

깊이영상에서 실시간 얼굴 검출을 위한 I-MCTBoost

주성일*, 원선희*, 최형일*

The I-MCTBoost Classifier for Real-time Face Detection in Depth Image

Sung-Il Joo *, Sun-Hee Weon *, Hyung-Il Choi *

요약

본 논문에서는 실시간 얼굴 검출을 위한 부스팅 기반 분류 방법을 제안한다. 제안하는 방법에서는 조명과 얼굴 크기 및 변형에 강건하게 얼굴을 검출하기 위해 깊이영상을 이용하고, 깊이차이특징을 사용하여 I-MCTBoost 분류기를 통해 학습 및 인식을 수행한다. I-MCTBoost는 약분류기로 구성된 강분류기들의 연결을 통해 인식을 수행한다. 약분류기의 학습 과정은 깊이차이특징을 생성하고, 이 중에서 8개의 특징을 조합하여 약분류기를 구성하며 이때 각 특징은 2진비트(binary bit)로 표현된다. 강분류기는 정해진 약분류기의 개수만큼 반복적으로 약분류기를 선택하는 과정을 통해 학습이 이루어지며, 학습 과정에서 학습 샘플의 가중치를 갱신하고 학습 데이터를 추가하여 강건한 분류를 수행할 수 있도록 한다. 본 논문에서는 깊이차이특징에 대해 설명하고 이를 이용한 I-MCTBoost의 약분류기 학습 방법과 강분류기 학습 방법에 대해 제안한다. 마지막으로 제안된 분류기를 기존 MCT를 이용한 분류기와 정성적, 정량적 분석을 통해 비교하고 제안한 분류기의 타당성과 효율성을 입증한다.

▶ Keywords : 얼굴 검출, 깊이영상, 깊이차이특징, I-MCTBoost(Improved Modified Census Transform), 부스팅

Abstract

This paper proposes a method of boosting-based classification for the purpose of real-time face detection. The proposed method uses depth images to ensure strong performance of face detection in response to changes in lighting and face size, and uses the depth difference feature to conduct learning and recognition through the I-MCTBoost classifier. I-MCTBoost performs recognition by connecting the strong classifiers that are constituted from weak classifiers. The learning process

•제1저자 : 주성일 •교신저자 : 최형일

•투고일 : 2014. 2. 8, 심사일 : 2014. 2. 21, 게재확정일 : 2014. 2. 25.

* 송실대학교 글로벌미디어학과(Dept. of Global Media, Soongsil University)

※ 이 논문은 2013년도 정부(미래창조과학부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임 (No.2013R1A1A2012012)

for the weak classifiers is as follows: first, depth difference features are generated, and eight of these features are combined to form the weak classifier, and each feature is expressed as a binary bit. Strong classifiers undergo learning through the process of repeatedly selecting a specified number of weak classifiers, and become capable of strong classification through a learning process in which the weight of the learning samples are renewed and learning data is added. This paper explains depth difference features and proposes a learning method for the weak classifiers and strong classifiers of I-MCTBoost. Lastly, the paper presents comparisons of the proposed classifiers and the classifiers using conventional MCT through qualitative and quantitative analyses to establish the feasibility and efficiency of the proposed classifiers.

▶ Keywords : Face detection, Depth image, Depth difference feature, I-MCTBoost (Improved Modified Census Transform), Boosting

I. 서론

최근 네트워크 환경이 급속히 발전하고 소셜 네트워크 서비스와 같은 사용자들 간의 정보 교류가 활발해지면서 사용자와 시스템 간의 지능형 인터랙션 기술이 각광받기 시작했다. 지능형 인터랙션 기술은 시스템이 사용자로부터 유용한 정보를 수집하고 가공하여 개인에 대한 정보를 파악하고 그에 따른 시스템의 서비스를 제공할 수 있는 기술로서, 상황에 적응적인 인터랙션 기술이 가장 주목받고 있다. 시스템이 사용자를 인식하고 추적하여 상황을 인식한다면 이를 위한 데이터를 수집하고 제공하기 위해 카메라를 이용한 컴퓨터 비전 기술이 주요 기술로서 적용될 수 있다. 그러나 기존의 연구들은 주로 컬러영상을 이용하여 사용자의 행위를 인식하는 기술들이 연구되었으며, 배경이 복잡하거나 조명에 민감하다는 특성과 함께 3차원 정보를 추출하는데 문제점들을 가지고 있다. 따라서 컬러영상 정보를 대체할 수 있는 깊이(depth)영상 정보를 이용하는 연구가 활발히 이루어지고 있으며, 이러한 기술들은 향후 영상기반 상황인식 시스템을 위한 원천기술로서 해당 연구 분야에서 크게 주목하고 있다.

상황인식 연구를 위해 가장 먼저 필요한 기술은 바로 사용자 인식 기술이 될 것이다. 사용자 인식에 대한 연구는 크게 영상으로부터 사용자의 얼굴을 판단하는 얼굴 검출 기술과 검출된 얼굴에 대해 사용자를 인지하는 얼굴 인식 기술로 구분할 수 있다. 즉, 영상에서 수많은 객체들로부터 얼굴을 판단

하여 분리한 후, 검출된 얼굴에 대해 특징을 추출하여 등록된 사용자 여부를 판단함으로써 사용자 인식을 수행하는 것이다. 이중 얼굴 검출 기술은 사용자 인식 성능을 좌우하는 상당히 중요한 기술로서 작용하게 된다. 영상에 존재하는 많은 객체들과 얼굴을 분리하기 위해 대부분의 연구에서는 피부 색상 모델을 이용하거나, 얼굴의 구성요소들에 대한 특징을 추출하는 방법을 사용한다. 그러나 이러한 연구는 대부분 컬러영상 정보나 그레이(gray)영상 정보에 의존하고 있기 때문에 조명이나 환경변화에 민감하게 반응한다는 문제점이 있다. 또한, 인종이나 나이, 성별에 따라 사용자 얼굴에 대한 정보를 일반화하기 어렵기 때문에 얼굴 검출 성능이 매우 낮아지게 된다는 문제점이 있어 이를 해결하기 위한 새로운 접근 방식의 연구가 필요하다.

II. 관련연구

영상을 기반으로 얼굴을 검출하는 연구는 크게 지식 기반 방법(Knowledge-based method)[1], 특징 기반 방법 (Feature-based method)[2-4], 템플릿 매칭 방법 (Template-matching method)[5], 외형 기반 방법 (Appearance-based method)[6-11]으로 구분된다.

