휴먼 행동 분석을 위한 3차원 제스처 데이터베이스의 설계 및 구축

노명철, 황본우, 김성민, 신호근, 박아연, 이성환
고려대학교 정보통신대학 인공시각연구센터/컴퓨터학과
{mcroh, bwhwang, smkim, hkshin, aypark, swlee}@image.korea.ac.kr

Design and Construction of 3D Gesture Database for Analyzing Human Behaviors

M.-C. Roh, B.-W. Hwang, S. Kim, H.-K. Shin, A-Y. Park and S.-W. Lee
Dept. of Computer Science and Engineering, Korea University

요 약

인간 행동 분석은 컴퓨터 비전 및 패턴인식 분야에서 활발하게 연구가 이루어지는 분야이다. 이러한 행동 분석하고 평가하기 위해서는 다양한 환경과 종류의 제스처를 포함하고 있는 데이터베이스의 구축이 필수적이다. 본 논문에서는 총 40명의 사람에 대하여 일상생활에서 일어날 수 있는 14개의 정상 제스처, 위급한 상황에서 발생할 수 있는 10개의 비정상 제스처 그리고 30개의 명령형 제스처를 수집한 KU(Korea University) 제스처 데이터베이스를 소개한다. 각각의 제스처는 스테레오 카메라를 통해 얻어진 2차원 제스처 동영상, 3차원 동작 카메라를 통해 얻어진 3차원 모델의 좌표 정보 그리고 2차원 실루엣 동영상을 포함하고 있다.

1. 서 론

휴먼의 행동 분석은 행동 인식, 비디오 보안, 휴먼-기계 상호 작용 등의 많은 분야에서 활발하게 연구되고 있는 분야이다. 효율적이고 신뢰성 있는 연구를 위해서는 제스처의 특징 분석, 제안된 알고리즘 평가 및 검증이 필요한데 이를 위해서는 잘 구축된 데이터베이스의 사용이 필수적이다.

제스처 연구는 목적에 따라서 몇 가지로 구분될 수 있다. 손동작, 수화 분석을 위한 손 제스처 연구[1], 휴먼 몸체의 추적 및 포즈 재구성을 위한 상체 제스처 연구[2]가 있고, 걸음걸이 인식을 위한 전신 제스처[3] 등이 있다. 이러한 기존의 연구를 위한 제스처 데이터베이스는 제한된 범위의 2차원 영상 또는 3차원 영상만을 제공하며, 동시에 2차원과 3차원 영상을 촬영하여 구축된 데이터베이스는 현재까지 존재하지 않는다. 그러나 다양한 환경에서 휴먼 각 요소를 추출, 추적하고 제스처 인식을 위한 연구를 위하여서는 다양한 데이터를 동시에 제공할 수 있는 데이터베이스가 필수적이고, 이를 위하여서 KU 제스처 데이터베이스를 구축하였다.

2. KU 제스처 데이터베이스 개요

KU 제스처 데이터베이스는 전신 제스처와 명령형 제스처의 두개의 데이터베이스로 구성되어있고, 2차에 걸쳐서 데이터가 수집되었다. 첫 번째 기간에는 일상생활에서 일어날 수 있는 14개의 정상 제스처를 정의하여 20명의 60~80세의 노인들을 대상으로 전신 제스처 데이터를 수집하였고, 두 번째 기간에는 20명의 일반인을 대상으로, 위급한 상황에서 발생할 수 있는 10개의 비정상 제스처를 정의하여 전신 제스처 데이터를 수집하고, 30개의 명령형 제스처를 정의하여 전신과 상반신 제스처 데이터를 수집하였다. 표 1은 수집된 인원의 연령대별 분포를 보여준다.

데이터베이스는 3차원 동작 데이터(1), 2차원 스테레오 비디오 데이터(2), 2차원 실루엣 데이터(3)의 3개의 데이터들로 구성되어 진다. 또한 전신 제스처 데이터베이스는 3개의 방향(정면을 기준으로 0도, +45도, -45도)에서 촬영된 2차원 스테레오 데이터를 포함하고 있으며, 명령형 데이터베이스는 정면에서 2대의 각각 다른 초점거리를 가지는 렌즈(6mm, 12mm)를 사용하여 전신과 상반신을 촬영한 데이터를 포함하고 있다. 그림 1은 첫 번째 기간 촬영을 위한 스튜디오 환경을 보여준다.

