

논문 2010-47SP-2-4

고성능 실시간 얼굴 검출 엔진의 설계 및 구현

(Design and Implementation of Real-time High Performance Face Detection Engine)

한 동 일*, 조 현 중**, 최 중 호**, 조 재 일***

(Dongil Han, Hyunjong Cho, Jongho Choi, and Jae-Il Cho)

요 약

본 논문에서는 로봇 시각 처리 활용을 위한 실시간 얼굴 검출 하드웨어 구조를 제안한다. 제안한 구조는 조명 변화에 강인하고 초당 60 프레임 이상의 속도로 처리된다. 조명 변화에 강인한 얼굴 특성 추출을 위해 MCT(Modified Census Transform) 변환을 이용하였다. 그리고 AdaBoost 알고리즘은 얼굴 특징 데이터의 학습 및 생성을 하며, 이 생성된 학습 데이터를 이용해 얼굴 검출을 하게 된다. 본 논문에서는 메모리 인터페이스부, 이미지 크기 조정부, MCT 생성부, 후보 얼굴 검출부, 신뢰도 비교부, 좌표 재조정부, 데이터 그룹화부, 검출 결과 표시부로 구성된 얼굴 검출 하드웨어 구조 및 Xilinx사의 Virtex5 LX330 FPGA를 이용한 하드웨어 구현 검증 결과에 대해 설명한다. 카메라로 부터 입력받은 이미지를 이용해 검증한 결과로 초당 최대 149프레임의 속도로 한 프레임 당 최대 32개 얼굴을 검출함을 확인하였다.

Abstract

This paper propose the structure of real-time face detection hardware architecture for robot vision processing applications. The proposed architecture is robust against illumination changes and operates at no less than 60 frames per second. It uses Modified Census Transform to obtain face characteristics robust against illumination changes. And the AdaBoost algorithm is adopted to learn and generate the characteristics of the face data, and finally detected the face using this data. This paper describes the face detection hardware structure composed of Memory Interface, Image Scaler, MCT Generator, Candidate Detector, Confidence Comparator, Position Resizer, Data Grouper, and Detected Result Display, and verification Result of Hardware Implementation with using Virtex5 LX330 FPGA of Xilinx. Verification result with using the images from a camera showed that maximum 32 faces per one frame can be detected at the speed of maximum 149 frame per second.

Keywords : Modified Census Transform, Adaboost Algorithm, Face Detection, Real-Time, FPGA Implementation, Hardware Design

I. 서 론

사람의 신체 정보를 이용하는 생체 인식 기술의 발전에 따라 얼굴, 홍채, 지문, 망막 등을 이용한 신원 인증 방법은 기존의 열쇠, 번호 등을 이용한 방법을 대체해 가고 있는 실정이다. 특히 얼굴 인식의 경우 성형, 얼굴 표정 변화 등의 가변적 요소가 커서 다른 생체 정보를 이용한 신원 인증 방법에 비해 상대적으로 인식률이 낮다는 단점에도 불구하고, 비접촉식의 편의성과 낮은 비용으로 설치 및 운용이 가능하다는 점 등의 장점을 바탕으로 현재 일부 정부 기관 및 기업, 은행 등 높은 보안을 요구하는 곳에서 다른 생체 정보 기반 인증 기술

* 정회원, ** 학생회원, 세종대학교 컴퓨터공학과
(Dept. of Computer Engineering,
Sejong University)

*** 정회원, 한국전자통신연구원 로봇연구부
(Robot Research Department, ETRI)

※ 본 연구는 지식경제부 산업원천기술개발사업[과제번호: KI001813, 과제명: u-로봇 HRI 솔루션 및 핵심 소자 기술 개발]과 한국연구재단의 기초연구사업(No. 2009-0072846)에서 지원되었습니다.

접수일자: 2009년8월30일, 수정완료일: 2010년2월16일

과 함께 사용되고 있으며^[1] 개인용 노트북 컴퓨터에서도 얼굴 인식을 통한 사용자 인증을 함으로서 그 활용 범위가 점차 넓어지고 있다. 이런 얼굴 인식이 신뢰할 만한 인식률을 얻기 위해선 영상에서 정확히 얼굴 부분만을 추출해 낼 수 있는 얼굴 검출 기술이 필수적이라고 할 수 있다.

지금까지 얼굴 검출 알고리즘은 주로 PC 기반 환경에서 동작하도록 개발되어 왔으며, 이 기술을 바로 임베디드 시스템에 적용할 경우 상대적으로 부족한 리소스와 성능 때문에 실시간으로 얼굴 검출이 불가능하거나 낮은 검출률을 보임으로서 효율적인 얼굴 검출이 힘들었다. 그러나 서비스 로봇 분야의 확대로 인해 일상생활 속에 청소 로봇, 엔터테인먼트용 장난감 로봇 등 가정용 로봇 시장이 커지고 또한 휴대 전화, 디지털 카메라 등의 휴대 기기 등에서 인간의 얼굴 정보를 사용할 수 있는 활용 분야가 늘어나면서 임베디드 시스템에서의 고성능 실시간 얼굴 검출 기술의 필요성은 더욱 더 커져가고 있다.

한편 이러한 얼굴 검출 성능에 가장 영향을 많이 미치는 요소로 조명 변화, 얼굴의 회전, 얼굴 표정 변화 등이 알려져 있다. 이 중에서 얼굴의 회전이나 표정의 변화는 사용자가 얼굴 검출 시스템에 협동적일 때 그 영향력이 줄어들 수 있지만, 조명 변화의 경우 사용자의 의지가 아니라 실내/실외 등의 주변 환경, 시간, 빛의 반사등의 이유로 인해 발생할 수 있기 때문에 조명 변화에 강인한 얼굴 검출은 여전히 어려운 부분으로 남아 있다.

본 논문에서는 조명 보상 작업에 필요한 비용을 줄이기 위해 사물의 구조적 정보만을 추출함으로써 조명 변화의 영향을 줄이는 MCT (Modified Census Transform)^[2] 변환을 이용해 영상을 변환하였으며, 이렇게 변환된 영상을 이용해 임베디드 시스템에서 조명 변화에 강인한 실시간 얼굴 검출을 수행하는 하드웨어 구조를 제시하였다. 또한 Vritex5 LX330 FPGA를 이용해 제시된 하드웨어 구조를 구현한 후 다양한 조명 변화를 갖는 실생활 환경에서 실험함으로써 그 성능을 검증하였다.

본 논문의 구성은 다음과 같다. II장에서는 본 논문에서 사용된 MCT 변환과 이렇게 변환된 영상을 이용해 학습 데이터를 만드는 Adaboost 알고리즘에 대해 설명하고, III장에서 본 논문에서 제안하는 전체 하드웨어 구조 및 세부 구조에 대해 설명한다. IV장에서는 이

하드웨어 구조를 PC환경에서 동작시킨 결과와 FPGA에서 구현한 후 실생활 환경에서 실험한 결과를 분석하고, 마지막으로 V장에서 결론 및 향후 연구 방향에 대해 논한다.

