

## 다중 특징을 이용한 견고한 손추적 및 인식 시스템

천성용<sup>1</sup>, 박신원<sup>1</sup>, 장호진<sup>1</sup>, 이찬수<sup>1</sup>, 손명규<sup>2</sup>, 이상현<sup>2</sup>

<sup>1</sup>영남대학교 전자공학과, <sup>2</sup>대구경북과학기술원

{whiteyongi,psw0083,hojin7942@ynu.ac.kr}@ynu.ac.kr, {smk,pobbylee}@dgist.ac.kr

### Robust Hand Tracking and Recognition System Using Multiple Feature Data Fusion

Sung Yong Chun<sup>1</sup>, Shin Won Park<sup>1</sup>, Ho Jin Jang<sup>1</sup>, Chan-Su Lee<sup>1</sup>,  
MyoungGyu Sohn<sup>2</sup>, Sang-Heon Lee<sup>2</sup>

<sup>1</sup>Electronics Engineering Yeungnam University

<sup>2</sup>Daegu Gyeongbuk Institute of Science and Technology

#### 요 약

본 연구에서는 효과적인 손 제스처 인식을 위하여 다중 특징을 이용한 견고한 손 추적 방법을 제시한다. 기존의 많은 손추적 장치들이 칼라 정보나 모션 정보와 같은 단일한 정보를 바탕으로 손을 검출하고, 이를 바탕으로 손의 추적하는 방법들을 제시하고 있다. 이러한 방법들의 경우에는 손 추적 중에 환경이나 상황이 변하게 되면, 손추적의 정확도가 현저하게 떨어지게 된다. 본 연구에서는 이러한 문제점들을 보완하기 위하여, Adaboost를 이용한 손 검출, 역투영을 기반으로 손 색상을 이용한 추적, KLT를 바탕으로 한 모션 추적을 이용한 검출을 동시에 수행하며, 각 센서의 추적 결과에 대한 칼만 필터 적용 뿐 아니라, 각 센서 정보를 통합하여 견고한 결과를 얻기 위한 방법을 제시한다. 이를 바탕으로 손제스처 인식 시스템을 개발하였으며, 개발된 제스처 인식을 바탕으로 비디오 플레이를 제어하는 시스템을 구현하였다.

Keywords : Data Fusion, Gesture recognition, Adaboost, KLT, Back Projection

## 1. 서론

최근 컴퓨터의 연산 처리 능력이 좋아지면서 컴퓨터 비전 기술을 이용하여 사람을 찾아내고, 그 사람의 행동이 어떤 의미를 갖는지 분석하는 제스처 인식 연구가 활발하게 이루어지고 있다. 특히 Microsoft 사의 Natal 과 같이 게임에 모션인식 기술이 접목되면서 제스처 인식은 많은 사람들의 주목을 받게 되었다. 게임뿐만 아니라 Smart Home 환경이나 IPTV 를 이용한 서비스가 새로운 가치를 창출하는 비중이 점점 커지면서, 기존의 리모컨이 아닌 사람과 조금 더 능동적인 상호작용을 할 수 있는 제스처 인식 기술에 대한 연구가 활발히 진행되고 있다[1][2].

과거부터 손제스처 인식에 대한 다양한 연구들이 많이 진행되었으나, 최근에 이르러서야 프로세서 성능 등이 크게 향상됨에 따라, 컴퓨터 비전 기술을 이용한 손 제스처 인식과 이를 이용한 자연스러운 상호작용 기술은 다양한 응용분야와 함께 주목을 받고 있다. 하지만, 기존 시스템의 경우에는 매우 제한된 환경에서만 동작하고 있으며, 사용자에게도 매우 제한된 동작만을 허용해 주고 있다. 특히, 손 추적의 경우에는 제한된 특징점이나 색상, 모양 등 하나의 방법에 의존하여 추적하려고 하기 때문에 많은 문제점을 발생하고 있으며, 추적 중에 손의 추적에 실패하는

경우가 자주 발생함으로, 손 제스처를 인식하는데 문제가 있다.

