없으며 첫 페이지에 본 문구와 출처를 반드시 명시해야 합니다. 이 외의 목적으로 복제, 배포, 출판, 전송 등 모든 유형의 사용행위를 하는 경우에 대하여는 사전에 허가를 얻고 비용을 지불해야 합니다.

정보과학회논문지 : 소프트웨어 및 응용 제36권 제12호(2009.12)

도(DOF: Degree of Freedom)에 따라 2D 모델과 3D 모델로 분류되며, 영상에서 사람의 자세 측정 범위에 따라 다양한 형태로 연구되고 있다. 2D 모델은 2차원 평면에서 x, y 축으로 회전이 가능한 모델로 영상에서 사람의 정면 자세와 제한된 범위의 측면 자세(사람의 걸음걸이) 분석에 이용되며, 3D 모델은 3차원 공간에서 x, y, z 축으로 회전이 가능하며 사람의 정면 자세 뿐만 아니라 복잡한 사람의 움직임에 적용된다. 2D 모델은 영상에서 사람의 정면 자세에 정확한 분석이 가능하며 빠른 수행시간의 장점이 있지만 측면 영상에 대해 취약하며, 3D 모델은 다양한 각도에서 사람의 자세 분석이 가능하지만 모델 생성과 수행시간이 느린 단점이 있다.

사람의 2D, 3D 모델은 분석하고자 하는 사람의 자세 형태에 따라 선택적으로 적용되며, 본 논문에서는 2D 모델을 적용범위로 한정하였다.

2. 기존 연구

모델을 이용한 사람의 자세 분석(human pose analysis)방법은 크게 두 가지로 분류될 수 있다. 사람의 자세 분석에 필요한 사전 정보(prior information)의 유무에 따라 모델-기반(model-based)분석 방법과 모델-자유(model-free)분석 방법으로 분류 된다[1]. 모델-자유 분석 방법은 분석하고자 하는 특정 객체를 포함한 이미지 데이터를 인지하기 위해 지역적인 에지(edge)나 영역(region)등과 같이 하위 레벨(low level)에서 분석하여 종합적인 그룹 데이터를 산출한다. Brand[2]는 HMM(Hidden Markow Model)을 이용하여 사람의 다양한 자세와 속도의 형태를 모델로 만들었으며, Albu et al.[3]은 Distance transform을 이용하여 사람의 각 부분을 추출하고 뼈대 형태(skeleton)의 모델을 생성하였다. 하지만, 모델-자유 분석 방법은 전역적인 모델(global model)이 없는 접근 방법으로 특정 객체를 분석할 때 잡음에 취약한 문제점이 있다[3].

모델-기반 분석 방법은 이미지에서 예상 되는 특정 모델의 사전 정보를 이용하며, 입력영상에서 분석하고자 하는 특정 객체(object)와 가장 적합하게 일치된 모델 정보를 찾게 된다. 모델의 사전 정보는 구조상의 잡음, 데이터의 손실등과 같은 구조적 복잡성에 의해 발생하는 잠재적인 문제점을 해결하는데 이용된다[1,4].

모델-기반 분석 방법을 이용하여 특정 객체를 분석하는 연구는 많은 분야에 응용되어 왔다. 의료 과학과, 사람의 얼굴 분석, 사람의 뼈대-구조(skeleton-type)를 이용한 자세의 인식 등 특정 객체의 통계적 분석을 기반으로 생성된 모델을 이용하는 실험은 모델의 형태에 따라 크게 닫힌-경계(closed-boundary) 모델과 열린-경계(open-boundary) 모델로 분류되며[4], 정확성 측면에

서 다른 분야에 적용되고 있다. 닫힌-경계 모델은 사람의 내부 조직과 같이 원의 형태에 적합한 모델이며, Kass et al.[5]이 제안한 Active Contour Models은 특정 객체의 윤곽선(boundary)을 이용하여 정보를 분석한다. 열린-경계 모델은 닫힌-경계 모델과 달리 모델의 형태가 뼈대-구조와 유사하며 특정 객체의 구조적 분석이나 실시간 추적(tracking) 연구 등에 이용된다. Navaratnam et al. [6]은 사람의 상체 부분에 대한 뼈대-구조 형태 모델 생성을 통해 사람의 자세를 인식하였으며, Shanon et al.[7], Haritaoglu et al.[8]과 Howe et al.[9]는 “cardboard person model”을 이용하여 사람의 걸음걸이와 자세를 인식하였다.

모델-기반 분석 방법 중에서 Kim et al.[10]은 능동형태모델(Active Shape Models)을 이용하여 사람의 움직임을 추적하는 연구를 하였다. 사람의 형태 모델을 닫힌-경계 형태로 생성하여 사람의 윤곽을 추출하는 연구로 두 팔과 두 다리를 이용하여 형성되는 다양한 자세를 정합하기에는 적합하지 않다. 기존 연구에서 능동형태모델이 사람의 몸과 관련된 분야에 적용된 것은 영상에서 사람의 윤곽을 추출하여 형태 모델을 생성하고, 새로운 입력 영상에서 사람을 추적하는 실험이었다[10,11]. 사람 모델을 윤곽 형태(닫힌-경계 형태)로 생성하여 적용 시킬 경우 사람의 자세에서 팔과 다리의 움직임에 따라 다양한 형태로 구분되는 자세 분석에 적합하지 않다.

본 논문에서는 다양한 형태의 사람 자세 정합을 위해 능동형태모델을 이용하여 뼈대-구조 형태 모델을 생성하고 정합하는 새로운 접근 방법을 제안한다[12]. Shi et al.[13]은 능동형태모델을 이용하여 중국의 문자인 한자를 인식하는 방법을 제안하였다. 이 연구는 한자 모델을 능동형태모델을 이용하여 생성하였으며, 본 논문에서 연구한 사람의 뼈대-구조와 유사한 형태라 할 수 있다.

제안된 방법은 모델 생성과 정합 과정에서의 빠른 수행 시간과 다양한 자세의 정확한 분석을 위해 기존 윤곽 형태의 사람 모델이 아닌 뼈대-구조 형태의 모델을 적용한다. 하지만, 능동형태모델을 뼈대-구조 형태로 사람의 자세 정합에 적용했을 경우 자세를 결정짓는 팔과 다리의 정확한 정합에 문제점이 있다. 사람의 몸은 얼굴과 다르게 다양한 픽셀 정보를 포함 하므로 기존 능동형태모델의 방법을 그대로 적용할 경우 정확한 정합이 이루어 지지 않는다. 사람 몸의 다양한 색상 정보와 전후(fore-rear direction)만을 고려한 특징점의 방향정보로 인해 발생하는 문제점은 입력 영상의 차영상 정보와 사람의 자세를 결정 짓는 팔과 다리의 중요 특징점의 측면 기준(profile normal)에 방향정보를 추가하여 해결하였다. 실험에 사용된 차영상은 정해진 배경 정보에서 사람의 자세 정보를 수작업으로 추출하여 이용하였다.

본 논문에서 제안된 방법은 모델 생성 단계(training process)와 모델 정합 단계(matching process)로 분류되며, 능동형태모델에 대한 자세한 사항은 Cootes et al. [14]에 설명되어 있다. 제안된 방법은 모델 정합 단계에서 최대 30번 이내 반복 수행으로 가장 적합한 형태의 모델 정보를 추출하였으며, 최대 수행 시간은 0.03초 이내의 빠른 수행 시간을 얻었다.

논문의 구성은 다음과 같다. 3장은 능동형태모델을 이용한 뼈대-구조 모델 생성을 설명한다. 4장은 생성된 형태 모델을 이용하여 새로운 입력 영상에서 사람 자세 정합 과정과 기존 능동형태모델의 정합 알고리즘에 추가된 방법을 설명한다. 5장은 뼈대 구조 형태 모델과 윤곽 형태 모델을 사람의 자세 정합에 적용한 실험을 비교 설명하며, 6장은 결론과 향후 연구 방향에 대해 설명한다.

