

모노 카메라 영상기반 시간 간격 윈도우를 이용한 광역 및 지역 특징 벡터 적용 AdaBoost기반 제스처 인식

황승준¹ · 고하윤¹ · 백중환^{1*}

AdaBoost-based Gesture Recognition Using Time Interval Window Applied Global and Local Feature Vectors with Mono Camera

Seung-Jun Hwang¹ · Ha-Yoon Ko¹ · Joong-Hwan Baek^{1*}

¹Department of Electronics and Information Engineering, Korea Aerospace University, Gyeonggi-do, 10540, Korea

요약

최근 안드로이드, iOS 등의 셋톱박스 기반의 스마트 TV에 대한 보급에 따라 제스처로 TV를 컨트롤 할 수 있는 새로운 접근을 제안한다. 본 논문에서는 모노 카메라 센서를 이용한 AdaBoost 기반 제스처 인식에 관한 알고리즘을 제안한다. 우선, 신체 좌표 추출을 위해 가우시안 배경 제거 및 Camshift 기반 자세 추적 및 추정 알고리즘을 사용한다. AdaBoost 학습 모델을 신체 정규화된 광역 및 지역 특징 벡터의 집합을 특징 패턴으로 하여, 속도가 다른 동작들을 인식할 수 있도록 하였다. 또한 속도가 다른 다양한 제스처를 인식하기 위해 다중 AdaBoost 알고리즘을 적용하였다. CART 알고리즘을 이용하여 성공적인 중요 특징 벡터를 확인하고 중요도가 낮은 특징벡터를 제거하는 방식을 적용하면서 분류 성공률이 높은 최적의 특징 벡터를 탐색하였다. 그 결과 24개의 주성분 특징 벡터를 찾았으며, 기존 알고리즘에 비해 낮은 오분류율(3.73%)과 높은 인식률(95.17%)을 지닌 특징 벡터 및 분류기를 설계하였다.

ABSTRACT

Recently, the spread of smart TV based Android iOS Set Top box has become common. This paper propose a new approach to control the TV using gestures away from the era of controlling the TV using remote control. In this paper, the AdaBoost algorithm is applied to gesture recognition by using a mono camera. First, we use Camshift-based Body tracking and estimation algorithm based on Gaussian background removal for body coordinate extraction. Using global and local feature vectors, we recognized gestures with speed change. By tracking the time interval trajectories of hand and wrist, the AdaBoost algorithm with CART algorithm is used to train and classify gestures. The principal component feature vector with high classification success rate is searched using CART algorithm. As a result, 24 optimal feature vectors were found, which showed lower error rate (3.73%) and higher accuracy rate (95.17%) than the existing algorithm.

키워드 : 3D 제스처 인식, 패턴 인식, 기계 학습, AdaBoost 알고리즘, 신체 분할

Key word : 3D Gesture Recognition, Pattern Recognition, Machine Learning, AdaBoost, Body Segmentation

Received 15 January 2018, Revised 17 January 2018, Accepted 24 January 2018

* Corresponding Author Joong-Hwan Baek(E-mail:jhbaek@kau.ac.kr, Tel:+82-02-300-0125)

Department of Electronics and Information Engineering, Korea Aerospace University, Gyeonggi-do, 10540, Korea

Open Access <http://doi.org/10.6109/jkiice.2018.22.3.471>

pISSN:2234-4772

©This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License(<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.
Copyright © The Korea Institute of Information and Communication Engineering.

I. 서론

사회 흐름이 계속해서 스마트화에 초점이 맞춰짐에 따라 구글, 애플, 삼성 등과 같은 글로벌 기업의 스마트 TV가 보편화되었다. 스마트 TV란 인터넷 접속 기능이 결합되어, 각종 애플리케이션을 설치해 TV 방송 시청 이외의 다양한 기능을 활용할 수 있는 다기능 TV를 말한다. 다기능을 갖는 여러 종류의 스마트 TV들에서 볼 수 있는 공통적인 특징으로 컨트롤 방식의 변화를 꼽을 수 있다. 그 중 카메라와 센서의 기술이 발전함에 따라 여러 HCI(Human Computer Interaction)기술이 등장하고 있다. 스테레오 카메라, TOF(Time of Flight) 카메라, 딥러닝 기반 모노카메라 등을 이용하여 사람의 손동작이나 포즈를 인식하는 기술들이 그 예이다[1][2]. 하지만 스마트 TV 성능 상 깊이 센서, GPU를 지원하지 않아 깊이 맵이나 딥러닝 기술의 적용의 어려움이 있다. 이러한 문제를 해결하기 위해 스마트 TV의 웹캠이나 휴대전화 카메라 기반의 TV 컨트롤 기술이 필요하다. 본 논문에서는 모노 카메라 영상 기반의 스마트 TV 제어 위한 제스처 인식 알고리즘을 제안한다.

