http://dx.doi.org/10.3745/KTSDE.2013.2.10.713

Real-time Hand Region Detection based on Cascade using Depth Information

Joo Sung Il† · Weon Sun Hee†† · Choi Hyung Il†††

ABSTRACT

This paper proposes a method of using depth information to detect the hand region in real-time based on the cascade method. In order to ensure stable and speedy detection of the hand region even under conditions of lighting changes in the test environment, this study uses only features based on depth information, and proposes a method of detecting the hand region by means of a classifier that uses boosting and cascading methods. First, in order to extract features using only depth information, we calculate the difference between the depth value at the center of the input image and the average of depth value within the segmented block, and to ensure that hand regions of all sizes will be detected, we use the central depth value and the second order linear model to predict the size of the hand region. The cascade method is applied to implement training and recognition by extracting features from the hand region. The classifier proposed in this paper maintains accuracy and enhances speed by composing each stage into a single weak classifier and obtaining the threshold value that satisfies the detection rate while exhibiting the lowest error rate to perform over-fitting training. The trained classifier is used to classify the hand region, and detects the final hand region in the final merger stage. Lastly, to verify performance, we perform quantitative and qualitative comparative analyses with various conventional AdaBoost algorithms to confirm the efficiency of the hand region detection algorithm proposed in this paper.

Keywords : Hand Region Detection, Depth Image, Kinect, Adaboost, Depth Feature

깊이정보를 이용한 케스케이드 방식의 실시간 손 영역 검출

주 성 일† · 원 선 희†† · 최 형 일†††

요 약

본 논문에서는 깊이정보를 이용하여 케스케이드 방식에 기반한 실시간 손 영역 검출 방법을 제안한다. 실험 환경 조명 조건의 변화로부터 빠르고 안정적으로 손 영역을 검출하기 위해 깊이정보만을 이용한 특징을 제안하며, 부스팅과 케스케이드 방법을 이용한 분류기를 통해 손 영역 검출 방법을 제안한다. 먼저, 깊이정보만을 이용한 특징을 추출하기 위해 입력영상의 중심 깊이 값과 분할된 블록의 평균 깊이 값의 차이를 계산하고, 모든 크기의 손 영역 검출을 위해 중심 깊이 값과 2차 선형 모델을 이용하여 손 영역의 크기를 예측한다. 그리고 손 영역으로부터의 특징 추출을 통한 학습 및 인식을 위해 케스케이드 방식을 적용한다. 본 논문에서 제안한 분류기는 정확도를 유지하고 속도를 향상시키기 위하여 각 스테이지를 한 개의 약분류기로 구성하고 검출율을 만족하면서 오류율이 가장 낮은 임계값을 구하여 과적합 학습을 수행한다. 학습된 분류기를 이용하여 손 영역을 분류하고, 병합단계를 통해 최종 손 영역을 검출한다. 마지막으로 성능 검증을 위해 기존의 다양한 아다부스트와 정량적, 정성적 비교 분석을 통해 제안하는 손 영역 검출 알고리즘의 효율성을 입증한다.

키워드 : 손 영역 검출, 깊이 영상, 키넥트, 아다부스트, 깊이 영상 특징

1. 서 론

최근 스마트 기기의 급속한 발전과 보급화로 인해 기기 조작을 위한 인터페이스에 대해 사용자들의 관심이 날로 급증하고 있다. 이러한 대중들의 심리를 반영하기 위해 가장 집중적으로 연구되고 있는 분야가 바로 지능형 사용자 인터페이스 분야이다. 사용자 인터페이스에 대한 관심과 연구는 이미 수십년 전부터 행해져 왔으나 근래에 들어 이러한 스마트 기기 시장이 활발해지면서 그 기술적인 요구가 증가하게 되었다. 지능형 인터페이스 기술 중에서도 사용자들의 편의성과 직관성을 최대로 반영할 수 있는 기술이 제스처 인터페이스 기술이다. 가장 대표적인 예로 마이크로소프트사에서 개발한 키넥트(Kinect) 센서는 RGB 카메라와 IR 카메라 센서를 결합하여 사용자들의 제스처 및 동작을 인식한 실시간 체감형 게임이 가능한 기술이다. 키넥트 센서의 저가형 하드

※ 이 논문은 서울시 산학연 협력사업(SS110013)의 지원을 받아 수행된 연구임.
※ 이 논문은 2013년도 정부(미래창조과학부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임(No. 2013R1A1A2012012).
† 준 회 원 : 숭실대학교 미디어학과 박사과정
†† 정 회 원 : 숭실대학교 미디어학과 Post Doc.
††† 종신회원 : 숭실대학교 미디어학과 교수
논문접수 : 2013년 5월 28일
심사완료 : 2013년 7월 6일
* Corresponding Author : Choi Hyung Il(hic@ssu.ac.kr)

웨어 공급과 공개된 라이브러리의 제공으로 인해 수많은 응용 가능한 제스처 인식 기술들이 개발 가능하게 되었다. 그러나 사실 이러한 제스처 인식은 컴퓨터 비전 분야에서 오랫동안 연구되어 온 분야이다. 크게 제스처 인식은 두 가지 종류로서 손을 검출하여 손 포즈와 같은 정적(static)인 제스처(posture)를 인식하는 기술과 손의 이동 궤적을 이용한 동적(dynamic)인 제스처(gesture)를 인식하는 기술들이 연구되었다. 하지만 이러한 제스처 인식 기술을 위해서는 영상으로부터 손 영역을 분할하고 검출하는 단계가 선행되어야한다. 이를 위한 연구 분야로는 크게 3가지로 구분할 수 있다. 먼저 색상정보를 이용한 연구[1,2]가 주를 이루고 있으며, 최근 들어 색상과 깊이정보를 결합한 연구와[3~5], 깊이정보만을 이용한 연구가[6~9] 활발히 진행되고 있다.