3. 뼈대 모델 생성

사람의 자세 정합을 위해 사용되는 모델-기반 분석 방법은 사람의 몸과 관련된 사전 정보를 갖게 된다. 사람의 자세 정합에 사용되는 모델은 응용 분야에 따라 막대-형태(stick-figure)와 체적-형태(volumetric-figure)로 분류된다[1]. 사람의 외형 윤곽 형태로 구성되는 체적-형태 모델은 사람의 위치 추적 관련 연구에 이용된다[10,11]. 이러한 형태의 모델은 사람의 위치 추적과 다르게 다양한 형태 정보를 분석해야 하는 사람의 자세 정합에 적합하지 않다.

능동형태모델의 일반적인 특징은 얼굴 관련 실험에 적용된 사례에서 열린-경계 형태와 닫힌-경계 형태를 모두 적용할 수 있다는 것을 알 수 있으며 눈과 눈썹, 입은 닫힌-경계에 해당하며, 코, 턱 선은 열린-경계에 해당된다.

Cheng et al.[15]은 능동형태모델과 HMM(Hidden Markov Model)을 이용하여 뼈대-구조 기반으로 사람의 걸음걸이 분석 방법을 제안하였다. 하지만, 이 방법은 능동형태모델을 이용하여 뼈대-구조 모델을 생성하여 사람의 자세 분석에 적용한 것이 아닌 사람의 걸음걸이에 제한적으로 적용한 것으로 본 논문에서 제안한 방법과 차이가 있다.

능동형태모델을 사람의 몸과 관련된 실험에서 닫힌-경계 형태 모델로 적용할 경우 사람의 자세에서 팔을 들었는지, 다리를 벌렸는지 등을 구분 짓기 위해 추가적인 작업이 필요하며 정확성 측면에서도 문제점이 있다.

사람은 두 팔과, 다리를 이용하여 다양한 자세를 한다. 막대-형태, 뼈대-구조 모델은 체적-형태 모델 보다 다양한 형태의 사람 자세 정합에 적합한 형태라 할 수 있다[16].

2D 이미지에서 뼈대-구조에 기반한 사람의 자세 정합은 뼈대-구조 모델을 생성하는 단계부터 시작 된다. 2D 이미지에서, 뼈대-구조 모델은 이미지에서 객체의 중간 축과 관련이 있다[17].

본 논문에서는 모델-기반 분석 방법을 이용하여 뼈대-구조 형태의 모델을 생성하고 사람 자세 정합을 수행한다. 뼈대-구조 모델은 Cootes et al.[14]이 제안한 능동형태모델을 이용한다. 능동형태모델은 새로운 입력 영상에서 특정 객체를 분석하기 위해 반복적으로 변형되는 형태 객체의 통계적인 모델이며, 훈련 단계에서 분석된 통계적 변형 정보는 정합 과정에서 수행 시간이 빠른 장점이 있다. 기존 연구에서 능동형태모델은 얼굴 영상과 의료 영상에서 닫힌-경계 형태 모델로 적용되어 왔으며 많은 연구 성과를 산출 하였지만, 사람의 자세 정합에 적용할 경우 기존 닫힌-경계 형태 모델은 특징점의 증가로 인해 모델 생성과 정합 과정에서 수행 시간이 오래 걸리는 문제점이 있다. 능동형태모델의 빠른 수행 시간과 사람 자세의 다양한 형태에 정확한 정합을 위해 본 논문에서는 열린-경계 형태 모델인 뼈대-구조 모델을 적용한다. 600개의 몸 이미지에서 사람의 몸에 각 관절과 사람 자세의 특징을 나타내는 17개의 특징점을 이용하여 뼈대-구조 모델 생성을 위한 정보를 얻는다. 그림 1은 사람의 몸에 지정되어 있는 전체 특징점과 평균 형태 모델을 보여 준다. 우리는 이미지에서 뼈대-구조 모델 정보를 얻기 위해 수작업으로 특징점을 지정하였다.

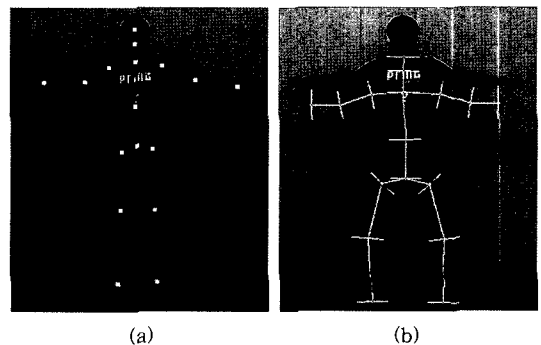


그림 1 특징점 정보를 이용한 평균 형태 모델(mean shape model) 생성: (a) 특징점 (b) 평균 형태 모델(mean shape model)

사람 자세의 다양성 정보는 600개의 전체 이미지에서 뼈대-구조를 생성하는 각 특징점으로 지정된 좌표 정보를 이용한 통계적 분석으로 훈련과정을 통해 획득되며, 모델의 형태 다양성(shape variation) 분석을 위해 각 형태 집합의 정렬(alignment)과정이 수행된다. 정렬은

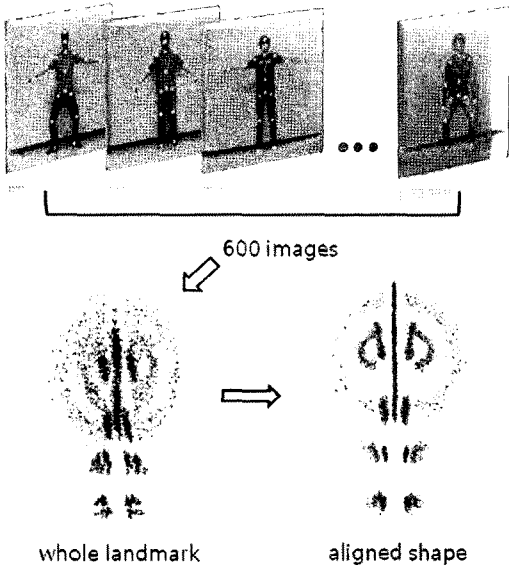


그림 2 600개 이미지: 특징점과 정렬 형태

형태 집합의 크기조정(scaling), 회전(rotation), 이동(translating) 연산과 거리의 최소화 과정을 통해 각 형태 집합을 가장 근접하게 일치시키는 과정이다. 그림 2는 600개 이미지의 전체 특징점 정보와 정렬 과정을 수행한 결과를 하나의 이미지로 보여준다.

정렬된 형태 집합은 특징점이 n 개인 벡터 X 로 표현할 수 있다. 우리는 17개의 특징점으로 하나의 형태 모델을 표현하고 각 단일 벡터는 $2n$ 개의 요소 벡터로 구성되며, 다음과 같다.

$$x_i = (x_{i,0}, y_{i,0}, x_{i,1}, y_{i,1}, \dots, x_{i,n-1}, y_{i,n-1})^T, i = 1, \dots, N \quad (1)$$

T 는 전치 연산을 의미하고, N 은 훈련 이미지 수를 의미한다. 평균 형태 벡터는 \bar{x} 다음과 같이 연산된다.

