

회전무관 3D Star Skeleton 특징 추출

(Rotation Invariant 3D Star Skeleton Feature Extraction)

전 성 국 [†] 홍 광 진 [†] 정 기 철 ^{‡‡}
 (Sungkuk Chun) (Kwangjin Hong) (Keechul Jung)

요약 포즈인식은 최근에 유비쿼터스 환경, 행위 예술, 로봇 제어 등에서 그 필요성이 증가되고 있는 분야로써, 컴퓨터비전, 패턴인식 등에서 활발히 연구되고 있다. 하지만 기존의 포즈인식 연구들은 사람의 회전이나 이동에 따라서 불안정한 인식률을 보인다는 단점을 갖고 있다. 이는 포즈 인식을 위해 추출한 특징이 사람의 회전, 이동 등의 다양한 변수에 영향을 크게 받기 때문이다. 이를 극복하기 위하여 본 논문에서는, 다 시점(multi-view) 환경에서의 3D Star Skeleton과 주성분 분석(principal component analysis: PCA)에 기반한 사람의 회전에 강건한 특징 추출을 제안한다. 제안된 시스템은 포즈의 특징 추출을 위해 다 시점 환경 기반의 visual hull을 생성하는 과정에서 획득 가능한 깊이 정보를 표현하는 8개의 projection map을 입력데이터로 사용한다. 이를 통해 포즈의 3D 정보를 반영하는 3D Star Skeleton을 구성하고 주성분 분석 기반의 회전에 강건한 특징을 추출한다. 실험결과에서는 다양하게 회전된 사람으로부터 생성된 3D Star Skeleton에서 특징을 추출하고 다양한 인식기를 통해 포즈인식을 해보았으며, 제안된 특징 추출 방법이 사람의 회전에 강건함을 알 수 있었다.

키워드 : 포즈 인식, 제스처 인식, 주성분 분석, 특징 추출, 군집화 알고리즘, 다 시점 환경

Abstract Human posture recognition has attracted tremendous attention in ubiquitous environment, performing arts and robot control so that, recently, many researchers in pattern recognition and computer vision are working to make efficient posture recognition system. However the most of existing studies is very sensitive to human variations such as the rotation or the translation of body. This is why the feature, which is extracted from the feature extraction part as the first step of general posture recognition system, is influenced by these variations. To alleviate these human variations and improve the posture recognition result, this paper presents 3D Star Skeleton and Principle Component Analysis (PCA) based feature extraction methods in the multi-view environment. The proposed system use the 8 projection maps, a kind of depth map, as an input data. And the projection maps are extracted from the visual hull generation process. Though these data, the system constructs 3D Star Skeleton and extracts the rotation invariant feature using PCA. In experimental result, we extract the feature from the 3D Star Skeleton and recognize the human posture using the feature. Finally we prove that the proposed method is robust to human variations.

Key words : posture recognition, gesture recognition, PCA, feature extraction, clustering algorithm, multi-view environment

· 본 과제는 한국소프트웨어진흥원의 SW공학 요소기술 개발과 전문인력 양성 사업의 결과물임을 밝힙니다.

· This research was supported by Korea SW Industry Promotion Agency (KIPA) under the program of Software Engineering Technologies Development and Experts Education.

[†] 학생회원 : 숭실대학교 미디어학과

k612051@ssu.ac.kr

hongmsz@gmail.com

^{‡‡} 종신회원 : 숭실대학교 미디어학과 교수

kejung@ssu.ac.kr

논문접수 : 2009년 7월 6일

심사완료 : 2009년 8월 25일

Copyright©2009 한국정보과학회 : 개인 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 제10호(2009.10)

1. 서 론

최근에 유비쿼터스 환경(ubiquitous environment), 행위 예술(performing art), 비디오 감시 시스템(video surveillance system), 로봇 제어(robot control), 인간과 컴퓨터의 상호작용(human computer interaction: HCI) 등의 필요성이 부각됨에 따라 이들의 실현을 위해 연구되고 있는 분야 중 하나가 포즈인식이다. 포즈인식은 카메라나 몸에 부착된 센서 등을 통해 사람의 포즈를 인식하는 연구로, 기계학습(machine learning), 컴퓨터 비전(computer vision), 패턴 인식(pattern recognition) 등의 분야에서 연구가 진행 중이다. 하지만 포즈인식의 결과가 다양한 변수들, 예를 들면 사람의 회전, 이동 등의 변화에 의해 민감하다는 점은 안정된 포즈인식을 위해 반드시 해결해야 하는 문제이다.

비전 기반의 포즈인식 연구는 2D 방법, 3D 방법으로 나눌 수 있다. 2D 포즈인식은 명시적으로 신체의 특정 부분을 찾아 인식하는 방법과 포즈의 모양이나 특징 추출 기반의 방법으로 나뉘지고, 3D 포즈인식은 사용한 카메라의 수에 따라 단일 시점(single-view) 기반의 방법과 다 시점(multi-view) 기반의 방법으로 나눌 수 있다.

명시적 2D 방법은 신체의 팔, 다리, 머리 등의 위치를 찾고 이를 통해 포즈를 인식하는 방법으로 영상에서 신체의 각 부분의 위치와 신체 부분간의 위치 관계에 대한 사전정보가 필요하다. Haritaoglu et al.[1]은 단일 실루엣 영상의 수직, 수평 히스토그램과 데이터베이스의 영상과의 비교를 통해 포즈를 인식한다. Wren et al.[2]은 포즈의 모양과 외곽선의 다중 통계 모델을 통해 영상 분할 후 신체의 각 부분을 찾는다. Iwasawa et al.[3]은 세 단계로 수행되는 포즈인식 방법을 제안하였다. 먼저 포즈의 중심을 찾은 후 상체의 방향을 계산하고, 마지막으로 실루엣 분석을 통해 발, 손, 팔꿈치, 무릎을 추정한다. Girondel et al.[4]은 비디오 영상에서의 정적인 포즈인식을 위해 얼굴과 손의 위치, 신장과 몸의 중심의 위치의 특징을 이용하여 포즈를 인식하는 시스템을 제안하였다. 이러한 방법들은 좋은 포즈인식률을 얻기 위해서 신체의 각 부분들이 정확히 찾아져야 한다.

특징 추출에 기반한 2D 방법은 신체의 각 부분의 위치를 고려하지 않는 대신, 포즈의 특징을 찾고 이를 통해 포즈를 인식하는 방법이다. Baumberger와 Hogg[5]은 실루엣의 edge 중 두드러진 점들을 eigenshape을 통해 찾아 인식하는 방법을 제안하였다. Fujiyoshi et al.[6]은 포즈에 대한 실루엣 영상에 별 모양의 Star Skeleton을 생성하여 포즈를 표현하였다. Chen et al.[7]은 포즈를 Star Skeleton을 통해 인식하고 HMM을 통해 동작 인식에 적용하였다. Panini와 Cuchiara[8]는

포즈의 실루엣을 수직, 수평으로 투영하여 확률 분포 맵을 만들고 이를 통해 포즈를 인식하였다. 이와 같은 특징 기반의 포즈인식은 많은 시스템 자원을 요구하지 않으므로 실시간 인식이 필요한 시스템에 적합하다. 하지만 인식률이 카메라의 위치에 의존적이라는 단점을 가지고 있다.

3D 포즈인식 방법은 사용한 카메라의 수에 따라 단일 시점 기반 방법과 다 시점 기반 방법으로 나눌 수 있다. 단일 시점 기반의 포즈인식 방법으로, Kameda et al.[9]은 영상에 투영된 포즈와 실루엣의 외곽선간의 관계를 통해 회전 각을 결정하고 이 각각의 회전 각을 통해 포즈를 인식한다. Mori와 Malik[10]은 2D 영상에서 추출한 특징을 다 시점에서 촬영한 데이터베이스의 영상들과 매칭하여 관절을 찾고 3D 포즈를 추정하는 방법을 제안하였다. Shakhnarovich et al.[11]은 frontal-view로부터 획득한 많은 데이터베이스로부터 local sensitive hashing을 통해 빠르게 상체 포즈를 추정하는 방법을 제안하였다. Micilotta et al.[12]은 단일 영상에서 3D 상체 포즈 추정을 위해 AdaBoost를 통해 신체의 각 부분을 찾고 확률 모델을 통해 신체의 위치를 추정한다. 찾아진 신체의 실루엣과 edge map을 데이터베이스의 인공의 3차원 모델과 chamfer distance를 통해 매칭하여 3D 포즈를 추정한다. Boulay et al.[13]은 가상 공간을 모델링하고 가상의 사람모델을 통해 공간에서의 가능한 모든 포즈에 대한 실루엣 영상을 데이터베이스에 저장한다. 그리고 인식할 포즈의 실루엣과 데이터베이스의 모든 실루엣과 비교하여 포즈를 인식한다. J.S. Hu et al.[14]은 푸리에 서술자(Fourier descriptor)를 통해 2D 영상의 포즈의 외곽선으로부터 특징을 추출하고 이를 통해 포즈 인식을 하였다. 이와 같은 방법들은 시점 의존적(viewpoint-dependent)인 단점을 극복하기 위해 많은 수의 샘플 데이터 확보가 요구된다. 이를 극복하기 위해 많은 수의 학습데이터를 줄이는 알고리즘 역시 제안되었다. Wu와 Huang[15]은 지도학습(supervised learning)과 자율학습(unsupervised learning)방법을 조합한 Discriminant-EM(DEM)을 제안하였다. DEM은 적은 양의 학습데이터를 통해 많은 양의 unknown데이터를 분류하고 이때 사용된 분류 모델을 EM 알고리즘을 통해 반복적으로 개선해 나가는 방법이다. Imai et al.[16]은 다양한 종류의 appearance데이터를 인식하기 위해 지역적으로 압축된 특징(locally-compressed feature)을 통해 각 클래스의 전형적인 데이터로부터 가능한 모양의 appearance데이터를 만들고 이를 학습한다. 하지만 위의 방법들은 하나의 카메라를 사용하므로 포즈가 가려짐이나 카메라의 시점에 따라서 불안정한 인식률을 보인다.