II. MCT 변환과 Adaboost 학습 알고리즘

조명에 의한 사물 외관의 왜곡 현상을 줄이는 문제는 얼굴 인식 분야를 포함한 거의 모든 패턴 인식 분야에서 해결해야 할 큰 문제로 남아있다. 얼굴 인식을 크게 나뉘보면 영상에서 얼굴 영역을 검출하는 얼굴 검출부와 이렇게 검출된 얼굴 영역을 이용해 신원을 인증하는 얼굴 인증 부분으로 나눌 수 있다^[1]. 이때 조명은 얼굴 검출부 뿐만 아니라 얼굴 인증부에도 크게 영향을 미치기 때문에 이를 해결하기 위한 연구가 많이 진행되고 있다. 하지만 이런 조명 보상은 때론 실제 얼굴 검출이나 인식에 필요한 시간 보다 더 많은 연산 능력이 필요할 경우도 있으며^[3], 흔히 사용되는 히스토그램 균등화 (histogram equalization)가 임베디드 시스템에서 적용될 경우 전체 연산능력의 10% ~ 15% 정도를 사용함으로써^[4] 실시간 얼굴 인식 및 검출에 큰 부담을 주고 있다.

본 장에서는 이런 조명 보상에 필요한 비용을 최소화 하면서 영상에서 인식에 필요한 특징들을 추출하는 MCT 변환과 이렇게 얻어진 특징들을 효과적으로 훈련시키는 Adaboost 학습 알고리즘에 대해 기술한다. 또한 임베디드 시스템에서 학습 데이터의 저장을 위해 사용되는 메모리 공간을 최소화하기 위해 기존의 캐스케이드(cascade) 구조가 아닌 단층 구조와 이로 인한 성능과 속도의 변화에 대해서도 기술한다.

2.1 MCT 변환

영상으로부터 얼굴의 구조적 특징 정보만을 추출한 후 이를 이용함으로써 얼굴 인식 또는 검출 시스템의 성능 향상을 기대할 수 있다. 구조적 특징정보를 얻기 위한 도구로 MCT 변환을 들 수 있다. 영상의 각 화소의 위치 $X = (x, y)$ 에 해당하는 영역의 영상 밝기 값 $I(X)$ 는 다음과 같이 정의될 수 있다.

$$I(X) = gL(X)R(X) + b \quad (1)$$

여기서 $R(X)$ 는 X 위치에서의 반사 값이며 $L(X)$ 는 조명 값이다. 또한 g 는 카메라의 비례 계수이며 b 는 바

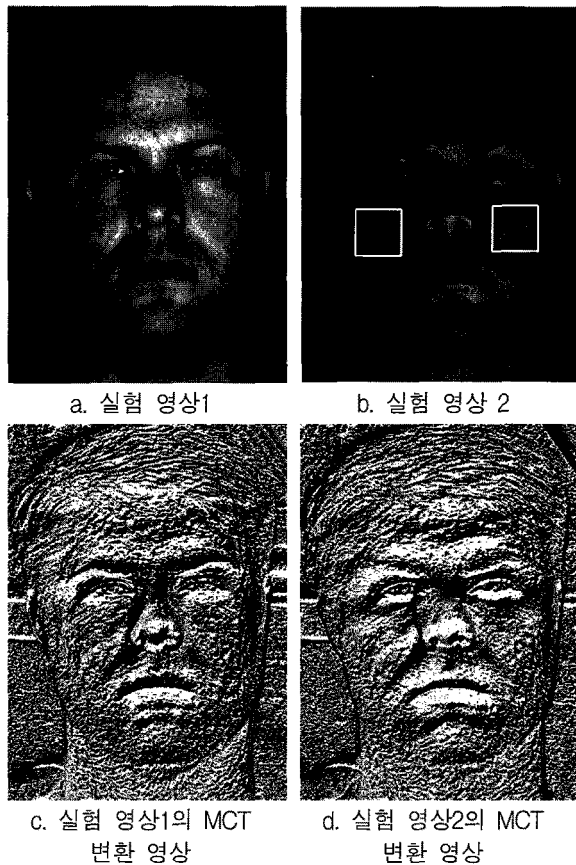


그림 1. 서로 다른 조명 조건을 갖는 두 영상의 MCT 변환
 Fig. 1. Gray level and MCT images in different illumination level.

이어스 값으로 이 두 값은 영상 평면에서 상수로 취급할 수 있다.

조명에 강인한 얼굴검출을 하기 위해서는 $L(X)$ 가 모든 X 에 대해 동일할 때 얼굴의 반사 성질 $R(X)$ 에 의해서만 결정되는 $I(X)$ 를 고려해야 하지만, 실생활의 영상에서는 조명의 위치나 밝기에 의해 X 가 변할 때 $L(X)$ 가 변하기 때문에 $L(X)$ 가 고정되어 있거나 계산해 낼 수 없다는 가정 없이는 조명에 강인한 $I(X)$ 를 얻을 수 없다. 하지만 영상의 작은 범위 내에서는 $L(X)$ 가 변하지 않거나 혹은 변하더라도 아주 작게 변하기 때문에 $I(X)$ 에 큰 영향을 미치지 않는다고 가정할 수 있다. 즉 그림 1의 b. 와 같은 영상에서 얼굴의 왼쪽 편과 오른쪽 편은 조명으로 인해 서로 다른 $L(X)$ 을 갖게 되지만, 각각의 사각형 영역 안에서는 $L(X)$ 가 거의 변하지 않기 때문에 지역적으로 조명에 강인한 $I(X)$ 를 얻을 수 있다.

MCT변환은 위와 같은 사실을 이용하여 영상을 변환

한다. 크기가 정해져 있는 것은 아니지만 여기서는 $L(X)$ 가 거의 동일하다고 가정할 수 있을 만큼 작은 3×3 크기의 윈도우를 영상 내에서 이동시키며 윈도우 내의 구조 정보를 $\{0, 1\}$ 의 이진 형식으로 나타낸다. 이때 이런 이진 형식 정보는 사물의 윤곽, 굴곡, 교차점 등의 정보를 담고 있다. 이런 MCT 변환은 다음과 같은 식으로 정의할 수 있다.

$$\Gamma(X) = \bigotimes_{Y \in N'} \zeta(\bar{I}(X), I(Y)) \quad (2)$$

여기서 X 는 영상 내의 화소의 위치를 나타내고 X 를 중심으로 하는 3×3 윈도우를 $W(X)$ 라 한다면, N' 는 $W(X)$ 내의 화소들의 집합이고 Y 는 이 윈도우 내의 9개의 화소 각각을 나타낸다. 또한 $\bar{I}(X)$ 는 윈도우 내의 화소들의 평균값이고 $I(Y)$ 는 윈도우 내의 화소들의 각각의 밝기 값이다. $\zeta(\cdot)$ 는 비교함수로서 만약 $\bar{I}(X) < I(Y)$ 일 경우 $\zeta(\bar{I}(X), I(Y))$ 는 1이 되고 그 이외의 경우 0 이 된다. \otimes 는 집합 연산자로 $\zeta(\cdot)$ 함수의 결과 이진 형식들을 연결해 주는 역할을 하며, \otimes 연산을 통해 9 개의 이진 형식이 연결되게 된다. 이론적으로 9개의 화소 값이 모두 1 이 되는 경우는 발생하지 않기 때문에 총 511개의 구조가 표현될 수 있다. 이렇게 연결된 이진 형식은 10진수로 변환된 후 MCT 변환 영상의 화소 값이 된다^[2]. 아래의 그림은 임의의 영상 내의 위치 X 에 중심을 둔 윈도우 $W(X)$ 와, 이를 MCT 변환 한 후 영상의 X 위치 값의 변화를 나타낸다.