지식 기반 방법은 얼굴의 눈, 코, 입 등의 얼굴구성요소의 일정한 거리와 위치 관계를 전제로 얼굴을 검출하는 연구로서, 간단한 연산을 통해 빠르게 얼굴 검출이 가능하지만, 복잡한 배경을 가지거나 다수의 얼굴 검출을 수행하기에는 오류

가 많으며, 사전에 규칙을 정하기가 매우 어렵다는 문제점이 존재한다.

특징 기반 방법은 얼굴 요소들의 크기나 모양, 상호 연관성, 얼굴의 색상과 질감 정보, 또는 이러한 요소들의 혼합된 형태의 정보를 이용해 얼굴을 검출하는 방법이다. 이 방식은 얼굴 부위의 잡음이 존재하거나 주변 환경의 밝기 변화에 민감할 뿐 아니라 다른 사물에 의해 가려지는 폐색(occlusion)이 발생할 경우 얼굴의 특징을 추출하기가 어렵기 때문에 얼굴 검출 성능이 낮아지게 된다는 문제점이 있다.

템플릿 매칭 방법은 얼굴에 대해 표준 템플릿을 생성한 후, 입력 영상과 비교하여 얼굴을 검출하는 방법으로 상대적으로 구현하기 쉬우나, 얼굴 영상의 크기, 모양, 포즈 등의 다양한 변화에 대해 효율적으로 처리하기 어렵다.

마지막으로 외형 기반 방법은 얼굴(face)데이터와 비 얼굴(non-face) 데이터로 학습된 얼굴 특징들을 이용하여 통계적인 분석 방법을 통해 얼굴을 검출하는 방법으로 대표적으로 주성분분석(PCA), 동적 형상 모델(ASM, AAM), SVM 또는 신경망, 아다부스트(Adaboost) 분류기 등을 이용하는 방법이 있다. 이 방법들은 얼굴 정면 검출에는 성능이 높지만 상대적으로 얼굴 측면에 대한 검출 성능이 낮다. 이는 학습한 데이터베이스의 대부분이 정면 얼굴 영상으로 구성되기 때문이다. 그러나 이 접근 방법은 앞서 설명한 세 가지 방법들과 비교하여 얼굴 검출 성능이 가장 우수한 방법이다.

본 연구에서는 특징 기반 방법의 MCT(Modified Census Transform) 특징을 이용하여 외형 기반 방법인 MCT를 이용한 분류기를 개선하여 깊이영상에 적용함으로써 실시간 얼굴 검출을 수행한다. MCT 특징은 Fröba[11]에 의해 제안되었으며, [그림 1]과 같이 얼굴 요소의 특징을 학습하기 위해 중심 화소의 밝기 값과 3x3 마스크의 주변 화소들과의 평균 밝기 값을 비교하여 평균값보다 크면 1을 작으면 0을 할당하여 511개의 패턴 중의 하나로 변환하는 지역적 패턴 특징이다. 이 MCT 특징을 이용하여 학습하기 위해 학습 샘플집합의 얼굴영상에서 기존의 아다부스트 분류기를 개선한 형태를 적용한다. Fröba가 적용한 얼굴 검출을 위해 [그림 2]와 같이 4단계의 캐스케이드(cascade) 구조로 구성되며, 각 단계에서는 모든 얼굴을 통과시키고 많은 비 얼굴을 제거하는 역할을 수행한다. 각 단계는 임계값에 따라서 현재의 분석 윈도우(analysis window)를 비 얼굴로 거절시키거나 다음 단계로 통과시킬 수 있다. 이때 각 단계에 속한 분류기는 하나의 강분류기(strong classifier)를 의미하는 것으로서 이는 MCT를 특징으로 하는 다수의 약분류기(weak classifier)로 구성된다. 즉, 흰색으로 표현된 점(white dot)

은 분석 윈도우 내의 약분류기의 위치를 나타내며 이 약분류기들을 이용하여 각 단계에서 얼굴과 비 얼굴 여부를 판단하도록 강분류기를 구성한다.

Fröba가 이용한 MCT 특징은 지역적 특징을 표현하기 때문에 조명의 변화에 매우 강건하게 작용하며, 캐스케이드 구조의 변형된 아다부스트 분류기는 거짓 긍정률(false positive rate)를 감소시키기 때문에 얼굴 검출 성능을 향상시킬 수 있다. 그러나 얼굴 검출을 위해 Viola와 Jones[10]에 의해 제안된 Haar-like 특징과 아다부스트 분류기를 결합한 방법과 비교하였을 때 입력 영상에 존재하는 얼굴 크기의 변화에 강건하게 검출할 수 있도록 마스크 크기를 변화하여 특징을 추출하는 Haar-like 특징과 달리 MCT 특징은 고정된 마스크 형태를 사용하여 영상을 다운 스케일링(down-scaling)하면서 검출하기 때문에 검출속도가 비교적 느리다는 문제점이 있다.

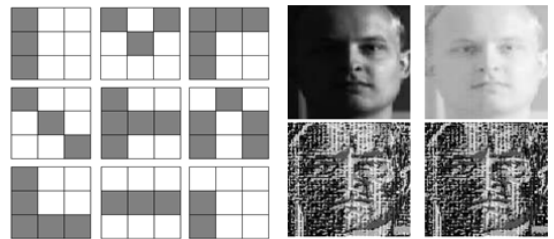


그림 1. MCT 패턴과 조명에 강건한 MCT 특징의 예[11]
Fig. 1. Example of the MCT patterns and illumination invariance of the MCT feature.

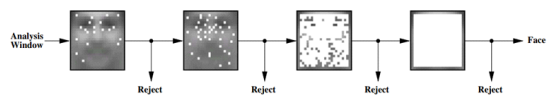


그림 2. 4개의 분류기로 구성된 캐스케이드 검출기[11]
Fig. 2. The four classifiers of Cascade Detector

본 논문에서는 얼굴 검출을 수행하기 위해 조명의 영향이나 폐색의 문제를 해결할 수 있도록 깊이영상을 이용하며, 크기 변화에 강건하도록 깊이차이특징과, 실시간 검출을 위해 인식 속도를 향상시키기 위한 I-MCTBoost 분류기를 제안한다.

논문의 구성은 3장에서는 기존 MCT 특징과 얼굴검출을 위한 깊이차이특징에 대해 설명하고 4장에서는 검출된 특징을 이용하여 학습 및 인식을 위한 I-MCTBoost 분류기에 대해 자세히 기술한다. 5장에서는 실험을 통한 타당성 입증 결과를 기술하며 6장에서는 결론과 향후 연구에 대해 기술한다.