따라서 본 연구에서는 기존의 손 검출 시스템의 제한점을 극복하고, 다양한 요소들을 고려함으로 효과적으로 손 검출과 추적을 할 수 있는 방법들을 위해 색상, 움직임, 모양을 모두 고려한 데이터 통합적 방법을 제안한다. 손 제스처 인식을 위해서 손 검출 단계, 다중 특징을 이용한 손 추적 단계, 손 제스처 인식 단계로 나누어 제스처를 인식을 이용한 사용자 인터페이스를 개발하였다. 손 검출 단계에서는 두 가지 손 모양에 대하여 Adaboost 를 이용한 검출 기법을 사용하였으며[3], 검출 효율을 높이기 위하여 히스토그램 정규화를 수행하였다. 다중 특징을 이용한 손 추적 단계에서는 Adaboost 를 이용한 손 검출, 역투영을 기반으로 손 색상을 이용한 추적[4], KLT 추적기를 이용한 검출 결과를 이용하여, 이를 통합한 추적 결과를 결합함으로 추적의 정확도를 향상시키며, 일부의 추적기가 추적을 할 수 없는 경우에 대해서도 효과적인 추적이 가능하다. 본 추적기를 바탕으로, 주어진 2 차원 위치 운동을 바탕으로 동작 상태 인식과 특징점 추출을 통하여 손 제스처를 인식할 수 있는 시스템을 개발하였으며, 10 가지 제스처에 대한 인식률은 80%에 달한다. 인식된 제스처는 IPTV 를 위한 미디어 플레이어를 제어하도록 하였다.

## 2. 손 검출 및 추적 기법

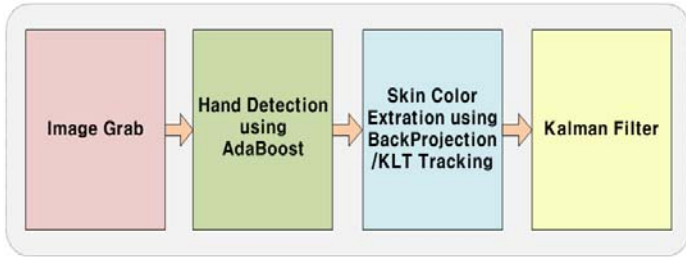


그림 1. System Overview

손 검출 및 추적 방법에 대한 전체적인 구조는 그림 1 에서 보여주고 있다. 카메라로부터 입력 받은 영상을 이용하여 손을 검출한다. 검출된 손으로부터 피부색을 추출하고, 움직임을 추적을 하게 된다.

### 2.1. Adaboost를 이용한 손 검출 및 추적

본 연구에서는 효과적인 손 검출을 위해, Viola & Jones 이 제안한 cascade Adaboost 기법을 이용하여 손 모양을 검출하도록 하였다. Adaboost의 경우에 처음에 효과적인 얼굴 검출을 위하여 적용되었으며, 본 연구에서는 고정된 손 모양을 검출하기 위하여 Adaboost를 사용하였다.

Adaboost는 높지 않은 분류 성능을 보이는 간단한 형태의 약한 분류기들을 선형적으로 결합하여 강한 분류기를 만든다. Adaboost에서는 Haar 특징값이 사용되었다. Haar 웨이블릿은 1910년에 Haar가 제안한 간단한 직교 웨이블릿이고, 영상에서 수평, 수직, 대각 방향의 에지를 각 스케일로 분해하는 것이다. 각 영역 안에 위치한 픽셀들의 값을 더하여 영역의 합을 구한 뒤 그 값들에 가중치를 곱한 값의 합을 구하여 특징 값을 계산한다. 이러한 특징 값을 바탕으로 설정된 임계값 (threshold value)과 비교하여 간단하게 원하는 데이터가 있는지를 분류할 수 있다. 다음의 식은 약한 분류기를 나타낸다.