동작 인식에 관련된 연구는 시, 공간적 변이를 가진 패턴을 추출하는 은닉 마르코프 모델(Hidden Markov Model), DTW(Dynamic Time Warping) 등을 통하여 많은 부분 연구가 되었다[3][4]. 본 논문에서는 배경 모델에 의해 생성된 신체 좌표 정보를 기계학습의 특징벡터로 하는 제스처 인식 알고리즘을 제안한다. 제안하는 알고리즘은 RGB영상 기반의 배경모델을 구성하고 신체 정보를 가져오는 방식으로 TOF카메라에 비해 떨어지는 정확도를 Camshift 기반의 추적, 보정을 수행하여 신체 좌표 추적의 정확도를 향상한다. 다음으로, 신체 좌표의 일정시간 연속적인 값을 추출하고 정규화한 후 제안한 광역 및 지역적 특징벡터를 생성한다. 제스처는 사용자의 신장이나 사용자와 카메라의 거리에 따라 값의 크기가 변하는데, 이를 해결하기 위해 인식된 사용자의 어깨넓이로 신체의 크기를 나누어 특징을 정규화한다. 또한 제스처는 사용자의 움직임 속도에 따라 변하는데, 이 때문에 시간 및 공간적인 변이를 가진 패턴 추출법이 필요해진다. 하지만 이러한 시, 공간적인 변이를 가진 패턴 추출법은 시작과 끝을 구분할 수 없는 제스처 인식에는 한계가 존재한다. 본 논문에서 제안한 시간간격 윈도우를 이용한 광역 및 지역특징 벡터 적용

AdaBoost 기반 제스처 인식기법은 이러한 시, 공간적인 변이를 가진 패턴 추출 알고리즘이 아닌 다른 학습 모델을 이용하여도 처음과 끝 구분에 상관없이 제스처를 인식할 수 있도록 한다.

본 논문의 구성은 다음과 같다. 2장에서는 제스처 취득 방법과 광역적 지역적 특징벡터에 관해 기술한다. 3장에서는 시간간격 윈도우와 샘플링을 적용한 제스처의 인식 및 학습에 관한 방법을 설명한다. 4장에서는 제안한 알고리즘의 구현과 실험 결과를 보이며, 5장에서 결론을 맺는다.

II. 제스처 취득 및 특징 벡터 정의

2.1. 배경 제거 기반 신체 좌표 추출

사용자의 신체 모션 데이터를 취득하는 기기로는 모노 카메라 기반, 구조광을 이용한 Kinect, 시간 지연 방식의 TOF, 레이저를 이용한 Lidar 등 카메라 방식과 Wii 컨트롤러를 이용한 모션 데이터를 취득하는 방법들이 있다. 본 논문에서는 신체의 각 관절 좌표를 취득하는 것에 있어 스마트폰, TV에서 사용 가능한 그림 1과 같은 가우시안 배경 제거 기법을 이용한 Camshift 기반 자세 추적 및 추정 알고리즘을 사용한다[5].

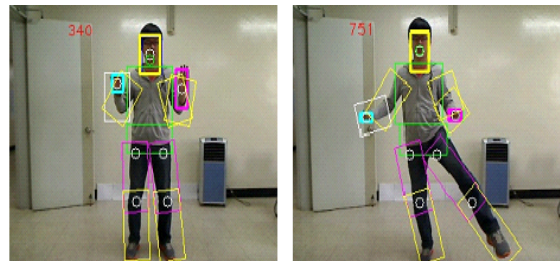


Fig. 1 Human Body Tracking and Pose Estimation Using Modified Camshift Algorithm

본 논문에서는 사람이 제스처를 취하는 시간이 다름에 따라 신체 좌표 60 프레임을 기준으로 다양한 제스처 속도에 인식되도록 하기 위해 신체 좌표를 샘플링 하여 학습한다. 또한 모노카메라 기반 신체 좌표 추출 알고리즘 특성상 깊이 값이 존재 하지 않는다. 신체 좌표의 깊이 값 추출을 위해 초기 탐색한 신체부위의 넓이를 기준으로 넓이의 변화량에 따라 깊이 값을 대체하는 방법으로 상대적 3D 제스처 인식을 수행하도록 하였다.

2.2. 특징 벡터(Feature Vector)

동작 인식을 사용하는 사람의 신체적 조건이 모두 동일할 수는 없다. 이러한 신체적 차이점에서 발생하는 문제점을 해결하기 위하여 광역적 및 지역적 특징 벡터를 제안한다. 신체의 비례를 이용한 정규화된 특징 벡터를 사용함으로써 사용자의 신체 변화에 따른 인식률을 향상시킨다.

