

헬스케어를 위한 영상기반 기절동작 인식시스템 개발

소인미*, 강선경*, 김영운*, 이지근*, 정성태**

Development of a Vision Based Fall Detection System For Healthcare

In-Mi So *, Sun-Kyung Kang *, Young-Un Kim *, Chi-Geun Lee *, Sung-Tae Jung **

요 약

이 논문은 스테레오 영상을 이용하여 응급상황을 인식하기 위하여 기절 동작을 인식하는 방법을 제안한다. 사람의 다양한 동작에서 학습과 인식에 필요한 영상 정보를 추출하기 위하여 3차원 정보를 사용하였고, 인식 알고리즘으로는 HMM을 이용하였다. 두 대의 카메라 영상에서 각각 배경을 생성한 다음에 배경 영상과 입력 영상의 차이를 이용하여 움직인 객체를 추출하였다. 그리고 움직인 객체를 포함하는 사각형을 생성한 다음 두 카메라의 캘리브레이션 정보를 이용하여 3차원 정보를 추출하였다. 3차원 공간상에서의 사각형의 너비와 높이의 변화량과 사각형 중심점 위치의 변화량 각각에 대하여 동작 인식률을 실험하였다. 실험 결과 너비와 높이의 특장값을 이용하는 것보다 중심점의 3차원 위치 변화량을 이용하는 것이 높은 인식률을 보였다.

Abstract

This paper proposes a method to detect fall action by using stereo images to recognize emergency situation. It uses 3D information to extract the visual information for learning and testing. It uses HMM(Hidden Markov Model) as a recognition algorithm. The proposed system extracts background images from two camera images. It extracts a moving object from input video sequence by using the difference between input image and background image. After that, it finds the bounding rectangle of the moving object and extracts 3D information by using calibration data of the two cameras. We experimented to the recognition rate of fall action with the variation of rectangle width and height and that of 3D location of the rectangle center point. Experimental results show that the variation of 3D location of the center point achieves the higher recognition rate than the variation of width and height.

▶ Keyword : Activity Recognition(동작인식), Segmentation(객체추출), Detection(동작감지), Calibration(캘리브레이션), HMM

• 제1저자 : 소인미

• 접수일 : 2006.11.15, 심사일 : 2006.11.18, 심사완료일 : 2006. 12.26

* 원광대학교 컴퓨터공학과 대학원

**원광대학교 컴퓨터공학과교수

※ 이 논문은 2006년 정부(교육인적자원부)의 재원으로 한국학술진흥재단의 헬스케어기술개발사업 지원을 받아 수행된 연구임.

1. 서론

인체의 동작 상태를 구분해 주는 것을 상황인식(context awareness)이라고 한다. 여기서 상황(context)이라는 용어를 Dey[1]는 '상황은 사용자와 시스템(응용 프로그램) 간의 상호 작용에 연관된 존재물(사람, 장소, 사물 등)을 특징 짓는 모든 형태의 정보'로 정의하였다. 1994년 Schilit와 Theimer은 상황인식 컴퓨팅 용어를 최초로 논의하고, 사용자의 작업과 관련 있는 적절한 정보 또는 서비스를 사용자에게 제공하는 과정에서 상황 즉, 사용자 위치, 사용자의 주변 사람 정보와 사용 가능한 물체 정보, 시간이 경과되면서 이러한 대상의 변화까지 수용할 수 있는 경우를 상황인식 시스템이라고 하였다. 일반적인 상황 정보는 사용자 상황(정상, 병환, 사고피해, 장애발생), 물리적 환경 상황(실내, 자동차, 실외, 야외), 컴퓨팅 시스템 상황(전원 On/Off, 인터넷 On/Off-line, 로그 In/Out), 사용자-컴퓨터 상호 작용 이력 등으로 분류될 수 있고, 유비쿼터스 컴퓨팅 시스템의 필수적인 기능으로 인식되고 있다(2).

2000년 고령사회의 진입과 가족 구성원의 보호를 받지 못하는 독거노인의 수가 급증함으로써 개인 중심의 예방 의료가 필요하게 되고 언제 어디서나 서비스를 받을 수 있는 건강관리기술이 필요하게 되었다. 따라서 대내 거주하는 구성원의 인체 동작 상태를 구분하는 상황인식 기술이 필요하게 된다. 지금까지 연구되고 개발된 상황인식시스템 들을 살펴보면 생체신호 센서를 장착하여 인체동작 상태를 구분 정보로 활용하거나 웨어러블 컴퓨터의 입력 장치를 활용하기도 하고 가속도계를 이용하여 인체동작상태의 상황인식을 하기도 한다(3,4,5,6). 컴퓨터가 사람의 몸동작이나 손짓을 자동으로 인식이 가능하다면 사람과 컴퓨터간의 편리한 통신 매체 수단이 될 수 있고 좀 더 편리한 인터페이스와 환경을 제공 받을 수 있을 것이다(7). 또한 사용자의 생체신호 계측 및 자동진단, 응급경보가 가능한 헬스케어형 의료 정보시스템의 일부분으로 응용되어 사고발생전의 예방 및 발생 직후의 즉각적인 대응을 가능하게 할 수 있다.