III. 특징 생성

1. MCT(Modified Census Transform) 특징

MCT 특징은 [11]에서 얼굴 검출을 위해 제안한 특징 추출 방법으로서 특정 위치의 화소 x 에 대해 주변 인접 화소와의 관계를 패턴으로 표현하는 방법이다. MCT 특징은 밝기값 자체를 그대로 사용하지 않고 주변 픽셀과의 상관관계를 패턴으로 표현하기 때문에 조명 변화에 강한 장점이 있다. 3x3 크기의 커널을 사용하여 MCT 특징을 구하는 방법은 아래 식 (1)과 같다.

$$\Gamma(x) = \otimes_{y \in N'} \zeta(\bar{I}(x), I(y)) \quad (1)$$

이때, N' 은 커널 내의 인접 화소들의 집합이며, $I(y)$ 는 y 의 밝기값, $\bar{I}(x)$ 는 커널 내 화소들의 평균 밝기값을 나타낸다. $\zeta(\cdot)$ 는 비교함수로서 $I(y)$ 가 $\bar{I}(x)$ 보다 크면 1, 작으면 0을 출력한다. \otimes 는 10진수 변환연산자로서 $\zeta(\cdot)$ 의 결과 값인 9자리 2진수 배열을 10진수로 변환시킨다. 즉, [11]에서는 3x3 크기의 커널로 정의함으로서 0부터 510까지 511개의 MCT 특징값을 추출하여 얼굴 검출에 적용한다. [그림 3]에서는 입력 영상 $I(x)$ 에 대해 표현된 MCT 특징 $\Gamma(x)$ 의 예를 나타낸다.

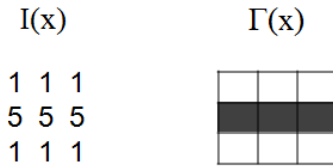


그림 3. MCT 특징의 예
Fig.3. Example of MCT feature

2. 깊이차이특징

본 논문에서는 깊이영상을 이용한 얼굴검출을 위해 MCT 특징을 개선한 지역적 특징 구조를 가진 깊이차이특징을 생성한다. 기존의 MCT 특징은 영상의 크기에 관계없이 고정적인 크기의 윈도우 커널을 사용함으로써 지역적 특징을 나타내었다. 따라서 모든 가능한 크기의 얼굴을 검출하기 위해 윈도우 커널을 조절하지 않고, 영상의 크기를 변화시켜야만 하는 단

점이 있다. 또한, 지역적으로 이웃한 화소들간의 관계를 이용해서 특징을 생성함으로써 특징의 개수가 비교적 적다. 컬러영상에서 얼굴은 얼굴 내부에 눈, 코, 입과 같은 밝기값의 차이가 뚜렷한 구성요소를 포함하고 있기 때문에 이러한 특성을 이용하여 Haar-like 특징은 좋은 결과를 보였다. 하지만 깊이영상의 경우에는 얼굴 내부에 이러한 특징이 명확하지 않아 컬러영상에서 사용하던 공통적인 특성을 사용할 수 없다.

따라서 본 논문에서 제안한 깊이차이특징은 얼굴 크기에 따라 윈도우를 동적으로 변화시킬 수 있으며, MCT 보다 많은 특징 집합을 구성할 수 있고, 깊이영상의 특성을 부각시켜 기존 MCT의 비트 단위 특징 표현법을 개선하였다.

[표 1]은 깊이차이특징 추출을 위한 알고리즘을 나타낸 것으로서, 영상 I 와 분할 개수 N_x, N_y 가 입력된다. 이때 I_{width}, I_{height} 는 입력 영상의 가로와 세로의 크기를 나타내며, $step_x, step_y$ 는 각각 특징을 구하기 위한 윈도우 블록들의 x 축과 y 축으로의 이동 변위를 의미한다. $block_w, block_h$ 는 윈도우 블록의 가로와 세로 크기를 나타내며, $Depth_c$ 는 입력 영상에서의 중심 위치의 깊이 값을 의미한다. 또한, End_x, End_y 는 입력된 영상으로부터 추출될 윈도우 블록의 가로와 세로 개수이다. 즉, 입력 영상에 대해 블록의 크기와 이동 변위에 따라 영상에서의 중심 깊이 값과 현재 윈도우 블록영역(ROI)의 평균값의 차이를 계산하여 이를 Fv 배열에

표 1. 깊이차이특징 추출 알고리즘
Table 1. The algorithm for depth difference feature extraction

Input :
I : Input image
N_x, N_y : the number of blocks in x and y axis
Definition :
$step_x = I_{width}/2N_x, step_y = I_{height}/2N_y$
$block_w = I_{width}/N_x, block_h = I_{height}/N_y$
$End_x = 2N_x - 1, End_y = 2N_y - 1$
Algorithm :
$i \leftarrow 0$
for $x = 0, \dots, End_x$
for $y = 0, \dots, End_y$
$ROI = RECT(x \times step_x, y \times step_y, block_w, block_h)$
$Fv[i] = Depth_c - Area(ROI) / (ROI_w \times ROI_h)$
$i \leftarrow i + 1$
end
end

저장한다. 이때 입력된 영역(ROI)의 합을 반환하는 $Area(\cdot)$ 함수가 많은 연산시간이 필요하므로 본 논문에서는 적분영상(Integral image)를 이용하여 3번의 산술 연산만으로 효율적인 연산을 수행한다.

얼굴 검출 시 검출하고자 하는 얼굴 크기를 사전에 정의하고 가능한 크기에 대해 모두 스캔하여 얼굴을 검출하던 기존의 방식과 달리 한 번의 스캔만으로 모든 크기의 얼굴을 검출하기 위해 중심 깊이 값을 이용한 크기 예측을 수행한다. 사람의 얼굴 크기는 개인적인 차이는 있을 수 있으나 일반적으로 유사하므로 그 분산이 크지 않다. 따라서 중심 깊이 값에 따른 영역 크기를 예측하기 위하여 2차 다항모델(2nd polynomial model)을 이용한다.

$$y = P\lambda \tag{2}$$

$$\lambda = (P^T P)^{-1} P^T y \tag{3}$$

$$P = \begin{bmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ 1 & x_3 & x_3^2 \\ \vdots & \vdots & \vdots \\ 1 & x_n & x_n^2 \end{bmatrix}, y = \begin{bmatrix} r_1 \\ r_2 \\ r_3 \\ \vdots \\ r_n \end{bmatrix} \tag{4}$$