위의 2D 방법과 단일 시점 기반의 3D 방법의 단점을 극복하고자, 다 시점 기반의 3D 포즈인식방법이 제안되었다. Ren et al.[17]은 3대의 카메라를 통해 실루엣 영상을 추출하고 추출된 영상을 데이터베이스의 영상들과 비교해 포즈를 추정한다. Delamarre와 Faugeras[18]는 다른 포즈임에도 불구하고 시점에 따라 비슷하게 보이는 포즈들에 대한 인식을 위해 여러 대의 카메라를 설치하여 포즈인식을 하였다. Cohen과 Li[19]는 4대의 카메라를 이용해 포즈의 visual hull을 생성하고 이를 통해 포즈인식을 하였다.

포즈의 연속된 동작, 제스처 인식에서는 포즈 인식에서 발생할 수 있는 시점 의존적 특성으로 인해 발생한 회전, 가려짐 등의 문제를 인식기의 성능에 의존하여 해결한다. 2D기반의 제스처 인식의 경우에서는, Mohan et al.[20]과 Moghaddam과 Yang[21]은 Support Vector Machine을 통한 방법을 제안하였다. Yang[22]은 신경망을 통한 해결방법, Wilson과 Bobick[23]과 Yeung과 Wang[24]은 퍼지 이론 기반의 해결방법을 제안하였다. Bicego와 Murino[25]는 가려짐, 영상분할 등의 문제를 Hidden Markov Model을 통해 해결하였다. 3D 기반의 제스처 인식의 경우에, Peursum et al.[26]은 4대의 카메라로부터 입력된 영상에서 4개의 Star Skeleton을 각각 생성하고 데이터베이스에 저장된 데이터와 가장 유사한 Star Skeleton을 선택하여 이를 기반으로 연속된 포즈의 제스처 인식을 수행한다. 이때 영상 분할, 가려짐 등에서 발생할 수 있는 문제를 Hidden Markov Model을 통해 해결한다.

본 논문에서는 다 시점 기반의 3D 방법과 특징 추출 방법의 하나인 Star Skeleton[6,7]을 결합한 포즈인식을 위한 3D Star Skeleton기반의 특징 추출을 제안한다. 제안된 방법은 다음의 특성을 지닌다. 첫째, 기존 단일 시점 기반 방법의 카메라 위치에 의존적인 단점을 극복하기 위해 다 시점 기반의 3D 방법을 적용한다. 이를 통해 회전된 동일 포즈에 대해 동일한 모양의 3D Star Skeleton이 생성된다. 둘째, 생성된 3D Star Skeleton에 주성분 분석(principal component analysis: PCA)을 적용하여 사람의 회전에 강건한 특징을 추출한다. 즉, 동일한 포즈에 대해 동일한 모양을 갖지만 사람의 회전에 따라 회전되어 있는 3D Star Skeleton에 주성분 분석을 적용하여, 회전에 강건한 특징을 추출한다.

논문의 구성은 다음과 같다. 2장에서는 기존 Star Skeleton, 3장에서는 제안된 3D Star Skeleton을 생성하고 특징을 추출하는 방법을 제안한다. 실험 및 결과는 4장에 서술되어 있다. 마지막으로 5장에서 결론을 서술한다.

2. Star Skeleton

Star Skeleton[6, 7]은 인간의 동작이나 포즈인식을 위한 skeleton생성 기술이다. Star Skeleton은 인간의 포즈를 묘사하기 위한 대표적인 특징 중 하나로써, 포즈의 중심점과 외곽선 위의 extremity들을 연결한 5개의 벡터로 구성된다. Star Skeleton은 많은 양의 픽셀 단위 연산을 필요로 하지 않으므로 실시간 생성이 가능하다.

Star Skeleton은 포즈의 중심점과 포즈의 외곽선에서 돌출된 점(salient point), extremity를 연결하여 생성된다. Extremity를 찾기 위해, 외곽선 추적을 통해 중심점과 외곽선 위의 점들과의 거리를 구한다. 그리고 중심점으로부터 거리가 먼 외곽 점들을 extremity로 결정한다. 마지막으로 extremity와 중심점을 연결하여 Star Skeleton을 생성한다. 이때 생성된 중심점과 extremity를 연결하는 벡터가 포즈인식을 위한 특징이 된다. 그림 1은 생성된 Star Skeleton을 나타낸다.

기존의 Star Skeleton은 간결하고 수행 시간 측면에서 장점을 가지고 있는 반면, 3차원의 인간 포즈를 2차원의 skeleton으로 표현하므로 다른 2D 기반의 포즈인식 방법과 같이 시점에 의존적인 특성을 가지고 있다. 예를 들어 그림 1(b)의 경우, “T pose”인지 “stand”인지 인식하기 힘들다.

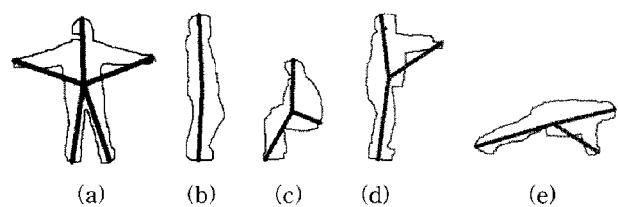


그림 1 기존 Star Skeleton의 결과([7])

3. 3D Star Skeleton 생성 및 특징 추출 방법

3.1 시스템 개요

3D Star Skeleton을 통한 특징 추출방법은 다 시점 기반의 3D 방법과 특징 추출 방법의 하나인 Star Skeleton[6,7]을 결합한 포즈인식을 위한 특징 추출방법이다. 제안된 방법은 기존 특징 추출 기반의 포즈인식방법과 같이 많은 시스템 자원을 요구하지 않고 간결한 방법을 통해 특징을 추출한다. 또한 기존 단일 시점 기반 방법의 카메라 위치에 의존적인 단점을 극복하기 위해 다 시점 기반의 3D 방법을 적용한다. 이를 통해 회전된 동일 포즈에 대해 유사한 모양의 회전된 3D Star Skeleton을 생성하고 주성분 분석(principal component analysis: PCA)을 적용하여 사람의 회전에 강건한 특징을 추출한다. 제안된 시스템은 projection map 생성, 후보군 검색, 군집화 및 3D Star Skeleton 생성 그리고

특징 추출의 4 부분으로 구성된다.

Projection map[27]은 depth map의 한 종류로 가상 카메라에서의 사람의 위치와 상대적인 거리 등의 정보를 담고 있다. Projection map은 visual hull[28]을 통한 3D 복원 과정 중에 획득할 수 있는 영상으로 이를 위해 캘리브레이션 된 8개의 카메라를 이용한다. Visual hull을 생성하기 위해 다 시점 카메라를 통해 입력 받은 영상 내 실루엣 영상의 외곽선을 캘리브레이션 데이터를 이용하여 3차원 공간으로 역 투영한다. 각 카메라로부터 역 투영된 영상들의 점들 중 공간상에서 교차하는 점을 계산하여 얻어지는 결과가 visual hull이다. 이 과

정에서 물체를 구성하는 모든 점을 하나의 2차원 평면 위로 투영한 것이 projection map이다. 제안된 방법은 8개의 projection map을 사용하며, 이를 통해 포즈의 3 차원 정보가 반영된 3D Star Skeleton을 생성한다.

후보군 검색은 8개의 projection map 각각에서 3D Star Skeleton의 최종 extremity가 될 수 있는 점을 추출하는 단계로 추출된 점을 후보군으로 정의한다. 후보군 검색에 대한 과정은 그림 2(b)에 묘사되어 있다.

군집화 및 3D Star Skeleton 생성과정에서는 각 projection map의 지역 좌표계에 존재하는 후보군들을 전역 좌표계로 변환한 후, 전역 좌표계 내에 산재되어 있는