따라서 본 논문에서는 비접촉형 데이터로 영상 데이터를 이용하여 기절동작을 인식함으로써 응급상황을 자동으로 감지하는 시스템을 제안한다.

다음 그림 1은 기절동작 인식 시스템의 전체적인 흐름을 나타내고 있다. 영상 데이터 획득을 위해 CCD 카메라를 통해 DVR 장치에 저장한 후 배경과 객체를 분리하고, 스테레오

오 카메라를 캘리브레이션하여 각 카메라를 통해 획득된 영상 데이터를 3차원 가상공간에 역 투영(back projection) 시켜 필요한 특징값들의 깊이 정보를 포함한 3차원 값을 얻을 수 있다.

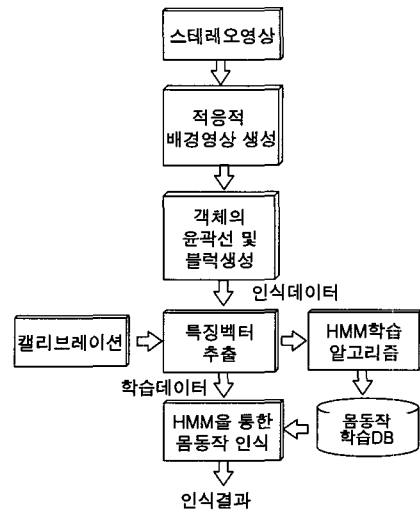


그림 1. 전체시스템 구성도
Fig 1. Overall System Configuration

스테레오 카메라 영상에서 각각 배경을 생성한 다음 배경 영상과 입력 영상의 차이를 이용하여 움직이는 객체를 배경영상에서 분리한다(8). 그리고 객체의 외곽선을 추출한 뒤 움직이는 객체를 포함하는 사각형 영역을 생성하여 객체를 추적하도록 하였다. 두 카메라의 캘리브레이션 정보를 이용하여 3차원 정보를 추출하였다. 3차원 공간상에서의 사각형 너비와 높이의 변화량과 사각형 중심점 위치의 변화량에 각각에 대하여 동작 인식을 실험을 하였다. 또한 본 논문에서는 학습과 인식 방법으로, 확률에 따른 패턴의 일련을 모델링한 HMM(Hidden Markov Model)을 이용하였고, HTK(Hidden markov Toolkit)를 사용하여 학습과 인식 실험을 하였다. 추출된 특징값들은 HTK에 적용할 수 있도록 데이터의 형식을 변형하여 사용하였다. 실험 결과 사각형의 너비와 높이의 특징값을 이용하는 것보다 사각형 중심점의 3차원 위치 변화량을 이용하는 것이 높은 인식률을 보였다.

본 논문의 2절에서는 특징벡터 추출과 캘리브레이션에 대하여 설명하고 3절에서는 HMM 알고리즘을 이용한 몸동작 학습과 인식에 관해 설명한다. 4절에서는 몸동작인식에 관한 실험결과를 비교분석하여 설명한다.

II. 특징벡터 추출

본 논문에서 제안하는 기절동작 인식 방법은 전체적으로 두 과정으로 나눌 수 있다. 첫 번째 과정은 영상을 획득하여 배경으로부터 몸동작 객체를 추출하는 과정이고 두 번째 과정은 특징벡터를 추출한 다음에 HMM을 이용하여 동작을 인식하는 것이다.

2.1 영상획득

그림 2와 같이 2대의 CCD 컬러 카메라로부터 352×240의 해상도로 획득되는 영상 데이터는 DVR(Digital Video Recorder) 시스템으로 전송되고 채널별 영상 정보로 저장된다. DVR 시스템은 카메라를 통해 입력된 영상 데이터의 아날로그 신호를 디지털 신호로 전환하여 국제적인 표준 방식인 MPEG 또는 H.263/H.264 등으로 영상을 압축 및 복원하여 녹화, 검색은 물론 저장된 영상을 고해상도로 재생하여 볼 수 있는 디지털 감시시스템이다.



그림 2. CCD 컬러 카메라
Fig 2. CCD Color Camera

2.2. 배경과 객체의 분리

2.2.1 배경 영상 준비

카메라를 통하여 초당 30프레임의 영상데이터를 획득할 수 있다. 영상데이터는 시간에 따라 미세한 잡음이 발생할 수 있고 조명의 변화로 인하여 배경영상이 변화될 수 있으므로 전처리 과정을 거쳐서 안정적인 배경영상을 생성해야 한다. 전처리 과정을 살펴보면 먼저 중간값 필터를 이용해 잡음을 제거한 후 컬러영상을 흑백영상으로 변환한다. 전처리된 영상을 N 프레임 까지 받아들여 각 픽셀 값들을 누적한다. 누적된 값을 N으로 나누어 배경영상의 각 픽셀에 대한 평균값을 구한다. 그림 3은 프레임을 누적해 평균값을 적용한 스테레오 배경 영상을 보이고 있다.