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i \quad (2)$$

형태 모델 특징점의 유기적인 움직임을 의미하는 변화량은 평균으로부터의 편차 Δx_i 에 주성분 분석(PCA: principal component analysis)을 적용하여 구할 수 있으며,

$$\Delta x_i = x_i - \bar{x} \quad (i = 1, \dots, N) \quad (3)$$

이러한 편차들로부터, $2n \times 2n$ 의 공분산 행렬(covariance matrix) S 는 다음과 같이 연산될 수 있다.

$$S = \frac{1}{N} \sum_{i=1}^N \Delta x_i \Delta x_i^T \quad (4)$$

형태 모델 point의 변화량은 공분산 행렬(covariance matrix) S 의 $2n$ 개 고유벡터(eigenvector), P_1, \dots, P_{2n} 와 일치하는 $2n$ 고유값(eigenvalue)으로 설명될 수

있다.

벡터 x_i 와 일치하는 k^{th} 번째 주성분(principal component)은 다음과 같은 선형조합에 의해 정의된다.

$$b_{i,k} = P_k^T (x_i - \bar{x}), i = 1, \dots, N, k = 1, \dots, 2n \quad (5)$$

주성분들은 훈련 과정에서 분석된 형태의 변화량에 대해 선형 독립(linear independent)를 나타낸다. 가장 큰 값을 갖는 고유값 λ_1 에 해당하는 첫 번째 고유벡터는 형태 변화량이 가장 큰 정보를 갖는다. k^{th} 주성분으로 설명되는 전체 형태 변화량의 비율은 λ_k 와 일치한다. 전체 형태 변화량은 $2n$ 보다 작은 수의 주성분으로 설명이 가능하며, 이를 t 라 한다($t < 2n$). t 값은 훈련 과정에서 분석되는 형태에 대해 97%이상의 전체 변화량을 포함하는 처음 t 개의 주성분을 선택함으로써 값을 할당하게 된다. 전체 변화량 λ_T 는 다음과 같이 정의되며,

$$\lambda_T = \sum_{k=1}^{2n} \lambda_k \quad (6)$$

실험을 통해 처음 3개의 주성분을 선택하게 된다.

훈련 과정에서 정렬된 형태 집합들은 평균 형태와 처음 t 개의 고유벡터와 관련된 고유값의 선형조합의 합으로 표현된다.

$$x_i = \bar{x} + P b_i, i = 1, \dots, N \quad (7)$$

여기서 $P = (p_1 p_2 \dots p_t)$ 는 처음 t 개의 주성분을 의미하고 $b_i = (b_{i,1}, b_{i,2}, \dots, b_{i,t})^T$ 는 다음 식으로 연산된 주성분을 의미한다.

$$b_i = P^T (x_i - \bar{x}), i = 1, \dots, N \quad (8)$$

식 (8)은 b_i 값에 따라 새로운 형태 모델이 생성됨을 의미한다.

4. 뼈대-구조 정합

능동형태모델은 형태 모델을 이용하여 새로운 입력 이미지에서 특정 객체를 해석(interpretation)하는 정합 알고리즘을 포함하고 있으며, 의료 과학과 같은 많은 연구 분야에서 능동형태모델의 모델 생성과 모델 정합의 특성을 응용하고 있다[18-20].

능동형태모델에서 형태 모델은 각 특징점에 측면 기준을 갖고 있으며 이미지에서 가장 적합한 위치를 찾는 데 활용된다. 측면 기준은 형태 모델의 각 관절 마디를 의미하는 특징점의 수직 방향에 대한 표면 픽셀 정보를 의미한다. 특징점의 측면 기준은 기존 능동형태모델의 정합 알고리즘에 이용된 픽셀 정보로 사람의 얼굴과 같이 비슷한 표면 픽셀 정보를 갖고 있는 객체에 적용할 때 가능한 방법이다(그림 1(b)). 얼굴 연구 분야에서는 새로운 입력 이미지에서 얼굴 영역을 분석하는 방법으로 능동형태모델 본래의 정합 방법을 활용하지만[21,22], 사람의 자세를 분석하는 연구에서 기존의 정합 방법인

특징점 주변 측면 기준을 이용하는 것은 다양한 색상의 옷을 입은 사람의 몸은 얼굴과 같이 비슷한 픽셀 정보를 포함하고 있지 않기 때문에 부정확한 정합결과를 산출한다. 또한, 측면 기준을 이용한 정합 방법은 뼈대-구조 형태 모델을 이용한 정합 실험에서 팔과 다리와 같은 몸의 끝부분에 대한 정합이 정확하게 수행되지 않는다. 얼굴 관련 연구에서는 형태 모델이 원의 형태(oval figure)로 형성되기 때문에 전후, 두 가지 방향성을 갖는 측면 기준을 이용한 정합이 가능하다.

각 특징점은 얼굴 관련 실험에서 사용되는 두 가지 방향 정보 이상(위, 아래)의 방향성을 고려해야만 한다. 이러한 문제점을 해결하기 위해, 우리는 기존 능동형태 모델의 정합 알고리즘에 두 가지 사항을 추가하여 사람의 자세 정합을 수행한다. 얼굴과 의료 영상 과학에서 사용되는 특징점 주변 픽셀 정보를 이용한 정합 방법을 사람의 몸에 적용하였을 경우 발생하는 다양한 색상 정보에 따른 문제점은 새로운 입력 이미지에서 측정하고자 하는 특정 객체(사람 몸 영역)에 대한 차영상 연산을 추가하여 해결하였으며, 원 형태에서 가능한 측면 기준의 두 가지 방향성 문제는 사람의 몸 뼈대-구조의 팔과 다리의 끝부분 특징점의 측면 기준에 위, 아래의 방향정보를 추가하여 해결하였다. 그림 3은 뼈대-구조 모델을

이용하여 새로운 입력에서 사람의 자세를 추출하는 과정을 보여준다. 표 1은 정합 과정에 대한 알고리즘이다.

5. 분석 및 실험 결과

기존 2D 영상에서 사람의 자세 정합에 관한 연구로 능동형태모델을 적용한 사례는 Kim et al.[7]의 사람의 움직임을 추적하는 연구로 뼈대-구조 모델이 아닌 단한-경계 형태 모델로 본 논문의 실험 결과 비교에서 기준 데이터가 다른 문제점이 있다. 기존 연구와의 비교 실험은 제안된 뼈대-구조 모델이 이미지에서 정합 과정을 수행할 때 수행 시간과 정합이 수행된 결과 이미지에서 사람의 자세와 뼈대-구조 모델간의 거리 값 측정으로 대체한다.

단한-경계 형태 모델과 열린-경계 형태 모델을 사람의 자세 정합에 적용하였을 경우 모델 생성과 정합 과정에서의 수행 시간과 정확도를 비교하여 열린-경계 형태 모델의 장점을 설명할 수 있다. 훈련 과정을 통해 모델을 생성할 때 단한-경계 형태 모델과 열린-경계 형태 모델의 생성 시간이 많은 차이가 있음을 알 수 있다. 사람의 윤곽을 중심으로 형태 모델을 생성할 경우 특징점 수가 상당 부분 증가 되므로, 모델 생성에 필요한 연산 시간이 늘어나게 된다. 실험을 통해 단한-경계 형태 모델의 평균 특징점 수는 약 60개 이상이였으며, 뼈대-구조 형태 모델의 특징점 수는 20개 이하로 많은 차이가 남을 알 수 있었다. 60개 이상의 특징점을 이용하여 생성된 단한-경계 형태 모델과 17개의 특징점을 이용하여 생성된 뼈대-구조 형태 모델을 이용하여 정합과정에서 반복횟수에 대한 정확도를 비교 분석하였다. 사람의 팔과 다리 영역을 8개의 블록으로 분할하였으며, 이 영역은 형태 모델이 정확하게 정합됐는지를 판단할 수 있는 영역으로 설정하였다. 60개 이상의 특징점을 이용하여 모델을 생성하였을 경우, 약 30개의 특징점이 팔과 다리, 8개의 영역에 포함되며, 17개의 특징점을 이용하여 모델을 생성하였을 경우, 8개의 특징점이 8개의 영역에 포함 되었다. 특징점 수에 따른 정확도 분석은 다음 식을 통해 산출된다.