식(2)는 깊이 값에 따른 영역의 반지름을 2차 다항모델을 이용하여 표현한 것으로서, λ 를 구하기 위해 식(3)으로 풀이할 수 있다. 식(3)에서의 $\lambda = [\lambda_1 \lambda_2 \lambda_3]^T$ 이며, y 는 식(4)에서와 같이 학습 데이터로부터 얻은 영역의 반지름(r)으로 이루어진 열벡터로 정의할 수 있다. 또한, P 는 얼굴 영역의 중심으로부터 추출된 깊이 값(x)으로 구성된 $n \times 3$ 행렬이

다. 즉, 얼굴 영역이 카메라로부터 거리에 따라 영역의 크기가 일정한 범위에 속한다는 사전 정보를 이용하여 얼굴 영역의 크기를 예측하기 위해 2차 다항모델을 이용한다.[13]

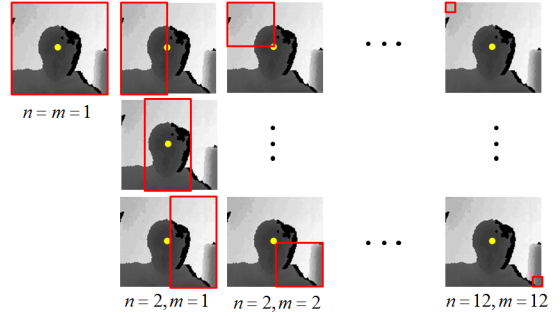
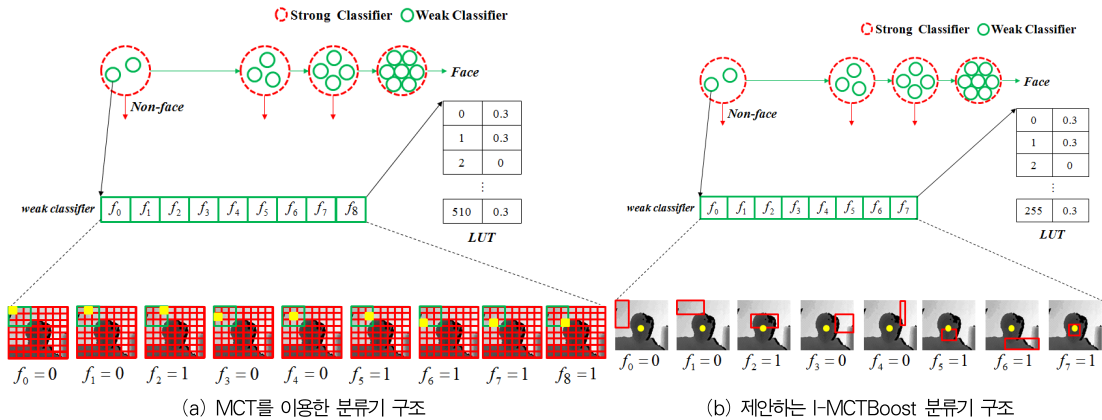


그림 4. 깊이차이특징 추출을 위한 블록 크기 예
Fig. 4. Example of block size for depth difference feature extraction

[그림 4]는 깊이차이특징을 추출하기 위해 블록의 크기를 변화시키는 과정을 나타낸 예로서 적색 사각형이 [표 1]에서 설명한 알고리즘에 의해 분할되는 블록을 의미한다. 황색 점은 학습 영상에서의 얼굴 영역의 중심점으로 특징 추출을 위해 블록의 평균과 차이를 구하기 위한 깊이 값을 얻는 위치를 의미한다.

IV. 학습 및 인식

학습 영상으로부터 특징을 추출하면 학습을 수행 할 수 있다. 본 논문에서는 III장의 2절에서 설명한 깊이차이특징을 이용하여 학습을 수행하며, 학습 과정은 크게 두 단계로 구성된



(a) MCT를 이용한 분류기 구조 (b) 제안하는 I-MCTBoost 분류기 구조

그림 5. MCT를 이용한 Boosting과 I-MCTBoost 구조
Fig. 5. The structure for Boosting with MCT and I-MCTBoosting

다. 첫 번째는 약분류기 생성 과정이며, 두 번째는 선택된 다수의 약분류기를 조합하여 강분류기를 생성하는 단계이다. 본 장에서는 먼저 제안하는 분류기 구조를 설명하고, 강분류기 학습 과정과 약분류기 학습 과정에 대해 설명한다.

1. I-MCTboost 구조

[그림 5]는 기존 MCT를 이용한 분류기와 제안하는 I-MCTBoost 분류기의 구조를 보여준다. 그림과 같이 기존 MCT에서 사용되는 약분류기는 지역적으로 이웃한 위치와 차이를 기준으로 분류 효율이 좋은 위치가 선택된다. MCT의 경우 이미지 화소값을 기준으로 계산되지만, 본 논문에서는 깊이영상을 이용하고 한 번의 스캔으로 검출 단계를 완료하기 위해 2차 다항모델에 의해 영역이 주어지면 해당 영역을 22x22로 분할하고 3x3 블록(녹색 사각형)의 깊이 값의 평균과 이웃된 블록 평균(황색 사각형)의 차이 값으로 0과 1을 결정한다.

반면에 본 논문에서 제안하는 I-MCTBoost는 중심의 깊이 값과 지역적으로 다양한 영역과 차이 값을 계산하고 임계 값 처리를 통해 0과 1이 결정된다. 또한 이들의 조합으로 약분류기가 구성된다. 따라서 기존의 MCT에 비해 좀 더 많은 특징들 중에 좋은 분류기를 선택하는 것이 가능하다.