색상정보를 이용한 손 영역 검출방법은 손의 피부 색상정보를 활용하는 연구들이 주를 이룬다. Suk[1]은 Haar-like 얼굴 검출기를 이용하여 얼굴을 검출하고, 검출된 얼굴의 색상 모델과 YIQ 색상 모델을 or로 조합하여 손 영역을 검출한다. 검출된 영역은 가우스 함수를 이용하여 추적하고, 추적된 궤적을 통해 제스처를 인식하였다. 이 방법은 손 검출을 위해 얼굴을 검출해야하는 선행 조건이 필요하며, 또한 조명의 변화에 민감한 단점이 있다. Bhuyan[2]은 RGB 공간의 피부 색상 분포와 전경과 배경의 조건부 확률을 이용하여 손 영역을 검출하고, 검출한 손 영역으로부터 손의 중심점과 주방향성을 추출하여 손과 팔 영역을 분할하며, 손 영역에서 기하학적 특징들을 추출함으로서 손 끝점 검출을 수행한다. 이 방법은 기존에 공개된 피부 색상모델을 이용하여 손 영역을 검출하였으며 실험 환경이 매우 제한적이며 주변 객체와의 폐색에 민감하다는 문제점이 있다.

색상정보의 가장 큰 문제점인 환경변화에 취약한 점을 보완하기 위해 깊이정보와 결합하여 해결하는 연구들도 많이 이루어졌다. Park[3]은 손이 몸체보다 앞에 있다고 가정하고, 키넥트의 깊이 영상으로부터 누적 히스토그램을 구하여 손의 후보 영역을 찾고, 손의 후보 영역 중 정확한 손 영역을 찾기 위해 베이시안(Bayesian) 규칙과 피부 색상을 이용하여 최종 손 영역 검출한다. 이 방법은 색상정보만을 단독으로 이용한 경우보다 훨씬 좋은 성능을 나타내었고, 깊이정보만을 이용한 연구보다 성능이 높았으나, 손이 항상 몸 앞에 위치한다는 가정과 색상을 사용하기 때문에 어두운 곳에서는 불가능한 단점이 있다. Van den Bergh[4]는 RGB 영상에서 얼굴을 검출하고, 검출한 얼굴의 거리 값을 임계값으로 적용하여 배경을 제거한 뒤, 나머지 영역에서 피부색으로 손 영역을 검출하는 혼합된 형태를 제안하였다. 이 방법 또한 단독으로 색상영상이나 깊이 영상만을 이용하는 방법보다 높은 정확성을 보였으나, 깊이정보만을 단독으로 이용한 것이 아니기 때문에 연산량이 더 높으며, 조도가 낮은 곳에서는 실행 불가능한 단점이 있다. Trindade[5]는 RGB-D 센서로부터 RGB 색상에 의한 피부 색상 필터링을 먼저 수행하여 몸체 부분과 얼굴, 손 영역 부분을 검출하고, 깊이축(depth axis)에 따라 히스토그램을 배치한 후 임계값으로 필터링한다. 이후 k-평균 클러스터를 이용하여 이상치(outlier)를 제거함으로서 손 영역의 중심점을 추정하며, 추정된 손 중심점을 기반으로 최종 손 영역을 검출한 후 포즈 인식을 수행한다. 이 방법은 손 영역 검출을 위해 색상정보와 깊이정보를 융합한 형태로 필터링 과정의 이상치 제거와 클러스터링 기법을 적용함으로써 검출 단계의 정확성을 높였으나, 이 또한 색상정보를 이용하기 때문에 이전 연구들과 같은 조도의 변화에 따라 문제를 발생시킬 수 있으며, 손 영역을 검출하는데 많은 단계를 거치므로 오류가 발생할 확률이 높을 수 있다.

색상정보와 깊이정보를 결합한 연구들의 성능이 손 영역 검출성능을 향상시켰으나 여전히 색상정보에 의존하여 발생하는 많은 문제점들이 남아있다. 이를 해결하기 위해 색상정보를 배제하고 깊이정보만을 이용하는 연구들이 이루어지고 있다. Mo[6]는 저해상도에서 손 형상인식을 위해 손 모델을 정의하고, 레이저 기반 카메라를 통해 입력된 깊이정보를 이용하여 카메라로부터 가장 가까운 영역이 사용자의 손이라고 가정한 후, 손과 손목, 배경 부분을 분할한다. 이 방법은 손이 몸체보다 뒤쪽에 위치하거나 카메라와 손 사이에 다른 물체가 있는 경우 검출에 실패하게 된다는 문제점이 있다. Liu[7]는 카메라로부터 일정 거리 안에 있는 객체를 사람이라 가정하고, 그 객체만 가로와 세로로 프로젝션하여 얼굴을 검출한다. 손동작 인식을 위해서는 일반적으로 손은 몸체와 떨어져있다고 가정하고 팔 부분과 손 부분을 분리하기 위해 비율적인 상수를 사용하여 손 영역만 추출한다. 하지만 이 방법은 얼굴이 영상 안에 존재해야하며, 여러 사람이 있는 경우 복잡해지는 문제가 발생할 수 있다. Malassiotis[8]은 팔 영역을 분할하기 위해, 깊이 영상을 순차적으로 스캔하여 초기 클러스터링을 수행하고 사전에 분류된 픽셀들과 그의 이웃들로부터 거리에 따라 각 픽셀을 분류한다. 그리고 인접한 클러스터들을 계층적으로 병합하여 최종적인 팔 영역을 검출한다. 손과 팔뚝을 분할하기 위해 3차원 공간에서 팔의 좌표들을 통계적으로 모델링하고, 혼합 가우시안 모델을 이용하여 3차원 x 좌표의 확률 분포를 산출하고 이에 따라 손과 팔뚝 영역을 분할한다. 이 방법은 정적인 포즈 인식을 위한 논문이기 때문에 동적인 제스처 인식 분야에서는 깊이정보의 변화가 발생할 경우 분포 모델을 갱신해야한다는 문제점이 있다. Suryanarayan[9]은 깊이정보를 이용하여 손 포즈 인식을 위해 2차원 형태정보와 압축된 3차원 형태 디스크립터, 3차원 볼륨매트릭 형태 디스크립터를 제안한다. 이를 위해 먼저 손을 찾아야 하는데, 손을 찾는 방법은 깊이 값으로 히스토그램을 생성하고 Otsu's의 임계값 방법으로 손과 나머지 부분을 분리한다. 이 또한 카메라와 손 사이에 다른 객체가 있는 경우 문제가 발생한다.

이러한 문제점들에 의해 본 논문에서는 실험 환경의 조명 조건에 영향을 받지 않고, 안정적으로 손 영역을 검출하기 위해 깊이정보만 이용하면서 속도도 빠른 특징을 제안하고, 부스팅(Boosting)과 케스케이드(Cascade) 방법을 이용하여 학습 및 인식을 통해 실시간 손 영역 검출 방법에 대해 제안한다. 또한 기존 다양한 아다부스트(Adaboost)와 케스케

이드가 결합된 분류기와 제안하는 분류기의 성능을 정량적, 정성적인 비교를 통해 타당성을 입증하고, 다양한 실험을 통해 속도 및 효율성에 대해 설명한다.