$$f_i = \begin{cases} +1, & v_i \geq t_i \\ -1, & v_i \leq t_i \end{cases}$$

$f_i$  : weak classifier,  $t_i$  : threshold,  $v_i$  : feature value

약한 분류기는 윈도우 내의 특징 점들 중 대상과 배경을 가장 잘 분류하는 하나의 특징을 선별한다. 이를 위해 적분이미지(integral image)를 통해 구한 픽셀 값의 합을 이용하여 에러가 최소화되는 문턱 값을 찾는다. 이렇게 구한 약한 분류기들을 선형적으로 결합하여 최종적인 강한 분류기를 구성한다. 다음 식은 강한 분류기를 나타낸다.

$$F = \text{sign}(w_1f_1 + w_2f_2 + \dots + w_nf_n)$$

이러한 Haar 특징값을 이용한 Adaboost의 경우에 정적이며 변화가 없는 물체에 대한 검출에는 매우 효과적이지만, 손과 같이 그 모양이 변하고 시점이 변함에 따라 다른 모양을 나타내는 객체를 검출하는 것은 매우 어려운 문제이다. 이러한 문제를 해결하기 위하여, 본 연구에서는 손 검출에 있어서는 고정된 손 모양을 종류별로 학습하도록 하였다. 또한 기존의 시스템의 경우에 작은 회전에도 검출률이 많이 떨어지며, 이를 보완하기 위하여, 기존의 Haar 특징 뿐 아니라, Lienhart 등이 제안한 [5] 확장된 Haar 특징이 사용되었다. 그림 2는 사용된 Haar-like 특징값을 보여주며 검은색과 흰색의 직사각형 형태로 이루어져 있으며 서로 인접하여 있다.

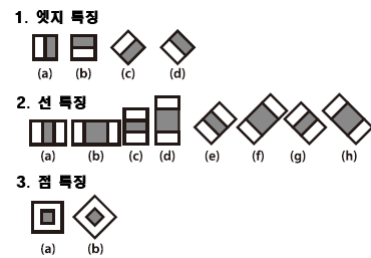


그림 2. 사용된 Haar-like Features

Adaboost는 하나의 고정된 손 모양에 대해서만 인식이 가능하기 때문에 본 연구에서는 그림 3과 같이 두 개의 손 모양에 대하여 개별적으로 학습하여 검출하도록 하였다. 그리고 각 손 모양에 대해서 Positive 이미지가 손바닥 3299개와 주먹 2670개가 각각 사용되었으며 Negative 이미지는 1942개를 공통으로 사용하여 학습을 하였다. 그림 4는 학습에 사용된 Positive 이미지의 예를 나타낸다.



그림 3. 사용한 두 가지 손 모양

Adaboost의 손 검출 기능을 이용하여 연속된 프레임에서 손 검출을 수행하여 손 추적에 사용할 수 있다. 그림 5는 다른 손모양에 대하여 검출을 이용한 추적의 예를 보여준다. 본 연구에서는 Adaboost만을 이용한 시스템에서 손추적에 대한 성능을 평가하기 위하여 거리가 최대 1.2 미터 떨어진 지점에서 해상도 320 x 240 에서 손검출 성능을 평가하였다. 사용한 Adaboost의 성능을 평가하기 위해 3800장의 테스트 이미지에 적용한 결과 약 3100장에서 손이 검출되었고, 81%의 검출 성능을

보여주었다. 하지만 Adaboost를 이용한 경우에는 원거리에 있는 경우나 손 모양이나 시점이 달라지는 경우에 검출율이 현저하게 떨어져서 Adaboost의 손 검출만으로는 손모양 추적을 통한 제스처 인식에 사용할 수 있는 견고한 손추적을 얻을 수 없다. 따라서 본 연구에서는 피부색등의 다른 추적방법을 병행하여 적용한다.