2.2.1. 신체 좌표 정규화

사람의 좌표는 제스처가 행해지는 장소나 제스처를 취하는 사람에 따라 그 크기는 다르다. 이러한 사람마다 다른 제스처를 인식하기 위해 손끝 좌표와 손목 좌표의 중점의 좌표를 정규화할 필요가 있다. 즉, 손끝 좌표와 손목 좌표의 중점과 척추 사이의 거리를 사용자의 어깨 넓이로 나누는 정규화 방식을 사용한다. 이렇게 정규화된 특징벡터는 사용자들의 신체 크기나 카메라와의 거리와 상관없는 유사한 값을 가지게 된다. 신체 좌표상의 척추, 손끝, 손목, 왼어깨, 오른어깨의 좌표를 P_{spine} , P_{hand} , P_{wrist} , $P_{lshoulder}$, $P_{rshoulder}$ 라 하면 정규화된 손의 위치 좌표 P_n 은 다음 식 (1) 과 같다.

$$P_n = \frac{P_{spine} - \left(\frac{P_{hand} - P_{wrist}}{2} \right)}{P_{rshoulder} - P_{lshoulder}} \quad (1)$$

2.2.2. Jia Sheng의 특징 벡터

3D 좌표의 특징 벡터를 추출하기 위해 Jia Sheng은 손끝의 변위, 제스처가 행해지는 최대, 최소의 좌표 값, 전체 구간의 각도변화 정보의 누적 값을 이용하였다[6]. Jia Sheng의 특징벡터 f_1 에서 f_{12} , f_{15} 는 위치 좌표의 변화량에 관한 정보이며, f_{13} , f_{14} 는 각도 정보의 누적 값이다. f_{15} 는 동작이 행해지는 시간 정보이다. 이러한 Jia Sheng의 특징 벡터는 광역 특징 벡터로만 이루어져 있으며, 지역적인 특징은 고려하지 않았음을 알 수 있다. 제스처의 종류에 따라 광역적인 특징벡터는 서로 같아질 수가 있으므로 지역적인 특징정보는 광역적인 특징 벡터를 보완하기 위해 필요한 정보이다.

2.2.3. 특징 벡터 정의

지역적 특징정보란 제스처의 부분적인 정보로 초기

에 어떤 움직임이 있는지와 중, 후반에는 어떤 변화가 있는지를 구간별로 자세하게 포함하고 있는 정보이다. 이러한 제스처 구간을 일정 크기로 분할함에 따라 특징 벡터의 수는 늘거나 줄어들 수 있다. 구간을 1프레임 단위로 정하면 모든 좌표 변화 값이 특징 벡터가 되고, 전체 프레임으로 정하면 시작과 끝 좌표의 변화 값이 특징 벡터가 된다. 그림 2에 제스처의 동작에 따른 지역적 특징 벡터 정의 방법을 보인다.

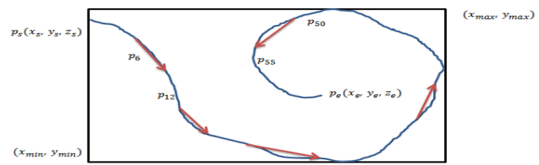


Fig. 2 Local Feature Vectors

본 논문에서는 5개로 나누어진 구간에서 x , y , z 각각의 변화량을 특징 1~15로 정의 하였다. f_1, f_2, f_3 은 초기 위치의 변화량이고, f_{13}, f_{14}, f_{15} 는 마지막 위치의 변화량이 된다. 관련된 지역적 특징벡터 식을 아래에 보인다. 식 (2), (3), (4)는 각각 x , y , z 의 지역적 특징 벡터가 된다. 다음의 특징벡터는 60프레임의 기준을 5개로 분할하였기에 12프레임의 간격을 보이지만, 3장에서 설명하는 시간간격 특징벡터의 샘플링단계에서 20프레임의 예상 동작에서는 4프레임의 간격으로 특징벡터를 생성하게 된다.

$$f_1 = x_{12} - x_1, f_4 = x_{24} - x_{13}, \dots, f_{13} = x_{60} - x_{49} \quad (2)$$

$$f_2 = y_{12} - y_1, f_5 = y_{24} - y_{13}, \dots, f_{14} = y_{60} - y_{49} \quad (3)$$

$$f_3 = z_{12} - z_1, f_6 = z_{24} - z_{13}, \dots, f_{15} = z_{60} - z_{49} \quad (4)$$

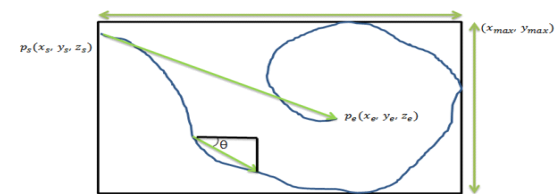


Fig. 3 Global Feature Vectors

그림 3은 제스처의 동작에 따른 광역적 특징벡터 정의 방법 표시도이다. 광역적 특징벡터는 전체 동작에 관련된 특징들로 전체적인 동작의 모양에 관한 정보를 담고 있다. 전체 동작의 x , y , z 의 최대 변화량, 처음과

끝의 x, y, z 변화량, x, y, z 변화량의 총 누적 값, x, y, z 변화량 중 최댓값에 대한 정보를 담고 있다. 정의한 광역적 특징벡터 18개를 아래에 보인다.