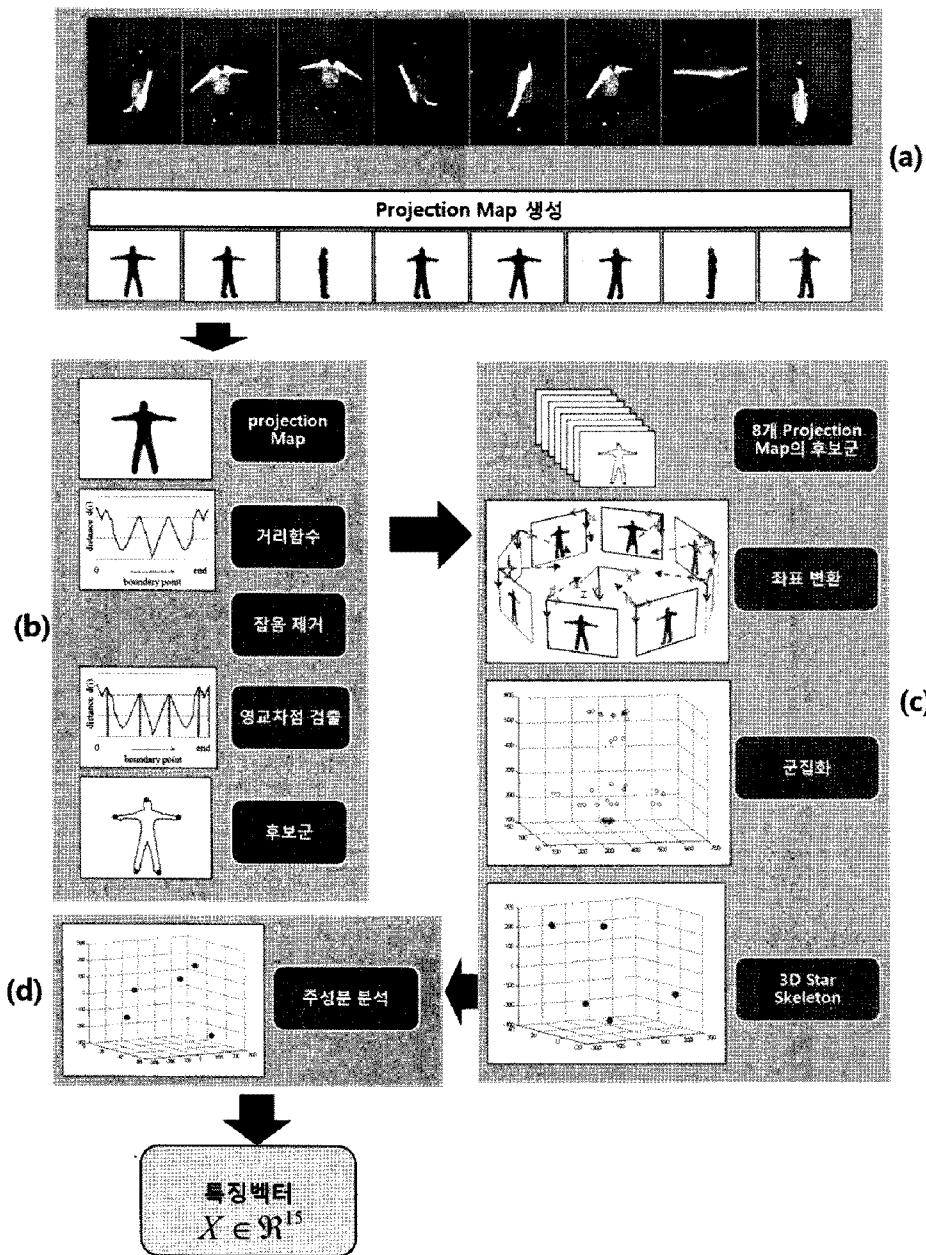


그림 2 제안된 시스템의 흐름도: (a) projection map 생성, (b) 후보군 검색, (c) 군집화 및 3D Star Skeleton 생성, (d) 특징 추출

후보군들을 K-평균 군집화 알고리즘으로 군집화한다. 이 과정을 통해, 모든 후보군들은 5개의 그룹으로 군집화가 되고, 각 군집의 평균은 제안된 3D Star Skeleton을 구성하는 extremity가 된다. 또한 3D Star Skeleton의 중심(centroid)을 좌표계의 중심으로 이동시켜 사람의 이동에 강건한 3D Star Skeleton을 생성한다. 그림 2(c)는 군집화 및 3D Star Skeleton 생성과정을 묘사한다.

특정 추출단계에서는 3D Star Skeleton으로부터 사람의 회전에 강건한 특징을 추출하는 단계이다. 이를 위해 주성분 분석을 통해 3D Star Skeleton의 좌표축을 변환하고 특징을 추출한다. 그림 2(d)는 특정 추출과정을 나타낸다.

3.2 Projection map 생성

Projection map[27]은 3차원 정보를 갖는 depth map의 한 종류로 사람의 위치와 카메라로부터의 거리 정보의 3차원 정보가 포함된 영상이다. Projection map은 visual hull[28]을 통한 3D 복원 과정 중에 획득할 수 있는 영상으로 이를 위해 캘리브레이션된 8개의 카메라를 이용한다. Visual hull을 생성하기 위해 다 시점카메라를 통해 입력 받은 영상 내 실루엣 영상의 외곽선을 추출하고 각 외곽선 상의 점을 캘리브레이션 데이터를 이용하여 3차원 공간으로 역 투영한다. 역 투영된 실루엣 영상은 공간상에서 원뿔의 형태를 띠게 되는데 공간 안에서 모든 실루엣 원뿔들이 교차하는 점을 계산된 결과가 visual hull이다. 이 과정에서 물체를 구성하는 모든 점을 하나의 2차원 평면위로 투영한 것이 projection map이다. Projection map은 입력 파라미터에 따라서 각기 다른 깊이 정보, 카메라로부터 사람의 가장 가까운 표면까지의 거리, 가장 먼 표면까지의 거리, 사람의 두 께를 포함하게 된다. 제안된 시스템은 사람의 위치 정보와 카메라로부터 사람의 가장 가까운 표면까지 거리를 포함하는 3차원 정보를 반영한 8개의 projection map을 입력 데이터로 사용한다.

Projection map은 3D visual hull을 생성하는 과정에

서 획득할 수 있다. 그림 3과 같이, 투영 면(projection plane)에 수직인 벡터 \vec{v} 를 통해 카메라와 가까운 면의 접촉점 P를 추출할 수 있다. Projection map의 생성은 다음과 같은 과정을 통해 가능하다. 먼저 3차원 공간의 사람을 촬영할 수 있도록 가상의 카메라를 설치하고 촬영한다. 이 과정에서 촬영된 사람은 카메라의 투영 면으로 투영되게 된다. 이 투영과정에서 각 픽셀마다의 방향과 접촉점이 정해지게 된다. 포즈의 가장 가까운 표면을 랜더링하는 과정에서, 접촉점 P가 쉽게 추출될 수 있으며, 버퍼에 저장되게 된다. 그림 3(c)는 우리가 사용하는 projection map의 한 예이다. 화소값이 카메라로부터의 거리를 의미하므로 projection map의 좌표계는 X, Y, Z의 세 축($640 \times 480 \times 256$)으로 구성된다.

3.3 후보군 검색

후보군 검색은 8개의 projection map 각각에서 3D Star Skeleton의 최종 extremity가 될 수 있는 점을 추출하는 단계로 추출된 점을 후보군이라 정의한다. 후보군은 각 projection map에 투영된 포즈의 외곽선 중 포즈의 중심과 거리가 먼 점에서 추출되는 데, 포즈 중심과의 거리 계산시에 projection map의 x, y축 상의 거리뿐만 아니라 projection map의 화소값인 카메라로부터 포즈까지의 거리 정보도 고려한다. 즉, 기존 2D Star Skeleton은 2차원의 외곽선으로부터 extremity를 검출하는데 반해 제안된 3D Star Skeleton은 입력데이터인 projection map이 포함하고 있는 깊이 정보를 고려하므로 3차원의 외곽선을 바탕으로 extremity를 계산한다. 따라서 포즈의 3차원 데이터를 반영한 3D Star Skeleton 생성이 가능해지며 이를 통해 2차원의 외곽선으로부터 생성되는 2D Star Skeleton의 시점 의존적인 단점을 극복한다.

그림 4는 후보군 검색의 과정을 나타낸다. 후보군의 검출을 위해 각 8개의 projection map에서 다음과 같은 수행과정을 거친다. 첫째, 외곽선 위에 존재하는 점들과 포즈의 중심점과의 거리를 외곽선 추적을 통해 얻는다.

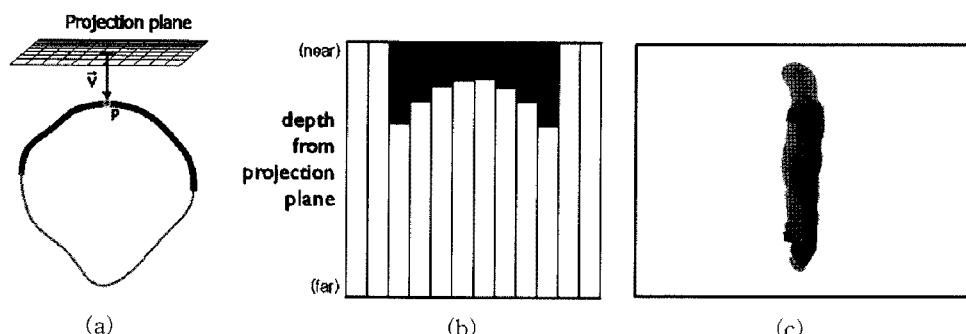


그림 3 Depth map을 이용한 projection map의 생성: (a) 투영면부터 사람의 가장 가까운 표면까지의 거리, (b) depth map은 사람의 가까운 표면의 깊이 정보를 저장, (c) 사용된 projection map ($640 \times 480 \times 256$)

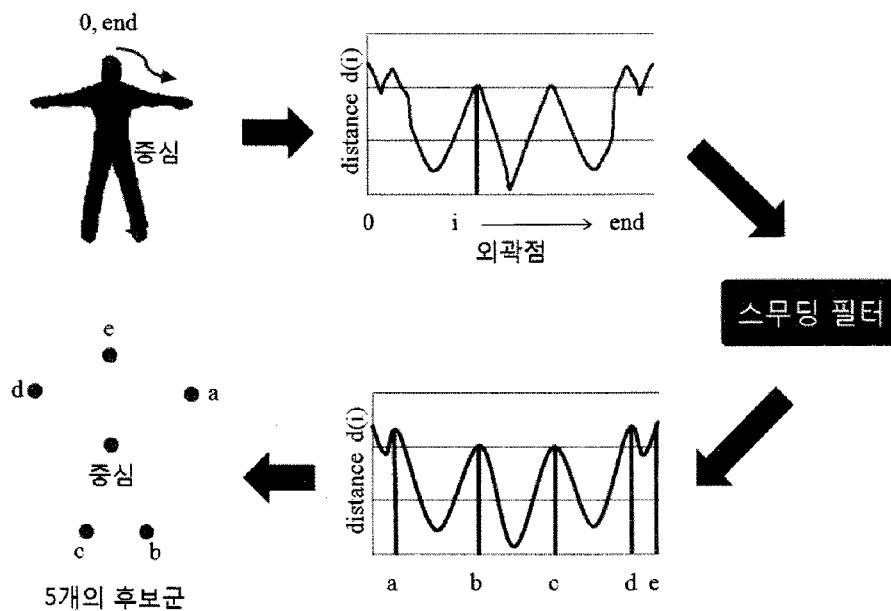


그림 4 외곽선의 extremity와 중심까지의 거리는 거리함수로 나타내진다. 잡음 제거 후, 후보군은 1차 미분을 통해 영교차점에서 추출된다.