그림 4. 학습한 Positive 이미지의 일부

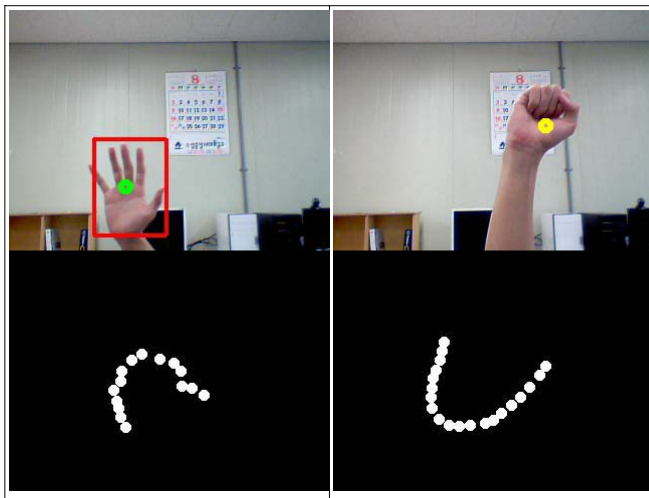


그림 5. Ada-Boost에 대한 움직임 추적

## 2.2. 피부색 검출 및 이를 이용한 손 운동 추적

피부색 검출을 위해 본 연구에서는 역투영 기법(back projection)을 사용하였다. 역투영은 특정 픽셀들 또는 픽셀들의 패치(patch)들이 히스토그램 모델로 표현된 픽셀들의 분포에 얼마나 잘 들어맞는지를 검사하는 방법이다. 역투영 영상의 픽셀 값은 피부색 히스토그램을 이용하여 영상에서 피부색일 확률이 높은 곳을 나타낸다.

처음에는 Adaboost에서 검출된 손 영역에서 피부색

의 영역을 추출하게 된다. 이 추출한 영역을 HSV색 공간에서 2차원 H-S 히스토그램으로 변환하여 역투영 기법을 적용하고, 이를 이용하여 손의 피부색을 구하게 된다. 만약 Adaboost가 현재 프레임에서 손을 검출하지 못 하더라도 역투영된 피부색을 사용하여 손 영역을 검출할 수 있게 되며, 이 영역의 무게 중심을 구하여 피부색을 이용한 손 추적을 하게 된다.



그림 6. 역투영을 이용한 손 영역 추출

## 2.3. KLT를 이용한 손 운동 추적

움직임을 이용한 손 운동의 추적을 위해 옵티컬 플로우에 기반한 Kanade - Lucas - Tomasi Feature Tracker (KLT) 추적을 이용하였다. 추적을 위한 다양한 지역 특징들이 존재하며, 효과적인 추적을 위해 어떠한 특징을 사용할 것인지를 결정하는 일은 매우 중요하다. 서로 직교하는 방향으로 미분값이 크게 나타나는 점들이 추적하기에 적합하며, 이러한 점들을 코너(conner)라고 부른다. 우리는 Shi와 Tomasi가 제안한 코너 검출[6]을 사용하여 코너를 추출하였다.

옵티컬 플로우는 이전 영상과 현재 영상간의 특징점들이 이동한 양을 계산해 준다. 특히 Lucas-Kanade Pyramid는 원본 영상으로부터 영상 피라미드를 구성하고, 상위 계층에서 시작하여 점점 하위 계층으로 추적을 수행한다. 영상 피라미드를 사용함으로써 커다란 움직임도 찾아낼 수 있다. 코너 검출을 이용한 좋은 특징점을 옵티컬 플로우에 사용하여 KLT를 구성하였다.

KLT는 이전 영상과 현재 영상에서 특징 점들의 이동한 변위에 의존한다. 그러므로 배경이 변하는 경우, 물체가 지나가는 경우에도 옵티컬 플로우가 발생하며, 이러한 손 이외의 영역의 플로우를 제거하지 않으면 올바른 결과를 얻을 수 없다. 이러한 문제를 해결하기 위하

여 KLT 추적기에 역투영 결과를 마스크로 갖도록 하였다. 그림 7은 마스크가 없는 경우에 배경 및 주변 환경의 움직임이 검출된 옵티컬 플로우를 보여준다. 이 경우에 옵티컬 플로우의 평균을 구한다면, 손의 움직임과는 무관한 다른 움직임으로 인한 영향을 받게 된다. 반면에, 역투영 결과를 바탕으로 한 마스크를 이용하여 그림 8과 같이 배경의 영향을 받지 않는 추적 결과를 얻을 수 있다.