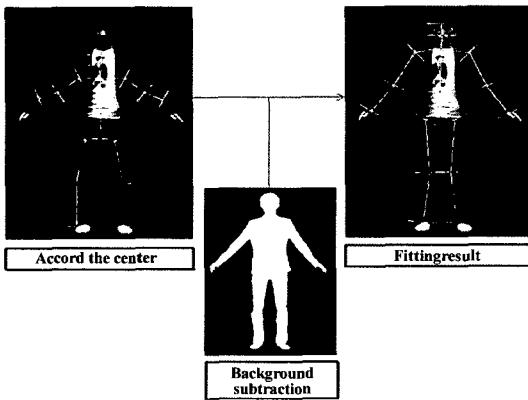


그림 3 정합 과정

표 1 정합 알고리즘

단계 1	형태 파라미터, b , "0" 초기화(평균 뼈대 모델(mean skeleton model)).
단계 2	중심 일치(차영상 영역과 평균 뼈대 모델(mean skeleton model)을 이용).
단계 3	모델 객체 생성, $x \cong \bar{x} + Pb$
단계 4	차영상 영역에서, x , 각 특징점의 위치를 찾는다 (팔,다리 8개 특징점 : 4 방향정보). x , 와 정합하기 위해 모델 파라미터를 갱신한다.
단계 5	$b = P^T(x - \bar{x})$
단계 6	b , 단계 3에 적용.
단계 7	정합되지 않을 경우, 단계 4 다시 수행.

$$1 \cong \frac{\text{정합된 영역의 갯수}}{\text{팔과 다리 8개 영역}} \quad (9)$$

8개의 팔과 다리 영역에 특징점이 모두 포함된 경우 "1"에 가까운 수치가 산출되며 이 경우 정확한 정합의 결과로 간주한다.

그림 6은 뼈대-구조 모델과 닫힌-경계 형태 모델이 사람의 팔과 다리, 8개 영역에 정합된 결과를 보여준다. 그림 6(b), (c)에서 알 수 있듯이, 모델의 형태에 따라 정확한 정합을 위한 팔과 다리 영역의 특징점 수가 많은 차이가 남을 알 수 있으며, 이는 정합 과정에서 반복 횟수에 영향을 미친다. 그림 4의 (a)는 사람의 팔과 다리에 뼈대-구조 모델이 정합되는 8개 영역을 나타내며, (b)는 17개의 특징점을 이용한 뼈대-구조 모델 정합을 보여 주며, (c)는 기존 닫힌-경계 모델정합 결과를 보여 준다.

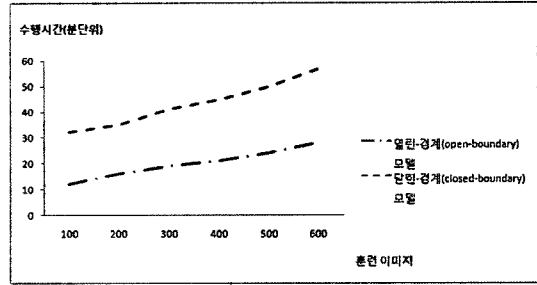
60개 이상의 특징점을 이용하여 윤곽 형태의 모델을 생성하였을 경우 정확한 정합이 이루어지기 위한 반복 횟수가 17개의 특징점을 이용하여 뼈대-구조 모델을 생성하였을 경우보다 증가됨을 알 수 있었다.

그림 5는 사람의 윤곽을 중심으로 닫힌-경계 모델과 뼈대-구조 형태 모델을 적용하였을 경우 모델 생성 시간과 정확한 정합이 이루어지기까지의 반복 횟수를 나타낸 그래프이다.

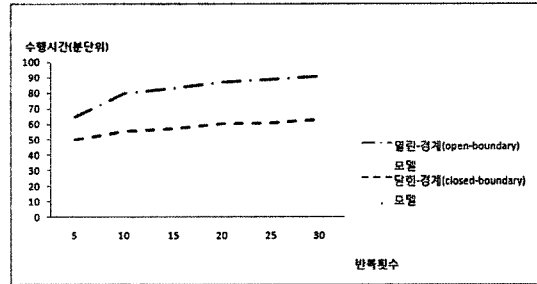
뼈대-구조 모델을 사람의 자세 정합에 적용하였을 경우 정합과정에서 각 반복 주기 별 정확도를 측정하기 위해 정합된 모델과 사람이 직접 산출한 모델의 위치 정보간 거리 값을 측정하여 유사 도를 비교하였다. 거리 값을 산출하는 식은 다음과 같다.

$$Distance(X,Y) = \sqrt{(X-Y)(X-Y)^T} \quad (10)$$

팔과 다리의 8개 영역에 해당되는 각 뼈대-구조 모델의 특징점이 포함된 경우 정합 과정에서 반복 주기의 종료료를 의미하며, 그림 6은 각 반복 주기 별 거리 값을



(a)



(b)

그림 5 모델 형태에 따른 생성 시간과 정확도 비교 그래프: (a)모델 생성 시간 비교, (b) 정합과정의 정확도 비교

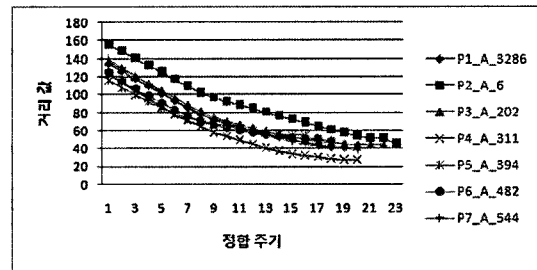
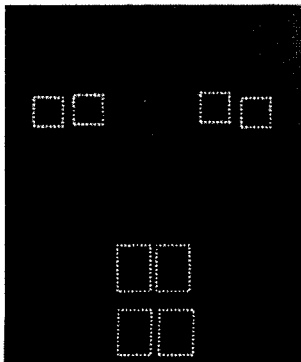
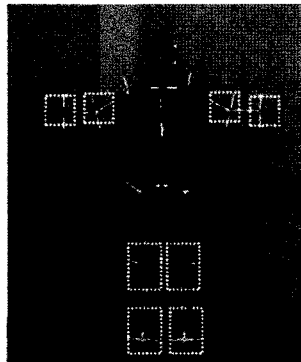


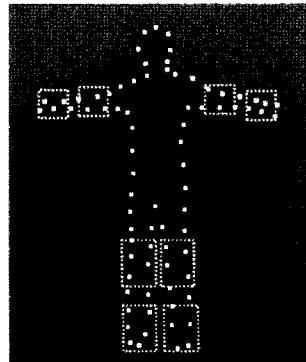
그림 6 정자세 정합과정에서의 반복 주기 별 유사도 측정



(a)



(b)



(c)

그림 4 팔과 다리, 8개 영역과 정합 결과

나타낸다. 실험을 통해 8개 영역에 각 특징점이 포함되어 정확한 정합이 이루어진 결과를 나타내는 거리 값은 60 이하임을 알 수 있었다. 그림 6(a)의 가로 항목은 7명의 사람(P1~P7)에 대한 자세(정자세: A)와 이미지 번호를 나타내며, 세로 항목은 정합 과정에서의 반복 주기를 의미한다.

그림 7은 본래 능동형태모델의 정합 알고리즘을 이용하여 뼈대-구조 형태로 사람의 자세에 적용하였을 경우 부정확한 실험 결과를 보여준다.