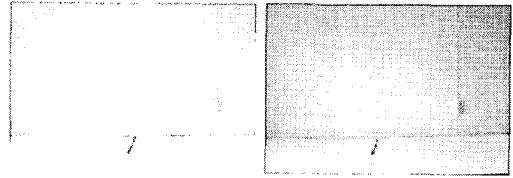


그림 3. 프레임을 누적해 평균값을 적용한 배경영상
Fig 3. Background images averaged for a sequence of frames

2.2.2 배경 영상과 객체분리를 통한 몸동작 정보 추출

배경 영상과 몸동작 정보를 분리 하기위해 미리 처리된 배경 영상과 몸동작 영상 정보가 사용된다. 몸동작 영상에서 정보를 추출하기 위해 몸동작 영상의 전처리가 필요하다. 배경 영상의 전처리 방법과 같이 먼저 중간값 필터를 이용해 잡음을 제거 한 후 컬러영상을 흑백영상으로 변화 시킨다. 컬러영상을 흑백으로 변환된 픽셀 값을 임계값과 비교하여 이진 영상을 생성하였다. 그림 4는 중간값 필터와 흑백 처리가 적용된 영상을 나타내고 있다.

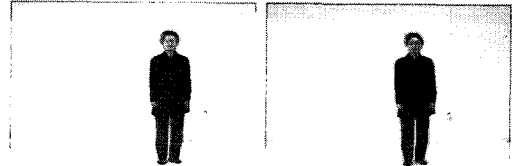


그림 4. 필터와 흑백처리가 적용된 영상 (Left / Right)
Fig 4. Result images after median filtering and color-to-gray conversion (Left/Right)

몸동작 객체를 추출하기 위해 배경영상에서 추출된 각 픽셀 값과 전 처리된 몸동작 영상의 픽셀 값을 이용해 몸동작 영역을 분리해 낸다. 이를 위해 먼저 식(1)과 같이 몸동작 객체 영상을 생성한다.

$$G_{(x,y)} = \begin{cases} 255, & \|Background_{(x,y)} - Object_{(x,y)}\| > TH \\ 0, & Otherwise \end{cases} \dots\dots (1)$$

식(1)과 같이 배경영상의 각 픽셀의 평균값에서 몸동작 영상의 각 픽셀 값을 뺀 후 기준 값 보다 크면 흰색(255)으로, 반대로 작거나 같으면 검정색(0)으로 설정한다. 이런 처리를 거쳐 두개의 값 검정색(0), 흰색(255)를 가진 이진 데이터 결과를 획득 하게 된다. 이중 흰색(0)값이 객체이며, 검정색(0)값이 배경이 된다. 그 결과 그림 5와 같이 배경 영상과 몸동작 객체가 분리되어 표시된다.



그림 5. 배경 영상과 분리된 몸동작 객체 (왼쪽 / 오른쪽)
Fig 5. Body objects segmented from background images (Left/Right)

위와 같이 획득한 몸동작 객체 이미지는 많은 잡음을 가지고 있으며, 이를 제거 할 수 있는 방법이 필요하다. 본 논문에서는 침식(Erode)과 팽창(Dilate)의 연산을 이용하여 불필요한 잡음을 제거했다. 다음으로 잡음이 제거된 몸동작 객체 이미지에서 윤곽선(Contour) 정보를 찾는다. 윤곽선이란 움직임 객체의 가장자리 선을 말하는데, 본 논문에서는 객체의 가장자리에 있는 한 픽셀로부터 시작하여 가장자리에 있는 픽셀을 계속 따라가다가 시작점에 도착하면 추적을 멈추는 방법을 사용하였다. 윤곽선 검출의 시작점은 영상의 맨 윗줄의 좌측에 있는 픽셀부터 시작하여 오른쪽방향으로 스캔하면서 다음 조건을 만족하는 픽셀 $P(x, y)$ 를 선택한다.

$$P(x, y) = 0 \text{ 그리고 } (P(x, y-1) = 255 \text{ 또는 } y = 0)$$

윤곽선 검출의 시작점으로부터 물체의 가장자리 픽셀을 한 방향으로 따라가다가 원래 자리에 돌아오면 탐색이 종료된다. 픽셀 $P(x, y)$ 가 가장자리 픽셀일 경우에 인접한 가장자리 픽셀은 픽셀 $P(x, y)$ 주위의 8개 픽셀 중에 존재하는데, 이를 찾는 방법은 다음과 같다.

- 1) 인접 픽셀중의 한 픽셀에서 검사를 시작한다.
- 2) 픽셀값이 배경색이면 계속해서 다음 픽셀을 검사한다. 픽셀값이 물체색이면 그 픽셀을 인접한 가장자리 픽셀로 정하고 검사를 멈춘다.

다시 말하면, 인접한 8개의 픽셀을 순서대로 검사하면서 물체색을 가지는 첫 번째 픽셀을 만나면 그 픽셀을 인접한 가장자리 픽셀로 정하는 것이다. 가장자리 픽셀 $P(x, y)$ 주위의 8개 픽셀에 시계 방향으로 번호를 부여했다.