2. 강분류기 학습

표 2. 정의
Table 2. Definition

$\cdot w = \{w^0, \dots, w^7\}, w^k = \{wb_k, wf_{basis}, w\theta_k\}$
$\cdot C = \{C^0, \dots, C^{n-1}\}, C^k = \{Cb_k, Cf_{basis}, C\theta_k\}$
$\cdot Comb(A) = \sum_{k=0}^{n-1} V_k(Ab_k, Af_{basis}, A\theta_k)$
$V_k(Ab_k, Af_{basis}, A\theta_k) = \begin{cases} 2^k & Af_{basis} - Ab_k \geq A\theta_k \\ 0 & otherwise \end{cases}$

[표 2]는 강분류기 및 약분류기 알고리즘에 사용되는 정의이다. 먼저 w 와 C 는 약분류기를 나타내며, w 는 8개의 특징으로 이루어져있고, C 는 n 개의 특징으로 구성된다. 이때 C 는 약분류기 생성 과정에서 만들어지기 때문에, 0~8개의 특징을 포함할 수 있다. 또한 특징 w^k 와 C^k 는 영상에서의 상대적인 블록 영역의 평균(wb_k, Cb_k)과 영상의 중심 화소 깊이 값(wf_{basis}, Cf_{basis}), 0과 1을 결정하기 위한 임계값

($w\theta_k, C\theta_k$)으로 구성된다. w 는 강분류기 학습 과정에서 사용되며, C 는 약분류기 학습 과정에서 사용된다. 마지막으로 $Comb(A)$ 는 조합함수로서 약분류기 A 에 포함된 특징의 수(n)에 따라 $0 \sim 2^n - 1$ 의 범위를 갖는다.

[그림 6]은 위에서 설명한 정의를 표현한 그림이다. 약분류기 w 는 8개의 깊이차이특징 $w^0 \sim w^7$ 을 포함한다. $wLut$ 는 룩업 테이블이며, w^k 는 0 또는 1을 출력하기 위해 $w^k = \{wb_i, wf_{basis}, w\theta_i\}$ 로 구성된다. wb 는 wf_{basis} 와 비교를 위한 블록의 평균 깊이 값을 의미하며, $w\theta$ 는 임계값을 나타낸다. 최종적으로 약분류기의 결과는 $Comb(.)$ 에 의해 룩업 테이블을 참조하여 결정된다. [표 2]에서 정의한 C 도 같은 구조를 가지지만 약분류기 생성 과정에서 사용되므로 w 와는 다르게 특징의 수가 유동적이다.

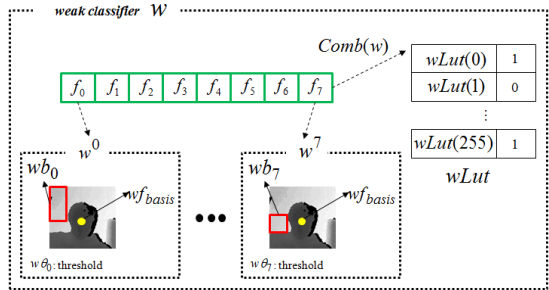


그림 6. 약분류기 구조
Fig. 6. The structure of weak classifier

강분류기와 약분류기에 대한 개념은 Freund[14]에 의해 Adaboost 분류기를 소개한 논문에서 언급되어 있다. [표 3]과 같이 분류기를 구성하는 스테이지(강분류기)를 구성하기 위한 알고리즘은 기존 MCT를 이용한 부스팅 방법과 유사하다 [11].

먼저, 학습을 위해 얼굴(Positive)과 비 얼굴(Negative) 영상으로 이루어진 학습 샘플 데이터베이스를 구성한다. 본 논문에서는 학습 데이터를 수집하기 위하여 다양한 거리에서 III장의 2절에서 구한 2차 다항모델을 이용하여 결정된 크기만큼의 영역을 분할하여 수집하였으며, 비 얼굴 영상은 얼굴을 포함하고 있지 않은 깊이영상으로부터 무작위로 추출하였다.

표 3. I-MCTBoost 강분류기 알고리즘(11)
Table 3. The algorithm for I-MCTBoost strong classifier

Input : $(\Gamma_1, c_1), \dots, (\Gamma_m, c_m)$ where $c_i = 0$ for $\Gamma_i \in F$ and $c_i = 1$ for $\Gamma_i \in B$ where F and B is the class of faces and non-faces.

Initialize : $D_1(i) = \frac{1}{2l}, \frac{1}{2n}$ for $c_i = 0, 1$ where l and n are the number of faces and non-faces.

Algorithm :

- For $t=1, \dots, T$:
 - Generate weak classifier : **Table 4**

$$w_t \leftarrow C, wLut_t \leftarrow CLut$$

- Calculate error ε_t :

$$G_t^P(r) = \sum_{i,r} D_t(i) I(\text{Comb}(w_t) = r) I(c_i = 0)$$

$$G_t^N(r) = \sum_{i,r} D_t(i) I(\text{Comb}(w_t) = r) I(c_i = 1)$$

$$\varepsilon_t = \sum_r \min(G_t^P(r), G_t^N(r))$$

- Choose $\alpha_t = \frac{1}{2} \ln \left(\frac{1 - \varepsilon_t}{\varepsilon_t} \right)$

- Update the distribution :

$$D_{t+1}(i) = \frac{D_t(i)}{Z_t} \times \begin{cases} e^{-\alpha_t} & \text{if } wLut_t(\text{Comb}(w_t)) = c_i \\ e^{\alpha_t} & \text{otherwise} \end{cases}$$

where Z_t is chosen for normalize that D_{t+1} .

- The strong classifier is combined the result of weak classifier with sample x :

$$h(x) = \sum_{t=1}^T \alpha_t wLut_t(\text{Comb}(w_t))$$

- The final strong classifier is based on the final face model :

$$H(x) = \begin{cases} non-face & \text{if } h(x) \geq T_{s\theta} \\ face & \text{otherwise} \end{cases}$$

(Γ_i, c_i) 는 i 번째 학습 데이터로서 Γ_i 는 i 번째 학습 샘플이며, c_i 는 해당 샘플의 클래스를 나타낸다. 학습 샘플이 주어

지면 샘플 가중치 D 는 얼굴과 비 얼굴 학습 샘플 수에 따라 초기화된다. 강분류기에 포함 될 약분류기 수 T 는 사용자에게 의해 입력되며, T 개의 약분류기를 선택할 때까지 반복된다.

먼저 각 학습 영상으로부터 추출된 특징을 이용하여 약분류기를 생성한다. 이때 생성되는 약분류기는 I-MCT 약분류기라 정의하고 깊이차이특징을 이용한다. I-MCT 약분류기는 [표 4]에 의해 생성된다. 약분류기가 생성되면, 생성된 약분류기 C 와 약분류기의 룩업 테이블 $CLut$ 는 $w_t, wLut_t$ 에 각각 대입된다. w_t 가 결정되면, G_t^P 와 G_t^N 을 이용하여 오류율(ε_t)을 계산하고 신뢰도 α_t 를 산출한다. 마지막으로 학습 샘플의 가중치(D)를 갱신하고 정규화를 수행한다.