이러한 문제점을 해결하기 위해 우리는 정합 과정에서 두 가지 사항을 추가 하였다. 사람의 몸은 다양한 색상 정보를 갖고 있으므로 모델 생성 단계에서 측면 기준의 주변 픽셀 정보를 이용하여 또 하나의 측면 기준 모델을 생성하지 않고, 입력 이미지에 대해 차영상을 추출하여 사람의 몸 색상 정보를 흰색과 검정색으로 단순화 시켰다. 정합 과정에서 이용되는 측면 기준은 차영상으로 추출된 사람 영역의 흰색과 검정색을 판별하여 특징점의 위치를 이동 시킨다. 새로운 입력 이미지에서 사람 영역의 차영상 정보를 이용할 경우 다양한 색상 정보에 정확한 정합을 수행 하지 못한 문제점은 해결

되지만, 다양한 자세에서 팔과 다리의 정확한 정합이 수행되지 않는 문제점이 있다. 이러한 문제점은, 17개의 특징점 중 팔과 다리 끝부분을 의미하는 8개의 특징점에 속한 측면 기준에 대해 위아래와 좌우, 즉 4가지 방향성을 고려 하도록 하여 팔과 다리의 정확한 정합이 이루어 지도록 하였다. 이러한 두 가지 추가 사항으로 본래 능동형태모델의 정합 알고리즘을 사람의 자세에 적용하였을 경우 발생하는 문제점을 동시에 해결하였으며, 모델 생성 과정에서 각 특징점의 측면 기준에 대한 추가적인 모델을 생성하지 않으므로 생성 시간을 단축 시켰다.

본 논문에서 제안된 방법은 Pentium(R) 4 CPU 3.00GHz, 2.00GB RAM을 이용하여 사람 몸의 정면 영상을 고려하여 실험하였으며, 훈련 과정에서 뼈대-구조 모델을 생성하기 위해 해상도 640*480, 600개의 이미지를 사용하였다. 모델 생성 실험을 통해 사람의 자세 변화 량을 표현 할 수 있는 97% 이상의 주성분이 3개의 축으로 축소됨을 알 수 있었다. 표 2는 주요 축(principal axis)의 비율을 보여 준다.

그림 8-그림 12까지는 다양한 사람의 자세에 대한 뼈

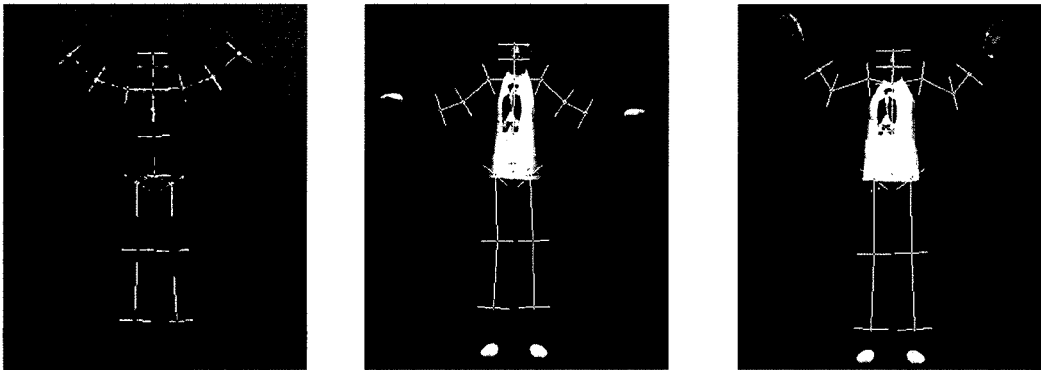


그림 7 부정확한 결과

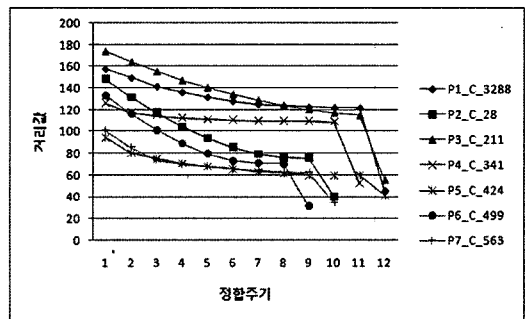
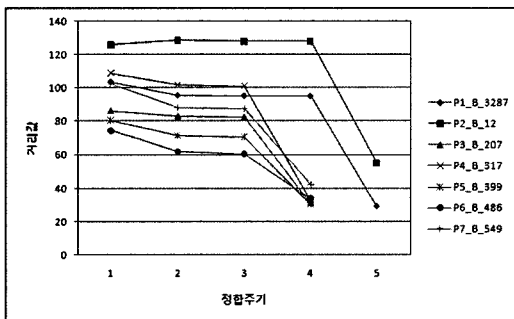


그림 8 양팔 벌린 자세 정합 과정의 정합 주기 별 유사도 측정

그림 9 양팔 올린 자세 정합 과정의 정합 주기 별 유사도 측정

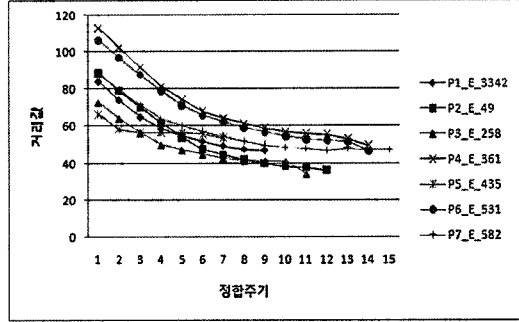
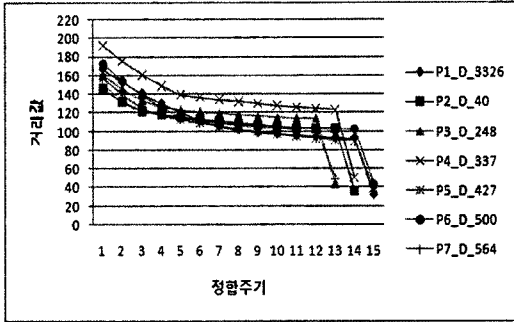
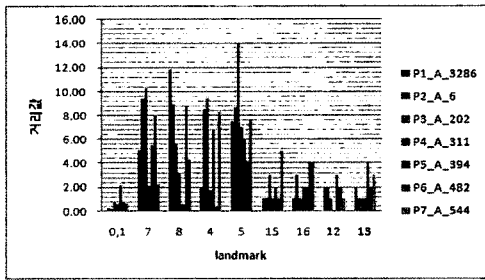
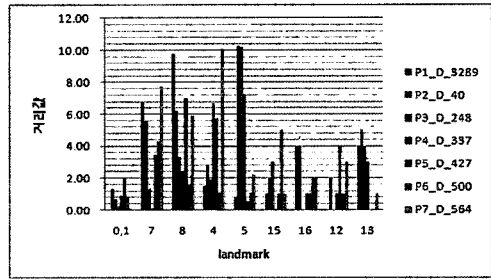


그림 10 양팔 머리 위로 올린 자세 정합 과정의 정합 주기 별 유사도 측정

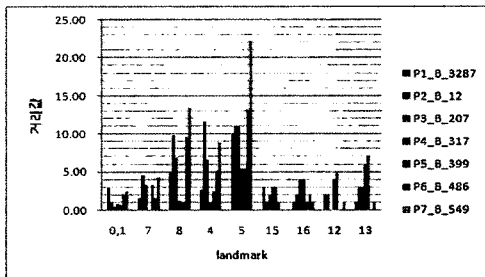
그림 11 양 팔 내린 자세 정합 과정의 정합 주기 별 유사도 측정



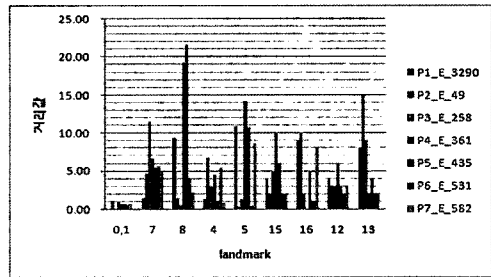
(a)