인접 픽셀 검사의 시작 위치는 현재까지 가장자리 픽셀을 탐색해온 방향에 따라 다르게 결정된다. 먼저 윤곽선 검출의 시작점에서는 1번 픽셀부터 검사를 시작하고 윤곽선을 따라가는 도중에는 현재의 탐색에서 가장자리 픽셀로 선택된 픽셀 번호와 그에 따른 다음의 탐색에서 인접 픽셀의 검사를 시작할 픽셀 번호가 표 1에 나타나 있다.

표 1. 가장자리 픽셀의 위치와 다음 검사 시작 위치
Table 1. Edge pixel and next probe location

가장자리 픽셀 발견 위치	다음 검사 시작 위치
0	7
1	7
2	1
3	1
4	3
5	3
6	5
7	5

가장자리 픽셀을 탐색해 나가다가 더 이상 연결된 가장자리 픽셀이 없는 경우 또는 탐색의 시작점을 만나는 경우에는 탐색을 종료한다. 그림 6은 몸동작 객체의 윤곽선 정보를 추출한 결과 화면이다. 마지막으로 그림 7과 같이 추출된 몸동작 객체의 윤곽선 정보를 이용해 사각형 영역을 추출해 한다. 이 사각형 영역 정보를 이용해 2차원 정보를 획득하고 캘리브레이션 과정을 통해 3차원 정보를 획득해 특징값으로 사용한다.



그림 6. 몸동작 객체의 윤곽선 추출 (왼쪽 / 오른쪽)
Fig 6. Contour extraction of body object (Left / Right)



그림 7. 객체 윤곽선의 사각형 영역 추출 (왼쪽 / 오른쪽)
Fig 7. Extraction of rectangle region of body object (Left / Right)

2.3 캘리브레이션

캘리브레이션은 카메라들간의 기하학적인 관계를 활용하여 정확한 3차원 정보를 얻기 위한 방법으로 외부 카메라 모델(extrinsic camera model)과 내부 왜곡 모델(intrinsic distortion model)을 결정하는 과정이다.

카메라는 3차원의 점들을 2차원의 이미지 평면위로 투영

시키는 장치이다. 3차원의 한 점 $V=(X, Y, Z)^T$ 가 2차원 이미지 평면의 한 점 $v=(x, y)^T$ 에 투영된다고 할 때에 이들의 관계는 식 (2)와 같이 나타낼 수 있다. 여기에서 f_x 와 f_y 는 카메라의 초점 거리를 나타내고 (p_x, p_y) 는 주점(principal point)을 나타낸다. 주점이란 카메라의 중심을 통과하는 선이 이미지 평면과 수직으로 만나는 점을 의미한다.

$$x = f_x \frac{X}{Z} + p_x, \quad y = f_y \frac{Y}{Z} + p_y \dots\dots\dots (2)$$

식 (2)에서는 카메라 좌표계와 세계 좌표계(world coordinate)가 일치하는 것을 가정하고 있다. 카메라가 움직여서 두 좌표계가 일치하지 않는 경우에, 카메라의 움직임은 회전과 평행 이동 두 가지의 조합으로 나타낼 수 있다. 카메라가 세계 좌표계에 대하여 R 만큼 회전하고 T 만큼 평행 이동한 경우에는, 3차원 좌표에 먼저 R^{-1} 와 T^{-1} 만큼 회전 변환과 이동 변환을 적용한 다음에 위의 식을 적용하면 2차원 이미지 평면에서의 좌표를 구할 수 있다. 종합하면 3차원 공간의 한 점은 일련의 변환, 즉 회전 변환, 평행 이동 변환, 카메라 모델에 의한 투영 변환에 의해 2차원 이미지 평면의 한 점으로 투영된다. 이들 변환은 하나의 카메라 행렬 P 로 표현될 수 있고 이를 이용하여 투영을 다음 식과 같이 표현할 수 있다.

$$v = PV$$

그림 8에 나타나 있듯이 두 대의 카메라 이미지가 주어지면, 서로 대응되는 점에 대하여 3차원 공간으로의 역투영을 수행하여 3차원 좌표를 구할 수 있다. 그림에서 C 와 C' 은 카메라의 렌즈의 중심을 나타내고 v 와 v' 은 두 이미지 평면에서의 3차원 공간의 점 V 에 대한 투영을 나타낸다.

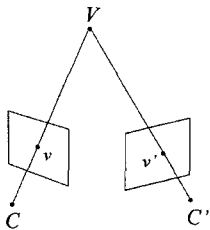


그림 8. 두 영상을 이용한 3차원 재구성
Fig 8. 3D reconstruction using stereo images

일반적으로, 카메라 캘리브레이션 과정은 다음의 세 단계

에 의해 수행된다. 우선, 3차원 실세계 좌표와 그에 대응되는 2차원 영상 좌표의 집합을 획득한다. 다음으로, 주어진 좌표 집합에 대해 일련의 방정식을 적용해 각각의 카메라에 대한 카메라 파라미터 값들을 계산한다. 마지막으로 투영 행렬(projection matrix)로부터 epipolar geometry를 구성할 수 있다(9). 본 논문에서는 Zhang의 캘리브레이션 알고리즘의 수정 버전에 근거하여 캘리브레이션 과정을 수행하였다(10). 캘리브레이션 과정은 최종적으로 world coordinate의 기준점으로부터 카메라까지의 회전 및 이동 행렬을 제공한다. 즉, 캘리브레이션 과정을 위한 입력으로 world coordinate로부터 두 카메라까지의 회전 및 이동 행렬 (R, T)right와 (R, T)left를 각각 얻을 수 있다. 이렇게 얻어진 카메라의 외부 파라미터를 이용하여 실세계에 대한 3차원 좌표를 얻을 수 있다.