2. 특 징

본 논문의 목적은 깊이 영상에서 손 영역을 빠르게 검출하기 위한 방법을 제안한다. 매우 간단한 특징을 사용하고, 이를 이용하여 부스팅과 케스케이드 방법으로 학습시켜 손 영역을 검출한다. 대표적으로 아다부스트를 이용한 검출 방법은 얼굴 검출에 사용되었다.[10] 얼굴 검출을 위해 하르 유사(Haar-like) 특징을 사용하였는데, 이는 얼굴 내부에 고정적인 특성을 가지고 있기 때문에 좋은 결과를 보여주었다. 하지만 손 영역 내부에는 뚜렷한 특성이 없으며, 깊이 영상의 경우 손의 형태 정보만 있을 뿐이다. 하지만 형태정보를 이용하기 위해서는 먼저 형태정보를 추출할 객체를 분할해야 하며, 이러한 방법은 매우 복잡할 수 있다. 따라서 본 논문에서는 다중 손 영역 추적을 위해 단일 영상 안에서 매우 간단하며, 효율적으로 손 영역을 검출하기 위한 특징을 제안한다.

2.1 중심과 차이를 이용한 특징

Table 1은 특징 추출 알고리즘을 보여준다. 영상 I와 분할하고자 하는 개수 Nx, Ny가 입력되면, 입력된 영상을 블록으로 분할하여 특징을 추출하게 된다. I_w, I_h는 입력영상의 가로와 세로의 길이이며, $Step_x, Step_y$는 각각 특징을 구하기 위한 블록들의 x축, y축 이동 변위를 나타낸다. 또한 $block_w$, $block_h$는 블록의 가로와 세로의 길이이며, $Depth_c$는 입력 영상 중심 위치의 깊이 값이다. 마지막으로 End_x, End_y는 입력된 영상에서 추출될 블록의 가로, 세로 개수이다. 영상이 입력되면, 블록의 크기와 이동 변위에 따라서 영상의 중심값과 현재 블록 영역(ROI) 평균값의 차이를 계산하여 특징값을 Fv 배열에 저장한다.

Table 1. The algorithm for feature extraction

- Input Image = I, Number of X= Nx, Number of y= Ny
- $step_x = I_w/2Nx,\ step_y = I_h/2Ny$
- $block_w = I_w/Nx,\ block_h = I_h/Ny$
- $End_x = 2Nx-1,\ End_y = 2Ny-1$
- $i \leftarrow 0$
- for $x = 0, \ldots, End_x$
  for $y = 0, \ldots, End_y$
  $ROI = RECT(x \times step_x,\ y \times step_y,\ block_w,\ block_h)$
  $Fv[i] = Depth_c - Area(ROI)/(ROI_w \times ROI_h)$
  $i \leftarrow i+1$
  end
end

Fig. 1은 제안하는 특징 $Nx=2, Ny=2$일 경우의 예를 보여준다. 중심의 빨간 점은 중심 깊이 값을 의미하며, 녹색 사각형 영역은 중심 깊이 값과 차이를 계산할 영역을 나타낸다. Nx와 Ny가 모두 2이기 때문에, 추출할 수 있는 특징의 수는 $End_x \times End_y = 9$개이다. 이와 같은 방법으로 학습 시 사용될 약분류기는 $Nx=\{1,\ldots,n\}$, $Ny=\{1,\ldots,m\}$인 모든 경우의 특징값을 사용한다. 특징값을 구하기 위해서는 [Table 1]의 입력된 영역(ROI)의 합을 반환하는 $Area$ 함수가 많은 연산 시간을 요구한다. 이러한 문제를 해결하기 위해 적분 영상(Integral Image)을[10] 이용하여 3번의 산술연산으로 효과적으로 해결하였다.

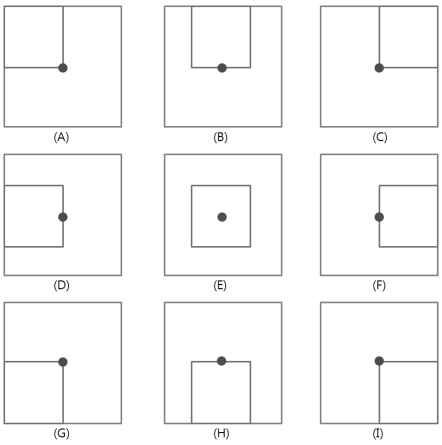


Fig. 1. The example of proposed feature

2.2 크기 불변

아다부스트를 이용한 검출 알고리즘의 경우 크기 불변을 해결하기 어렵다. 얼굴 검출의 경우에는 검출하고자하는 크기를 정하고 가능한 크기에 대해 모두 스캔하여 얼굴을 검출한다.[10] 이는 특징값 계산 시간을 적분 영상을 이용하여 줄이고, Positive와 Negative의 판단 과정을 케스케이드 방법을 사용하여 연산량을 줄임으로써 실시간 얼굴 검출이 가능하였다. 하지만 본 논문에서는 한 번의 스캔만으로 모든 크기의 손 영역을 검출하기 위해, 중심의 깊이 값을 이용하여 크기를 예측한다. 일반적으로 손 영역의 크기는 개인적인 차이는 있지만, 영역 크기의 분산이 크지 않다. 따라서 다음과 같이 2차 선형 모델을 통해 크기를 예측한다[11].

$$y = P\alpha \tag{1}$$

$$\alpha = (P^T P)^{-1} P^T y \tag{2}$$

$$P = \begin{bmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ 1 & x_n & x_n^2 \end{bmatrix}, \quad y = \begin{bmatrix} r_1 \\ r_2 \\ \cdot \\ \cdot \\ \cdot \\ r_n \end{bmatrix} \tag{3}$$

[수식 1]은 깊이에 따른 영역의 반지름을 2차 선형 모델을 사용하여 나타낸 수식이다. [수식2]의 $\alpha=[\alpha_1\ \alpha_2\ \alpha_3]^T$ 이며, y는 [수식 3]과 같이 각 학습 데이터로부터 얻은 영역의 반지름(r)으로 이루어진 열벡터이다. 또한 P는 손 영역의 중심으로부터 추출된 깊이 값(x)을 이용하여 만든 $n\times 3$ 행렬이다. 손 영역이 카메라로부터 거리에 따라서 영역의 크기가 일정한 범위에 속한다는 사전 정보를 이용하여 거리에 따른 영역의 크기를 유추하기 위해 2차 선형 모델을 사용하였다. 따라서 [수식 2]를 이용하여 α를 구하면 깊이 값에 대응하는 영역의 반지름을 산출할 수 있다.