이때 거리의 계산은 각 외곽점의 좌표와 외곽점의 화소값 즉, 카메라로부터의 거리 정보의 3차원 정보를 바탕으로 측정한다. 이 정보를 거리함수라 하며, 거리함수는 잡음을 포함하고 있을 수 있기 때문에 스무딩 필터를 적용하여 잡음을 제거한다. 마지막으로 거리함수를 1차 미분한 후 영교차점(zero-crossing point) 검색을 통해 지역적 최대값을 찾고 이를 후보군으로 추출한다. 후보군 추출 시 5개 미만의 extremity가 추출될 경우 추출되지 않은 extremity를 영벡터로 설정하여 5개의 extremity를 정의한다. 그림 5에서는 후보군 검색 과정을 서술하였다.

3.4 군집화 및 3D Star Skeleton 생성

제안된 3D Star Skeleton의 extremity를 결정하기 위해서, 후보군들을 전역 좌표계로 옮긴 후 군집화 한다. 좌표 변환을 위해 기준 전역 좌표계를 1번 projection map으로 강제 지정한 후, 각 projection map의 지역 좌표계에 존재하는 후보군들을 변환 행렬을 통해 전역 좌표계로 좌표 변환한다.

좌표 변환된 후보군들을 통해 3D Star Skeleton의 extremity를 결정하기 위해 군집화를 수행한다. 좌표 변환된 후보군들은 3D 전역 좌표계 내에 산재해 있다. 즉, 포즈의 같은 부분에서 추출된 후보군들이라 하더라도 신체의 두께와 projection map의 생성 위치에 따라서 전역 좌표계 내에서 같은 위치에 존재하는 것이 아니라 서로 인접해 있다. 따라서, 포즈의 동일한 부분에서 생성된 후보군들을 한 점으로 수렴시키기 위해 군집화가 수행된다. 군집화의 결과로 생성된 각각의 그룹의 평균은 3D Star Skeleton의 extremity가 된다.

군집화를 위해 K-평균 군집화 알고리즘이 사용된다. 이 과정을 통해 모든 후보군들은 5개($K=5$)의 군집으로 나눠진다. 3D Star Skeleton의 중심점은 후보군 추출을 위해 사용된 각 projection map의 포즈 중심점을 전역 좌표계로 변환한 후, 이들의 평균으로 정의한다. 마지막으로 3D Star Skeleton의 중심점이 전역 좌표계의 원점이 되도록 중심점과 각 평균의 좌표를 이동하여 3D Star Skeleton이 사람의 이동에 강건한 특징을 갖게 한다. 그림 6은 군집화를 통한 3D Star Skeleton 생성을 나타낸다.

3.5 특징 추출

기존의 포즈인식 연구들 중 시점 의존적인 방법들은 사람의 회전에 따라 같은 포즈임에도 불구하고 다른 포즈로 오인식하는 문제점을 가지고 있다. 이를 극복하기 위해 우리는 주성분 분석(principal component analysis: PCA)을 통해 3D Star Skeleton으로부터 회전에 강건한 특징을 추출한다.

주성분 분석은 일반적으로 두 가지 의미를 가지고 있다. 첫째, 데이터를 가장 잘 표현하는 축을 찾는 것이며, 둘째, 데이터를 표현하는 축으로 투영하여 데이터의 좌표계를 변환할 때 투영축의 개수를 조절하여 데이터의 차수를 조절할 수 있다. 후자의 경우는 데이터의 손실 압축을 위해 일반적으로 쓰이는 방법이다. 우리는 주성분 분석의 첫 번째 의미인 데이터인 생성된 3D Star Skeleton을 표현하는 축을 찾고 extremity를 투영하여 회전에 강건한 특징을 추출한다. 즉, x, y, z의 값들을 갖는 5개의 3차원 extremity를 가장 잘 표현하는 고유ベ

입력 데이터 : projection map 내의 포즈의 외곽선출력 데이터 : 후보군

1. projection map 내 포즈 외곽선의 중심점(x_c, y_c, z_c)을 구한다.

$$x_c = \frac{1}{N_b} \sum_{i=1}^{N_b} x_i$$

$$y_c = \frac{1}{N_b} \sum_{i=1}^{N_b} y_i$$

$$z_c = \frac{1}{N_b} \sum_{i=1}^{N_b} z_i$$

N_b : 외곽선 위의 점의 개수

(x_i, y_i, z_i) : 외곽선 위의 한 점

x_i, y_i : 영상 좌표계에서의 외곽점의 좌표

z_i : 확소값

2. 외곽선 위의 점(x_i, y_i, z_i)들과 중심값(x_c, y_c, z_c)과의 거리 d_i 를 구한다.

$$d_i = \text{dist}((x_i, y_i, z_i), (x_c, y_c, z_c)) = \sqrt{(x_i - x_c)^2 + (y_i - y_c)^2 + (z_i - z_c)^2}$$

이 연산은 1차원의 거리함수 $d(i) = d_i$ 로 선언된다.

3. 스무딩 필터를 통해 거리함수 $d(i)$ 의 잡음을 제거한다. $\hat{d}(i)$ 는 잡음 제거된 거리함수이다.

4. $\hat{d}(i)$ 의 지역적 최대값을 추출하여 최종 3D Star Skeleton의 extremity가 될 수 있는 후보군을 획득한다. 지역적 최대값은 $\hat{d}'(i)$ 의 영교차점을 통해 추출할 수 있다. 단, 지역적 최대값은 $\hat{d}(i)$ 이 임계값 T_{peak} 이상이고, 다른 지역적 최대값과의 projection map에서의 위치가 임계값 T_{adj} 이상인 외곽점에서 추출한다.

$$\delta(i) = \hat{d}'(i)$$

if $\delta(i) = 0, \hat{d}(i) > T_{peak}$ and $\sqrt{(x_i - x_m)^2 + (y_i - y_m)^2 + (z_i - z_m)^2} > T_{adj}$

then i 번째 외곽점은 지역적 최대값으로 결정

(x_m, y_m, z_m)은 m번째 외곽점에 있는 지역적 최대값

실현을 통해서 T_{peak} 은 100, T_{adj} 은 150으로 할당한다.

그림 5 후보군 검색 알고리즘

터 3개를 주성분 분석을 통해 구하고 구해진 고유벡터 축으로 5개의 extremity를 투영한다. 이를 통해 투영된 5개의 3차원 extremity를 구할 수 있고 이를 일렬로 나열하여 15차의 특징벡터를 추출한다. 그림 7은 주성분 분석을 통한 특징 추출 과정이다.

그림 8은 1번 projection map을 향한 T pose의 3D Star Skeleton과 회전된 T pose의 3D Star Skeleton, 그리고 두 3D Star Skeleton을 주성분 분석을 통해 축 변환한 후의 비교를 나타낸다. 그림 8(a)에서 알 수 있듯이 3D Star Skeleton은 사람의 회전을 반영하기 때문에 동일한 T pose에 대해서 회전에 따라 유사한 모양을 갖지만 회전되어 생성된다. 이는 3.4의 전역 좌표계로 좌표 변환 시 사람의 회전을 고려하지 않고 강제

로 1번 projection map의 좌표계를 기준 전역 좌표계로 하여 좌표 변환하였기 때문이다. 즉, 동일한 포즈에 대해서 모양은 유사하지만 사람의 회전에 따라 회전된 3D Star Skeleton이 생성된다. 따라서 생성된 3D Star Skeleton 자체를 특징으로 사용할 경우 사람의 회전에 대해 강건한 인식 결과를 얻을 수 없을 것이다.

반면에 그림 8(b)의 경우, 주성분 분석을 통해 T pose 3D Star Skeleton의 주축을 구하고 주축을 기준으로 좌표변환을 하면 사람의 회전 정도가 다른 두 3D Star Skeleton의 대응되는 extremity가 유사한 위치에 존재함을 알 수 있다. 이는 사람이 회전하더라도 포즈의 3D Star Skeleton을 나타내는 주축 역시 사람의 회전에 따라 회전하기 때문이며, 이 주축을 기준으로 3D