그림 9와 같은 평면 패턴을 사용하여 Zhang의 알고리즘을 통해 카메라 파라미터를 얻을 수 있다.

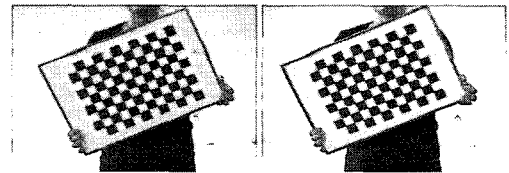


그림 9. 캘리브레이션을 위한 평면 패턴(왼쪽 / 오른쪽)
Fig 9. Flat pattern for calibration

먼저 그림 9와 같은 30mm 크기의 정사각형이 배열된 패턴을 스테레오 카메라를 통해 16개 촬영한 후 Zhang의 알고리즘을 통해 스테레오 카메라 각각의 내부 파라미터인 초점거리와 주점(principle point)을 획득하고, 외부 파라미터인 회전 벡터와 이동 벡터를 구할 수 있다.

카메라 행렬이 각각 P 와 P' 이라 할 때 그림 8의 투영 관계는 $v = PV$, $v' = P'V$ 와 같이 나타낼 수 있다. v 와 v' 의 위치가 정확하면 카메라의 중심 C 와 C' 으로부터 v 와 v' 을 통과하는 두 선은 3차원 공간에서 만나게 된다. 두선이 3차원 공간에서 만나는 경우에는 방정식을 이용하여 V 의 좌표를 간단하게 구할 수 있다. 그러나 카메라 이미지의 노이즈나 해상도의 한계로 인하여 v 와 v' 의 위치를 완벽하게 구할 수는 없고 결과적으로 두 선은 대부분의 경우에 공간 상에서 만나지 않는다. 이런 경우에는 V 의 좌표를 구하는 것은 복잡한 문제가 된다. 이 문제를 풀기 위해 여러 가지 방법이 제안되었는데 본 논문에서는 선형 최적화 방법을 사용한다. 이 방법에서는 $v = PV$, $v' = P'V$ 를 하나의 식 $AV=0$ 의 형태로 결합한다. v 와 v' 이 정확하다면 $AV=0$

이 만족되고 연립 방정식을 풀어서 V 의 값을 구할 수 있다. 그러나 v 와 v' 의 값에 노이즈가 있으면 $AV=0$ 을 만족하는 V 는 존재하지 않고 $\|AV\|$ 값을 최소화 하는 V 가 최적의 해가 된다. 이러한 선형 방정식의 최적화 해를 구하는 방법으로 여러 가지 방법이 있는데, 본 논문에서는 SV D(Singular Value Decomposition)를 사용하였다.[10]

2.3 특징벡터 추출

일본 논문에서는 2 종류의 특징값을 추출하여 실험하였다. 추출된 객체의 사각형 영역의 너비와 높이 및 중심점의 변화량을 특징값으로 활용 하였다. 즉, 첫 번째 특징값은 사각형 영역의 너비와 높이의 크기 변화량에 대한 값이다. 두 번째 특징값은 사각형 영역의 중심점(x, y, z)의 이동 변화량에 대한 값이다. 2종류의 특징값 모두 3이상의 변화량을 가질 때에만 특징값으로 인식하였다. 이는 잡음으로 인한 미세한 변화에 대해서는 특징값을 취하지 않도록 하여 오류를 처리하기 위함이다.

다음 그림 10은 기절동작에 대한 사각형 영역을 표시한 그림이다. 시간의 흐름에 따라 사각형 영역의 너비와 높이 변화가 있고 중심점이 이동됨을 알 수 있다.

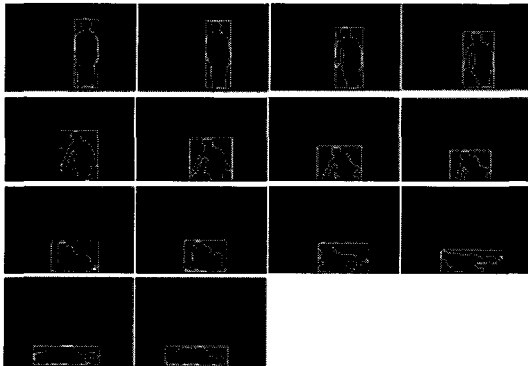


그림 10. 기절동작
Fig 10. Faint action

2.4.1 사각형 영역의 너비와 높이의 크기 변화량

두 대의 카메라를 통해 입력 받는 영상에서 학습과 인식 실험에 사용되는 3차원 정보를 얻을 수 있다. 그림 12는 2차원 좌표점을 이용해 3차원 특징점인 너비와 높이를 계산하는 방법을 나타내고 있다. 3차원 특징 정보는 좌측 영상의 사각형 선분의 중심점 4개(P1, P2, P3, P4)의 x, y 좌표점과 우측 영상의 사각형 선분의 중심점 4개(P'1, P'2, P'3, P'4)의 x, y 좌표점을 캘리브레이션한 후 4개(C

1, C2, C3, C4)의 3차원 좌표점 (x, y, z)값을 획득한다. 이 좌표점을 이용해 사각형 영역의 너비(C1, C3)와 높이(C2, C4)의 변화량을 계산할 수 있다.