$$f_{16} = \overline{p_s - p_e}, f_{17} = x_s - x_e, f_{18} = y_s - y_e, f_{19} = z_s - z_e \quad (5)$$

$$f_{20} = \sum_{i=start}^{end} (\arccos((y_{i+1} - y_i)/(x_{i+1} - x_i))), (\theta > 0) \quad (6)$$

$$f_{21} = \sum_{i=start}^{end} (-\arccos((y_{i+1} - y_i)/(x_{i+1} - x_i))), (\theta < 0) \quad (7)$$

$$f_{22} = \sum_{i=start}^{end} (\arccos((z_{i+1} - z_i)/(y_{i+1} - y_i))), (\theta > 0) \quad (8)$$

$$f_{23} = \sum_{i=start}^{end} (-\arccos((z_{i+1} - z_i)/(y_{i+1} - y_i))), (\theta < 0) \quad (9)$$

$$f_{24} = \sum_{i=start}^{end} (\arccos((z_{i+1} - z_i)/(x_{i+1} - x_i))), (\theta > 0) \quad (10)$$

$$f_{25} = \sum_{i=start}^{end} (-\arccos((z_{i+1} - z_i)/(x_{i+1} - x_i))), (\theta < 0) \quad (11)$$

$$f_{26} = \sum_{i=start}^{end} \theta, f_{27} = \sum_{i=start}^{end} \theta^2, (\theta = \angle(p_{i+1} - p)) \quad (12)$$

$$f_{28} = \overline{p_{max} - p_{min}} \quad (13)$$

$$f_{29} = (x_{max} - x_{min})/f_{28}, f_{30} = (y_{max} - y_{min})/f_{28}, \quad (14)$$

$$f_{31} = (z_{max} - z_{min})/f_{28}$$

$$f_{32} = \sum_{i=start}^{end} \overline{p_{i+1} - p_i}, f_{33} = \max(\overline{p_{i+1} - p_i}) \quad (15)$$

식 (5)의 f_{16} 은 처음과 끝 좌표의 삼차원 위치 변화 크기이며 f_{17}, f_{18}, f_{19} 는 처음과 끝의 x, y, z 변화량을 나타낸다. 식 (6), (7)의 f_{20}, f_{21} 은 2차원 xy 평면에 대한 제스처의 각도 변화량의 누적 정보를 나타낸다. 제스처의 각도 변화가 시계 방향, 반시계 방향과 크기에 따라 f_{20}, f_{21} 의 누적 값이 변하게 된다. 식 (8), (9)의 f_{22}, f_{23} 은 yz 평면에 대한 각도 정보이며, 식 (10), (11)의 f_{24}, f_{25} 는 xz 평면에 대한 각도 정보이다. 식 (12)의 f_{26}, f_{27} 은 3차원 좌표상에서 각도 변화량의 누적 값과 각도 변화량의 제곱의 누적 값이다. 식 (13)의 f_{28} 은 최댓값과 최솟값 좌표의 위치 변화 크기이며, 식 (14)의 f_{29}, f_{30}, f_{31} 은 최댓값과 최솟값의 x, y, z 변화량을 나타낸다. 식 (15)의 f_{32} 는 각 좌표의 변화 크기의 누적 값이며, f_{33} 은 그 중 최대 변화 크기 값이다.

III. 제스처 학습 및 인식

3.1. 시,공간적 변이를 가진 패턴 추출법

제스처를 인식하기 위해서는 미리 지정된 동작인 참조 패턴과 사용자의 입력동작인 입력 패턴과의 패턴 매칭이 필요하다. 더불어, 사용자에게 따라 변하는 제스처를 인식하기 위해서는 시간 및 공간적인 변이를 가진 패턴을 추출하는 기법이 필요하다. 시간 및 공간적인 변이를 가진 패턴을 추출하는 기법으로는 DTW, 신경망, 은닉 마르코프 모델 등을 들 수 있다[7][8][9]. 하지만 이러한 기법은 신호의 시작과 끝을 지정해야하는 문제가 존재하며, 제스처의 특성상 시작과 끝 동작을 지정하는 부분에 어려움이 존재한다. 본 연구에서는 일반적인 기계학습 알고리즘에서 사용 가능한 제스처 학습 및 인식 모델을 제안한다.

3.2. AdaBoost 알고리즘

AdaBoost 알고리즘은 분류 규칙을 순차적으로 생성하고, 이전의 분류 규칙을 적용하여 얻은 관측 값에서 샘플 데이터의 분포를 재조정한다. 학습 초기에 샘플 데이터의 가중치는 동일한 상태에서 시작하지만, 매 라운드가 진행될수록 이전의 분류 규칙을 적용하여 얻은 관측 값으로 오 분류 데이터에는 높은 가중치를 부여하고, 반대로 정 분류된 데이터에는 낮은 가중치를 부여하는 방식으로 샘플 데이터의 분포를 재조정한다. 이러한 과정은 분류하기 힘든 데이터에 우선권을 주기 위한 전략이다.

AdaBoost 알고리즘은 특성상 동적인 특징 보다는 정적인 특징을 가진 자세 인식에 많이 사용한다. 본 논문에서는 AdaBoost 알고리즘과 같이 시, 공간적인 변이를 가진 패턴추출 방법이 아닌 기계학습 알고리즘에 동적인 특징을 인식하기 위한 제안한 정규화된 특징 벡터를 이용한다[10].