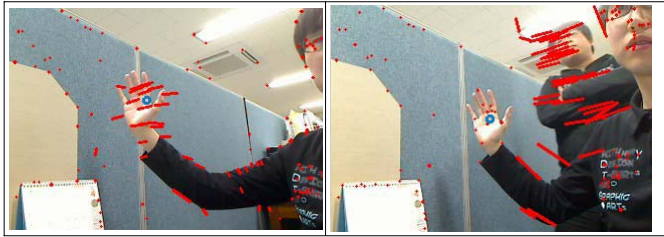


그림 7 마스크가 없는 경우의 KLT 결과



그림 8. KTL 추적 결과들

### 3. 노이즈 제거 및 추적 데이터 통합(fusion)

#### 3.1. Kalman 필터를 이용한 노이즈 제거

Adaboost를 이용한 손추적, 피부색을 이용한 손추적 및 KLT를 이용한 손 추적의 경우에 각 추적값에는 노이즈 성분이 많이 존재하게 된다. 따라서 부드러운 움직임에도 불구하고 검출된 데이터 값들은 부드러운 움직임을 만들어내지 못한다. 따라서 노이즈 성분을 제거하고 더 부드러운 추적 결과를 만들기 위해 Kalman 필터를 이용하였다.

Kalman 필터는 선형 동적 상태 모델과 측정값에 있어서의 불확실성이 존재할 때, 이 불확실성을 가우시안 노이즈로 모델링하고, 예측값과 측정값을 바탕으로 최적의 값을 계산할 수 있다. 모델 불확실성을 가지는 이전 값과 불확실성을 가지는 새로운 측정을 모두 고려하여, 이전의 정보와 새로운 정보의 확실성에 근거하여 각각의 정보에 가중치를 부여함으로써 최적의 값을 얻을 수 있다.

다음은 세 가지 검출 방식에 대하여 원래의 측정값과 Kalman 필터를 적용하여 얻은 새로운 값들을 보여준다. 그림9와 같이 Kalman 필터를 적용하면 움직임에 노이즈 성분이 많이 제거되고 매끄러워진 것을 알 수 있다.

각각의 그림에서 붉은색은 Kalman 필터를 적용하기 전 추적 결과이고, 푸른색은 Kalman 필터를 적용한 추적 결과이다. 그림의 x축과 y축은 이미지 좌표 공간에서 손이 움직인 좌표에 해당한다.