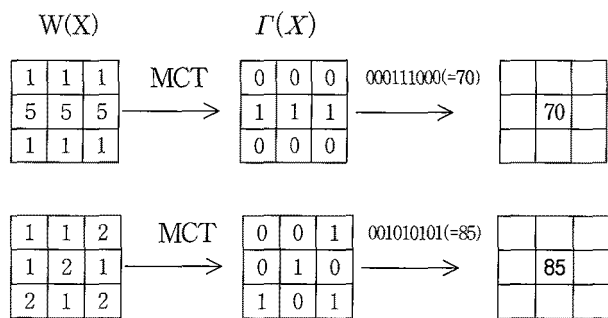


그림 2. MCT 변환의 예
 Fig. 2. Transform example of image to MCT value.

2.2 Adaboost 학습 알고리즘

본 논문에서는 얼굴 검출을 위한 사전 단계로서 먼저 MCT 변환 된 학습 영상들을 이용해서 얼굴 특징 데이터를 만든다. 이후 이 데이터를 이용하는 얼굴 검출기

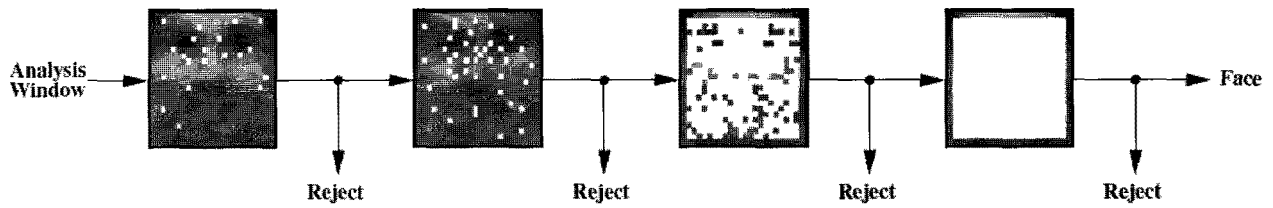


그림 3. MCT를 이용한 얼굴 검출기의 4단계 캐스케이드 구조[2]

Fig. 3. The four classifiers of Cascade Detector [2].

(the white dots mark positions of elementary classifiers in the analysis window)

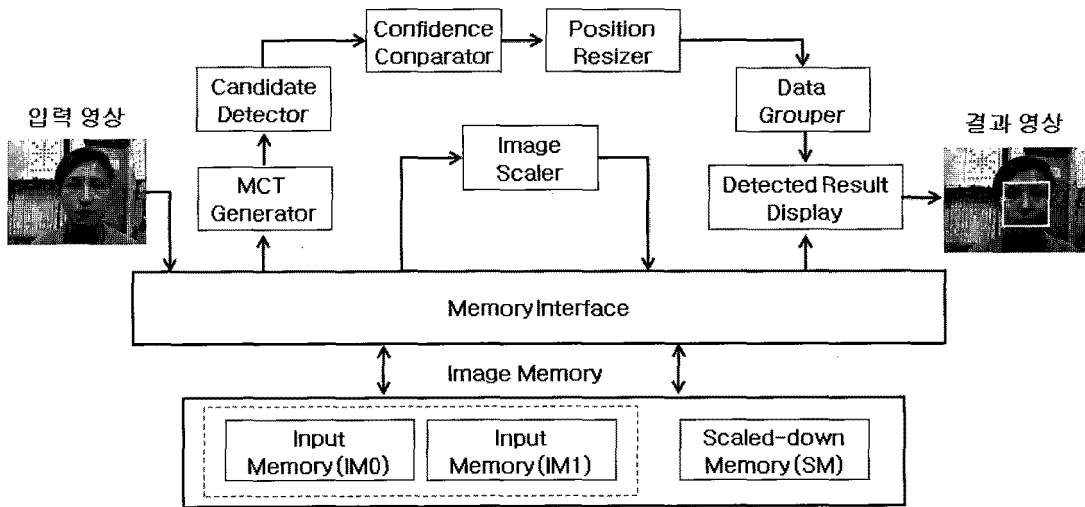


그림 4. 제안된 하드웨어 전체 구조도

Fig. 4. Proposed Block Diagram of Face Detection Engine.

가 실험 영상을 이용하여 얼굴인지 여부를 판단하게 된다. Viola 와 Jones^[5]는 Freund와 Schapire^[6]에 의해 제안된 Adaboost 학습 알고리즘을 이용해 효과적으로 얼굴을 구분하기 위한 특징을 추출하고 이 특징들을 38단계의 캐스케이드 구조로 구성함으로써 빠르면서도 높은 성능을 보이는 얼굴 검출 알고리즘을 제안하였다. 또한 Fröba와 Ernst는 MCT 변환 된 영상들과 Adaboost 학습 알고리즘을 이용해 그림 3과 같이 4단계로 이루어진 캐스케이드 구조로 된 얼굴 검출기를 소개했다^[2]. 이런 캐스케이드 구조는 초기 단계에서 변별력이 높은 소수의 특징들을 사용함으로써 명확히 얼굴이 아니라고 판단되는 부분 영상을 초기에 검출 대상에서 제거하고, 후기 단계로 갈수록 변별력은 상대적으로 낮지만 더 많은 특징들을 사용함으로써 판단이 어려운 부분 영상을 집중적으로 검사하게 된다. Fröba와 Ernst가 제안한 4단계의 캐스케이드 구조에서는 처음 1단계에 모든 얼굴 부분 영상은 통과시킴과 동시에 비얼굴 부분 영상을

99% 이상 제거함으로써 높은 속도 향상을 보였다.

이렇게 캐스케이드 구조를 갖기 위해서는 각 단계에서 필요한 특징들에 대한 위치 정보와 함께 그 특징 값에 대응되는 얼굴 신뢰도 값을 룩업 테이블의 형태로 저장하고 있어야 한다. 하지만 PC 환경에 비해 리소스의 제약이 있는 임베디드 시스템에서 이 모든 정보를 담고 있기에 메모리의 양이 일반적으로 충분하지 않다.

또한 하드웨어 구현을 이용할 경우 여러 작업의 동시 수행으로 소프트웨어에 비해 빠른 수행이 가능하다는 장점이 있지만 기존 다층 구조의 장점은 없어지고 필요 자원만 증가한다는 단점이 발생한다^[7]. 따라서 본 논문에서는 Fröba와 Ernst가 제안한 캐스케이드 구조의 마지막 4번째 단계만을 사용하는 단층 구조로 얼굴 검출기를 구성하였다.

검출 방식은 이미지 축소방식과, 검출 윈도우의 크기를 조절하는 방식이 있는데, 본 연구에서는 이미지 축소 방식을 이용하였다. 영상을 스케일 다운해가며

MCT 변환하고, 영상에 20x20 탐색 윈도우를 통과 시켜 얼굴 후보 영역 검출을 시도 하였으며, 검출 윈도우는 좌상단부터 행방향 우선 순으로 이동하며, 윈도우 안의 모든 화소의 MCT 값에 따라 얼굴인지 아닌지 여부를 판단하게 된다. 실험 결과 이렇게 단층 구조를 사용하더라도 검출성능에 큰 변화가 없어 충분한 검출 성능을 보임을 확인할 수 있었다. 검출 성능에 대한 자세한 내용은 IV장에서 논하기로 한다.