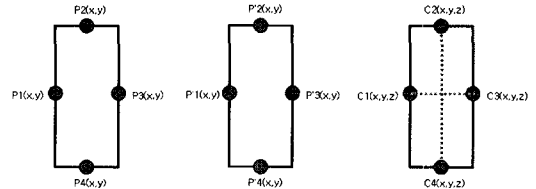


그림 11. 2차원 좌표점을 이용한 3차원 특징값(너비, 높이) 추출
Fig 11. 3D feature(width, height) extraction based on 2D feature points

2.4.2 사각형 중심점(x, y, z)의 이동 변화량

두 대의 카메라를 통해 입력 받는 영상에서 학습과 인식 실험에 사용되는 3차원 정보를 얻을 수 있다. 그림 12는 2차원 좌표점을 이용해 3차원 특징점인 (x, y, z)값을 계산하는 방법을 나타내고 있다. 3차원 특징 정보는 좌측 영상의 사각형 중심점 P(x,y)와 우측 영상의 사각형 중심점 P'(x,y)를 캘리브레이션 처리를 통해 C(x, y, z)와 같은 3차원 좌표값을 얻을 수 있다. 이와 같은 3차원 좌표값을 이용해 x축, y축, z축의 객체 이동에 따른 변화량을 계산해 특징값을 획득 하였다.

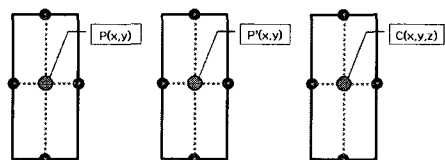


그림 12. 2차원 좌표점을 이용한 3차원 특징값(x, y, z) 추출
Fig 12. 3D feature(x, y, z) extraction based on 2D feature points

2.5 HMM 알고리즘을 이용한 몸동작 학습 및 인식

HMM은 다중 확률구조를 갖는 프로세스를 모델링 하는데 매우 적합한 방법으로 HMM 파라미터들의 정밀한 계산을 위한 효과적인 알고리즘이 존재한다는 장점 때문에 근래에 음성 인식 뿐만 아니라 얼굴 인식, 제스처 인식, 영상 인식 등 패턴 인식 영역에서 전반적으로 널리 사용되고 있는 방법이다.

본 논문에서는 HMM 알고리즘이 적용된 HTK(Hidden Markov Toolkit)을 이용하여 학습과 인식을 하였다. HTK는 HMM 인식 알고리즘을 수행할 수 있도록 HMM 모델

생성, 모델의 재추정과 학습 및 인식 그리고 인식결과 분석에 필요한 라이브러리를 지원하고 있는 인식관련 툴킷이다. 인식에 필요한 특징 데이터를 HTK 파일 형식에 맞게 변환하여 이용하였고 사용된 HMM 모델은 3 상태 left-to-right 전이형태의 모델을 사용하였다.

그림 14는 본 논문에서 사용된 학습과 인식과정을 잘 나타내고 있다. 먼저 특징값을 추출해 HTK 파일을 생성하고 학습에 필요한 레이블 파일, 스크립트 파일 등을 만들어 준다. 만들어진 HTK 파일과 레이블 파일 등을 이용해 HCompv, HERest 처리 과정을 거쳐 학습을 수행하고 이 과정의 결과로 몸동작 데이터베이스가 만들어진다. 학습이 완료되면 인식 테스트를 한다. 테스트를 수행하기 위해 학습된 몸동작 데이터베이스와 동작 사전, 스크립트 파일 등이 필요하며, HParse, HVite, HResults 처리 과정을 통해 최종 인식결과를 얻을 수 있다.

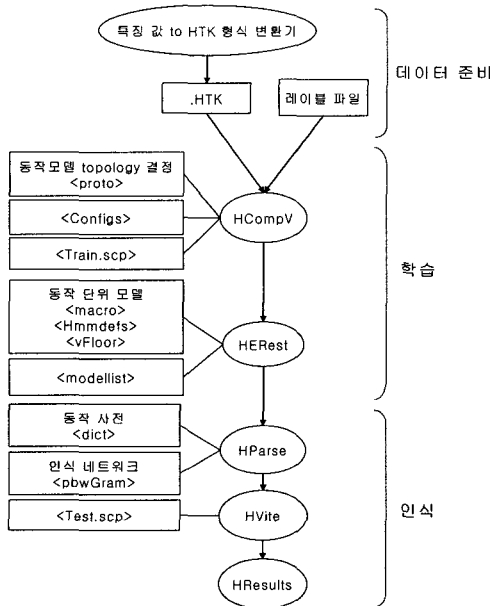


그림 13. HTK를 이용한 학습과 인식
Fig 13. Training and recognition by using HTK

III. 실험 결과

본 논문에서 실험한 실험환경은 일반적인 사무실의 자연 조명에서 영상을 촬영 했으며, DVR장비를 이용해 스테레오 카메라로부터 입력을 받아 실시간으로 영상 데이터를 획득 하였다. 기질동작 인식시스템은 입력 받은 영상에서 특징값