3. 케스케이드를 이용한 학습

케스케이드를 이용한 강분류기를 만드는 과정은 크게 3단계가 필요하다. 먼저 학습영상 집합(Training set)을 생성하고, 학습영상으로부터 특징을 추출한 뒤, 마지막으로 추출된 특징 중에 좋은 특징을 선택하여 분류기를 만들어야 한다.

손은 자유도가 매우 높기 때문에, 다양한 형태를 갖는다. 실제 제스처 인식에 적용하는 경우에는 손을 찾는 것도 중요하지만 사용자가 명령을 내리기 위한 손동작을 찾는 것이 중요하다. 따라서 인위적인 손동작을 찾기 위해 손바닥이 카메라를 향해 있으며 기울어지지 않은 손바닥을 찾는 것을 목표로 한다.

3.1 학습영상 수집

특징을 추출하여 케스케이드를 학습시키기 위해서는 학습 데이터가 필요하다. 일반적으로 학습영상의 경우 같은 크기의 정규화된 데이터를 사용한다. 하지만 본 논문에서 제안하는 특징의 특성 상 크기 정규화 과정은 불필요하다. 따라서 손 영역의 카메라로부터의 거리에 따라서 Positive 학습영상을 수집하였으며, Negative는 무분별하게 추출하였다.

Fig. 2는 수집한 학습영상의 예를 보여준다. 본 논문에서는 손을 정면으로 하며, 똑바로 편 손바닥을 검출하고자 하기 때문에 손모양의 변화는 크지 않다. 따라서 깊이 변화에 따른 학습영상과 좌우 손바닥 영상에 대해서 수집하였다. 또한 그림과 같이 깊이에 따라서 학습영상의 영역의 크기가 달라짐을 볼 수 있는데, 이는 2.2절에서 구한 α값을 이용하여 학습을 위한 손바닥 영역의 크기를 예측한 결과이다. 실제 학습에 사용한 Positive 영상은 201개이며, Negative 영상은 12,967개이다.

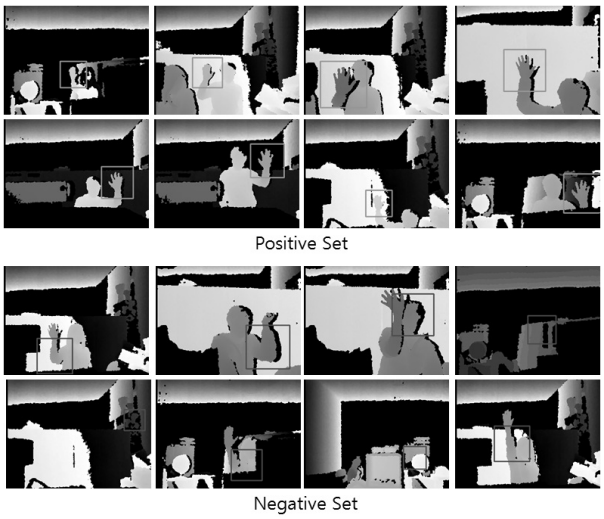


Fig. 2. Training set

3.2 특징 추출

학습영상과 학습하고자 하는 영역이 주어지면, 영역으로부터 특징을 추출해야한다. 특징은 Nx와 Ny에 의해 특징이 정해진다. Fig. 3은 특징 추출의 예를 보여준다. 입력 영상과 학습 영역이 주어지면, 주어진 영역으로부터 $Nx=1$, $Ny=1$ 부터 $Nx=n$, $Ny=m$이 될 때까지 특징을 추출한다.

$$N_{feature}=\sum_{Ny=1}^{m}\sum_{Nx=1}^{n}(2Nx-1)(2Ny-1) \tag{4}$$

[수식 4]는 n, m이 주어졌을 때 특징의 총 개수를 나타낸다. 본 논문에서는 $n=m=10$로 실험하였다. 따라서 본 연구에서는 10,000개의 특징을 사용하였다.

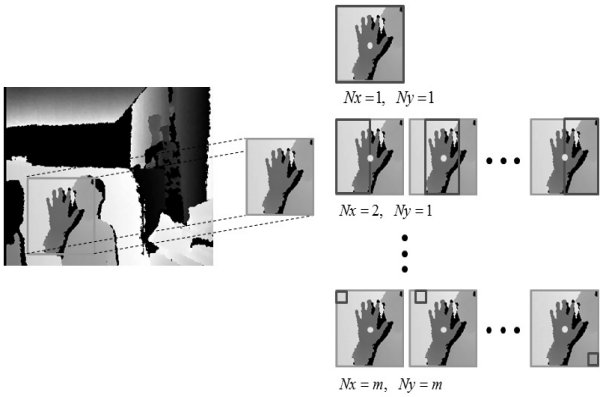


Fig. 3. The example of feature

3.3 학습

학습영상과 특징값이 주어지면 아다부스트를 이용하여 학습과정을 수행할 수 있다. Viola와 Jones이 제안한 얼굴 검출에 적용한 아다부스트의 경우에는 약분류기인 하르 유사 특징을 학습 집합에 대해 오류율이 가장 낮은 임계값을 구한다.[10] 하지만 본 연구에서는 검출율(Detection Rate)을 만족하면서 오류율이 가장 낮은 임계값을 구한다. 이는 검출하고자 하는 손 영역이 비교적 규칙적으로 단순한 패턴을 가지고 있기 때문이다. 이러한 이유로 검출 속도를 빠르게 하기위해 케스케이드 방법과 결합하고 좀 더 과적합(Overfitting) 학습을 위해 각 스테이지(Stage)를 하나의 약분류기로 구성하여 검출율를 유지하면서 연산 속도 또한 빠르게 하였다.