III. 하드웨어 구조

본 논문에서 제안된 전체 하드웨어 구조는 그림 4와 같다. 영상을 저장하기 위해 사용되는 메모리와 데이터의 흐름을 제어하는 역할을 메모리 인터페이스(Memory Interface) 블록이 하고 있으며 영상 크기 조정(Image Scaler) 블록은 영상의 크기를 단계적으로 줄이는 역할을 수행한다. MCT 생성(MCT Generator) 블록은 영상을 MCT 변환하는 역할을 수행하며 후보얼굴 검출(Candidate Detector) 블록은 윈도우가 현재 위치하고 있는 20x20 부분에 해당하는 화소 400개에 대한 MCT 값을 신뢰도 비교부(C Confidence Comparator) 블록에 전달하는 역할을 한다. 신뢰도 비교부는 오프라인 상에서 미리 학습되어 롬 테이블에 저장되어 있던 신뢰도 값을 이용해 그 화소가 얼굴 영역에 해당하지 않을 신뢰도 값을 구하고 400개의 신뢰도 값을 누적해서 더해간다. 이렇게 더해진 총 신뢰도 값은 미리 정해져 있던 임계값과 비교 되어 얼굴일 가능성이 있는 후보 영역을 찾는 역할을 한다. 좌표 재조정부(Position Resizer) 블록은 후보 얼굴 검출부와 신뢰도 비교부 블록을 통해서 얼굴 후보로 판단된 영역이 320x240 크기의 원 영상에서 어느 곳에 위치하는 지를 계산해내는 블록이다. 신뢰도 비교부와 좌표 재구성부에 의해 찾아진 얼굴 영역은 얼굴일 가능성이 있는 후보영역이며 이 가운데 얼굴이 아닌 배경영역을 찾아서 제거하여 최종 얼굴 영역을 결정하는 동작이 데이터 그룹화(Data Grouper) 블록에서 수행되며 마지막으로 검출 결과 표시부(Detected Result Display) 블록은 데이터 그룹화부에 의해 얼굴 영역이라고 판명된 영역에 해당하는 곳을 원 영상에 □ 모양의 박스를 덮어씌운 후 디스플레이 장치로 출력하는 역할을 수행한다. 다음 절에 각 세부 블록의 구조 및 설명을 기술하였다.

3.1 메모리 인터페이스

메모리 인터페이스 블록은 카메라로부터 들어오는 입력 영상을 처리하기 위한 두 개의 입력 메모리와, 영상 크기조절 된 영상을 처리하는 크기조정 메모리에 영상 데이터를 입출력하는 역할을 수행한다. 또한 이 블록은 영상을 처리하기 위해 필요한 영상크기 정보와 동기신호 정보 등을 생성하고 다른 블록으로 전송하는 역할을 수행한다. 그림 5는 이 블록에서 입력 메모리에 영상 데이터를 입출력하는 타이밍을 나타낸다.

카메라로부터 연속적으로 영상 프레임이 들어올 때 (F0~F4) 각 프레임은 입력 메모리0 (IM0)와 입력 메모리1 (IM1)에 번갈아가며 입력되게 된다. 이렇게 입력된 프레임은 다음 프레임이 IM0와 IM1에 입력되기 이전에 출력됨으로서 결과적으로 연속적으로 입력되는 모든 영상 프레임들이 처리되게 된다.

입력 메모리에 저장된 영상은 영상 크기 조정 블록과

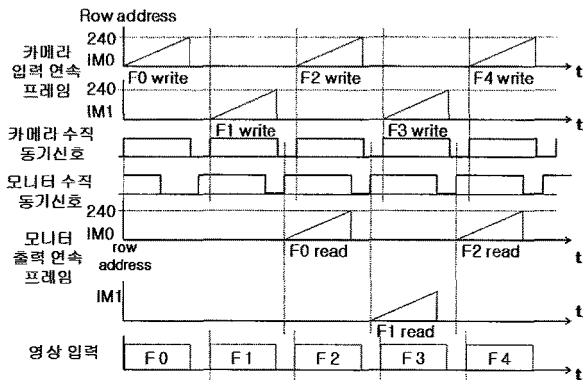


그림 5. 입력 메모리의 타이밍도
Fig. 5. Input Memory write and read Timing Diagram.

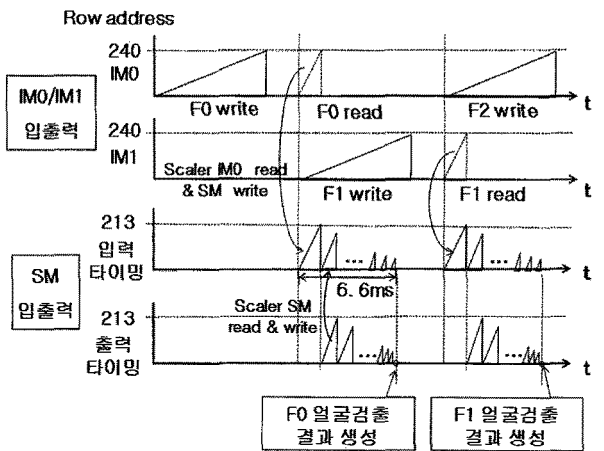


그림 6. 크기 조정 메모리의 타이밍도
Fig. 6. Scale-down Memory write and read Timing Diagram.

MCT 생성 블록으로 인가된다. 이 때 영상 크기 조정 블록으로 입력된 영상은 크기 축소 작업이 수행되며, 그 결과 영상이 크기조정 메모리로 저장되게 된다. 크기조정 메모리의 데이터 입출력 타이밍은 그림 6과 같다.

크기조정 메모리에 저장된 영상 크기 조정 블록의 결과 영상은 크기조정 메모리로부터 출력되어 영상 크기 조정 블록의 입력 영상으로 다시 사용되는 과정이 반복된다. 결과적으로 하나의 크기조정 메모리를 이용해서 영상 크기 조정 블록에서 수행되는 13번의 크기 축소 결과 영상들을 저장하게 된다.

3.2 영상 크기 조정

영상 크기 조정 블록은 카메라로부터 입력된 영상의 크기를 단계적으로 줄이는 역할을 하며 총 13번의 영상 크기 축소 작업이 수행된다. 영상의 크기를 줄이는 이유는 영상에서 얼굴을 찾는 윈도우의 크기가 20x20으로 고정되어 있기 때문에 영상에서 나타날 수 있는 다양한 크기의 얼굴을 찾기 위해서 원 영상의 크기를 단계적으로 줄여가며 각 단계마다 윈도우로 얼굴을 찾는 것이다. 결과적으로는 영상에서 차지하는 얼굴의 크기에 상관없이 얼굴 검출이 가능하게 한다.

영상 크기 조정 블록에서 영상의 축소과정은 양선형(Bilinear) 보간법을 이용한다. 이 블록에 입력된 영상은 우선 수평 방향으로 보간법이 수행된다. 이렇게 줄어든 영상은 모든 수평 라인 각각이 라인 메모리에 번갈아가

표 1. 단계별 변환되는 크기 및 소요시간 (54MHz 사용시)

Table 1. Size changes and Required time by Phases : when using 54 MHz clock.