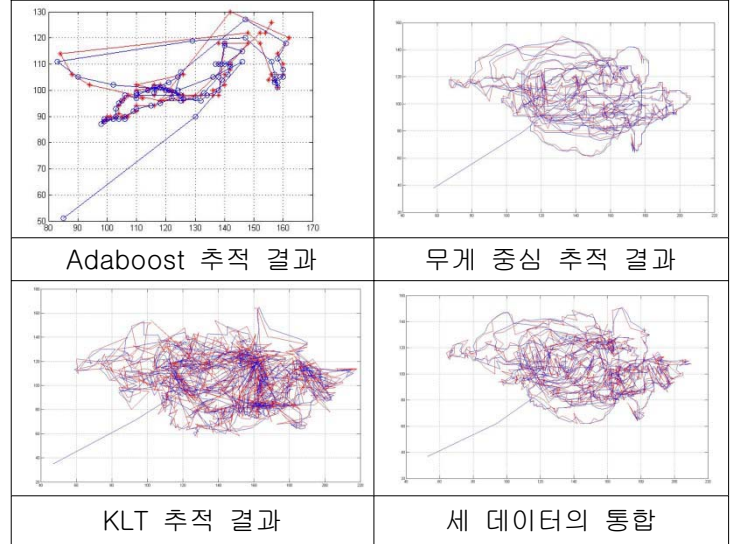


그림 9. Kalman 필터 적용 예

#### 3.2. 가중치 기반 데이터 통합(fusion)

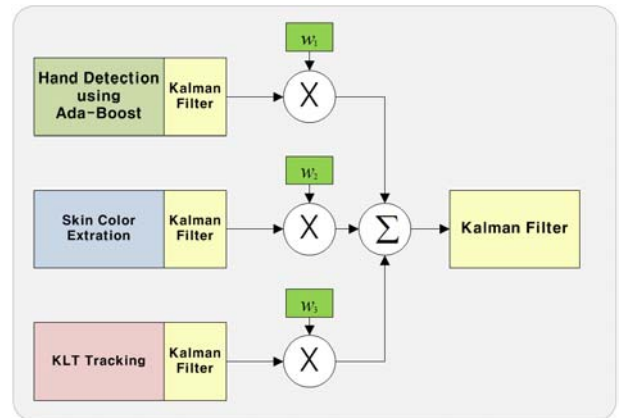


그림 10 세가지 데이터의 통합

세 가지 추적 데이터들은 추적에 대한 정확도가 다르며 때로는 세 가지 추적 데이터가 모두 존재하지 않고 일부의 데이터만이 존재하게 된다. 따라서 견고한 추적 결과를 얻기 위해서는 이러한 다른 측정데이터에 대한 효과적인 통합이 필요하다. 본 연구에서는 세 가지 추적 데이터 각각에 대한 상대적인 정확도를 바탕으로 서로 다른 고정 가중치를 주고 움직임을 보상하게 만들었다. 또한 추적 데이터를 얻을 수 없는 손실 부분은 유동적인 가중치 변화로 서로 보상할 수 있게 만들었다. 그리고 가중치를 적용시킨 세 추적 데이터를 통합하여 Kalman 필터를 적용하였다. 그림10은 이러한 데이터 통합과정을 보여준다. 그림 9에서 Adaboost의

경우에는 동일한 데이터임에도 일부영역에서 추적이 일어나지 않고 있으나, 통합된 추적 결과를 보여주는 그림 11에서는, 데이터가 끊어지지 않고 부드러운 손 움직임 추적 결과를 보여준다. 또한 가중치 값에 다시 Kalman 필터를 적용한 푸른색 선이 조금 더 매끄러운 움직임을 보여주었다.

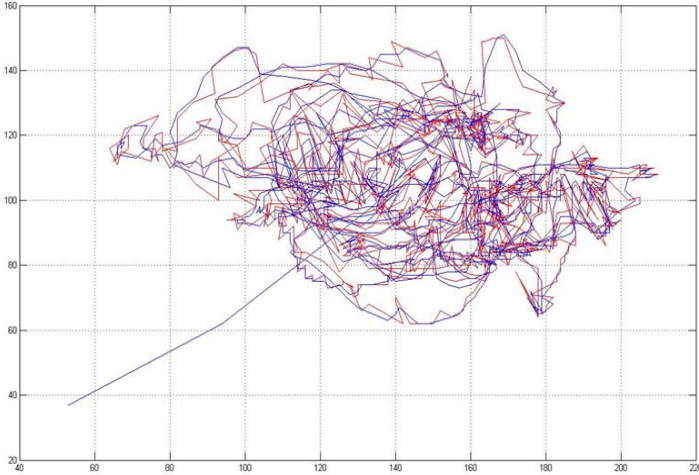


그림 11. 세 가지 검출 데이터에 대한 통합 결과

## 4. 제스처 인식 시스템

### 4.1. 제스처 인식을 위해 사용된 특징들

추적된 손 모양을 바탕으로, 손의 움직임을 파악하여, 손 제스처를 인식하도록 한다. 손 제스처 인식을 위하여 다음과 같은 특징들이 사용되었다.

현재 프레임  $t$ 와 이전 프레임  $(t-1)$ , 그리고 초기 프레임  $t_0$ , 현재좌표  $(R_x(t), R_y(t))$ 로 정의했을 때, 특징을 추출하기 위한 식들은 다음과 같다.