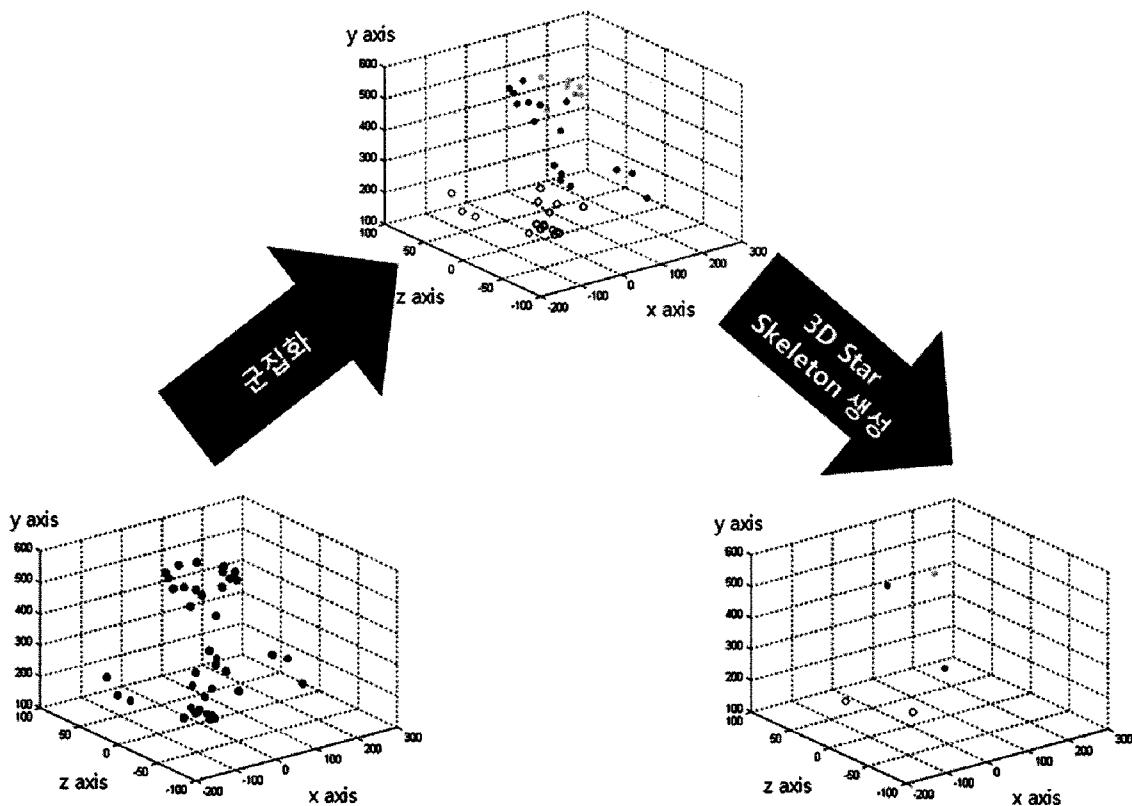


그림 6 군집화와 3D Star Skeleton 생성 과정

Star Skeleton을 좌표 변환하면 사람의 회전에 강건한 특징을 추출할 수 있다.

4. 실험 및 결과

이 장에서 포즈인식을 위한 3D Star Skeleton기반 특징추출을 평가하기 위한 실험을 하였다. 먼저 자유롭게 회전된 7개의 포즈에 대한 인식률을 제안된 특징추출방법과 다양한 인식기를 통해 측정하였다(4.2). 그리고 제안된 방법의 회전에 강건한 특성을 실험하기 위해 3D Star Skeleton의 raw data(SRD)와 3D Star Skeleton으로부터 주성분 분석을 통해 추출한 특징벡터(PBF)의 사람의 회전에 대한 인식률 비교를 하였다(4.3). 또한 PBF의 사람의 회전에 강건한 특성을 시각적으로 보여주기 위한 실험을 하였다(4.4). 그리고 기존의 2D/3D 기반의 포즈인식방법과 제안한 방법의 평균인식률에 대한 비교를 서술하였다(4.5). 실험 결과에 대한 서술을 하기 전에 4.1절에서 실험 환경과 실험에 사용된 데이터 베이스에 대해 설명한다.

4.1 실험 환경 및 데이터베이스의 구성

3D Star Skeleton을 생성하기 위한 입력데이터인 projection map을 생성하기 위해 8대의 카메라를 천정에 설치하고 원형의 촬영공간을 마련하였다. 또한 배경을 파란색으로 동일화하여 배경 제거(background subtraction)

처리에 대한 부담을 줄였다. 다음 그림 9는 촬영 공간을 나타낸다.

촬영된 결과 영상으로부터 projection map을 생성하기 위해 2.13GHz CPU와 2Gbyte 메모리, 그리고 nVidia GeForce 8800GTX 그래픽 카드로 구성된 시스템을 사용하였으며, Direct3D와 고수준 쉐이더 언어(High-Level Language:HLSL)을 이용하여 projection map을 생성하였다. 8대의 카메라는 하나의 좌표계로 정확하게 캘리브레이션 되어 있고 일정한 간격으로 배치되어 있다.

우리는 포즈인식을 위해 7개의 포즈를 정의하였다. 그림 10은 우리가 정의한 7개의 포즈를 보여준다. 데이터 베이스의 구축을 위해 20명의 사람(남자 12명, 여자 8명)을 6주에 걸쳐서 7개의 포즈에 대해 촬영하였다. 처음 4주는 10명을 대상으로 테스트데이터를 만들기 위한 촬영을 하였고, 마지막 두 주에 4주에 참가하지 않은 10명을 대상으로 학습데이터를 위한 촬영을 하였다. 테스트 데이터는 각 포즈당 200개, 총 1400번(200개×7포즈) 촬영을 하였다. 또한 한 사람당 한 포즈에 대해서 20번의 촬영을 하고, 총 140번(20번×7포즈)의 촬영을 하였다. 또한 회전에 대한 인식률 변화를 측정하기 위해서 4개의 회전 범위($315^{\circ} \sim 45^{\circ}$, $45^{\circ} \sim 135^{\circ}$, $135^{\circ} \sim 225^{\circ}$, $225^{\circ} \sim 315^{\circ}$)를 정하고 한 범위 내에서 무작위로 회전하면서 촬영하였다.

입력 데이터 : 3D Star Skeleton의 5개의 extremity.

$$S = \{s_i \mid s_i \in R^D, 1 \leq i \leq N, N: extremity의 수, D: extremity의 차원\}$$

출력 데이터 : 특징벡터 $X, X \in R^{15}$

1. extremity의 평균 벡터를 구한다.

$$\bar{s} = \frac{1}{N} \sum_{i=1}^N s_i$$

2. extremity에서 평균벡터를 뺀다.

$$s'_i = s_i - \bar{s}, \text{ for } i = 1 \text{ to } N$$

3. $s'_i, 1 \leq i \leq N$ 의 공분산 행렬(covariance matrix) Σ 을 구한다.

4. Σ 의 고유벡터와 고유값(eigenvalue)을 구한다. 그리고 고유값의 내림차순으로 정렬하고 고유벡터(eigenvector) u_i 를 고유값 기준으로 정렬한다.

5. 주축 변환 행렬 $U = [u_1, \dots, u_D]^T$ 를 만들고 s'_i 를 주축 변환 행렬을 통해 주축으로 투영한다.

$$S' = [s'_1, \dots, s'_N]$$

$$P = US' \quad P = [p_1, \dots, p_N], P \in R^{D \times N}, p_i \text{는 주축에 투영된 } s_i$$

6. 특징벡터 구성 시, N개의 extremity를 표현하는 최적의 축에 투영된 값의 순서로 나열하기 위해 투영된 벡터의 집합인 행렬 P 의 첫 번째 행을 기준으로 내림차순 정렬하여 행렬 P' 를 만들고 P' 의 첫 번째 행부터 마지막 행까지 순차적으로 나열하여 특징벡터 X 를 구성한다.

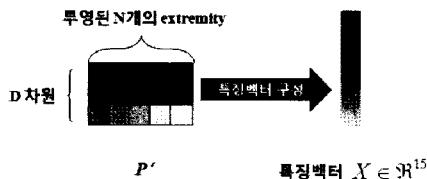
for $r = 1$ to D

for $c = 1$ to N

$$X_{r \times 5+c} = p'_{rc}$$

where X 는 특징벡터, $X \in R^{15}$

덧붙여 특징벡터의 구성은 다음의 그림과 같다.



즉, 고유벡터로 투영된 extremity의 집합 행렬 P' 를 최적의 고유벡터에 투영된 값의 순서로 특징벡터를 구성하는 것이다.

그림 7 특징 추출 알고리즘

학습데이터 수집을 위해서 전방 1번 카메라를 중심으로 대략 $-15^\circ \sim 15^\circ$ 의 회전 범위 내의 회전만 허용하고 촬영하였다. 이는 제안된 방법의 사람의 회전에 강건한 특성을 증명하기 위한 것으로, 만일 학습데이터가 $-180^\circ \sim 180^\circ$ 의 자유롭게 회전된 포즈에 대한 것일 경우, 인식률이 높다 하더라도 높은 인식률이 인식기의 성능 때 문인지 아니면 특징의 회전에 강건한 특성 때문인지 증명하기 힘들기 때문이다. 다시 말하면, 학습데이터는 전

방을 향한 데이터들로만 수집하고 이를 자유롭게 회전된 테스트데이터를 인식해 봄으로써, 제안된 특징이 회전에 강건하다는 것을 증명하기 위함이다. 학습데이터를 위해서 한 사람당 한 포즈에 대해 5번의 촬영을 하였고, 7개의 포즈에 대해서 총 35번의 촬영을 하였다. 따라서 한 포즈당 10명이 50번의 촬영을 하고, 총 350개의 데이터베이스를 구축하였다. 표 1은 데이터베이스에 대한 설명이다.