표 1. 정상 제스처 수집 인원 분포

나이 / 성별	60~69세(50%)	70~79세(35%)	80~세 (15%)
남자(10명)	5	4	1
여자(10명)	5	3	2

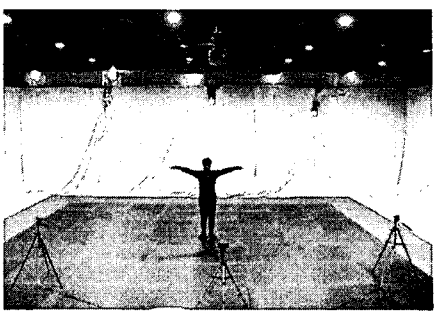


그림 1. 스튜디오 환경

3. 촬영 환경

3차원 동작 데이터를 얻기 위한 동작 카메라는 Motion Analysis 사의 Eagle Digital System을 이용하였다. 첫 번째 기간 촬영을 위해 사용된 스튜디오는 10.1m(넓이)x11.1m(길이)x5.1m(높이)의 크기로 각 면에 3개씩의 3차원 동작 카메라를 설치하였고, 정면을 기준으로 0도, 45도, -45도 방향에 스테레오 카메라를 설치하였다. 두 번째 기간 촬영을 위해서는 6m(넓이)x9m(길이)x4m(높이)의 스튜디오를 사용하였으며, 각 면에 4개씩 총 12개의 3차원 동작 카메라와 정면을 기준으로 0도, 45도, -45도 방향에 스테레오 카메라를 설치하였다. 명령형 동작 촬영의 촬영은 두 번째 기간에 같이 이루어졌으며, 동일한 조건하에, 방향별 3대의 스테레오 카메라를 이용하는 대신 정면 3.6앞에 초점 거리가 각각 6mm, 12mm인 스테레오 카메라 두 대를 사용함으로, 전신과 상반신 스테레오 영상을 촬영하였다. 그림 2는 두 기간에 사용된 두개의 스튜디오 구조를 보여준다.

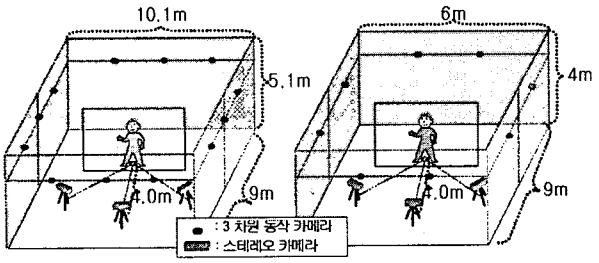


그림 2. 스튜디오 구조 설계

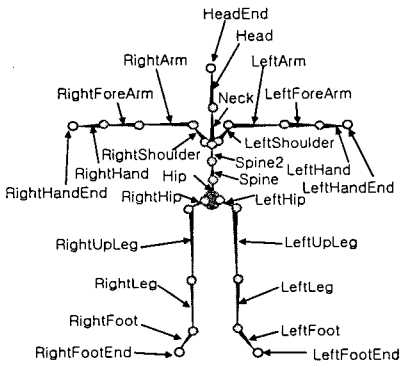


그림 3. 제안된 3차원 신체 모델

그림 4. 3차원 제스처 데이터의 예
(정면 영상(위), 좌측 영상(아래))