을 추출하고 학습과 인식에 필요한 데이터를 생성해 주며, 최종 인식의 결과 값을 확인 할 수 있도록 되어있다. 개발 환경은 Microsoft Visual C++ 6.0 언어를 사용 하였고 Intel Pentium IV 3.0GHz CPU, 1GB RAM, 120GB HDD 시스템과 Microsoft WindowsXP Platform 환경에서 실험이 수행되었다.

실험에 사용된 몸동작은 총 9개(서 있다가 기절하기, 걷기, 뛰기, 의자에 앉기, 의자에서 일어서기, 허리 굽혀 굽기, 제자리에서 점프하기, 서 있다가 바닥에 앉기, 앉아 있다가 일어서기)이다. 사람은 7명분이며 각각 9동작씩 총 63개의 동작이 실험에 사용 되었다. 또한 같은 동작이 10번 반복해 총 630개의 동작으로 구성되어 있다. 10번 반복 중 첫 번째부터 일곱 번째는 학습 데이터로 사용되었으며, 여덟 번째에서 열 번째는 인식 실험 데이터로 사용되었다. 실험은 특징값 추출방식에 따라 3가지로 구분해 진행하였다. 첫 번째는 객체의 사각형 너비와 높이에 대한 크기 변화량으로 5회 실시 하였으며, 인식 결과는 표 2와 같이 평균 90.47%의 인식률을 보였다. 두 번째는 사각형의 중심점 좌표(x, y)의 변화량으로 5회 실시 하였으며, 인식 결과는 표 3과 같이 평균 91.42%의 인식률을 보였다. 세 번째는 사각형의 중심점 좌표(x, y, z)의 변화량으로 5회 실시 하였으며, 인식 결과는 표 4와 같이 평균 94.28%의 인식률을 보였다. 실험 결과에서 보논바와 같이 몸동작에 따른 움직임 객체의 사각형 중심점 좌표(x, y, z)의 변화량 대한 특징값을 사용하는 것이 더 좋은 인식률을 보였음을 알 수 있었다. 이런 결과는 캘리브레이션을 통한 3차원 위치 정보를 이용해 x축, y축 특징값 이외의 z축의 특징 정보(움직임 객체의 깊이 정보)를 추가해 더 좋은 결과를 얻을 수 있음을 알 수 있다. 또한 움직임 객체의 사각형 너비와 높이를 특징값으로 사용 한 것 보다 움직임 객체의 사각형 중심점을 특징값으로 사용하는 것이 좋은 결과값을 얻을 수 있었다.

표 2. width, height 값의 변화량에 따른 인식률
Table 2. Recognition rate for the variation of width and height

학 습							테 스트				인식률
L1	L2	L3	L4	L5	L6	L7	L8	L9	L10	95.24%	
L3	L4	L5	L6	L7	L8	L9	L10	L1	L2	95.24%	
L5	L6	L7	L8	L9	L10	L1	L2	L3	L4	85.71%	
L7	L8	L9	L10	L1	L2	L3	L4	L5	L6	80.95%	
L9	L10	L1	L2	L3	L4	L5	L6	L7	L8	95.24%	
평균 인식률										90.47%	

표 3. 중심점(x, y) 이동의 변화량에 따른 인식률
Table 3. Recognition rate for the variation of the movement of center point (x, y)

학 습							테 스트			인식률
L1	L2	L3	L4	L5	L6	L7	L8	L9	L10	100%
L3	L4	L5	L6	L7	L8	L9	L10	L1	L2	80.95%
L5	L6	L7	L8	L9	L10	L1	L2	L3	L4	85.71%
L7	L8	L9	L10	L1	L2	L3	L4	L5	L6	100%
L9	L10	L1	L2	L3	L4	L5	L6	L7	L8	90.48%
평균 인식률										91.42%

표 4. 중심점(x, y, z) 이동의 변화량에 따른 인식률
Table 4. Recognition rate for the variation of the movement of center point (x, y, z)

학 습							테 스트			인식률
L1	L2	L3	L4	L5	L6	L7	L8	L9	L10	100%
L3	L4	L5	L6	L7	L8	L9	L10	L1	L2	85.71%
L5	L6	L7	L8	L9	L10	L1	L2	L3	L4	95.24%
L7	L8	L9	L10	L1	L2	L3	L4	L5	L6	95.24%
L9	L10	L1	L2	L3	L4	L5	L6	L7	L8	95.24%
평균 인식률										94.28%

본 논문에서는 스테레오 영상으로부터 캘리브레이션 하여 3차원 정보를 추출하여 특징값으로 이용하고 이를 HTK를 통하여 HMM 학습을 하고 인식하는 방법을 제안하였다. 캘리브레이션을 통한 3차원 위치 정보를 이용해 x축, y축 특징값 이외 z축의 특징 정보(움직임 객체의 깊이 정보)를 추가해 동작에 따른 움직임 객체의 사각형 중심점 좌표(x, y, z)의 변화량을 특징값으로 사용함으로써 인식률이 향상됨을 보였다. 이것은 3차원 특징값을 이용하는 것이 견고하고 안정된 인식을 한다는 것을 보이고 있다.