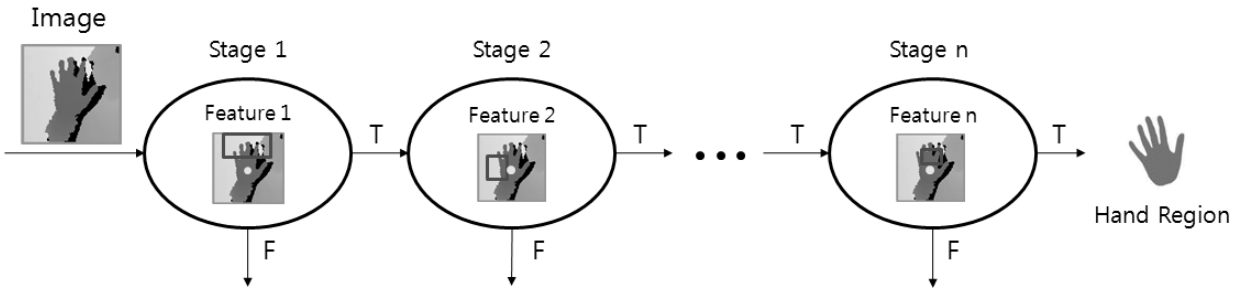


Fig. 4. The structure of classifier using cascade

$$d_r = \max((T^+ - S^+)/T^+,\ S^+/T^+) \quad (5)$$

$$e_r = \min(S^+ + (T^- + S^-), S^- + (T^+ - S^+)) \quad (6)$$

[수식 5]와 [수식 6]은 약분류기의 검출율과 오류율 계산식이다. S^+와 S^-는 임계값 이하의 각 Positive 샘플과 Negative 샘플의 가중치의 합이며, T^+와 T^-는 각각의 Positive와 Negative 샘플들의 전체 가중치 합을 의미한다. 따라서 각 스테이지를 하나의 약분류기로 구성하여 검출율을 유지하기 위해서는 임계값 결정 시 검출율을 만족하는 약분류기를 생성해야한다. 그러므로 각 특징의 임계값은 검출율 D_{target}을 만족하면서 오류율이 최소가 되는 위치로 정한다.

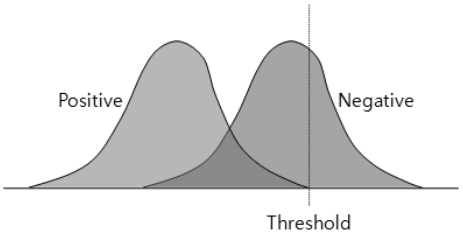


Fig. 5. The Method of threshold

Fig. 5는 임의의 특징이 주어졌을 경우 해당 특징에 대해 검출율이 최대이며 오류율이 최소인 임계값 위치를 보여준다. 녹색은 Positive 구간이며, 적색은 Negative 구간을 나타낸다. 그림과 같이 Positive 구간을 모두 포함하면서 오류율이 가장 적은 위치가 특징의 임계값으로 결정된다. 일반적인 아다부스트와 케스케이드의 결합을 이용한 분류기는 여러 개의 스테이지로 구성된다. 학습 단계에서 스테이지의 구성 조건은 검출율과 오긍정률(False Positive Rate)의 조건을 만족해야하므로 복수개의 약분류기가 포함될 수 있다.

Fig. 4는 제안하는 방법의 분류기 구조를 보여준다. 검출하고자하는 손 영역의 형태는 손바닥을 편 상태의 정면 영상이므로 크기에 따른 변화를 제외하면 대부분 유사한 형태를 지닌다. 이러한 이유로 좀 더 분류기를 과적합시키기 위해 모든 스테이지에 하나의 약분류기를 할당한다. 선택된 약분류기는 위에서 설명한 임계값 결정 규칙에서 검출율을 만족시키기 때문에 대부분의 Positive 샘플들을 받아들이면서 Negative를 분류한다. 따라서 여러 개의 스테이지를 지나도 검출율은 유지된다. 반면에 오긍정률은 스테이지 생성 조건에 사용되지 않기 때문에 매우 높은 오긍정률이 발생할 수 있다. 하지만 선택된 약분류기의 오긍정률이 0.7이하만 된다면 20개의 스테이지만으로도 검출율을 유지하면서 FPR=0.000798(0.7^{20})의 결과를 얻을 수 있다.

Table 2는 본 연구에 적용된 분류기 알고리즘을 간단하게 보여준다. F_i는 현재 오긍정률을 나타내며, F_{target}과 D_{target}은 사용자가 정하는 상수로 목표 오긍정률과 검출율이다. N은 학습영상 집합이다. 알고리즘을 보면 위에서 설명한 임계값 결정 방법에 의해 획득한 모든 특징들 중에서 $d_r \geq D_{target}$을 만족하면서 가장 e_r값이 작은 특징을 선택한다. 하지만 D_{target}이 1일 경우 e_r이 0.5이상 되는 상황이 발생할 수 있다. 이러한 문제를 해결하기위해 다음 단계에서 e_r의 값이 0.5이상이라면 D_{target}을 감소한 후 다시 최적 분류기를 선택하게 된다. 이렇게 특징이 선택되었다면, 선택된 특징과 이전에 선택된 특징들의 조합인 케스케이드 분류기를 이용하여 F_i을 계산한다. 선택된 특징은 케스케이드를 구성하는 하나의 스테이지로 생성되며, 다음 스테이지를 생성하기 위해 현재 구성된 케스케이드로 모든 Positive 샘플과 오검출된 Negative 샘플만으로 학습 샘플 N을 구성하고 F_i가 F_{target}보다 작을 때까지 반복한다.

Table 2. The classifier algorithm

- *While* $F_i > F_{target}$
  1. Select the best classifier
     $d_r \geq D_{target}$ and $\min(e_r)$
  2. if $e_r \geq 0.5$ then decrease D_{target} and go step 1.
  3. Evaluate current cascaded classifier and Update F_i
  4. $N \leftarrow \varnothing$
  5. Put false detections and positive samples into the set N