강분류기 학습 과정을 수행하면 T 개의 약분류기가 선택되며, 이를 통해 강분류기가 생성된다. $h(x)$ 는 샘플 x 가 주어졌을 때 얼굴이 아닌 정도를 나타낸다. 결과적으로 $H(x)$ 는 샘플 x 를 평가함수 h 에 대입하여 선택된 모든 약분류기의 결과를 합하여 $T_{s\theta}$ 이상이라면 비 얼굴로 판단하고 반대의 경우라면 얼굴로 인식한다. $T_{s\theta}$ 는 모든 Positive 학습 샘플을 통과 시킬 수 있는 최소값으로 정한다.

3. 약분류기 학습

[그림 5]의 (b)와 같이 하나의 약분류기는 8개의 깊이차이특징으로 구성된다. 각각의 깊이차이특징은 0 또는 1을 출력하고, 이들의 조합으로 약분류기의 출력이 결정되므로 $0 \sim 2^8 - 1$ 까지의 범위를 갖는다. 약분류기를 구성하는 깊이차이특징은 영상의 중심 깊이 값과 블록 영역의 차이 값을 계산하여 임계값 이상이면 1, 그렇지 않으면 0을 출력하여 특징을 표현한다. 이때 사용되는 임계값은 모든 학습 샘플들과 샘플들의 가중치를 고려하여 오류율이 최소가 되는 임계값으로 설정된다.[10]

I-MCTBoost의 약분류기 생성과정은 [표 4]와 같다. 결과적으로 약분류기 생성 알고리즘은 수많은 특징들 중에서 효율적인 분류가 가능한 8개 깊이차이특징의 조합을 찾는 알고리즘이다. 먼저 현재 스테이지로부터 각 학습 샘플의 가중치 D 를 입력받아 tD 에 대입한다. tD 는 특징을 선택하는 시점에서만 사용되는 임시 가중치로서 각 샘플의 가중치를 나타내며, 8개의 깊이차이특징들이 저장될 C 를 초기화한다.

g_k^P 와 g_k^N 은 특징의 오류값을 계산하기 위한 배열이다. 또한 $I(x)$ 는 판단함수로 x 가 참인 경우에는 1, 거짓인 경우에는 0을 나타낸다. 따라서 $g_k^P(r, f)$ 는 임의의 특징 $f = \{fb, f_{basis}, f\theta\}$ 가 주어졌을 때 현재 분류기 C 와 새로

선택될 특징 f 에 의해 Positive 학습 샘플의 가중치가 누적 되는 히스토그램이다. 여기서 r 의 범위는 $0 \leq r < 2^{k+1}$ 이다. g_k^N 은 g_k^P 와 같이 Negative 학습 샘플에 대한 가중치가 누적되는 히스토그램이다. 따라서 주어진 특징에 대해 g_k^P 와 g_k^N 을 구함으로써 오류값 ϵ_k 를 구할 수 있다. 결국 모든 특징들 중에 현재 분류기 C 에 추가되었을 때 오류값이 가장 작은 특징을 C^k 에 추가한다. k 번째 특징이 추가되면 선택된 특징이 포함된 분류기 C 를 이용하여 $CLut$ 를 갱신하고 신뢰도 α_k 를 계산한다. 계산된 신뢰도를 바탕으로 다음 특징 선택을 위해 가중치 tD 를 갱신하며, 8개의 특징이 선택될 때까지 반복한다. 최종적으로 8개의 특징을 포함하는 약분류기 C 와 2^8 크기의 룩업 테이블 $CLut$ 를 출력한다.

표 4. I-MCTBoost 약분류기 학습 알고리즘
Table 4. Training of a I-MCTBoost weak classifier

Initialize :

$$tD \leftarrow D, C \leftarrow \emptyset$$

Algorithm :

- For $k = 0, \dots, 7$
 - Generate tables of local weighted kernel indices from faces and non-faces:

$$g_k^P(r, f) = \sum_{i, r, f} tD_k(i) I(Comb(C) + V_k(fb, f_{basis}, f\theta) = r) I(c_i = 0)$$

$$g_k^N(r, f) = \sum_{i, r, f} tD_k(i) I(Comb(C) + V_k(fb, f_{basis}, f\theta) = r) I(c_i = 1)$$

- Calculate error ϵ_k for each look-up table (f):

$$\epsilon_k(f) = \sum_r \min(g_k^P(r, f), g_k^N(r, f))$$

- Select the best classifier :

$$C^k = f | \epsilon_k(f) = \min_f \{\epsilon_k(f)\}$$

- Update look-up table for weak classifier :

$$CLut(r) = \begin{cases} 0 & \text{if } g_k^P(r) > g_k^N(r) \\ 1 & \text{otherwise} \end{cases}$$

- Choose $\alpha_k = \frac{1}{2} \ln \left(\frac{1 - \epsilon_k}{\epsilon_k} \right)$

- Update the distribution :

$$tD_{k+1}(i) = \frac{tD_k(i)}{Z_k} \times \begin{cases} e^{-\alpha_k} & \text{if } CLut(Comb(C)) = c_i \\ e^{\alpha_k} & \text{otherwise} \end{cases}$$

where Z_k is chosen for normalize that tD_{k+1} .

Output :

$$C = \{C_0, \dots, C_7\}, CLut$$

V. 실험 결과

본 논문에서는 Intel(R) Core2 Quad CPU 3.20GHz, 4G RAM 환경 하에서 실험되었으며, 마이크로소프트사의 키넥트(kinect) 센서로부터 획득한 320x240 크기의 깊이영상을 이용하여 실시간 얼굴 검출에 대해 실험하였다. I-MCTBoost 분류기의 학습에 사용된 깊이차이특징은 $n = 12, m = 12$ 로 설정하여 총 20,736개의 특징을 사용하였으며, MCT+Boost 분류기는 22x22 블록으로 나누어 총 400개의 MCT 특징을 사용하였다. 또한 학습 샘플은 Positive 약 3000장과 Negative 약 8000장으로 학습을 수행하였다. 학습 과정 중 각 스테이지 학습 과정이 종료되면, Negative 영상 집합으로부터 현재 분류기로 검출을 수행하여 Positive로 인식한 샘플을 추가 수집함으로써 학습 샘플을 갱신하였다. 몇 번의 실험결과 팔을 움직일 경우, 팔 부분에서 오검출이 빈번하게 발생하여 다양한 동작의 학습 데이터에 라벨을 하고 학습에 이용하였다.