본 논문에서는 시, 공간적인 변이를 갖지 않은 AdaBoost 학습 모델을 사용하기 때문에 위의 특징 패턴의 단일 학습은 제스처의 속도 변화 인식에는 적당하지 않다. 속도가 다른 제스처를 인식하기 위해 다중 AdaBoost 분류기가 필요하며 그에 따른 특징 벡터 샘플링 과정이 필요하다. 본 연구에서는 60 프레임의 제스처를 기본 특징 패턴으로 정한다. 이 특징 패턴은 인식할 수 있는 가장 느린 제스처이며, 샘플링을 통하여 20

프레임의 제스처까지 인식할 수 있게 한다.

3.3. 다중 AdaBoost 분류기 설계

속도가 다른 제스처를 인식하기 위해 다중 AdaBoost 분류기의 설계 및 설계된 분류기를 이용한 입력된 특징 벡터의 제스처 인식 과정을 그림 4에 보인다[11]. 위의 제스처 취득과정을 통해 생성한 제스처의 특징 벡터를 이진 분류 방식을 이용하여 AdaBoost 분류기를 설계한다. 이 AdaBoost 분류기는 이진 분류 방식이기에 인식하고자 하는 제스처의 수만큼의 다중 AdaBoost 분류기가 된다. 설계된 다중 AdaBoost 분류기에 최근 20 프레임에서 60 프레임까지의 취득된 제스처 특징 패턴을 입력하면 각 제스처 분류기에 따라 참, 거짓의 결과를 내면서 제스처 인식이 가능하게 된다.

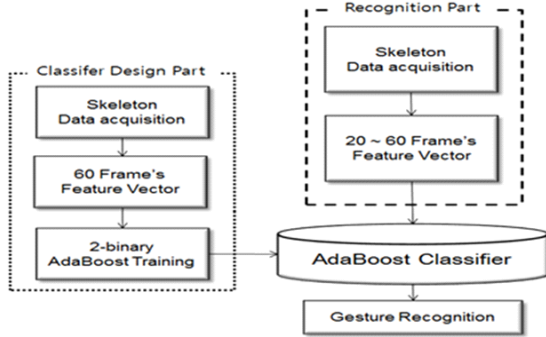


Fig. 4 Gesture training and classifier structure Diagram

3.4. 다중 AdaBoost 분류기 학습

본 연구에서는 AdaBoost 분류기를 다수 사용하여 입력 받은 동작을 구분하는 다중 AdaBoost 분류기를 사용한다[12][13]. 분류 방식은 1:N의 이진 분류 방식을 사용하여 학습하고자 하는 동작을 제외한 나머지 동작들은 모두 네거티브로 학습한다. 따라서 동작을 제스처 1, 2, 3, 4로 구분했을 때, 제스처 1 동작을 훈련하고자 할 때는 제스처 1을 제외하고 나머지 3개의 동작의 데이터는 전부 네거티브로 학습 되도록 한다. 제스처 2, 3, 4 또한 마찬가지로 이진 분류 방식을 통해 학습되는 방법을 그림 5에 보인다. 동작을 위와 같은 이진 분류 방식으로 학습하면서 같이 학습한 다른 동작으로 인식되는 오인 식률을 줄일 수 있다.

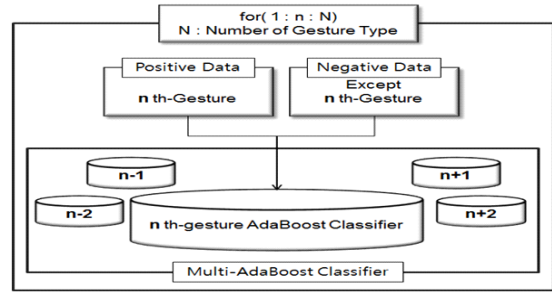


Fig. 5 Multiple AdaBoost classifiers in binary classification

3.5. 프레임 간의 관계를 이용한 군집 결정

매 프레임에 진행되는 제스처 인식의 순서도를 그림 6에 보인다. 본 논문에서는 속도가 다른 제스처를 인식하기 위해 최근 20 프레임부터 60 프레임까지의 특징 벡터를 AdaBoost 분류기에 입력한다. AdaBoost 분류기는 매 프레임 별 샘플링 된 수만큼의 제스처 인식을 수행하며, 다양한 속도의 제스처를 분류한다.

위와 같이 매 프레임 별, 샘플링 수에 따라 제스처 인식을 수행하는데, 경우에 따라 동일 시간에 복수의 인식결과가 발생될 수도 있다. 이러한 경우는 각 분류기의 결과값을 합산하여 최대치 값으로 최종 제스처를 결정한다.

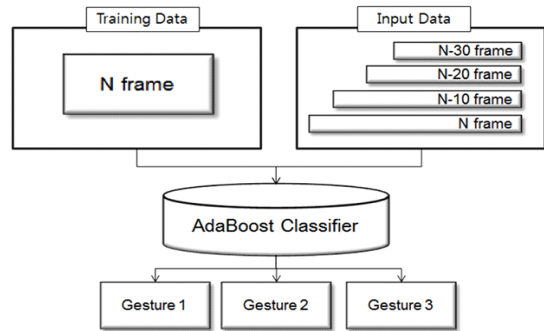


Fig. 6 Algorithm for multi-variable frame classification

3.6. CART 알고리즘

CART 알고리즘은 데이터 마이닝 기법 중의 하나인 의사결정나무로, 종속변수가 범주형이면 분류나무를, 연속형이면 회귀나무를 그리는 알고리즘이다. 나무모델은 시각화 되어 있어 결과 해석에 용이하다.