4. 손 영역 검출

본 논문의 목적은 하나의 프레임만으로 매우 빠르게 손 영역을 검출하는 것이다. 이를 위해 3장에서 손 영역을 판

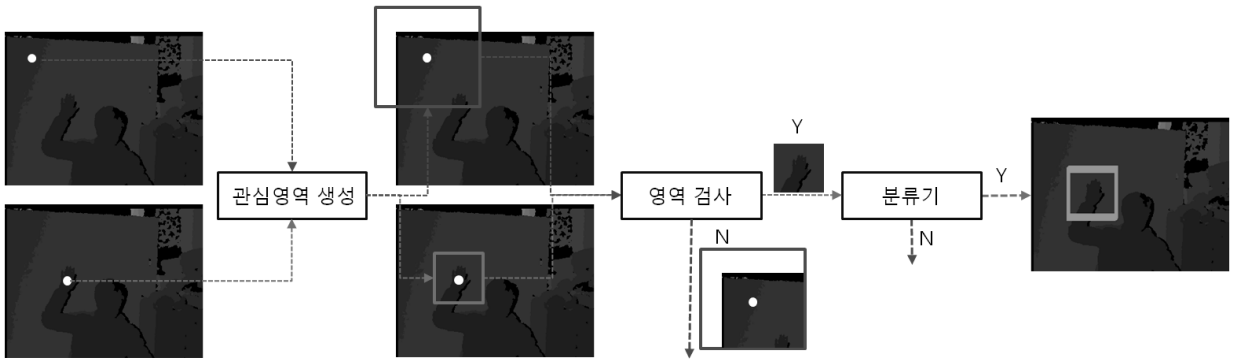


Fig. 6. The process of hand region detection

단하기 위한 분류기 생성 과정에 대해 설명하였다. 이렇게 생성된 분류기를 이용한 손 영역 검출은 크게 2단계로 구성된다. 먼저 영상 전체에 대해서 케스케이드 분류기를 이용하여 손 영역과 손 영역이 아닌 영역을 분류하는 과정을 수행하고, 검출된 손 영역들 간의 병합 과정을 통해 최종 손 영역 결정한다.

4.1 검출

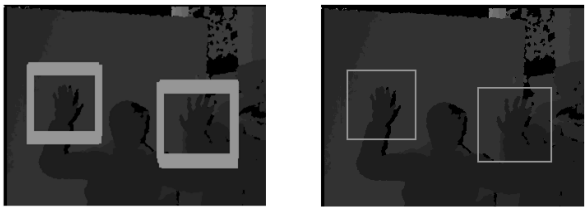
3장에서 손 영역 검출을 위한 케스케이드 분류기를 생성하였다. 생성된 분류기는 어떤 영역이 주어졌을 경우 손 영역인지 아닌지를 반환한다. 이러한 분류기를 이용하여 크기에 불변한 손 영역을 검출해야하며 크기 불변을 만족시키기 위해 2.2절에서 설명한 2차 선형 모델을 생성하였다.

Fig. 6은 2차 선형 모델과 3장에서 생성한 케스케이드 분류기를 이용한 손 영역 검출 과정을 보여준다. 먼저 분류하고자 하는 화소의 위치가 주어지면 화소의 깊이 값으로 2차 선형 모델에 적용하여 관심영역을 생성한다. 생성된 관심영역은 입력 영상의 범위 안에 포함 여부를 판단하는 영역 검사를 수행한다. 이렇게 영역 검사까지 통과한 화소는 분류기로 보내지며, 3장에서 생성한 분류기를 구성하는 스테이지를 모두 통과할 경우 손 영역으로 분류된다.

4.2 병합

Fig. 7은 병합 전의 결과와 병합 후 결과를 보여준다. Fig. 6의 과정을 마치면 Fig. 7의 (A)와 같은 결과를 얻을 수 있다. 그림에서 볼 수 있듯이 손 영역의 근처에 매우 많은 사각형들이 겹쳐 있는 것을 볼 수 있다. 이러한 사각형들은 하나의 손 영역일 경우 하나의 영역으로 병합되어야 한다.

Table 3 병합 알고리즘을 보여준다. $distance(A,B)$는 A사각형의 중심점과 B사각형의 중심점 간의 깊이 차이를 나타내며 $intersect(A,B)$는 A사각형과 B사각형의 겹치는 영역의 크기를 나타낸다. 또한 Th_d은 깊이 차이에 대한 임계값이며, Th_r은 겹치는 영역에 대한 임계값이다. 실험에서는 각각 50과 0.6을 사용하였다. 또한 $TempR$은 여러 개의 사각형 정보가 저장될 수 있는 메모리 공간이다. 즉, 손 영역의 후보를 뜻하는 사각형들의 중심점 간의 깊이 차이와 겹치는 영역을



A : result of classifier　　B : result of merge

Fig. 7. Detection result

검사하여 조건을 통과한 사각형을 메모리 공간 $TempR$에 추가하고, 다른 사각형을 검사할 때는 $TempR$의 평균 사각형과 깊이 차이, 겹침의 정도를 비교하여 병합을 수행한다. 이렇게 병합작업이 완료되면 하나의 평균 사각형으로 만들어, 평균 사각형을 최종 손 영역으로 결정한다. 평균 사각형을 구하는 방법은 사각형들의 각 꼭지점들 간에 평균 좌표로 평균 사각형을 획득한다. Fig. 7의 (B)는 병합 알고리즘을 수행한 후 결과를 보여준다.

Table 3. Merge algorithm

```
for i = 0, i < N_hand
  if R_i is not merged
    TempR ← Insert R_i
    for j = 0, j < N_hand
      if i ≠ j
        if R_j is not merged
          MR ← mean of TempR
          if distance(MR, R_j) < Th_d
            if intersect(MR, R_j)/(MR_w × MR_h) > Th_r
              TempR ← Insert R_j
            endif
          endif
        endif
      endif
    end
    Result ← Add the meaned rectangle of TempR
    TempR = ∅
  endif
end
```

5. 실험 결과

본 연구에서는 입력장치로 MS사의 Kinect를 사용하여 얻은 320*240 크기의 깊이 영상을 사용하였다. 또한 Intel(R) Core(TM) Quad CPU 2.66Ghz와 3GB 메모리에서 손 영역 검출에 대해 실험하였다.

Fig. 8은 10000개의 특징 중 선택된 21개의 특징을 보여준다. 이미지 하단의 숫자는 특징의 임계값을 의미하며, 녹색 사각형은 Parity가 1인 경우를, 적색 사각형은 Parity가 -1인 경우를 나타낸다. 대부분의 영역에서 Parity가 1인 것을 볼 수 있는데, 이는 중심과 사각형 영역 평균과의 차이가 주어진 임계값보다 작아야한다는 것을 의미한다. 즉 녹색 사각형 영역은 중심 화소보다 카메라로부터 멀리 떨어져 있어야 한다는 것을 의미한다. 적색 사각형의 경우 대부분 손 내부에 위치 한 것을 알 수 있는데, 이는 반대로 중심과 사각형 영역 평균의 차이가 임계값보다 커야 한다는 것을 의미한다. 하지만 Fig. 8의 11번째 특징의 경우 적색인데도 불구하고, 손 영역 외부에 위치한 것을 볼 수 있는데, 이는 임계값이 -6292.3으로 매우 작기 때문에 영역을 찾는데 문제를 발생시키지 않는다. 또한 선택된 21개의 특징을 보면 대부분 손의 형태와 유사하게 특징이 추출되어 매우 직관적으로 분류기의 의미를 파악 할 수 있다.