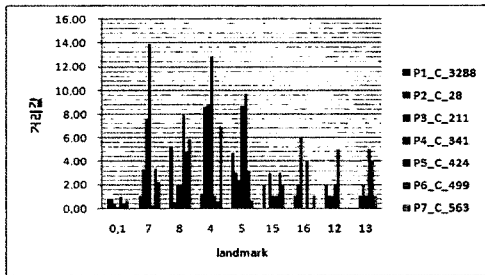
(d)



(b)



(e)



(c)

그림 12 정합 후 각 관절 별 유사도 측정

대-구조 능동형태모델의 정합 과정에서 각 주기 별 거리 값(pixel 단위)을 측정된 결과이다. 정합과정에서 반복 주기 별 거리 값이 60이하로 감소 될 때 사람의 자세에 정확한 정합이 이루어진다는 사실을 알 수 있으며, 이때의 반복 주기는 최대 30 이하이다. 그림 8은 정합과정에서 반복 주기가 5이하 이며, 그림 9-그림 11은 15 이하이다. 자세 정합에 적용된 뼈대-구조의 평균 모델과 유사한 자세에서는 정합과정에서의 반복 주기가 적음을 알 수 있다. 그림 12는 정합된 뼈대-구조 모델과 사람이 손으로 측정된 뼈대-구조 모델(사람의 자세 특징을 나타내는 특징점의 위치 정보)의 각 관절 별(머리, 팔, 다리, 몸) 거리 값을 측정된 결과 이다. 그림 12의 (a)에서 가로축은 17개 특징점에서 머리(0,1), 오른쪽 팔꿈치

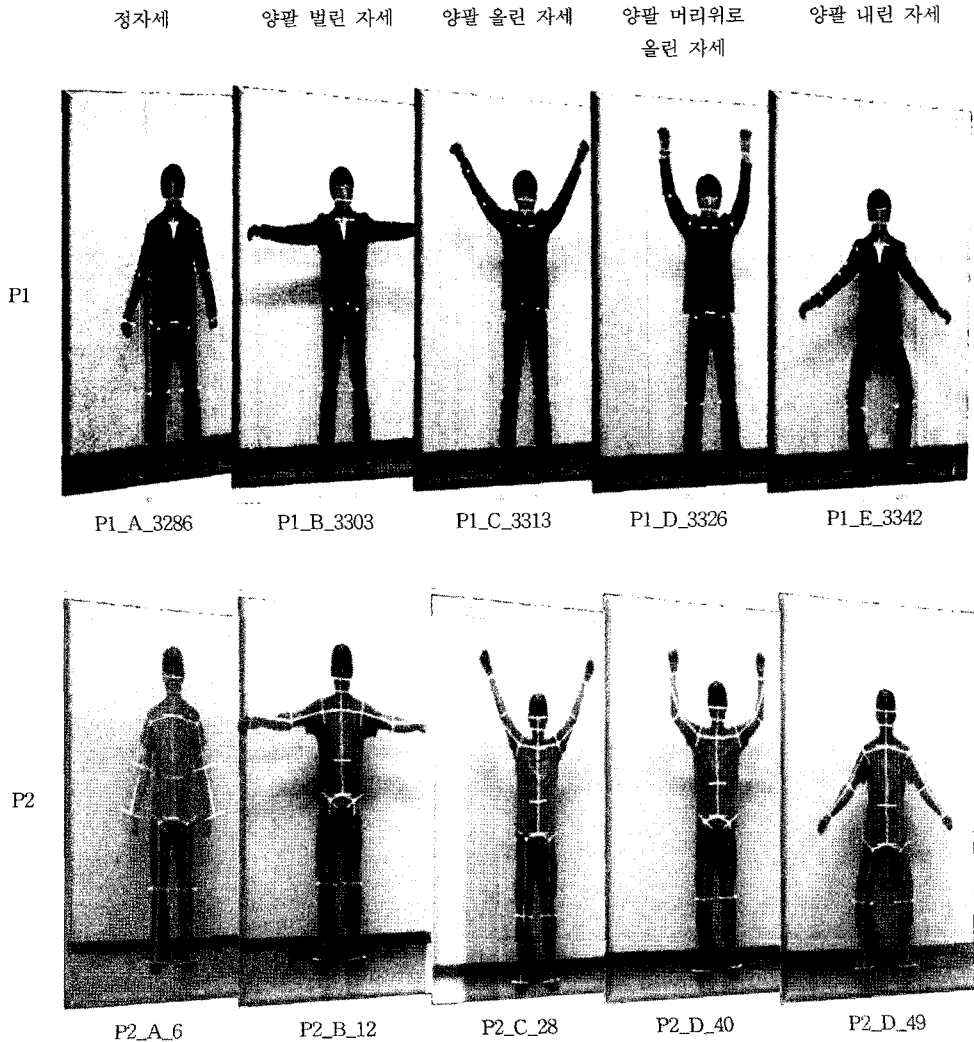
표 2 주성분

Shape mode variation		
1.	80.62%	(80.62%)
2.	12.68%	(93.3%)
3.	4.02%	(97.32%)

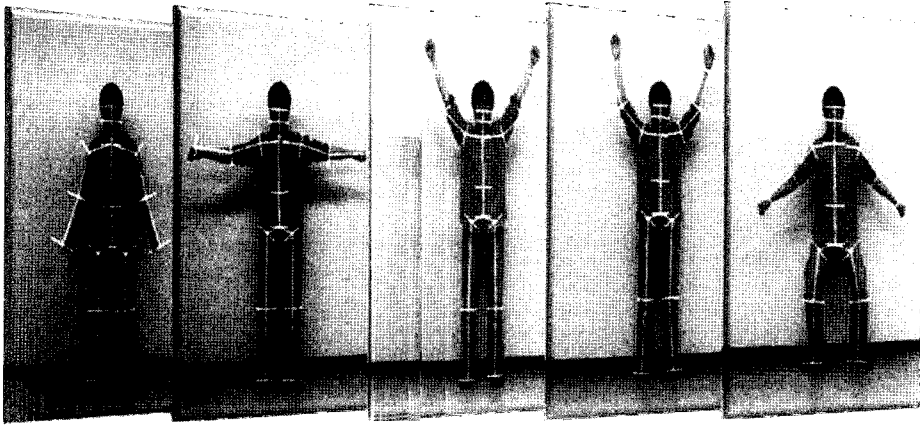
(7), 오른 손(8), 왼쪽 팔꿈치(4), 왼손(5), 오른쪽 무릎(15), 오른쪽 발(16), 왼쪽 무릎(12), 왼쪽 발(13)과 전체 평균을 나타낸다. 그림 12의 (a)-(e)는 그래프를 통해 사람의 팔을 나타내는 특징점(7, 8, 4, 5)에서의 거리 값이 사람 몸의 다른 부분 보다 큰 것을 알 수 있다. 이것은 사람의 자세 정합에서 팔의 움직임에 대한 정합이 다른 부분과 비교해 큰 오차가 발생하는 것으로 팔에 대한 정확한 정합이 사람의 자세 정합에서 가장 중요한 요소라는 것을 알 수 있다. 그림 8-그림 11에서 사람의 자세에 따른 다양한 정합 주기는 이미지에서 사람의 크기와 밀접한 관계가 있으며, 실험에 사용된 평균 모델과 유사한 크기의 이미지에서는 적은 수의 반복 주기를 갖는다.