하위 그룹으로의 분할은 간단한 조건식에 따라 명료하고 논리적인 규칙을 통해 이루어진다. 이때, 규칙에

의한 답이 “Yes”인 경우 왼쪽으로 분류, “No”인 경우 오른쪽으로 분류한다. 분류 나무에서 기본이 되는 2가지 핵심내용은 독립변수의 차원 공간을 반복적으로 분할하는 것과 평가용 데이터를 사용하여 가지치기를 하는 것이다.

본 논문에서는 8개의 제스처를 분류를 하는데 제한한 33가지의 특징벡터로 형성된 CART트리 기반으로 그 수를 줄여 사용한다. CART 트리를 형성하는 과정의 첫 번째는 루트 노드의 특징과 분할 값을 선정하는 것이다. 이는 연속된 변수들의 중간 값을 선택하고, 이러한 값을 이용하여 불순도를 얼마나 많이 제거하는가에 따라 순위가 매겨진다. 불순도를 측정하는 방법에는 여러 가지 방식이 있다. 가장 대표적인 방법으로 지니 지수가 있다. 반응 변수의 m 개의 집단은 $k = 1, 2, \dots, m$ 으로 표기된다고 가정할 경우, 직사각형 A에 대한 지니 불순도 지수는 다음 식 (16)과 같다.

$$I(A) = 1 - \sum_{k=1}^m p_k^2 \quad (16)$$

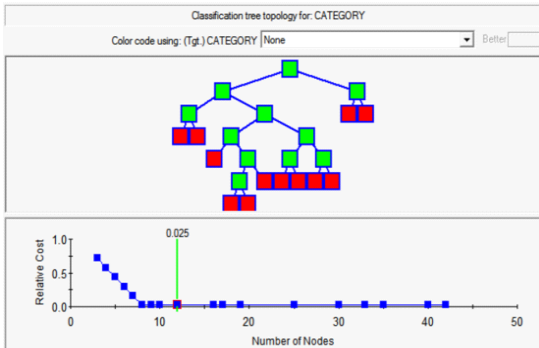


Fig. 7 CART analysis of Salford prediction model

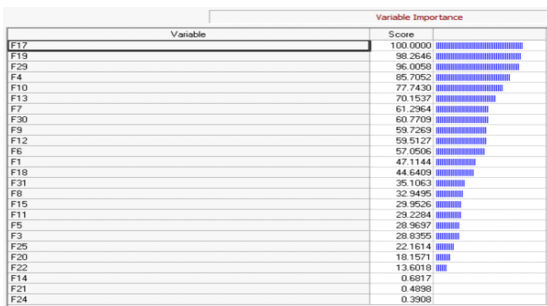


Fig. 8 Variable importance analysis of CART analyzer

p_k 는 직사각형 A에서 k 집단에 속하는 관찰치들의 비율이다. 이 측정치는 0 과 $(m-1)/m$ 사이의 값을 갖는다. 이진 분류에서 지니 지수는 $p_k = 0.5$ 에서 최대가 된다. CART 분류기를 이용하여 적정수의 노드를 설정할 수 있으며, 각 변수의 사용빈도, 분류한 양에 따른 중요도를 판별할 수 있다. 최소 에러에 따른 분류노드 생성 모습을 그림 7에 보이며, 각 변수의 중요도를 그림 8에 보인다. 변수의 중요도 수치는 전체 데이터 중 현재 변수로 분류되는 데이터의 확률을 의미한다. 그림 8에서 24개의 특징벡터 변수만으로도 전체 제스처를 구분함을 확인할 수 있다.

본 논문에서는 이러한 가지치기 방법을 이용하여 필요 없는 특징벡터를 줄이고 필요한 특징 벡터만을 사용하는 방법으로 학습용 데이터 피드백 알고리즘을 제안한다. 이는 기계학습 알고리즘의 기본이 되는 특징벡터의 차원의 저주문제를 해결하고 차원에 따른 최적의 성능을 보이는 지점을 찾을 수 있는 방법이다.

IV. 실험

4.1. 실험 데이터

제스처 모델의 실험 데이터의 수는 8개의 각각의 동작마다 세 명의 실험자로부터 얻은 총 450개이다. 이중 학습용 데이터로 300개, 평가용 데이터로 150개의 데이터를 사용하였다.

실험에 사용한 제스처는 아래의 그림 9와 같이 정하였다. 상, 하, 좌, 우로 직선으로 움직이는 제스처와, 위에서 아래로 둥그렇게 내려오는 제스처, 뒤에서 앞으로 둥그렇게 나오는 제스처 총 8가지의 기본 제스처이다. 이와 같은 기본 제스처의 연속적인 인식은 더욱 복잡한 제스처의 기본 토대가 되므로 이러한 기본 제스처의 인식은 중요하다.