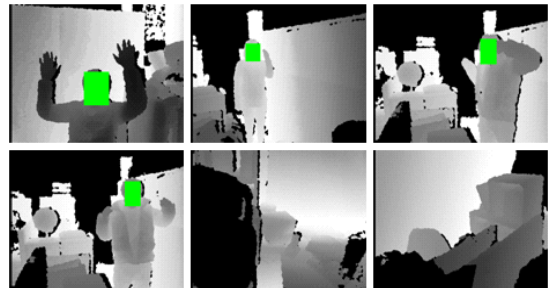


그림 7. 학습에 사용된 Negative 영상 수집
Fig. 7. Collecting negative sample images for training

[그림 7]은 추가적인 학습 데이터에 사용된 Negative 영

표 5. 분류기 성능 비교
Table. 5. Performance of Fröba's boosting and I-MCTBoost

Method	Stage	Weak Classifier	TPR	FDR	Calculation time(ms)
Fröba's boosting(11)	4	5-10-20-40	1.000	0.1452	89.04
		5-10-20-100 (*)	1.000	0.1304	92.82
		5-10-20-200	1.000	0.1397	100.33
Fröba's boosting(11)	5	5-10-20-40-211	0.997	0	101.48
	I-MCTBoost	3	2-3-3 (#)	0.995	0
1-2-1			0.968	0	17.08

상이다. 각 스테이지의 학습이 끝나면 기존 학습 샘플 중에 정상적으로 분류된 Negative 샘플은 제거하고 사전에 준비된 Negative 영상을 현재까지 학습된 분류기로 검출을 수행하였다. 만약 얼굴로 인식된다면 라벨링된 영상에서 인식된 위치가 녹색 사각형 내부에 있는지 확인하고, 내부에 없다면 Negative 학습 샘플에 추가하여 학습을 수행하였다.

[표 5]는 MCT를 이용한 Fröba의 방법(11)과 제안된 I-MCTBoost 방법의 오류발견율(False discovery rate, FDR)과 검출율(True positive rate, TPR), 연산속도를 비교하여 나타낸 것이다. 이때, 검출율은 $TPR = TP / (TP + FN)$, 오류발견율은 $FDR = FP / (FP + TP)$ 으로 계산하여 분류기의 검출 성능을 측정하였다.

[표 5]의 두 번째 열은 스테이지 즉, 강분류기 개수를 의미하고, 3번째 열은 각 강분류기마다 구성된 약분류기의 개수를 나타낸다. Fröba의 방법의 경우 전반적으로 검출성능은 좋았으나 오검출이 발생하는 경우가 많았으며, 5개의 강분류기로 분류기를 생성한 경우의 성능이 가장 뛰어났다. 제안한

방법의 경우 3개의 강분류기로 구성하고 각 단계에서 2-3-3개의 약분류기로 구성된 경우가 가장 좋은 결과를 나타내었음을 확인할 수 있다. 그러나 5개의 강분류기로 구성된 Fröba의 방법의 경우 본 논문에서 제안한 방법과 검출율은 비슷하였으나, 연산속도는 약 3배정도 느린 결과를 보였다. 이는 제안한 분류기보다 많은 특징을 사용했으며 각 특징을 계산하는데 좀 더 많은 시간이 필요하기 때문이며, 따라서 실시간 얼굴 검출에 적용하기에는 다소 무리가 있다.

[그림 8]은 Fröba의 방법(표 5의 * 표시된 분류기)과 I-MCTBoost 방법(# 표시된 분류기)을 이용하여 테스트 영상에서 얼굴 검출을 수행한 결과를 나타낸다. 연속 영상에서 동일한 프레임은 무작위로 선정하여 비교하였다. Fröba의 방법은 얼굴 회전이나 거리에 따라 오검출이나 미검출이 발생하는데 반해, 본 논문의 I-MCTBoost 방법은 오검출이 발생하지 않기 때문에 얼굴 검출의 정확도 면에서 성능이 훨씬 우수한 것을 확인할 수 있다. 또한 다중 사용자에 대해 실험한 결과 I-MCTBoost는 오검출이 거의 발생하지 않고 검출 성능

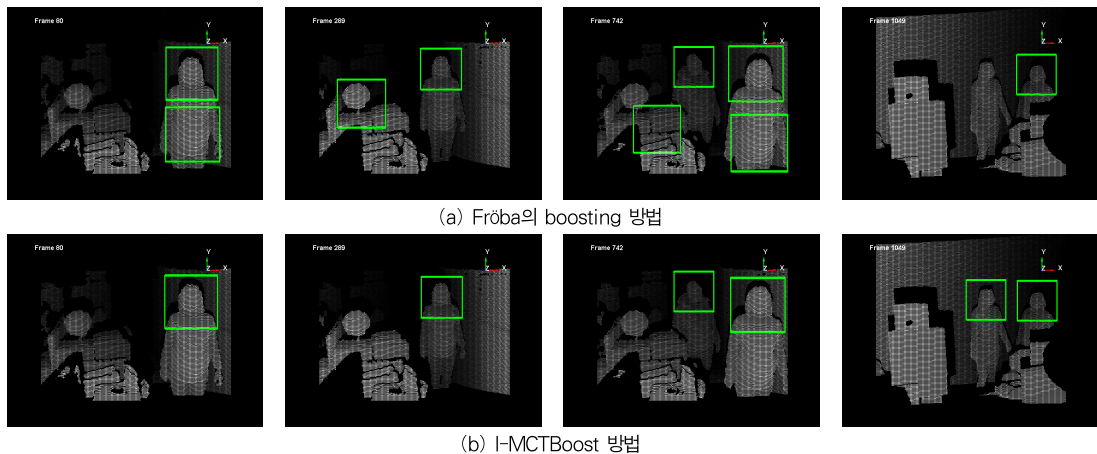


그림 8. Fröba의 boosting과 I-MCTBoost 얼굴 검출 결과
Fig. 8. Result of face detection with Fröba's boosting and I-MCTBoost

이 우수하여 사용자 인식을 위한 실시간 얼굴 검출이 가능한 것을 확인할 수 있다.