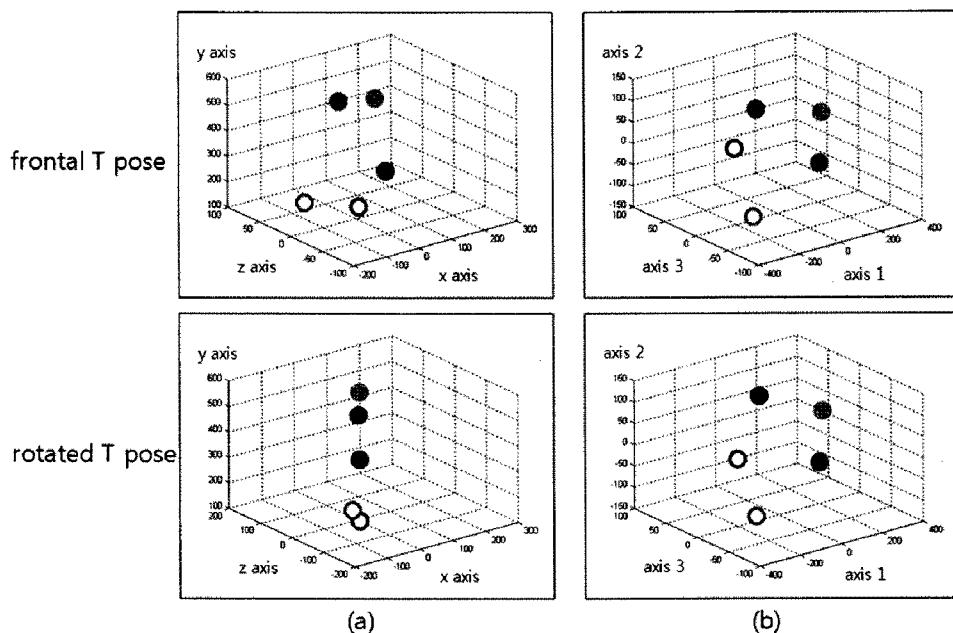


그림 8 주성분 분석을 통한 축 변환 결과: (a) 3D Star Skeleton, (b) 주성분 분석 적용 후의 3D Star Skeleton

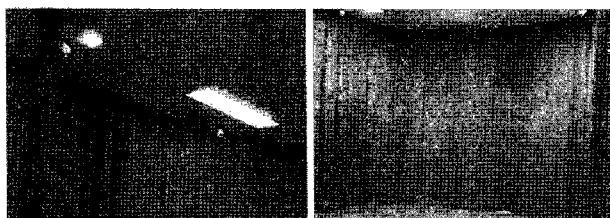


그림 9 실험을 위한 환경

표 1 데이터베이스의 설명

카테고리	DB명	회전 범위	개수
Test data	F	315° ~ 45°	350개
	R	45° ~ 135°	350개
	B	135° ~ 225°	350개
	L	225° ~ 315°	350개
Training data	Train	약 335° ~ 15°	350개

4.2 다양한 인식기를 통한 포즈인식률

본 단락에서는 우리는 먼저 다양한 인식기를 통한 포즈인식률 결과를 서술한다. 포즈인식을 위한 학습데이터로 위의 표 1에서 명명한 DB Train을 사용하였고, 학

습된 인식기를 각 DB F, R, B, L을 통해 테스트 하였다. 본 인식률 측정을 위해 우리는 Decision Tree(DT), Bayes Network(BN) 그리고 Multi-layer Perceptron (MLP)의 세 종류의 인식기를 이용하였다. 이는 제안한 방법을 통한 인식률이 인식기의 성능으로부터 독립적인 결과임을 보여주기 위함이다. 또한 3D Star Skeleton으로부터 주성분 분석을 이용해 추출한 특징(PBF)을 통한 포즈인식률과 3D Star Skeleton의 raw data(SRD) 즉, 3D Star Skeleton의 5개 extremity의 3차원 공간상의 좌표를 이용한 포즈인식률을 실험하였다. 편의를 위하여 표 2에 전체적인 인식률을 정리하였다.

각 인식기 별 평균 인식률은 PBF가 SRD에 비해 높은 인식률을 보였다. SRD를 통한 인식률은 각 인식기 별 평균 47.14%(DT), 63.81%(BN) 그리고 65.69%(MLP)의 인식률을 나타내었다. PBF를 통한 인식률의 경우 평균 인식률은 85.04%(DT), 89.23%(BN) 그리고 86.74%(MLP)를 보였다. 이는 SRD를 이용한 포즈인식보다 각각 37.9%(DT), 25.42%(BN), 21.05%(MLP)의 차이를 보이며 좋은 성능을 보였다.

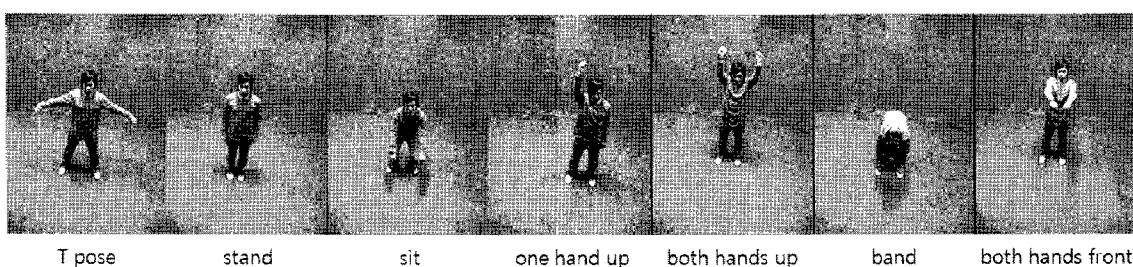


그림 10 사용된 7개의 포즈

표 2 다양한 인식기를 통한 포즈인식률(%)

특징벡터	인식기\포즈	T pose	stand	sit	OHU	BHU	band	BHF	Avg.	Var.	Max-Min
SRD	DT	29.70	52.47	53.15	52.37	50.70	46.70	44.88	47.14	69.0	23.45
	BN	44.00	46.62	78.75	52.73	78.88	93.17	52.55	63.81	374.97	49.17
	MLP	42.80	76.42	75.82	57.67	78.65	71.98	56.47	65.69	183.09	35.85
PBF	DT	83.65	91.62	87.28	79.83	85.05	90.00	77.86	85.04	25.59	13.76
	BN	90.33	92.50	90.85	84.20	89.95	94.13	82.68	89.23	17.88	11.45
	MLP	88.73	92.50	83.35	83.93	88.15	91.28	79.25	86.74	22.54	13.25

SRD: 3D Star Skeleton Raw Data; PBF: PCA Based Feature; OHU: one hand up; BHU: both hands up; BHF: both hands front

또한 평균 인식률뿐만 아니라 각 인식기의 포즈 별 인식률에서 PBF가 SRD보다 좋은 성능을 보였다. 인식률의 분산과 최고-최소 인식률의 편차는 각각 그 값이 적을수록 안정적인 포즈인식을 보임을 나타내는데, 세 인식기 모두에서 PBF를 사용한 포즈인식이 SRD기반의 인식보다 분산과 최고-최소 인식률 편차에서 적은 값을 보여주었다.

위의 실험결과에서 알 수 있듯이 PBF의 포즈인식률이 SRD를 통한 포즈인식률에 비해 3개의 인식기 모두에서 높은 인식률을 보였다. 이는 PBF의 사람의 회전에 강건한 특성에서 비롯된 결과이며, 이를 증명하기 위해 4.3절에서 PBF의 이러한 특성에 대해 실험한다.

4.3 회전에 강건한 특징

우리는 3D Star Skeleton에서 추출한 PBF(PCA Based Feature)가 사람의 회전에 강건한 특성을 갖고 있음을 본 실험에서 증명한다. PBF가 사람의 회전에 강건한 특징임을 증명하기 위해, SRD(3D Star Skeleton Raw Data)와 PBF와의 인식률을 비교 하였으며 학습데이터로는 DB Train($335^{\circ} \sim 15^{\circ}$)을 사용하고, 실험데이터로는 DB F, R, L, B를 사용하였다. PBF가 회전에 강건한 성질을 가지고 있다면 정면을 향하고 있는 포즈의 학습데이터를 통해 학습된 인식기의 자유롭게 회전된

포즈로 구성된 테스트데이터에 대한 인식률은 안정적일 것이다.

그림 11은 SRD와 PBF의 각 인식기별 회전에 따른 인식결과를 보여준다. 그림 11에서, SRD(그림 11(b))는 사람의 회전에 따라 편차가 큰 인식률을 보이는 반면, PBF(그림 11(a))를 통한 인식은 사람의 회전에 관계없이 다양한 인식기에서 SRD에 비해 안정적인 인식률을 보임을 알 수 있다. PBF의 인식률의 경우, 각 DB F, R, L, B에서 세 인식기 모두에서 81.13%(DT의 DB B에서의 인식률) 이상의 인식률을 보였다. 이는 PBF가 사람의 회전에 강건한 성질을 갖기 때문에 나타난 결과임을 알 수 있다.

반면에 SRD의 경우에는 사용된 테스트데이터 DB에 따라서 편차가 심한 인식률을 보였다(DT: 3.8%~88.66%, BN: 34.01%~95.24%, MLP: 41.73%~93.61%). 주목해야 할 점은 정면을 향한 테스트데이터 DB F의 인식률이 3개 인식기 모두 가장 높다는 점이다. 이는 정면 포즈로 구성된 학습데이터와 DB F의 포즈의 회전의 정도가 크지 않기 때문이다. 하지만 학습데이터와 다른 회전된 포즈로부터 생성된 DB R, L, B의 인식률은 상대적으로 낮았다. 특히 DB B의 경우 DT와 BN을 사용한 인식에서 가장 낮은 인식결과를 보여주었는데 이는 정

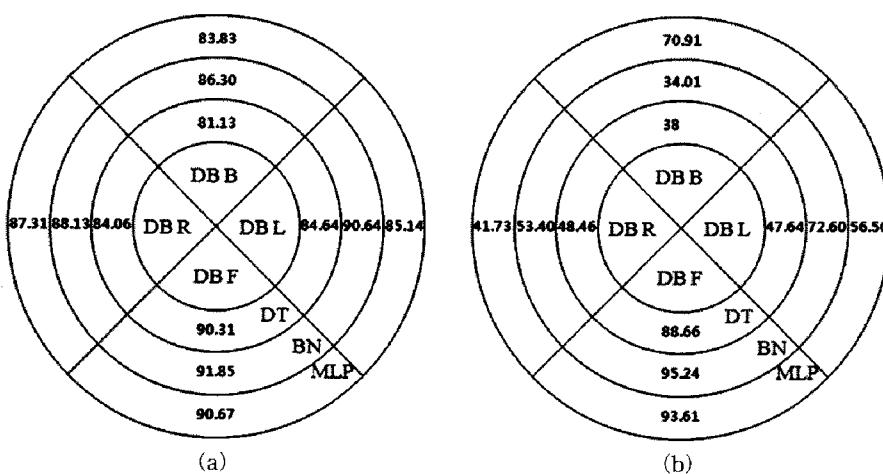


그림 11 회전에 따른 인식기 별 주성분 분석 기반의 특징과 3D Star Skeleton raw data의 인식률 비교:
(a) 주성분 분석 기반 특징의 인식률, (b) 3D Star Skeleton raw data의 인식률

면 포즈로 구성된 데이터로 인식기를 학습한 데 반해, DB B는 학습데이터와 반대방향의 포즈로 구성되었기 때문이다. 즉, SRD의 경우 회전을 고려하지 않은 특정 벡터이기 때문에 DB B에 대해 좋지 않은 성능을 보였다.