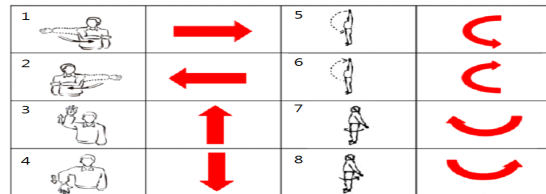


Fig. 9 Gesture to classify

4.2. 학습 및 성능 평가 방법

제안한 알고리즘의 성능을 평가하기 위해 AdaBoost 기계학습 알고리즘으로 분류기를 생성하여 오분류율과 정확도를 측정하였다. 다중 AdaBoost 분류기 설계는 Opencv 기반으로 구현하였다. CART 알고리즘을 이용한 불필요 특징벡터 분석은 Salford Predictive Modeler 7.0을 사용하였으며, 그 기능 중 변수 중요도 조사를 통해 특징벡터 중요도를 판단할 수 있다.

4.3. 비교 평가

기존 특징 벡터와 제안한 특징 벡터의 성능 평가 및 제안한 가지치기 피드백 프로세스의 성능 검증을 위해 비교 평가를 수행한다. 표 1은 다양한 속도에서 인식되는 제스처의 인식률을 제스처 종류에 따라 분류하였으며, 표 2는 8가지의 제스처 인식률을 속도에 따라 분류하였다.

Table. 1 Accuracy and error rates of existing and proposed feature vectors Comparison analysis table according to gesture type

Gesture Type	Jia-Sheng Feature		Proposed Feature			
	Accuracy (%)	Error (%)	Accuracy (%)	Error (%)	Accuracy (%)	Error (%)
1	99.00	0.67	97.66	3.03	98.17	2.83
2	95.50	2.83	95.11	1.78	95.83	0.83
3	95.33	3.33	94.43	5.90	94.83	5.50
4	99.00	1.00	98.00	1.90	98.50	1.67
5	91.33	6.00	90.21	3.35	90.83	3.00
6	96.50	2.83	96.47	3.97	97.00	3.67
7	85.83	16.50	90.41	8.07	90.83	7.67
8	92.50	5.33	94.82	5.77	95.33	4.67
Mean	94.38	4.81	94.64	4.22	95.17	3.73

Table. 2 Accuracy and error rates of existing and proposed feature vectors Comparison analysis table according to speed

Speed (sec/gesture, frame)	Jia-Sheng Feature		Proposed Feature			
	Accuracy (%)	Error (%)	Accuracy (%)	Error (%)	Accuracy (%)	Error (%)
0.66 (20)	90.25	8.92	90.92	5.94	92.83	5.83
1.00 (30)	94.5	4.08	94.32	3.55	94.92	3.42
1.33 (40)	95.25	3.83	95.96	4.59	95.17	4.17
2.00 (60)	97.5	2.42	97.36	2.8	97.75	1.5
Mean	94.38	4.81	94.64	4.22	95.17	3.73

제스처 종류나 제스처 동작 속도에 따라 차이가 일부 존재하지만, 평균적으로 제안한 알고리즘의 성능이 우수함을 알 수 있다. 기존 특징은 비교적 높은 정확도(94.38%)를 지녔지만, 또한 높은 오분류율(4.81%)을 보였다. 제안한 알고리즘은 광역 및 지역적 특징벡터의 추가와 더불어 CART의 가지치기 피드백을 통한 불필요 특징벡터를 제거함으로써 기존 특징 보다 낮은 오분류율(3.73%)과 보다 높은 정확도(95.17%)를 얻을 수 있었다.

V. 결론

본 논문에서는 모노 카메라를 이용한 AdaBoost 기반 제스처 인식에 관한 알고리즘을 제안하였다. AdaBoost 학습 모델에 신체 정규화된 시간 간격 특징 벡터의 집합을 특징 패턴으로 하여, 속도가 다른 동작들을 인식할 수 있도록 하였다. CART 알고리즘을 이용하여 중요 특징 벡터를 확인하고 중요도가 낮은 특징벡터를 제거하는 방식을 적용하면서 최적의 특징 벡터를 탐색하였다. 그 결과 최적의 24개의 특징벡터를 찾았으며, 기존 알고리즘에 비해 낮은 오분류율(3.73%)과 높은 인식률(95.17%)은 지닌 특징벡터 및 분류기를 설계하였다. 또한 시간 간격 윈도우를 적용하여 실제 환경에서 테스트한 결과 느린 속도 동작에서는 97.5%의 인식률과 2.42%의 오분류율의 결과를 보였다. 20~30 frame의 빠른 속도에서는 92.4%의 평균 보다 2.7% 낮은 인식률을 보였다. 이는 Camshift 기반 신체 모델 추정에서 추적 알고리즘 사용에 따른 인식률 저하로 보인다. 하지만 시간 간격 윈도우를 사용하여 서로 다른 속도에서도 제스처 인식이 90%이상으로 원활하게 이루어지는 결과를 보였다.