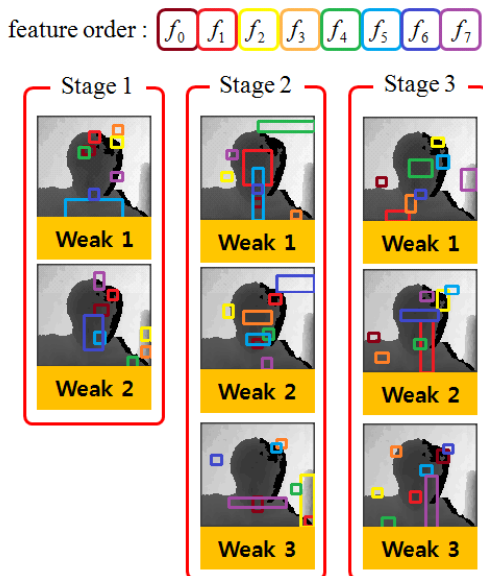


그림 9. I-MCTBoost 약분류기 예
Fig. 9. Example of I-MCTBoost weak classifier

[그림 9]는 I-MCTBoost를 이용하여 실제로 학습 과정에서 생성된 분류기를 보여준다. 각각의 영상 내부에 표시된 8개의 사각형은 약분류기의 학습과정을 통해 추출된 깊이차이 특징을 의미하고, 색상은 특징의 순서를 나타낸다. 각 특징들은 다양한 크기의 사각형에 의해 $f_0 \sim f_7$ 로써, 총 8개의 비트로 표현된다.

그림에서와 같이 특징들의 위치는 대부분 얼굴 영역 내부에 많이 분포하며, 배경을 판단할 수 있는 외부와 경계 부근에 존재함을 확인하였다.

VI. 결 론

본 논문에서는 실시간 얼굴 검출이 가능하도록 깊이영상을 이용하는 깊이차이특징을 이용하고 빠르고 정확한 성능을 위해 I-MCTBoost를 제안하였다. 얼굴의 크기변화나 회전 등에 강건하도록 2차 다항모델을 이용한 깊이차이특징을 적용하였고, 이를 비트로 표현함으로써 약분류기를 학습하고 강분류기를 구성하는 I-MCTBoost를 통해 기존의 MCT를 이용한 부스팅 기반 분류기보다 훨씬 적은 연산시간으로 높은 검출 성능을 확인하였다. 또한 조명의 변화와 무관한 깊이영상

을 활용함으로써 실제 사용자 인식을 위한 기반 연구로서의 가능성을 입증하였다. 향후 상황 인식을 위해 검출된 얼굴에 의해 해당 사용자의 몸체를 분할하기 위한 연구가 추가적으로 필요할 것이다.

참고문헌

- [1] G. Yang and T. S. Huang, "Human face detection in a complex background", The Journal of Pattern Recognition Society, Vol. 27, No. 1, pp. 53-63, January 1994.
- [2] D. Ghimire and J. W. Lee, "A Robust Face Detection Method Based on Skin Color and Edges", The Journal of Information Processing Systems, Vol. 9, No. 1, pp. 141-156, March 2013.
- [3] H. C. Vijaya Lakshmi, and S. PatilKulkarni, "Segmentation Algorithm for Multiple Face Detection in Color Images with Skin Tone Regions using Color Spaces and Edge Detection Techniques", The International Journal of Computer Theory and Engineering, Vol. 2, No. 4, pp. 553-558, August 2010.
- [4] T. Ahonen, A. Hadid and M. Pietikäinen, "Face Description with Local Binary Patterns: Application to Face Recognition", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 28, No. 12, pp. 2037-2041, December 2006.
- [5] A. Lanitis, C.J. Taylor and T. F. Cootes, "An Automatic Face Identification System Using Flexible Appearance Models", The Journal of Image and Vision Computing, Vol. 13, No. 5, pp. 393-401, June 1995.
- [6] T. F. Cootes, C. J. Taylor, D. H. Cooper and J. Graham, "Active shape models - their training and application", The Journal of Computer Vision and Image Understanding, Vol. 61, No. 1, pp. 38-59, January 1995.
- [7] T. F. Cootes, G. J. Edwards, and C. J. Taylor, "Active appearance models", IEEE Transactions on Pattern Analysis and Machine Intelligence,

Vol. 23, No. 6, pp. 681-685, June 2001.

[8] L. Wiskott, J. M. Fellous, N. Krüger, and C. Malsburg, "Face Recognition by Elastic Bunch Graph Matching", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 19, No. 7, pp. 775-779, July 1997.

[9] M. A. Turk and A. P. Pentland, "Face Recognition Using Eigenfaces", IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), pp. 586-591, June 1991.

[10] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features", In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Vol. 1, pp. I-511-I-518, December 2001.

[11] B. Fröba and A. Ernst, "Face Detection with the Modified Census Transform", In Proceedings of the Sixth IEEE International Conference on Automatic Face and Gesture Recognition (FGR'04), pp. 91-96, May 2004.

[12] D. H. Kim, W. H. Yun, Y. J. Cho and J. E. Lee, "Detection of Faces Located at a Long Range with Low-resolution Input Images for Mobile Robots", The Journal of Korea Robotics Society, Vol. 4, No. 4, pp. 257-264, November 2009.

[13] S. I. Joo, S. H. Weon and H. I. Choi, "Real-time Hand Region Detection based on Cascade using Depth Information", KIPS Transactions on Software and Data Engineering, Vol. 2, No. 10, pp. 713-722, October 2013.

[14] Y. Freund and R. E. Shapire, "A short introduction to boosting", In Journal of Japanese Society for Artificial Intelligence, Vol. 14, No. 5, pp. 771-780, September 1999.

저 자 소 개



주 성 일
 2008 : 한국산업기술대학교
 컴퓨터공학과(공학사)
 2010 : 송실대학교
 미디어학과(공학석사)
 현 재 : 송실대학교
 미디어학과 박사과정
 관심분야: Image Processing,
 Computer Vision,
 Pattern Recognition,
 Machine Learning.
 Email : sijoo82@ssu.ac.kr



원 선 희
 2005 : 한경대학교
 컴퓨터공학과(공학사)
 2007 : 송실대학교
 컴퓨터학과(공학석사)
 2012 : 송실대학교
 미디어학과(공학박사)
 현 재 : 송실대학교
 미디어학과 Post Doc.
 관심분야: Image Processing,
 Computer Vision,
 Pattern Recognition,
 Machine Learning.
 Email : nifty12@ssu.ac.kr



최 형 일
 1972 : 연세대학교
 전자공학과(공학사)
 1982 : 미시간대학교
 전자공학과(공학석사)
 1987 : 미시간대학교
 전자공학과(공학박사)
 1995 : 퍼지 및 지능시스템학회 이사
 1996 : 정보과학회 컴퓨터비전 및
 패턴인식 연구회 위원장
 1997 : IBM Waston Lab 방문연구원
 2005 : 한국정보과학회 이사
 현 재 : 송실대학교 미디어학과 교수
 관심분야: 컴퓨터공학
 Email : hic@ssu.ac.kr