결과를 보여준다. 제안된 방법은 정면에서 사람의 움직임만을 고려하여 실험 하였으며, 성공적인 정합 결과를 얻었다. 또한, 정합 과정에서 사람 몸의 표면 색상 정보를 이용하지 않고, 차영상과 추가된 측면 기준을 이용하여 빛의 영향을 받지 않는다. 정합 과정에서 최대 반복 주기는 30회 이하이며, 최대 수행 시간은 0.03초 이하로 매우 빠른 수행 시간결과를 얻을 수 있었다. 하지만, 우리는 한쪽 팔은 올리고 다른 쪽 팔은 내리는 자세와, 한쪽 다리만 구부리는 등의 자세에 대한 고려를 하지 않았기 때문에 모든 자세에 대한 실험 결과를 얻지는 못하였다. 사람의 다양한 자세에 대한 실험과 정면뿐만 아니라, 측면을 고려한 실험은 향후 개선된 알고리즘을 적용하여 발전시키겠다.

그림 13은 능동형태모델을 이용한 사람의 자세 측정



P3



P3_A_202

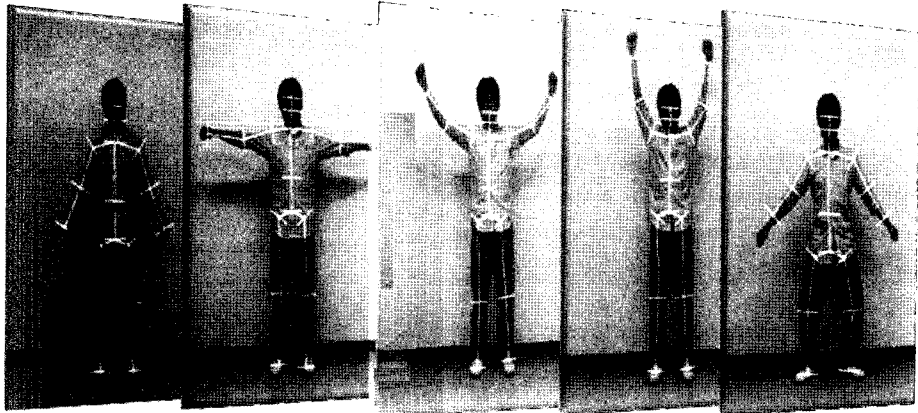
P3_B_207

P3_C_211

P3_D_248

P3_E_258

P4



P4_A_311

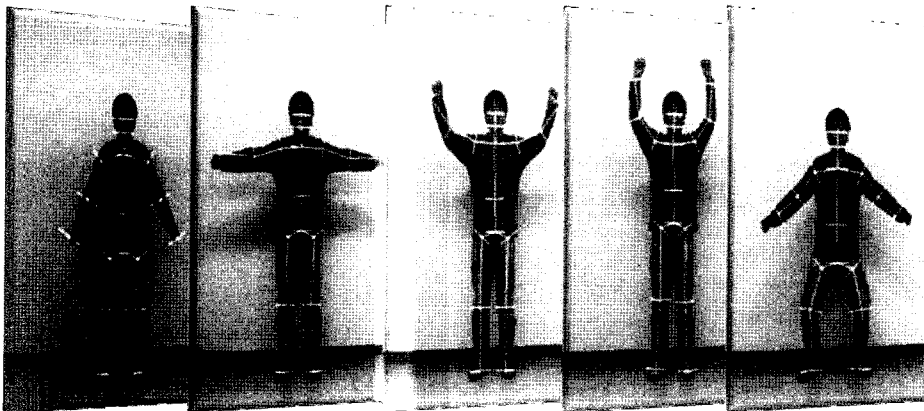
P4_B_317

P4_C_337

P4_D_341

P4_E_361

P5



P5_A_394

P5_B_399

P5_C_424

P5_D_427

P5_E_435

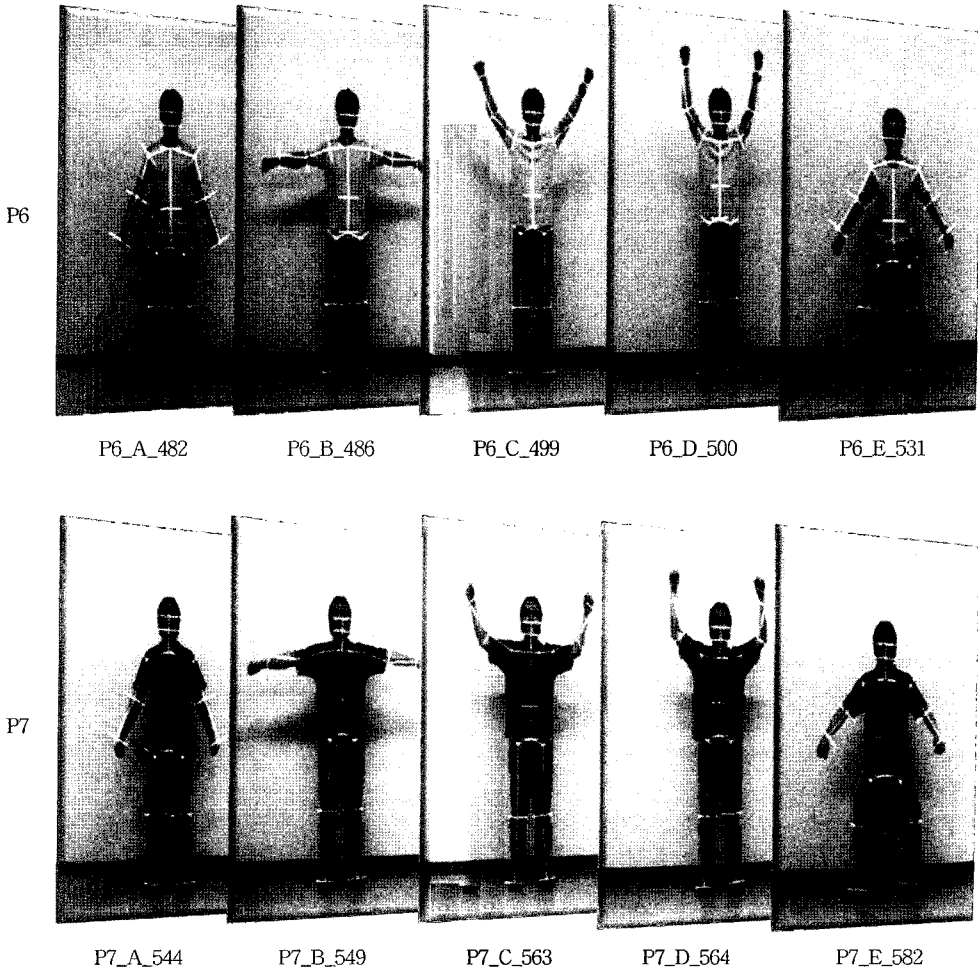


그림 13 다양한 자세에 대한 정합 실험 결과

6. 결론

본 논문은 뼈대-구조 형태의 능동형태모델을 이용하여 사람의 자세를 정합하는 모델-기반의 새로운 접근 방법을 제안한다. 뼈대-구조 모델은 사람의 자세 정합을 위한 특징점 정보를 포함한다. 사람은 두 팔과 다리를 이용하여 다양한 형태의 자세를 한다. 닫힌-경계 형태 모델을 사람의 자세에 적용할 경우 팔을 올린 자세인지, 다리를 구부린 자세인지에 대한 정확한 측정이 어려우며, 정확성을 높이기 위해 모델 생성에서 사람의 윤곽을 중심으로 특징점수를 증가시킬 경우 모델 생성과 정합 과정에서 연산이 늘어나게 되어 수행시간이 오래 걸리는 문제점이 있다. 이러한 문제점을 해결하기 위해 뼈대-구조 형태 모델을 사람의 자세 정합에 적용하였다. 사람의 윤곽에 따라 생성되는 닫힌-경계 모델과 비교해

상당히 적은 수의 특징점 정보를 이용하여 모델을 생성하고 정합할 수 있으며 모델 생성 시간과 정합 수행 시간에서 매우 효과적인 결과를 산출한다.