최근의 연구에서는 딥러닝이나 깊이 센서를 사용하는 연구가 주를 이룬다. 하지만 스마트 TV를 위한 상용 세톱박스의 성능은 최근의 연구 결과를 수행 할 만큼 컴퓨팅 파워가 충분하지 않다. 본 연구에서는 모노 카메라 웹캠 기반으로 신체를 인식하고 제스처 인식을 수행하여 저비용의 하드웨어 환경에서도 가능하게 하였다.

ACKNOWLEDGEMENT

The research was supported by 'Software Convergence Technology Development Program' (the Ministry of Science, ICT and Future Planning, No. 1711039588) and 'GRRRC program of Gyeonggi province' (GRRRC Aviation 2017-B04, Development of Interactive VR Player and Service with Space Media Convergence)

REFERENCES

- [1] L. Chen, H. Wei, and J. Ferryman. "A survey of human motion analysis using depth imagery," *Pattern Recognition Letters*, vol. 34, no. 15, pp. 1995-2006, Nov. 2013.
- [2] K. Lee, Y. Shin, Y. Lee, and S. Seol, "A Study on User Interface and Control Method of Web-based Remote Control Platform," *Asia-pacific Journal of Multimedia Services Convergent with Art, Humanities, and Sociology*, ISSN:2383-5287, vol. 7, no.6, pp. 827-837, June 2017.
- [3] K. J. Lee, "A Study on Gesture Recognition using Edge Orientation Histogram and HMM," *Journal of the Korea Institute of Information and Communication Engineering*, vol. 15, no. 12, pp. 2647-2654, Dec. 2011.
- [4] H. Duan and Y. Luo. "A Gestures Trajectory Recognition Method Based on DTW," in *Proceedings of the 2nd International Conference on Computer Science and Electronics Engineering*, pp. 364-366, 2013.
- [5] S. J. Hwang, et al. "Human Body Tracking and Pose Estimation Using Modified Camshift Algorithm," *Journal of Software Engineering and Applications*, vol. 6. no. 5B, pp. 37-42, May 2013.
- [6] S, Jia. "A study of adaboost in 3D gesture recognition," Department of Computer Science, University of Toronto, Technical Report, 2003.
- [7] Patsadu, Orasa, C. Nukoolkit, and B. Watanapa. "Human gesture recognition using Kinect camera," *Computer Science and Software Engineering, 2012 International Joint Conference on*. IEEE, Bangkok, Thailand, pp. 28-32, 2012.
- [8] Arici, Tarik, et al. "Robust gesture recognition using feature pre-processing and weighted dynamic time warping," *Multimedia Tools and Applications*, vol. 72, no. 3, pp 3045-3062, Oct. 2013
- [9] Takimoto, Hironori, J. Lee, and A. Kanagawa. "A robust gesture recognition using depth data," *International Journal of Machine Learning and Computing*, vol. 3, no. 2, pp 245-249, Apr. 2013.
- [10] Freund, Yoav, R. Schapire, and N. Abe. "A short introduction to boosting," *Journal-Japanese Society For Artificial Intelligence* vol. 14, no. 5, pp 771-780, Sep. 1999.
- [11] J. Zhu, S. Rosset, H. Zou and T. Hastie. "Multi-class adaboost," Technical Report 430, Department of Statistics, University of Michigan, 2009.
- [12] Hoffman, Michael, P. Varcholik, and Joseph J. LaViola. "Breaking the status quo: Improving 3d gesture recognition with spatially convenient input devices," *Virtual Reality Conference*, Waltham, MA, USA, pp. 59-66, Mar. 2010
- [13] S. J. Hwang, G. P. Ahn, S. J. Park and J. H. Baek. "AdaBoost-Based Gesture Recognition Using Time Interval Trajectory Features," *Journal of Advanced Navigation Technology*, vol. 17, no. 2, pp 247-254, Apr. 2013.



황승준 (Seung-Jun Hwang)

2012년 2월 : 한국항공대학교 정보통신공학 (공학사)
2014년 2월 : 한국항공대학교 일반대학원 정보통신공학 (공학석사)
2014년 3월 ~ 현재 : 한국항공대학교 일반대학원 항공전자정보공학 (박사과정)
※ 관심분야 : 영상처리, 패턴인식



고하윤 (Ha-Yoon Ko)

2017년 2월 : 한국항공대학교 정보통신공학 (공학사)
2017년 3월 ~ 현재 : 한국항공대학교 일반대학원 항공전자정보공학 (석사과정)
※ 관심분야 : 영상처리, 패턴인식



백중환 (Joong-Hwan Baek)

1981년 2월 : 한국항공대학교 항공통신공학 (공학사)
1987년 7월 : 오클라호마주립대학원 전기 및 컴퓨터공학 (공학석사)
1991년 7월 : 오클라호마주립대학원 전기 및 컴퓨터공학 (공학박사)
1992년 ~ 현재 : 한국항공대학교 항공전자정보공학과 교수
※ 관심분야 : 영상처리, 패턴인식, 영상압축, 멀티미디어