위의 실험결과를 통해서, 제안된 방법의 특징을 통한 포즈인식이 사람의 회전에 영향을 적게 받는다는 특성을 나타내었다. 이는 어느 특정 인식기에서만 발견할 수 있는 특성이 아니라 사용한 인식기 모두에서 나타나는 특징이었다. 따라서 PBF를 통한 포즈인식이 사람의 회전에 따라 안정적인 인식률을 보임을 통해서, PBF가 사람의 회전에 강건한 특징임을 결론 내릴 수 있다.

4.4 Visualization of PBF and SRD samples

이번 실험에서 우리는 위의 인식률을 실험을 통해 나타난 PBF의 특성인 사람의 회전에 강건한 성질을 더 직관적으로 보기 위해 특징공간상에서의 PBF와 SRD의 특징 분포를 시각화 해보았다. 이와 같이 임의의 특징공간상에 특징의 분포를 시각화 하여 실험하는 방법은 얼굴인식의 성능을 높이기 위해 Singular Value Decomposition(SVD)기반 잡음제거방법을 제안한 Lui et al. [29]가 활용한 방법이다. 실험을 위해 T pose와 stand에 대한 SRD와 PBF의 샘플을 테스트데이터 DB F, R, B, L에서 각각 1개 그리고 학습데이터 DB Train에서 2개 무작위로 선택하고, 임의의 특징공간 $X_6 - X_7$ 에서 추출된 샘플들의 분포를 그림 12에 나타내 보았다.

그림 12에서 (a)는 특징공간에서 SRD의 분포, (b)는 특징공간에서 PBF의 분포를 나타낸다. 우리는 그림 12에 분포되어 있는 특징벡터의 intra-class와 inter-class를 통해 PBF가 SRD에 비해 인식률이 높은 이유를 알 수 있었다. 먼저 SRD의 경우 같은 클래스의 학습데이터끼리 뭉쳐있고 다른 클래스와는 떨어져있음을 알 수 있다. 하지만 테스트데이터의 경우, 같은 클래스의 데이터끼리 모여있지 않고 넓은 범위에 존재하며 다른 클래스와 섞여있음을 알 수 있다. T pose의 경우 X_6 에서의

범위는 -161~218이고, stand는 -48~40의 범위를 갖는다. X_7 축을 기준으로 보면 T pose클래스는 -182~172의 범위를 갖고, stand는 -165~71의 범위를 갖는다. 즉, 두 포즈의 SRD의 분포는 양 축 모두에서 중복된 범위에서 존재한다.

PBF의 경우에 학습데이터는 SRD와 마찬가지로 각각 동일 클래스끼리 모여있고 다른 클래스와는 떨어져있다. 반면에 테스트데이터의 분포는 SRD와 다르게 회전에 관계없이 같은 클래스끼리 좁은 범위에 모여 있고 다른 클래스끼리는 특징공간상에서 분리되어 있다. 테스트데이터의 경우 X_6 에서 T pose와 stand가 각각 -1~60, 10~41의 중복된 범위에 존재하나 그 범위가 SRD보다 상대적으로 좁은 것을 알 수 있다. 또한 X_7 의 경우에 각 포즈마다 -185~142, 30~80의 좁은 범위에 존재하면서 클래스끼리 분리되어 분포되었음을 알 수 있다.

또한 그림 12를 통해서 알 수 있는 사실은, SRD는 테스트데이터와 다른 클래스의 학습데이터와의 거리가 같은 클래스의 학습데이터보다 가까운 경우가 발생해 오인식 될 확률이 높은 반면에 PBF는 같은 클래스의 학습데이터와 테스트데이터끼리 모여있고 다른 클래스의 학습데이터와 테스트데이터는 상대적으로 멀리 떨어져 있음을 알 수 있다. SRD의 경우, Euclidean distance로 특징공간에서의 거리를 측정했을 때 T pose의 DB B의 가장 가까운 학습데이터는 62.24가 떨어진 stand의 학습데이터이다. 또한 T pose의 DB R의 경우에도 가장 가까운 T pose의 학습데이터와의 거리가 341.67인데 반해, stand의 학습데이터와의 거리는 108.37로 T pose의 학습데이터보다 더 가깝다. Stand의 DB B와 R의 데이터는 각각 21.63, 73.76 떨어진 T pose의 학습데이터와 가장 가깝다. 이는 T pose나 stand의 테스트데이터가 각각 stand, T pose로 오인식 될 확률이 높음을 의미한다.

반면에 PBF의 경우에 SRD와는 달리 같은 클래스의

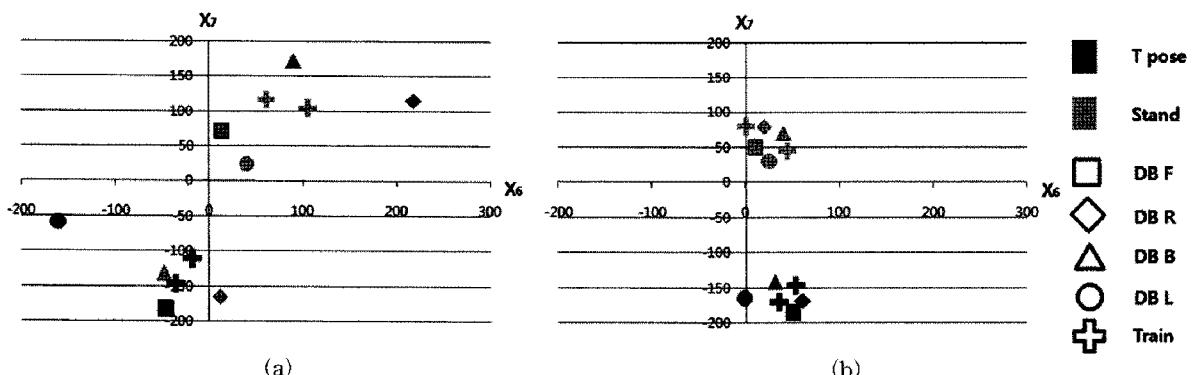


그림 12 T pose와 stand의 특징공간에서의 분포: (a) SRD의 T pose와 stand의 분포, (b) PBF의 T pose와 stand의 분포

학습데이터와 테스트데이터끼리 모여있고 다른 클래스의 학습데이터와 테스트데이터는 상대적으로 멀리 떨어져있다. 이는 SRD에 비해 오인식이 발생할 가능성이 적음을 의미한다. 예를 들어 T pose의 테스트데이터 DB F, R, B, L의 T pose 학습데이터와의 거리는 각각 22.36, 19, 29.83, 32.57이고 stand 학습데이터와의 거리는 205.19, 232.47, 219.91, 182.39로 T pose의 학습데이터와 상대적으로 가까움을 알 수 있다. 또한 stand의 테스트데이터 DB F, R, B, L의 stand 학습데이터와의 거리는 각각 31.32, 46.09, 26.93, 29.19인데 반해, T pose 학습데이터와의 거리는 185.07, 209.44, 228.04, 282.29로 상대적으로 멀리 떨어져있다. 즉, T pose와 stand의 테스트데이터가 각각 동일 클래스의 학습데이터와 가까움을 나타낸다.

위의 실험을 통해 우리는 PBF의 5가지 성질을 알 수 있다. (1) 같은 클래스에 속한 학습데이터끼리 compactness가 높다. (2) 다른 클래스의 속한 학습데이터끼리는 상대적으로 떨어져있다. (3) 테스트데이터의 경우 포즈가 회전되더라도 같은 클래스의 데이터끼리 compactness가 높다. (4) 테스트데이터에서 포즈가 회전되더라도 다른 클래스의 데이터끼리 상대적으로 떨어져있다. (5) 포즈의 회전에 관계없이 같은 클래스의 학습데이터와 테스트데이터끼리 compactness가 높다. 이와 같은 사실을 통해 우리는 PBF가 사람의 회전에 영향을 적게 받는 강건한 특성을 가지면서 안정적인 포즈인식을 가능하게 함을 확인할 수 있다.

4.5 기존 포즈 인식 방법과의 인식률 비교

제안된 포즈 인식 시스템의 성능을 증명하기 위하여 기존의 2D 기반 포즈 인식 시스템과 3D 기반 포즈인식 시스템의 인식률을 비교 분석하였다. 시스템간의 인식대상이 되는 포즈가 각각 다르기 때문에 평균인식률의 비교를 통해 성능을 비교하였다.

표 3은 기존 포즈인식 방법과 제안된 PBF와 BN을 통한 포즈인식 시스템의 평균인식률 비교를 나타낸다. 표 3에서 알 수 있듯이, 제안된 방법의 평균인식률이 기

존 2D/3D 기반의 포즈 인식 방법들과의 평균인식률 차이가 크지 않거나 우월하였다.