향후 본 시스템은 사고 발생 직후의 즉각적인 대응이 필요한 헬스케어형 의료정보시스템의 일부분으로 응용되어 적용이 가능하리라 본다.

참고문헌

[1] Dey, A.K., "Understanding and Using Context", Personal and Ubiquitous Computing Journal, Vol. 5 (1), pp. 4-7, 2001.
[2] 이선우, "장착형 센서를 이용한 사용자 상황 인식", 한국정

보처리학회지, Vol.10, No.4, pp. 66-78, 2003
[3] Baihua Li, Holstein, H., and Qinggang Meng, "Articulated Point Pattern Matching in Optical Motion Capture System", Control, Automation, Robotics and Vision, ICARCV 2002. 7th International Conference, Vol. 1, pp. 298-303, 2002
[4] Barnes, N.M., Rose, D.A.D., Garner, P.: Lifestyle Monitoring: Technology for Supported Independence. IEE Comp. and Control Eng. J. pp. 169-174, 1998
[5] Bonner, S.: Assisted Interactive Dwelling House: Edinvar Housing Assoc. Smart Tech. Demonstrator and Evaluation Site In: Improving the Quality of Life for the European Citizen (TIDE), pp. 396 - 400, 1997
[6] McKenna, S.J., Marquis-Faulkes, F., Gregor, P., Newell, A.F.: Scenario-based Drama as a Tool for Investigating User Requirements with Application to Home Monitoring for Elderly People. In Proc. of HCI, 2003
[7] 한탁돈, 정철호, 김기섭, 이남규, 이충규, "정보화 사회와 착용형 컴퓨터의 전망", 한국멀티미디어학회지, 제6권, 제1호, pp.88-104, 2002년
[8] 강현중, 이광형, "이동 객체 감시를 위한 실시간 객체추출 및 추적시스템", 한국컴퓨터정보학회, 제10권2호, pp.59-68, 2005
[9] W. Woo, N. Kim and Y. Iwadate, "Stereo Imaging Using a Camera with Stereoscopic Adapter", Proc. of IEEE - Systems, Man, and Cybernetics (SMC) 2000, vol.2, pp.1512-1517. Oct. 2000
[10] Z. Zhang, "Flexible camera calibration by viewing a plane from unknown orientations," Proc. of the Seventh IEEE Int'l Conf., vol. 1, pp.666-673, 1999
[11] Richard Hartley and Andrew Zisserman, Multiple View Geometry in computer vision, Cambridge University Press, 2003

저자 소개



소인미
 1994년 2월 : 원광대학교 컴퓨터공학과(공학사)
 2002년 8월 : 원광대학교 정보컴퓨터교육 전공(교육학석사)
 2004년 3월 ~ 현재 : 원광대학교 컴퓨터공학과 박사과정
 관심분야 : 영상인식 컴퓨터 비전, 영상처리



강선경
 2000년 2월 : 원광대학교 전기·전자공학부 졸업
 2004년 2월 : 원광대학교 정보·컴퓨터공학과 석사학위 취득
 2006년 2월 ~ 현재 : 원광대학교 컴퓨터공학과 박사과정
 관심분야 : 휴먼 컴퓨터 인터페이스, 영상처리, 영상인식, 임베디드시스템



김영운
 2003년 2월 : 원광대학교 전기전자및정보공학부(공학사)
 2005년 8월 : 원광대학교 컴퓨터 공학과(공학 석사)
 2006년 2월 ~ 현재 : 원광대학교 컴퓨터공학과 박사과정
 1994-1999 영원한친구 개발팀장
 2000-현재 파라(PARA) 대표
 2003-현재 원광보건대학 정보컨텐츠과 겸임교수
 관심분야 : 컴퓨터그래픽스, 영상처리, EAI, XML



이지근
 2000년 2월 : 원광대학교 컴퓨터공학과(공학사)
 2002년 2월 : 원광대학교 컴퓨터 공학과(공학 석사)
 2002년 3월 ~ 현재 : 원광대학교 컴퓨터공학과 박사과정
 관심분야 : 영상인식 영상인식, 컴퓨터 비전, 영상처리



정성태
 1987년 2월 서울대학교 컴퓨터공학과(공학사)
 1989년 2월 서울대학교 컴퓨터공학과(공학 석사)
 1994년 8월 서울대학교 컴퓨터공학과(공학 박사)
 1995년 3월~현재 원광대학교 전기전자및정보공학부 교수
 관심분야 : 영상인식, 영상처리, 컴퓨터 그래픽스