Fig. 9는 분류기의 연산량 측면의 성능을 보여주는 그림이다. (A)는 성능 비교 검사에 사용된 원본 영상이며, (B)는 Viola와 Jones[10]에서 제안된 Adaboost와 Cascade의 결합으로 분류한 결과이다. 마지막으로 (C)는 제안한 방법으로 분류한 결과이다. (B), (C)의 적색 영역은 4.1절에서 설명한 영역 검사 조건을 통과하지 못하여 분류 검사를 수행하지 않은 영역을 나타내며, 255로 표현된 화소는 손 영역이라 판단된 영역의 중심을 나타낸다. 나머지 영역의 화소값은

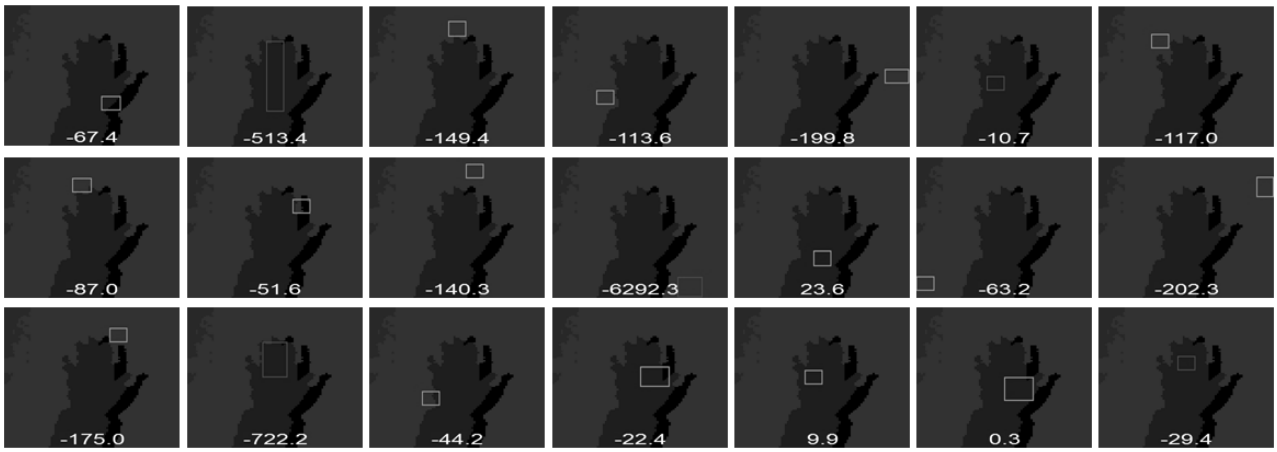


Fig. 8. Selected Features of 21

Table 4. The comparison of detection rate and false positive rate of various AdaBoost

		The maximum acceptable false positive rate							
		0.7	0.6	0.5	0.4	0.3	0.2	0.1	0
Discrete Adaboost	DR	1	0.99	0.98	1	1	0.99	0.99	0.92
	FPR	0.67	0.62	0.54	0.91	0.84	0.82	0.84	0.39
	Computation	25	25	25	28	27	31	39	106
Real Adaboost	DR	1	0.89	0.96	0.99	0.96	0.99	0.99	0.94
	FPR	0.33	0.41	0.41	0.28	0.43	0.53	0.39	0.07
	Computation	13	12	12	11	15	19	23	65
Gentle Adaboost	DR	1	0.99	0.98	0.99	0.98	0.98	1	0.85
	FPR	0.64	0.56	0.33	0.41	0.27	0.37	0.51	0
	Computation	13	11	12	13	15	17	20	77
The proposed method	DR	0.97							
	FPR	0							
	Computation	8							

(A) (B) (C)

Fig. 9. The performance of classifier

(A) (B) (C) (D)

(E) (F) (G) (H)

Fig. 10. The result of detection

분류기의 검사 횟수를 의미한다. 즉, 어두운 화소는 그만큼 분류기를 덜 통과하여 손 영역이 아니라고 판단된 영역을 나타낸다. 그림에서 보듯이 (C)가 (B)보다 255인 영역도 실제 손 영역에서만 나타났으며, 나머지 영역에 대해서도 (B)보다 훨씬 더 어두운 것을 볼 수 있다. 결론적으로 제안한 방법이 (B)의 방법보다 속도와 정확도면에서 좋은 것을 알 수 있다.

Table 4는 Discrete Adaboost, Real Adaboost, Gentle Adaboost와 제안한 방법의 실험결과를 보여준다[12]. 다양한 Adaboost에 대한 실험은 제안한 방법과 같은 학습 집합을 이용하여 학습을 수행하였으며, 허용 FPR을 0.7~0까지 0.1 단위로 줄여가면서 실험하였다. 먼저 DR은 검출율을 나타내며, FPR를 오긍정률을 나타낸다. Computation은 하나의 프레임을 기준으로 검출 과정이 완료되기까지 시간(ms)이다. 스테이지 허용 FPR은 각 부스팅 알고리즘으로 분류기 학습 시 현재 스테이지의 학습 단계를 완료 할 수 있는 최대 오긍정률을 나타낸다. 제안한 방법에서는 다른 분류기들과 다르게 스테이지를 완료하는 오긍정률 범위가 없으므로 하나의 실험 결과만 얻을 수 있었다. 전체적으로 다른 부스팅 방법은 대부분 좋지 않은 결과를 보였다.

그 중 가장 좋은 부스팅 결과는 Real Adaboost의 스테이지 허용 FPR을 0으로 하고 실험한 결과이다. 표와 같이 0.94의 검출율과 0.07의 오긍정율을 보였다. 하지만, 검출 속도는 다른 분류기에 비해 6배 정도 느려짐을 확인하였다. 이는 스테이지 허용 FPR을 0으로 설정하고 분류기를 학습시켜, 하나의 스테이지에 모든 약분류기가 포함되었으며, 이로 인해 케스케이드의 장점인 속도를 보상받지 못한 결과였다. 이에 반해, 실험 데이터에서 제안한 방법은 97%의 검출율로 하나도 잘못된 손 영역을 찾지 않았으며, 속도 또한 가장 빠른 결과를 보였다. 이는 오직 21개의 특징으로만 분류를 수행하고, 이러한 분류기들이 케스케이드 구조를 구성하기 때문이다.