4. 데이터 구성

4.1. 전신 제스처 데이터

일상생활의 제스처가 다양하지만, KU 제스처 데이터베이스에서는 노인들을 대상으로 (1)의자에 앉기, (2)의자에서 일어나기, (3)제자리 걷기, (4)무릎과 허리 짚기, (5)오른손 들기, (6)손 앞으로 뻗기, (7)허리 숙이기, (8)바닥에 앉기, (9)바닥에 무릎 꿇기, (10)바닥에 눕기, (11)손 흔들기, (12)제자리 뛰기, (13)앞으로 걷기, (14)원형으로 돌기의 14가지 제스처를 정상 제스처로 정의하였다. 비정상 제스처로는 바닥에 서있는 상태에서 (1)앞으로 쓰러지기, (2)뒤로 쓰러지기, (3)좌로 쓰러지기, (4)우로 쓰러지기와 바닥에 앉은 상태에서 (5)좌로 쓰러지기, (6)우로 쓰러지기, (7)뒤로 쓰러지기와 의자에 앉은 상태에서 (8)좌로 쓰러지기, (9)우로 쓰러지기, (10)앞으로 쓰러지기의 10가지 제스처를 정의하였고, 노인들은 부상이 생길 수 있으므로, 젊은 사람들을 대상으로 데이터를 수집하였다.

4.2. 3차원 제스처 데이터

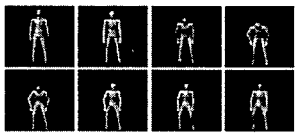
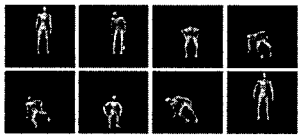
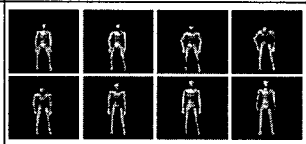
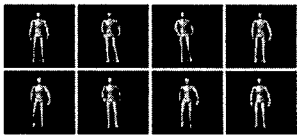
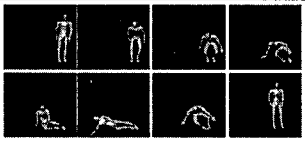
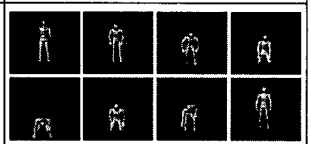
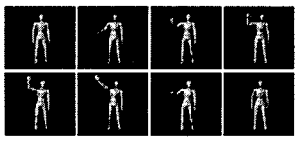
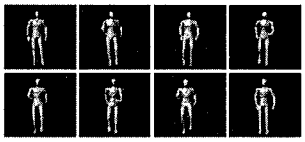
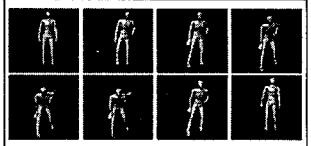
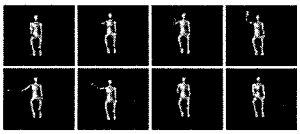
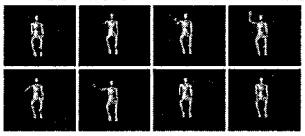
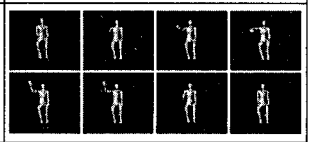
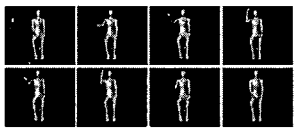
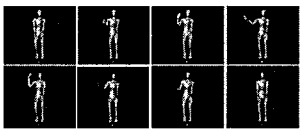
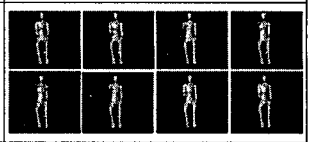
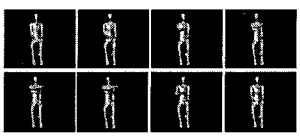
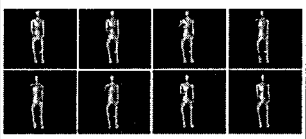
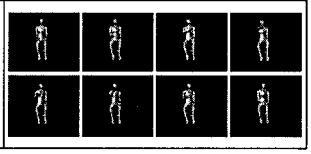
3차원 제스처들을 나타내기 위하여 각 부분을 21개로 나누어 3차원 신체 모델을 구성하였다. 그림 3은 3차원 신체 모델의 구성을 보여준다. 이 모델의 움직임을 저장하기 위하여 3차원 신체 모델의 분류 체계를 포함하고, 각 부분의 위치와 인접 프레임사이의 3차원 이동, 회전차로 나타내어지는 HTR 형식(Hierarchical Translation Rotation Format)을 사용하였다. 그림 4는 HTR형식으로 저장된 '바닥에 눕기' 제스처를 정면과 왼쪽 방향에서 보이는 형태로 재구성한 모델의 움직임을 보여준다.