기존 원 형태 모델에 사용되는 정합 알고리즘을 그대로 뼈대-구조 모델에 적용 하였을 경우 정합 과정에서 정확한 정합이 이루어지지 않는 단점이 있다. 원의 형태와 뼈대 형태는 고려해야 할 방향 정보가 다르며, 기존 능동형태모델에서 사용된 측면 기준의 두 가지 방향성은 팔과 다리의 움직임에 의해 형성되는 위아래 방향에 대한 정확한 정합이 이루어 지지 않는 문제점이 있다. 또한, 기존 능동형태모델의 정합 알고리즘은 모델 생성 단계에서 추가적으로 각 특징점 주변에 수직으로 일정 길이만큼 픽셀 정보를 모델로 생성하여 정합 단계에서 새로운 입력 이미지의 특정 객체 경계 부분의 픽셀 정보를 확인 하는 방법으로, 사람의 얼굴과 몸 내부 조직

과 같이 일정한 색상 정보를 갖는 객체의 분석에는 적합하지만, 사람의 몸과 같이 다양한 색상 정보를 갖는 분석에는 적합하지 않다. 이러한 문제점들을 해결하기 위해 능동형태모델을 이용한 사람의 자세 정합 과정에서 기존 알고리즘에 두 가지 사항을 추가하였다. 새로운 입력 이미지의 특정 객체에서 경계부분 픽셀 정보를 이용할 수 없는 문제점을 해결 하기 위해 입력 이미지의 차영상 정보를 이용하여 색상 정보를 흰색과 검정색으로 단순화시켰으며, 두 가지 방향성만 고려한 기존 능동형태모델의 측면 기준에 팔과 다리 8개의 특징점 측면 기준에 위아래의 추가적인 방향성을 고려하게 함으로써 팔과 다리 끝부분에 대한 정합 정확도를 높였다. 제안한 방법은 2D 이미지에서 생성된 모델을 이용하여 새로운 입력 이미지에서 특정 객체를 정합하는 방법이다. 사람의 자세 정보 추출을 위해 복잡하지 않은 배경 이미지에서 차영상을 추출하였으며, 영상의 정면이 아닌 측면에서 사람의 자세 정합에는 문제점이 있다. 향후 측면 영상을 고려한 실험을 통해 하나의 카메라를 이용하여 다양한 각도의 사람 자세를 정합할 수 있는 빠대-구조 모델을 생성하고, 빠른 수행 시간의 장점을 이용하여 실시간으로 사람의 자세를 추적할 수 있게 함으로써 3D 사람 모델 정보를 추출할 수 있는 연구로 발전시킬 계획이다.

참 고 문 헌

- [1] R. Poppe, "Vision-based human pose analysis: An overview," *Computer Vision and Image Understanding*, vol.108, pp.4-18, 2007.
- [2] M. Brand, "Shadow puppetry," *Proceedings of the International Conference on Computer Vision (ICCV99)*, vol.2, pp.1237-1244, 1999.
- [3] A. B. Albu, R. Bergevin, and S. Quirion, "Generic temporal segmentation of cyclic human pose," *Pattern Recognition*, vol.41, pp.6-21, 2008.
- [4] T. F. Cootes, C. J. Taylor, "Statistical Models of Appearance for Computer Vision," http://personalpages.manchester.ac.uk/staff/timothy.f.cootes/Models/app_models.pdf.
- [5] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active contour models," *In 1st International Conference on Computer Vision*, pp.259-268, London, June 1987.
- [6] R. Navaratnam, A. Thayananthan, P. H. Torr, and R. Cipolla, "Hierarchical part-based human body pose estimation," *Proceedings of the British Machine Vision Conference(BMVC'05)*, Oxford, United Kingdom, pp.479-488, 2005.
- [7] X. Shanon, J. B. Michael, and Y. Yacoob, "Card-board people: A parameterized model of articulated image pose," *Proceedings of the International Conference on Automatic Face and Gesture Recognition(FGR 96)*, Killington, VT, pp.38-44, 1996.
- [8] I. Haritaoglu, D. Harwood, and L. S. Davis, "W4s: A realtime system detecting and tracking people in 2 1/2D," *Proceedings of the European Conference on Computer Vision(ECCV'98)*, LNCS, vol.1(1406), pp.877-892, 1998.
- [9] N. R. Howe, M. E. Leventon, and W. T. Freeman, "Bayesian reconstruction of 3D human pose from single-camera video," *Advances in Neural Information Processing Systems(NIPS) 12*, Denver, CO, pp.820-826, 2000.
- [10] D. Kim, V. Maik, D. Lee, J. Shin, and J. Paik, "Active Shape model-based object tracking in panoramic Video," *Proc. ICCS, LNCS*, vol.3994, pp.922-929, 2006.
- [11] A. M. Baumberg, and D. C. Hogg, "An Efficient Method for Contour Tracking using ASM," *Pose of Non-Rigid and Articulated Objects, Proceedings of the 1994 IEEE*, pp.194-199.
- [12] C. H. Jang, and K. C. Jung, "Human pose estimation using ASM," *Proceedings of World Academy of Science, Engineering and Technology (WASET)*, vol.34, pp.312-316, 2008.
- [13] D. Shi, S. R. Gumm, and R. I. Damper, "Hand-written Chinese Radical Recognition Using Non-linear ASM," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.25, no.2, pp.277-280, 2003.
- [14] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, "ASM-Their Training and Application," *Computer Vision and Image Understanding*, vol. 61, pp.38-59, 1995.
- [15] I-C. Chang, and C.-L. Huang, "Skeleton-based Walking Pose Analysis Using Hidden Markov Model and ASM," *Journal of Information Science and Engineering*, vol.17, pp.371-403, 2001.
- [16] H. Sunder, D. Silver, N. Gagvani, and S. Dickinson, "Skeleton Based Shape Matching and Retrieval," *Shape Modeling International*, pp.130-139, 2003.
- [17] H. Blum, "A Transformation for Extracting New Descriptors of Shape," *Computer Vision and Image Understanding: MIT press*, pp.362-380, 1967.
- [18] B. V. Ginneken, A. F. Frangi, J. J. Staal, B. H. Romeny, and M. A. Viergever, "Active Shape Model Segmentation With Optimal Features," *IEEE Transactions on Medical Imaging*, vol.21, no.8, pp.924-933, 2002.
- [19] H. H. Thodberg, and A. Rosholm, "Application of the active shape model in a commercial medical device for bone densitometry," *Image and Vision Computing*, vol.21, pp.1155-1161, 2003.
- [20] T. F. Cootes, A. Hill, C. T. Taylor, and J. Haslam, "The use of ASM For Locating Structures in Medical Images," *Image and Vision Computing*, vol.12, no.6, pp.355-366, 1994.

- [21] W. Wang, S. Shan, W. Gao, B. Cao, and B. Yin, "An Improved Active Shape Model for Face Alignment," *Proceedings of the Fourth IEEE International Conference on Multimodal Interfaces*, pp.523-528, 2002.
- [22] S. Romdhani, S. Gong, and A. Psarrou, "A Multi-View Nonlinear Active Shape Model Using Kernel PCA," (*BMVC'99*), pp.483-492.



장 창 혁

2005년 숭실대학교 미디어학부 공학사
2009년 숭실대학교 미디어학과 공학석사
2009년~현재 숭실대학교 미디어학과 박사과정. 관심분야는 영상처리, 패턴인식, 컴퓨터비전