Fig. 10은 실시간 키넥트 카메라 입력으로 실험한 결과 영상을 보여준다. 녹색 사각형은 검출된 사각형들을 병합 작업까지 완료된 후의 사각형이다. 상단의 (A)~(D)까지의 결과는 모두 정확하게 검출된 것을 볼 수 있다. 학습 시 사용한 손 영역은 손을 활짝 편 상태의 정면 영상만을 학습에 사용하였다. 하지만 특징 자체가 형태의 세부적인 표현보다 중심과 영역 평균의 차이를 이용하므로 (A)의 경우와 같이 주먹을 쥔 상태로 검출한 것을 볼 수 있다. 반면 (E)와 (F)

는 왼손은 모두 검출 하였지만, 오른쪽 손은 검출하지 못하였다. 이는 분류기가 허용할 수 있는 손 형태의 범위를 벗어났기 때문이었다. (G)와 (H)에서는 손바닥 뿐 아니라 팔꿈치 있는 부분을 함께 검출되었는데, 이는 좀 더 다양한 Negative 샘플을 수집하여 학습한다면 충분히 개선할 수 있을 것으로 생각된다.

6. 결 론

본 논문에서는 깊이 영상만을 이용하여 매우 빠르고 정확하게 손 영역을 검출하는 방법에 대해 제안하였다. 깊이 영상의 특성을 고려한 제안하는 특징을 이용하여 기존 아다부스트와 달리 각 스테이지에 하나의 약분류기만으로 케스케이드를 구성하여 속도와 정확도 모두 좋은 성능을 나타냄을 보였으며, 학습된 분류기를 통해 한 번의 스캔만으로도 다양한 크기의 손 영역을 매우 신속하게 검출함을 입증하였다. 또한, 제스처 인식을 위한 기반 기술인 손 영역 검출을 시간적 정보가 불필요한 하나의 프레임만으로 검출이 가능한 방법을 제시함으로써 다중 손 영역 추적 시스템에 응용 가능할 것으로 생각된다. 본 논문에서는 제스처 인식을 위한 선행연구로 손 영역 검출연구를 수행하였으므로 향후 다중 손 영역 추적 및 제스처 인식에 대한 연구가 추가적으로 필요할 것이다.

참 고 문 헌

[1] H. I. Suk, and B. H. Sin, “Dynamic Bayesian Network based Two-Hand Gesture Recognition”, Journal of KIISE : Software and Applications, Vol.35, No.4, 2008.

[2] M. K. Bhuyan, D. R. Neog and M. K. Kar, “Fingertip Detection for Hand Pose Recognition”, International Journal on Computer Science and Engineering (IJCSE), Vol.4 No.3, pp.501-511, March, 2012.

[3] M. S. Park, Md. M. Hasan, J. M. Kim and O. S. Chae, “Hand Detection and Tracking Using Depth and ColorInformation”, IPCV′12, Vol.2, pp.779-785, 2012.

[4] M. Van den Bergh, and L. Van Gool, “Combining RGB and ToF Cameras for Real-time 3D Hand Gesture Interaction”, 2011 IEEE Workshop on Application of Computer Vision (WACV), pp.66-72, January, 2011.

[5] P. Trindade, J. Lobo and J. P. Barreto, “Hand gesture recognition using color and depth images enhanced with hand angular pose data”, IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI), pp.71-76, September 13-15, 2012.

[6] Z. Mo, U. Neumann, “Real-time Hand Pose Recognition Using Low-Resolution Depth Images”, Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06), Vol.2, pp.1499-1505, 2006.

[7] X. Liu and K. Fujimura, “Hand gesture recognition using depth data”, Proc. 6th. International Conf. on Automatic Face and Gesture Recognition, pp.529-534, Seoul, Korea, 2004.

[8] S. Malassiotis, M.G. Strintzis, “Real-time hand posture recognition using range data”, Image and Vision Computing, Vol.26, Issue 7, pp.1027-1037, 2 July, 2008.

[9] P. Suryanarayan, “Dynamic Hand Pose Recognition using Depth Data”, In 2010 International Conference on Pattern Recognition, pp.3105-3108, 2010.

[10] P. Viola and M. Jones, “Robust Real-time Face Detection”, International Journal of Computer Vision Vol.57, No.2, pp.137-154, 2004

[11] J. Sung-il, W. Sun-hee, C. Hyung-il, “Real-time Hand Region Detection and Tracking using Depth Information”, KIPS Transactions on Software and Data Engineering, Vol.1, No.3, pp.177-186, 2012.

[12] J. Friedman, T. Hastie, R. Tibshirani, “Additive logistic regression : a statistical view of boosting”, Technical report, Department of Statistics, Sequoia Hall, Stanford University, 1998.

주 성 일

e-mail : sijoo82@ssu.ac.kr

2008년 한국산업기술대학교 컴퓨터공학과 (공학사)

2010년 숭실대학교 미디어학과(공학석사)

2010년~현 재 숭실대학교 미디어학과 박사과정

관심분야 : Image Processing, Computer Vision, Pattern Recognition, Machine Learning

원 선 희

e-mail : nifty12@ssu.ac.kr

2005년 한경대학교 컴퓨터공학과(공학사)

2007년 숭실대학교 컴퓨터학과(공학석사)

2012년 숭실대학교 미디어학과(공학박사)

2012년~현 재 숭실대학교 미디어학과 Post Doc.

관심분야 : Image Processing, Computer Vision, Pattern Recognition, 3D Modeling

최 형 일

e-mail : hic@ssu.ac.kr

1972년 연세대학교 전자공학과(공학사)

1982년 미시간대학교 전자공학과(공학석사)

1987년 미시간대학교 전자공학과(공학박사)

1995년~1997년 퍼지 및 지능시스템학회 이사

1996년~1998년 정보과학회 컴퓨터비전 및 패턴인식 연구회 위원장

1997년 IBM Waston Lab 방문연구원

2005년~2006년 한국정보과학회 이사

1987년~현 재 숭실대학교 미디어학과 교수

관심분야 : Computer Vision, Pattern Recognition, Fuzzy & Neural Network, Machine Learning