단계	영상 크기	소요 시간	누적 시간
0 (원영상)	320 x 240	1.464 ms	1.464 ms
1	284 x 213	1.152 ms	2.616 ms
2	252 x 189	0.910 ms	3.526 ms
3	224 x 168	0.718 ms	4.244 ms
4	199 x 149	0.561 ms	4.805 ms
5	176 x 132	0.440 ms	5.245 ms
6	156 x 117	0.344 ms	5.589 ms
7	138 x 104	0.268 ms	5.857 ms
8	122 x 92	0.210 ms	6.067 ms
9	108 x 81	0.166 ms	6.233 ms
10	96 x 72	0.129 ms	6.363 ms
11	85 x 64	0.101 ms	6.464 ms
12	75 x 56	0.078 ms	6.542 ms
13	66 x 49	0.060 ms	6.602 ms

며 저장되게 되고, 이 라인 메모리에 저장된 영상이 수직 방향 크기 조정에 의해서 보간 됨으로서 전체적으로 수평, 수직 방향으로 크기가 줄어든 영상이 만들어지게 된다. 다음의 표 1은 각 단계마다 줄어든 영상의 크기 변화와 이를 위해 소요되는 시간을 나타낸다.

3.3 MCT 생성

본 블록은 내부적으로 크게 윈도우 인터페이스부와, 계산부로 나눌 수 있다. 윈도우 인터페이스부는 영상 내에서 3x3 윈도우를 이동시키며 윈도우 내의 9개의 화소를 동시에 계산부로 전송하는 블록으로 이를 위해 두 개의 라인 메모리를 사용하고 있다. 이렇게 인터페이스부에서 추출된 3x3 윈도우는 계산부로 전달되어 계산부에서 MCT 연산을 수행하게 된다. 본 블록의 블록도는 그림 7과 같다.

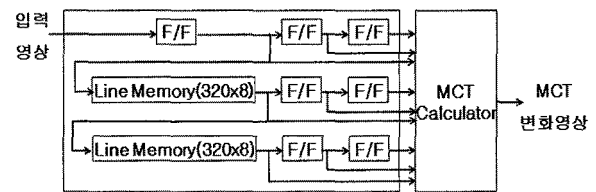


그림 7. MCT 생성 부 블록도
Fig. 7. MCT Generator block diagram.

3.4 후보 얼굴 검출 및 신뢰도 비교

변환된 MCT 영상은 20x20 탐색 윈도우를 통해 얼굴 후보 영역 검출부로 인가된다. 검출 윈도우는 좌 상단부터 행방향 우선 순으로 한 화소씩 이동시키며, 윈도우 안의 모든 화소는 MCT 값에 따라 각 화소 고유의 비얼굴 신뢰도를 가지게 되는데, 모든 점(400개)의 비얼굴 신뢰도 총합이 임계치 이상이 되면 얼굴이 아닌 것으로 판단하는 과정을 거치게 된다. 이 과정은 후보 얼굴 검출과 신뢰도 비교부의 분할 모듈 형태를 취하여 설계하였다.

후보 얼굴 검출은 순차적으로 입력되는 MCT 변환 영상을 모아서 20x20 윈도우 영상을 동시에 제공하는 역할을 한다. 구현을 위해 총 19개의 라인 메모리와 별도의 지연도구를 사용하였다.

표 2와 같이 AdaBoost로 학습된 데이터를 토대로 20x20 윈도우 내의 각 화소는 위치별 MCT값에 따른 고유 신뢰도 값을 가지게 되는데 신뢰도 비교부에서는 이 400개의 비얼굴 신뢰도 값을 합산하고, 총합이 임계치 이하인 경우의 좌표와 신뢰도를 출력한다.

표 2에 윈도우 내의 (1, 1), (14, 10) 위치에서의 각

표 2. 신뢰도 규칙 테이블 고정 소수점 변환 예
Table 2. Confidence Rule Table using Fixed point transform.

(a) 점 (1, 1)과 (14, 10)에서의 신뢰도 예
(a) Confidence value example at Coordinate (1, 1) & (14, 10)

MCT 값	(1, 1)	(14, 10)
1	1.302622	0.315946
2	1.772520	0.000000
3	0.487639	0.000000
4	0.937634	0.831502
5	0.155170	0.361624
6	0.316182	0.793199
7	0.308015	0.469133
8	1.77252	0.839976
9	1.154538	0.686302
10	0.358852	0.000000
11	0.000000	0.993162
}}	}}	}}
511	0.153845	0.485727

(b) 점(1, 1) Q 8.8 고정 소수점 형식 변환 (16bits)
(b) Q 8.8 Fixed Point Format(16bits) transform example.

점 (1, 1)의 신뢰도 Q 8.8 Format (16bits) 변환 예			
MCT 값	실수값	정수부	소수부
1	1.302622	00000001	01001101
2	1.772520	00000001	11000101
3	0.487639	00000000	01111100
4	0.937634	00000000	11110000
5	0.155170	00000000	00100111
6	0.316182	00000000	01010000
7	0.308015	00000000	01001110
8	1.772520	00000001	11000101
9	1.154538	00000001	00100111
10	0.358852	00000000	01011011
11	0.000000	00000000	00000000
}}	}}	}}	}}
511	0.153845	00000000	00100111

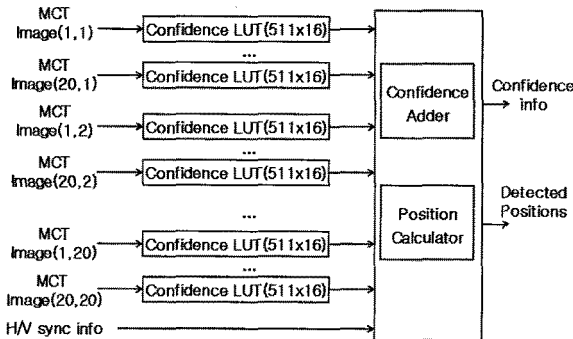


그림 8. 신뢰도 비교 블록 내부 구조
Fig. 8. Confidence Mapper Block Diagram.

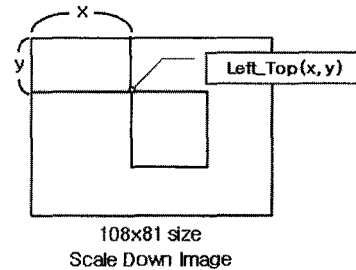
MCT 값에 대한 비얼굴 신뢰도 값 예를 나타내었다. 예를 들면, 윈도우 내 (1, 1) 위치에서 MCT값이 3일 경우 얼굴이 아닌 신뢰도 값이 0.487639 이며, (14, 10) 위치에서 MCT값이 3일 경우는 얼굴이 아닌 신뢰도가 0이므로 얼굴일 가능성을 높게 판단하게 된다.

이러한 신뢰도 값 400개를 모두 합하여 해당 윈도우 내에서의 비얼굴 신뢰도를 총합해서 후보 검출 영역을 판단하게 된다.

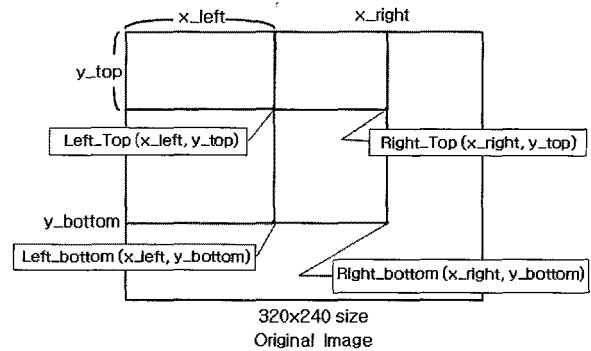
하드웨어 구현을 위해 표 2와 같이 실수 범위의 확률 값을 Q 8.8(16bits) 고정 소수점 형태로 변환하여 16비트 형태로 신뢰도 룩업테이블을 구성하였으며 총 400개의 신뢰도 룩(511x16bit)을 생성하여 신뢰도 비교부를 구성하였다.