5. 결 론

본 논문에서, 우리는 포즈인식을 위한 3D Star Skeleton을 통한 특징 추출방법을 제안하였다. 제안된 방법은 포즈의 3차원 정보를 반영하는 다 시점 기반의 특징 추출 방법으로 사람의 회전 등에 대한 안정적인 포즈인식 성능을 보여준다. 3D Star Skeleton은 depth map의 한 종류인 8개의 projection map을 입력정보로 사용한다. Projection map이 가지고 있는 포즈의 3차원 정보를 통해 8개 Projection map 각각에서 5개의 extremity 후보군을 추출하고, 각 후보군을 전역 좌표계로 좌표변환 하였다. 이를 K-평균 군집화 알고리즘을 통해 5개의 그룹으로 군집화 하고 각 군집의 평균을 3D Star Skeleton의 extremity로 정하였다. 회전에 강건한 특징 추출을 위해 3D Star Skeleton을 주성분 분석을 통해 구한 좌표축으로 투영하고, 투영된 3D Star Skeleton으로부터 포즈의 특징을 추출하였다. 실험결과에서는 Decision Tree, Beyes Network, Multilayer Perceptron의 3가지 인식기를 사용하여 자유롭게 회전된 7개의 포즈를 인식해보았다. 인식기들을 통해 각 포즈의 인식률과 회전에 따른 인식률을 측정하여 우리가 제안한 방법을 통한 포즈인식이 사람의 회전에 대해 안정된 인식률을 보임을 증명하였다. 또한 추출한 특징을 특징공간상에 시각화하여 추출한 특징이 회전에 강건하고 효율적인 포즈인식이 가능하게 함을 분석하였다.

우리는 앞으로 제안된 시스템에 대한 4가지 연구를 진행할 것이다. 먼저, 우리는 3D Star Skeleton을 구성하는 말단 점의 적절한 개수를 정의할 것이다. 비록 extremity가 증가함에 따라 생성된 skeleton의 정확성이 명백해지나, 수행 시간 역시 비례적으로 증가한다. 따라서 수행 시간과 skeleton의 정확성을 고려하여 합리적인 말단 점의 개수가 명확히 정해져야 한다. 둘째로, 기존의 포즈인식 기술과의 비교를 좀더 상세하게 할 것이다. 이

표 3 기존 포즈인식방법과 제안된 방법의 평균인식률 비교(%)

포즈 인식 방법	특징벡터	인식기	평균인식률
2D 기반 포즈 인식 (Girondel et al.[4])	VBB and PAB based feature	restricted plausibility-based classifier	88.1
2D 기반 포즈 인식 (Panini and Cucchiara [8])	H.&V. projections	최소 거리 분류기	77.53
3D 기반 포즈 인식 (Boulay et al.[13])	H.&V. projections	최소 거리 분류기	82.00
3D 기반 포즈 인식 (Cohen and Li [19])	shape descriptor	HMM	87.82
3D 기반 포즈 인식	PBF	Beyes Network	89.23

를 통해 우리의 시스템의 효율성을 정확하게 평가할 수 있게 될 것이다. 셋째, 인식을 위한 더 많은 포즈를 정의 할 것이다. 현재 시스템의 7개의 포즈인식을 확장하여 다양한 포즈인식이 가능한 시스템으로 개선할 것이다. 마지막으로, 3D Star Skeleton을 통해 인간의 동작인식에 대한 연구를 진행할 것이다. 동작 인식을 위해 Hidden Markov Model(HMM)을 적용하도록 하겠다.

참 고 문 헌

- [1] I. Haritaoglu, D. Harwood and L.S. Davis, "Ghost: A human body part labeling system using silhouettes," *Proc. of the 14th International Conference on Pattern Recognition*, pp.77-82, 1998.
- [2] C. Wren, A. Azarbayejani, T. Darrell and A. Pentland, "Pfinder: Real time tracking of the human body," *IEEE Transactions on Pattern Analysis Machine Intelligence*, pp.780-785, 1997.
- [3] S. Iwasawa, K. Ebihara, J. Ohya, S. Morishima, "Real-Time Estimation of Human Body Posture from Monocular Thermal Images," *Proc. of 1997 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp.15-20, 1997.
- [4] V. Girondel, A. Caplier, L. Bonnaud and M. Rombaut, "Belief theory-based classifiers comparison for static human body postures recognition in video," *International Journal of Signal Processing*, vol.2, pp.29-33, 2005.
- [5] A. Baumberg and D. Hogg, "An adaptive eigen-shape model," *Proc. of British Machine Vision Conference*, vol.1, pp.87-96, 1995.
- [6] H. Fujiyoshi, A.J. Lipton and T. Kanade, "Real-time human motion analysis by image skeletonization," *IEICE Transactions on Information Systems*, pp.113-120, 2004.
- [7] H.S. Chen, H.T. Chen, Y.W. Chen and S.Y. Lee, "Human Action Recognition Using Star Skeleton," *Proc. of the 4th ACM international workshop on Video surveillance and sensor networks*, pp. 171-178, 2006.
- [8] L. Panini and R. Cucchiara, "A machine learning approach for human posture detection in domestic applications," *Proc. of the 12th International Conference on Image Analysis and Processing*, pp. 103-108, 2003.
- [9] Y. Kameda, M. Minoh and K. Ikeda, "Three dimensional pose estimation of an articulated object from its silhouette image," *Proc. of the Asian Conference on Computer Vision*, pp.612-615, 1993.
- [10] G. Mori and J. Malik, "Estimating Human Body Configurations Using Shape Context Matching," *Proc. of the 7th European Conference on Computer Vision Copenhagen, LNCS*, vol.2352, pp. 150-180, 2002.
- [11] G. Shakhnarovich, P. Viola, T. Darrell, "Fast Pose Estimation with Parameter-Sensitive Hashing," *Proc. of the 9th IEEE International Conference on Computer Vision*, vol.2, pp.750-757, 2003.
- [12] A.R. Micilotta, E.J. Ong and R. Bowden, "Real-time upper body detection and 3D pose estimation in monoscopic images," *Proc. of the 9th European Conference on Computer Vision, LNCS*, vol.3953, pp.139-150, 2006.
- [13] B. Boulay, F. Bremond and M. Thonnat, "Applying 3d human model in a posture recognition system," *Pattern Recognition Letter*, vol.27, pp. 1788-1796, 2006.
- [14] J.S. Hu, T.M. Su, P.C. L., "3-D Human Posture Recognition System Using 2-D Shape Features," *IEEE International Conference on Robotics and Automation*, pp.3933-3938, 2007.
- [15] Y. Wu and T.S. Huang, "View-independent Recognition of Hand Postures," *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, vol.2, pp.88-94, 2000.
- [16] A. Imai, N. Shimada and Y. Shirai, "3-D Hand Posture Recognition by Training Contour Variation," *Proc. of the 6th IEEE International Conference on Automatic Face and Gesture Recognition*, pp.895-900, 2004.
- [17] L. Ren, G. Shakhnarovich, J. Hodgins, H. Pfister and P. Viola, "Learning Silhouette Features for Control of Human Motion," *Proc. of International Conference on Computer Graphics and Interactive Techniques*, vol.24, no.4, p.129, 2005.
- [18] Q. Delamarre and O. Faugeras, "3D articulated models and Multi-View tracking with silhouettes," *Proc. of the 7th IEEE International Conference on Computer Vision*, pp.716-721, 1999.
- [19] I. Cohen and H. Li, "Posture and Gesture Recognition using 3D Body Shapes Decomposition," *Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, p. 69, 2005.
- [20] A. Mohan, C. Papageorgiou, and T. Poggio, "Example-Based Object Detection in Images by Components," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.23, no.4, pp. 349-361, 2001.
- [21] B. Moghaddam and M.H. Yang, "Learning Gender with Support Faces," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.24, no.5, pp.707-711, 2002.
- [22] M.H. Yang, "Extraction of 2D Motion Trajectories and Its Application to Hand Gesture Recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.24, no.8, pp.1061-1074, 2002.
- [23] A.D. Wilson and A.F. Bobick, "Configuration States for the Representation and Recognition of

- Gesture," *Proceeding on International Workshop Automatic Face and Gesture Recognition*, pp. 129-134, 1995.
- [24] D.S. Yeung and X.Z. Wang, "Improving Performance of Similarity-Based Clustering by Feature Weight Learning," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.24, no.4, pp.556-561, 2002.
- [25] M. Bicego and V. Murino, "Investigating Hidden Markov Models' Capabilities in 2D Shape Classification," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.26, no.2, pp.281-286, 2004.
- [26] P. Peursum, H.H. Bui, S. Venkatesh and G. West, "Robust recognition and segmentation of human actions using HMMs with missing observations," *EURASIP Journal on Applied Signal Processing, Special Issue on Advances in Intelligent Vision Systems*, vol.13, pp.2110-2126, 2005.
- [27] K. Hong, C. Lee, K. Jung and K. Oh, "Real-time 3D Feature Extraction without Explicit 3D Object Representation," *Proc. of World Academy of Science, Engineering and Technology*, vol.34, pp. 283-288, 2008.
- [28] A. Laurentini, "The Visual Hull Concept for Silhouette-based Image Understanding," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.16, pp.150-162, 1994.
- [29] J.Liu, S.Chen and X.Tan, "Fractional Order Singular Value Decomposition Representation for Face Recognition," *Pattern Recognition*, vol.41, pp.378-395, 2008.

정기철

정보과학회논문지 : 소프트웨어 및 응용
제 36 권 제 4 호 참조



전 성 국

2009년 2월 숭실대학교 미디어학과 공학 학사. 2009년 3월~현재 숭실대학교 미디어학과 석사과정. 관심분야는 HCI, 컴퓨터 비전, 패턴인식, 제스처인식



홍 광 전

2004년 2월 숭실대학교 컴퓨터학과 공학 학사. 2006년 2월 숭실대학교 미디어학과 공학석사. 2009년 2월 숭실대학교 미디어학과 공학박사. 관심분야는 HCI, 증강현실, 카메라 기반 3차원 물체 모델링