Cumulative length of motion:

$$TL_x(t) = TL_x(t-1) + |R_x(t) - R_x(t-1)| \quad (1)$$

$$TL_y(t) = TL_y(t-1) + |R_y(t) - R_y(t-1)| \quad (2)$$

Cumulative orientation change:

$$TC_{\theta}(t) = TC_{\theta}(t-1) + \left| \tan^{-1} \left( \frac{R_y(t) - R_y(t-1)}{R_x(t) - R_x(t-1)} \right) - \tan^{-1} \left( \frac{R_y(t-1) - R_y(t-2)}{R_x(t-1) - R_x(t-2)} \right) \right| \quad (3)$$

Relative position:

$$EP_x = R_x(T) - R_x(0) \quad (4)$$

$$EP_y = R_y(T) - R_y(0) \quad (5)$$

Relative orientation:

$$EA_x = \tan^{-1} \left( \frac{R_y(t) - R_y(0)}{R_x(t) - R_x(0)} \right) \quad (6)$$

이동한 거리의 누적 값인 식(1),(2)로 각 제스처에서 손이 움직일 수 있는 거리를 구하였고, 식(3)에 의한 이동 거리 동안 누적된 각도의 값, 식(4)와 식(5)에 의한 제스처의 시작 위치를 기준으로 한 제스처의 종료 위치, 식(6)에 의한 제스처의 시작 위치를 중심으로 한 제스처 끝 위치의 각도를 구하였다. 이렇게 구한 값들은 모두  $[0, 1]$ 로 정규화 하여 인식기에서 사용되었다.

### 4.2. 제스처 인식을 위한 상태 다이어그램

제스처 인식을 위해 스테이트 머신을 만들어서 의미 있는 움직임과 의미 없는 움직임을 구별하고, 의도한 움직임이라면 어떤 움직임인지 판단했다.

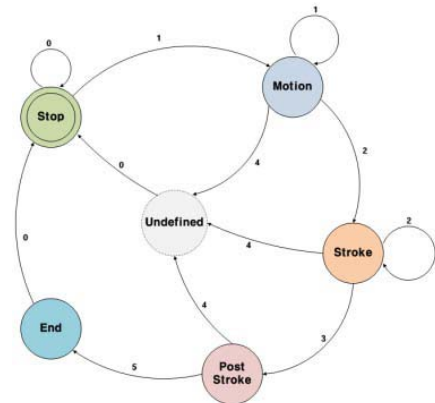


그림 12. 의미 있는 움직임과 의미 없는 움직임을 판단하는 스테이트 머신

움직임의 속도와 가속도를 계산하고, 동작의 움직임에 따라 Stop에서 Motion 상태로 이동하며, 속도와 가속도가 일정 이상이 되면 스테이트 머신의 상태가 Stroke로 바뀌어 의미있는 제스처로 인식하게 된다. 또한 Motion 상태부터 인식에 위한 움직임에 대한 특징점을 추출한다. 움직임 도중 속도가 감소하게 되면 움직임이 끝나가는 것으로 판단하고 스테이트 머신이 PostStroke로 변경되고, 일정 이하로 내려가면 End 스테이트로 바뀐다. 스테이트 머신이 End 스테이트가 되면 계산한 특징점들을 Support Vector Machin(SVM)에 입력하여 어떤 제스처인지 인식하게 된다.

## 5. 실험 결과

본 연구의 실험을 위해서 제스처 기반의 Media Player를 개발하여 테스트 하였다. Media Player는 인식된 제스처를 통해 제어가 가능하며, 맥내 IP-TV를 제어하는 기능과 비슷한 환경을 구현하도록 구성하였다. 그림 13과 같은 Media Player를 이용하여 동영상을 재생, 정지, 전체화면, 다음 영상, 이전 영상 등 총 10가지의 기능에 각각 제스처를 연결하여 테스트를 하였다. 여기에 사용된 제스처는 그림 14와 같다. 이 10개의 제스처

에 대해서 50개의 데이터를 학습하고, 15개의 테스트 데이터를 인식 시켰을 때, 80%의 인식률을 나타내었다. 보다 많은 데이터에 대한 학습을 통하여 인식율을 높일 수 있을 것이다.