3.5 좌표 재구성

이 블록은 영상 크기 조정된 영상에서 얼굴 후보영역을 찾았을 경우, 이를 원 영상에서의 대응 위치로 계산해 주는 역할을 담당한다. 그림 9 (a) 와 (b) 를 보면 영상 크기 조정 영상에서 검출된 후보 영역과 이에 따른 원 영상에서의 대응 영역의 좌표관계를 도식화 했고,



(a) 영상 크기 조정된 영상에서 검출좌표 및 영역



(b) 원 영상에서의 대응좌표 및 영역

그림 9. 검출 좌표의 원영상과 scale-down 이미지 간의 상관관계 예

Fig. 9. Detected position correspondence between original and scale-down image example.

식(3)에서는 그에 관한 수식관계를 나타내고 있다.

$$\begin{aligned}
 &Scalefactor = OriginalHsize / ScaleDownHsize \\
 &X_{left} = (x) * Scalefactor + 0.5 \\
 &Y_{top} = (y) * Scalefactor + 0.5 \\
 &X_{right} = x_{left} + (winsize(20) * Scalefactor + 0.5) \\
 &Y_{bottom} = y_{top} + (winsize(20) * Scalefactor + 0.5)
 \end{aligned}
 \tag{3}$$

제시된 수식관계가 비록 단순하지만 하드웨어 구조에 그대로 적용할 경우 복잡한 실수 연산과 반복적인 곱셈연산의 사용으로 구현의 최적화를 도모할 필요가 있다. 따라서 원활한 하드웨어 구현 및 실시간 처리 동작을 위해서는 곱셈 연산 및 실수 연산을 줄이거나 없애야 한다. 본 연구에서는 실시간 처리에 적합하게 복잡한 실수 연산 시간을 줄이기 위한 방법으로 320x240 원 영상 기준으로 총 13단계의 영상 크기 조절을 걸쳐 생성되는 이미지들에 대해 모든 좌표에 원 영상 대응 좌표를 식(3)에 따라 먼저 계산하고 정수화시켜 이 값을 룬 테이블에 저장, 검출위치의 좌표 좌측 상단(left-top)에 들어 올 때, 영상 크기 조정 단계 정보를 이용해 주소를 배정 하는 방법을 사용하였다.

그림 10에 좌표 재조정 부의 블록도를 나타내었으며

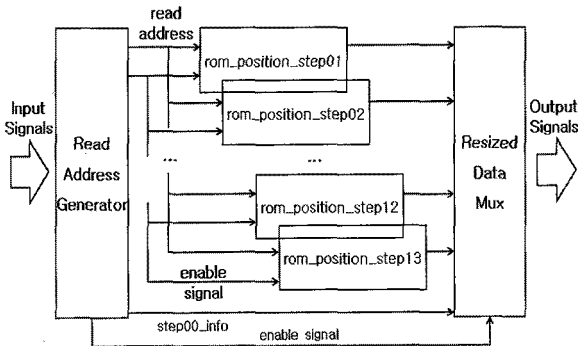


그림 10. 좌표 재조정 부 블록 내부 구조
Fig. 10. Position Resizer Block Diagram.

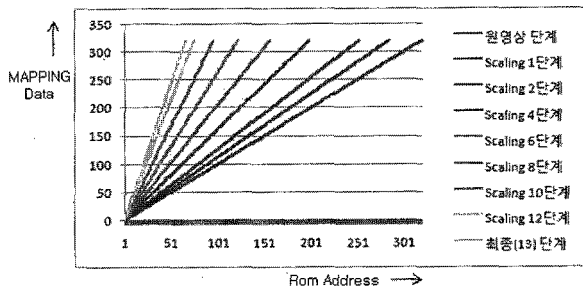


그림 11. 각 단계별 ROM 데이터 정보 예
Fig. 11. Graph of ROM addressing & mapping data.

각각의 룬에 저장된 단계별 배정되는 주소 정보의 예를 그림 11에 나타내었다.

3.6 데이터 그룹화

후보 얼굴 검출부 와 신뢰도 비교부를 거쳐 검출되는 후보영역들은 최종 검출영역 주위에서 여러 번(4~5회 이상) 검출 되게 되는데, 최종 얼굴 검출영역을 결정하기에 앞서 동일 얼굴로 판단되는 중복 영역들을 그룹화 하는 과정이 필요하다.

그림 12와 같이 겹침 영역이 발생할 때 동일 영상 크기에서 뿐만이 아니라, 탐색 윈도우 크기가 다른 영상에서도 중복 검출 될 수 있다. 겹침 판단 여부는 기존 영역과 새 검출 영역의 크기가 1/4 이상 겹칠 경우 중복 검출이라고 판단하며, 비얼굴 신뢰도 값을 비교하여 둘 중 낮은 경우(얼굴에 가까운)를 선택한다. 실험결과 그림 12. (a)에서 볼 수 있듯이 오검출 경우에는 겹침이



(a) 그룹화 미적용 (b) 그룹화 적용

그림 12. 데이터 그룹화 적용 전과 적용 후 결과
Fig. 12. Output Image before and after grouping.

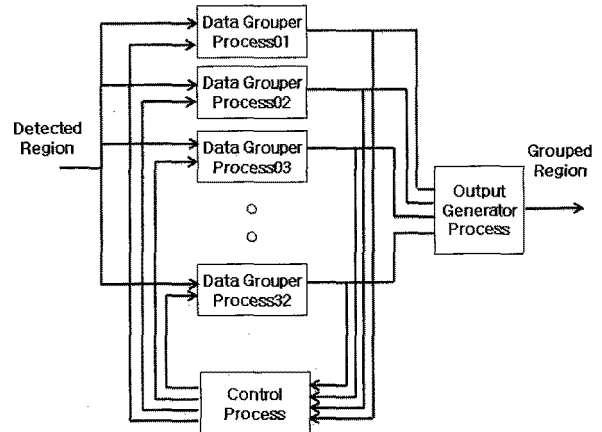


그림 13. 데이터 그룹화 부 블록 내부 구조
Fig. 13. Data Grouper Block Diagram.

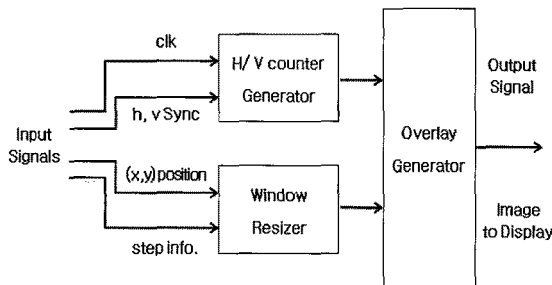
없거나 1~2회 이하인 경우가 대부분이었다. 이를 바탕으로, 오검출률을 낮추고 검출 신뢰도를 높이기 위해 겹침 횟수가 3회 이상인 경우만 최종 검출 영역으로 판단하도록 하였다.