표 2. 파일의 이름 체계

형식	Inx-E-Sag_Gi_CdV_TFram.Ext

기호	의 미
Inx	Index : {000, ..., 999}
E	Session : {1, 2, ..., 9}
S	Sex : {M,F}
Ag	Age : {00, ..., 99}
Gi	Gesture index : {00, ..., 99}
Cd	Camera direction : 00 : 3D motion Data, 01, 02, 03 : 스테레오 비디오 데이터 (각각 0도, 45도, -45도)
V	Camera view : D : 3차원 제스처 데이터, L : 왼쪽 옆모습 R : 오른쪽 옆모습
T	File Type D : 3차원 제스처 데이터 V : 2차원 스테레오 비디오 데이터 S : 2D 실루엣 데이터
Fram	Frame number : dddd : 3차원 제스처 데이터 vvvv : 비디오 데이터 {0000 , ... , 9999} : 영상 데이터
Ext	File extension {HTR, AVI, BMP}

표 3. 전신 제스처 영상과 영령형 제스처 영상의 예제

제스처	예제 영상들	제스처	예제 영상들	제스처	예제 영상들
의자에 앉기		바닥에 앉기		의자에서 일어나기	
제자리 걷기		바닥에 눕기		바닥에 무릎 꿇기	
오른손 들기		제자리 뛰기		무릎과 허리 짚기	
'3'자 쓰기		'8'자 쓰기		오라는 손짓하기	
뒤쪽으로 손 흔들기		오른쪽으로 손 흔들기		정중앙 가리키기	
왼쪽 중앙 가리키기		오른쪽 중앙 가리키기		가운데 위 가리키기	

4.3. 스테레오 비디오 데이터

전신 제스처 데이터베이스를 위해서 정면을 기준으로 0도, 45도, -45도 방향에 스테레오 카메라를 설치하여 촬영하여 3방향에서 바라본 제스처 데이터를 수집하였고, 영령형 제스처 데이터베이스를 위해서는 정면에 각각 6mm, 12mm의 초점 거리를 가지는 두 대의 스테레오 카메라를 설치하여 전신과 상반신 영상을 촬영하였다. 각 스테레오 영상은 320x240의 크기의 24비트 컬러 영상이다.

4.4. 실루엣 비디오 데이터

2차원 스테레오 영상 촬영 시에 사람이 없는 영상을 촬영하여 얻어진 배경 화면을 이용하여 2차원 스테레오 영상의 전경(사람)을 추출하고, 2진 영상으로 저장하여 실루엣 비디오 데이터를 생성하였다.

4.5. 파일 이름 규칙

사용자의 편의를 위하여 3차원 모션 데이터, 2차원 스테레오 데이터, 실루엣 데이터를 쉽게 구분하고, 파일 이름으로부터 대상 비디오의 내용을 알 수 있는 규칙을 만들었다. 표 2는 제스처 데이터베이스에서 사용되는 파일 이름 규칙을 보여준다.

표3은 구축된 3차원 전신 제스처 데이터와 영령형 제스처 데이터로 재구성된 모델의 예를 보여 준다. 보다 자세한 사항들은 홈페이지(http://GestureDB.korea.ac.kr)를 통해 참고 할 수 있다.

참고 문헌

[1] A. Martinez, R. Wilbur, R. Shay and A. Kak, "Purdue RVL-SLLL ASL Database for Automatic Recognition of American Sign Language," Proc. of IEEE International Conference on Multimodal Interfaces, Pittsburgh, USA, pp. 167 -172, October 2002.

[2] R. Bowden, R. Mitchell and M. Sarhadi, "Non-Linear Statistical Models for the 3D Reconstruction of Human Pose and Motion from Monocular Image Sequences," Image and Vision Computing, Vol. 18, No. 9, pp. 729-737, 2000.

[3] A. F. Bobick and A. Y. Johnson, "Gait Recognition Using Static, Activity-Specific Parameters," Proc. of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Vancouver, Canada, pp. 423-430, July 2001.