그림 13. 테스트를 위해 개발한 Media Player

이 실험에서 사용한 영상의 크기는 320×240이며, Intel Quad 2.5Ghz의 CPU의 컴퓨터에서 15FPS 정도의 처리속도를 보여주었다.

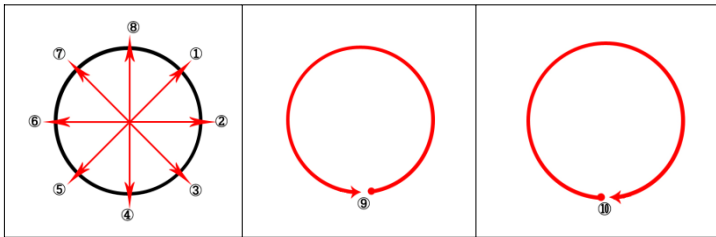


그림 14. 사용된 제스처의 종류



그림 15. 제스처 동작의 예

## 6. 결론

본 연구는 강인한 손추적을 바탕으로 한 손 제스처 인식과 이를 기반으로 한 제어 기술을 개발한 것으로 IPTV 등의 인터페이스에 적용될 수 있다. 기존의 방법 처럼 모양, 색 등 하나의 방법으로는 강인한 검출 및 추적이 어렵기 때문에 여러가지 특징을 사용하여 제스처를 추적하고 이를 통합하여 보다 견고한 제스처 추적을 얻는 방법을 제안했다.

제안한 알고리즘은 실험 결과에서 보였듯이 강인한 손 검출 및 추적 결과를 보였고, 처리시간도 15FPS 정도로 실시간 처리가 가능함을 보여주었다. 하지만 인식률이 80% 정도로 제스처 인식이 실패하는 경우가 있다. 따라서 향후 과제로 제스처 인식의 신뢰도를 향상시키기 위한 인식기 알고리즘을 개선하는 연구를 해야할 필요성이 있다. 또한 추적데이터의 통합에 있어서, 사용 가능한 추적 데이터 및 데이터의 신뢰도를 바탕으로 한 고정된 가중치 값이 아니라, 각 추적기의 Kalman에서 추적된 데이터의 정확도를 바탕으로 한 변동 가중치의 적용이 가능하며 이러한 방법에 대한 추가적인 연구가 필요하다.

## 7. 참고문헌

- [1] 김대환, 조상호, 천영재, 김대진, "스마트 홈에서의 TV 제어 시스템을 위한 손제스처 인식 방법", 한국정보과학회 가을 학술 발표 논문집, vol. 34, no. 2, 2007
- [2] 노명철, 장혜민, 강승연, 이승환, "휴먼-로봇 상호작용을 위한 비전 기반 3차원 손 제스처 인식", 한국정보과학회 가을 학술대회 발표 논문집, vol. 33, no. 2, 2006
- [3] Paul Viola, Michael Jones, "Rapid Object Detection using a Boosted Cascade of Simple Features", IEEE CVPR 2001, pp. 511-518, 2001
- [4] N. Tanibata, N. Simada, and Y. Shirai, "Extraction of hand features for recognition of sign language words" in Proc. of Int. Conf. on Vision Interface, pp. 391-398, 2002
- [5] R. Lienhart and J. Maydt, "An extended set of Harr-like features for rapid object detection", IEEE ICIP 2002, Vol. 1, pp. 900-903, Sep. 2002
- [6] Shi and Tomasi, "Good Features to Track", CVPR, pp.593-600, 1994
- [7] Chan-Su Lee, SangWon Ghyme, ChanJong Park, KwangYun Wohn, "The Control of Avatar Motion Using Hand Gesture," VRST'98, pp. 59-65