그림 13과 같이 본 구현에서는 한 영상 안에 각기 다른 32개까지의 얼굴을 검출 할 수 있도록 설계 하였다. 얼굴 검출 정보는 검출 결과 표시부로 보내질 뿐만 아니라 향후 얼굴 인식 등에 필요한 사전 정보로 사용되기 위해 호스트 인터페이스를 통해 프로세서로도 동시에 제공된다.

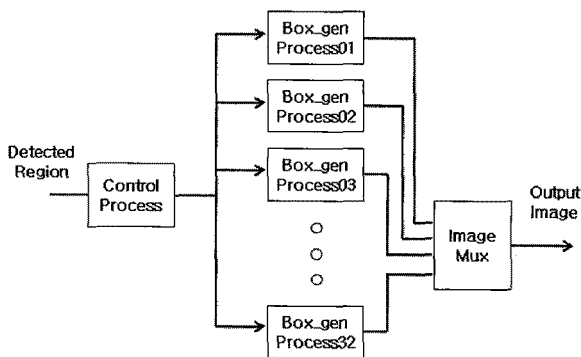
3.7 검출 결과 표시

데이터 그룹화부에서 제공되는 최종 검출 얼굴 영역을 메모리 인터페이스부로부터 받은 영상위에 추가하여 최종적으로 디스플레이 하는 역할을 수행하는 블록이다. 이 블록을 통해서 다른 처리 장치의 도움 없이 실시간으로 얼굴 검출 결과를 확인 할 수 있게 해준다.

그림 14와 같이 32개까지 검출된 영역 좌표 데이터는 각각 32개의 네모 상자 영상 생성 처리부로 보내어져 검출 영역에 네모상자 영상을 만들고 이 32개의 영상들은 영상 멀티 플렉서에서 종합되어 최종 출력되게 된다.



(a) 검출 결과 표시부 블록도



(b) 검출 결과 표시부 내부의 네모상자 생성 처리부

그림 14. (b)는 검출 결과 표시부의 세부 동작을 나타내고 있는데, 입력된 검출영역의 좌상단 좌표는 영상 크기 조절 단계정보를 이용하여 네모상자의 크기를 결정하고 수직, 수평 정보와 함께 32개의 네모상자 생성 처리부로 보내지게 됨을 알 수 있다.

IV. 설계 검증 및 성능 분석

제안된 구조의 검증을 위하여 Virtex5 lx330 FPGA를 이용하여 시스템을 구현하였고 QVGA급 카메라의 영상 (320x240 size)을 흑백 영상으로 변환하여 실시간 검증을 하였으며 그림 15에 검증 환경 사진을 나타내었다. Virtex5 FPGA에서 13.5MHz 클럭의 속도로 동작을 확인 하였으며, 53.01KB 메모리(31.6KB의 RAM 메모리, 21.4KB의 ROM 메모리) 와 80,796(38%)의 룩업테이블사용하고 있다. FPGA 합성 결과는 표 3에 자세히 나타내었다.

표 4와 같이 검증한 구조는 한 프레임을 처리하는데 약 361,800 클럭 사이클이 소요되며, 최대 54MHz로 동작이 가능하도록 설계되었다. 이 경우 최대 149fps를 처리 속도로 얼굴 검출이 가능함을 알 수 있으며 일반 카메라 인터페이스 시 30fps 처리가 가능함을 알 수 있다.

얼굴 검출 성능은 MIT+CMU에서 제공한 frontal face test set(Rowley et al., 1998)^[8](Test Set A, B, C, Rotated test set)과 Yale Test set(Georghiadis, A., 1997, 각각 1개의 얼굴을 포함한 165장의 이미지(감정 표현(분노, 놀람 등) 및 조명 변화))^[9]를 이용하여 검증 하였으며, MIT+CMU의 A,B,C 테스트 셋은 총 130장의



그림 15. 검증 환경
Fig. 15. Verification Environment.

그림 14. 검출 결과 표시부 블록 내부 구조
Fig. 14. Detected Result Display Block Diagram.

표 3. FPGA(Xilinx Virtex5 LX330) 합성 결과 요약
Table 3. FPGA Synthesis Result Summary.

Slice Logic Utilization	Used	Available	Utilization
Number of Slice Registers	18,551	207,360	8%
Number of Slice LUTs	80,796	207,360	38%
Number used as logic	26,295	207,360	12%
Number used as Memory	54,288	54,720	99%
Number of route-thrus	1,762	414,720	1%
Slice Logic Distribution			
Number of occupied Slices	29,450	51,840	56%
Number of LUT Flip Flop pairs used	90,133		
Number with an unused Flip Flop	71,582	90,133	79%
Number with an unused LUT	9,337	90,133	10%
Number of fully used LUT-FF pairs	9,214	90,133	10%
Number of unique control sets	5,243		
Number of slice register sites lost to control set restrictions	7,192	207,360	3%

표 4. 입력 영상에 따른 메모리 사용 및 연산 시간 (320x240 size 영상 기준)

Table 4. Memory requirement & Processing Time by Input Image (320x240 size image unit).

구분	사용량 (처리량)
Processing cycle (1 Frame)	361,800 clock
RAM 사용	31.6 KB
ROM 사용	21.4 KB
LUT	80,796 LUTs
최대 동작 클럭	54 MHz
최대 처리	149 frame/sec

정면 이미지(506개의 얼굴 포함)와 50장의 회전된 (Rotated Test Set) 이미지(223개 얼굴 포함)를 포함하고 있다. 각각의 영상들은 신문, 잡지 영상 및 다양한

표 5. 테스트 이미지(MIT+CMU, Yale Test set)에 대한 검출 및 오검출률

Table 5. Face detection result. (MIT+CMU and Yale test set).

구분	검출률 (Detection rate)	오검출 횟수 (False-positives)
Test A	70.41% (119/169)	1
Test B	64.94% (100/154)	0
Test C	85.25% (156/183)	0
Sum (A, B, C)	74.11%(375/506)	1
Yale Test set	100 % (165/165)	0
Average	80.48% (540/671)	-



그림 16. 얼굴 검출에 성공한 예
Fig. 16. Successful result from MIT+CMU test set.

조도 하에서의 영상들을 포함하고 있어 검출이 쉽지 않은 영상들이며 이 테스트 set 에 대한 검출 성능은 표 5 와 같이 오검출이 거의 없이 평균 80.48 % 이상의 매우 높은 검출률(오검출률 0.9 %)을 나타내었다.

또한 구현된 시스템을 이용하여 실내 환경에서 얼굴 검출 테스트를 수행한 결과 사무실 조도 하에서 정면 영상의 경우 100%의 검출률, 0%의 오검출률을 나타내었다.

그림 16은 테스트 이미지에 대해 검출 성공한 예이다. 그림 17에서는 정면 얼굴뿐만이 아니라 회전된 영



그림 17. 회전된 얼굴 검출에 성공 및 실패한 예
Fig. 17. Results from rotated Image.
(Success and Failure)

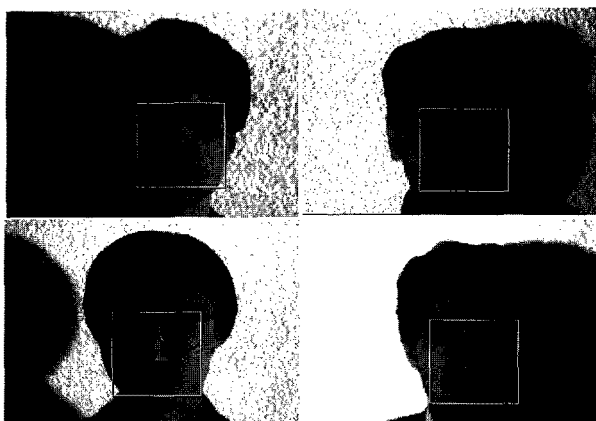


그림 18. 다양한 조명 환경에서 검출 결과
Fig. 18. Output Result in various light intensity.

표 6. 기존 FPGA 구현 사례와 성능 비교
Table 6. Comparison to previous FPGA Implementation

	Logic elements	Video rate(fps)	Detection accuracy
McCready System [10]	89,856	30 fps	80.9 %
Duy Nguyen System [11]	15,050	41.7 fps	86.6 %
Our System	26,295	30 fps	100 %

※ 320x240 frame size & 12.5 MHz clock 동작 기준.

※ 검출률은 Yale Test set 기준.

굴에 대해서도 좌우 12° 상하 15° 범위까지 검출 가능함을 보여주고 있다.

주목할 만한 것은, 이러한 일반적인 영상 이외에도 그림 18과 같이 조명에 의해 심하게 가려지거나 밝기

차이가 나지 않는 밝은 얼굴에서도 강인한 검출 성능을 나타내었다.

표 6은 기존 FPGA 구현 사례와의 성능 비교를 나타내었다.

V. 결론 및 향후 연구 방향

이미 얼굴 검출 알고리즘에 대해서는 다양한 형태의 연구 보고 사례가 있지만 본 연구에서는 MCT 변환 기법을 통해 조명에 의한 검출률 저하를 극복한 고성능의 얼굴 검출 하드웨어 구조를 개발하고, 초당 30 프레임(최대 149 프레임) 이상의 고성능 실시간 검출 엔진을 구현하였다.

본 연구에서 개발된 고성능 얼굴 검출기를 통한 정확한 얼굴 위치의 획득은 얼굴 인식 기술의 성능향상에도 크게 기여 할 수 있다. 두 기술의 융합을 통해 가정용 시각처리 기능으로 탑재 되어질 때, 사용자인식, 휴먼 추종 등의 다양한 응용 기술로 발전 될 수 있다.

또, 향후 본 연구의 하드웨어 구조가 ASIC으로 제작 되어 저가로 공급이 가능하게 되면 휴대폰과 같이 QVGA급 이상의 카메라를 포함한 임베디드 시스템에 활용되어 다양한 응용 분야를 창출할 수 있을 것으로 기대된다.

뛰어난, 검출률에도 불구하고 본 연구의 추후 과제로서 회전된 얼굴에 대한 검출 문제를 해결할 필요가 있다. 보안 시스템과 달리 다양한 임베디드 시스템의 응용 분야에서 얼굴 영상은 사용자 얼굴 방향이 시스템 협조적으로 정면으로 향하지 않을 수 있다.

또한 표 5에서 소개된 얼굴 검출률의 저하는 100% 회전 얼굴에 대한 미검출로 발생하였으며 회전 얼굴 문제만 추가적으로 해결되면 매우 이상적인 얼굴 검출 엔진으로 판단된다.

따라서 추후 연구 방향으로, 상하 좌우로 회전된 얼굴에도 검출 가능한 얼굴 검출 모델의 구축을 목표로 보안을 진행 하고자 한다.

참 고 문 헌

[1] Samir Nanavat, Michael Thieme and Raj Nanavati. "Biometrics", Wiley, pp.63-75, 2002.
[2] Bernhard Fröba and Andreas Ernst, "Face detection with the Modified Census Transform", IEEE International Conf. On Automatic Face and

Gesture Recognition(AFGR), pp. 91-96, Seoul, Korea, May. 2004.

[3] Ming-Hsuan Yang, Dan Roth and Narendra Ahuja. "A snow-based face detector". In Advances in Neural Information Processing Systems 12 (NIPS 12), pp.855-861. MIT Press, 2000.

[4] Najwa Aaraj, Srivaths Ravi, Anand Raghunathan and Niraj K. Jha, "Hybrid architectures for efficient and secure face authentication in embedded systems", IEEE Transaction on VLSI Systems, Vol.15, no. 3, pp.296-308, March 2007.

[5] Paul Viola and Michael J. Jones, "Robust real-time face detection" In International Journal of Computer Vision, pp. 137-154, 2004.

[6] Yoav Freund and Robert E. Schapire. "A decision-theoretic generalization of on-line learning and an application to boosting" in Journal of Computer and System Sciences, pp. 119-139, 1997.

[7] 이수현, 정용진, "얼굴 검출을 위한 SoC 하드웨어 구현 및 검증", 전자공학회 논문지 44권 SD 편 제 4호, 대한전자공학회, April, 2007.

[8] CMU/VASC Image Database, <http://vasc.ri.cmu.edu/idb/html/face/index.html>

[9] Georghiadis, A. : Yale Face Database, Center for computational Vision and Control at Yale University, <http://cvc.yale.edu/projects/yalefaces/yalefaces.html>

[10] R. McCready, "Real-time face detection on a configurable hardware platform," M.S. thesis, Dept. Elect. Comput. Eng., Univ. Toronto, Toronto, On, Canada, 2000.

[11] Duy Nguyen, David Halupka, Parham Aarabi, and Ali Sheikholeslami, "Real time Face detection and Lip feature extraction using Field-Programmable Gate Arrays", IEEE Transactions on SYSTEMS, MAN AND CYBERNETICS-ART B: CYBERNETICS, Vol. 36, no. 4, pp.902-912, AUGUST 2006.

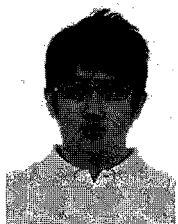
— 저 자 소 개 —



한 동 일(정회원)-교신저자
 1988년 2월 고려대학교 전자전산 공학과 졸업(학사)
 1990년 2월 한국과학기술원 전기 및 전자공학과 졸업(석사)
 1995년 한국과학기술원 전기 및 전자공학과 졸업(박사)
 1995년 2월~2003년 2월 LG전자 디지털TV연구소 책임연구원.
 2003년 3월~현재 세종대학교 컴퓨터공학과 정교수
 <주관심 분야 : 영상 처리, 컴퓨터 비전, 디지털 TV, 디스플레이 화질 개선, 시스템 온 칩>



조 현 중(학생회원)
 2008년 2월 세종대학교 컴퓨터 공학과 졸업(학사)
 2008년 3월~현재 세종대학교 컴퓨터 공학과 석사과정
 <주관심분야 : 컴퓨터 비전, HCI, 패턴 인식, 로보틱스>



최 중 호(학생회원)
 2009년 2월 상지대학교 컴퓨터 공학과 졸업(학사)
 2009년 3월~현재 세종대학교 컴퓨터공학과 석사과정
 <주관심 분야 : 영상처리, 신호처리, 컴퓨터 비전>



조 재 일(정회원)
 1990년 2월 성균관대학교 전기 공학과 졸업(학사)
 1992년 2월 성균관대학교 전기 공학과 졸업(석사)
 2009년 3월~성균관대학교 전자 전기컴퓨터공학과 박사과정
 1992년 1월~현재 한국전자통신연구원
 <주관심분야 : 로봇, 영상처리, Human-Robot Interaction, ASIC Design>