

현금 인출기 사용자의 선글라스 및 마스크 인식 시스템

임동악[†], 고재필[‡]

요약

본 논문에서는 현금인출기 사용자의 선글라스 및 마스크를 인식하는 시스템을 설명한다. 제안하는 시스템은 얼굴윤곽을 먼저 추출하고 이로부터 눈과 입의 위치를 추정한다. 마지막으로, 눈과 입 영역에 대해 각각 히스토그램 인덱싱 기법을 적용하여 선글라스 및 마스크를 인식한다. 눈이나 입 영역이 가려진 상태에서 얼굴윤곽 추출과 눈 및 입의 위치 추정을 위해 얼굴모양모델을 도입한다. 얼굴모양모델의 정합성능을 향상시키기 위해 2단계에 걸친 얼굴후보영역 검출을 도입하고 모델의 초기위치를 다변화하여 반복 정합을 실시한다. 배경모델에 기반한 얼굴후보영역 검출 방법의 성능을 보장하기 위해서 배경모델을 자동으로 갱신할 수 있도록 시스템을 구성한다. 실험에서는 연구실에서 획득한 영상에 대하여 시스템의 설정에 관한 실험을 제시하고 마스크 및 선글라스 인식 결과를 보인다.

A System for Recognizing Sunglasses and a Mask of an ATM User

DongAk Lim[†], Jaepil Ko[‡]

ABSTRACT

This paper presents a system for recognizing sunglasses and a mask of an ATM (Automatic Teller Machine) user. The proposed system extracts firstly facial contour, then from this extraction results it estimates the regions of eyes and mouth. Finally, it recognizes sunglasses and a mouth using Histogram Indexing based on those regions. We adopt a face shape model to be able to extract facial contour and to estimate the regions of eyes and mouth when those regions are occluded by sunglasses and a mask. To improve the fitting accuracy of the shame model, we adopt 2-step face detection method and conduct fitting several times by varying the initial position of the model instance. To achieve a good performance of the face detection method based on a background model, we enable the system to automatically update the background model. In experiment, we present some experiments on setting parameters of the system with images taken from in our laboratory, and demonstrate the results of recognizing sunglasses and a mask.

Key words: ATM security(ATM보안), Face Detection(얼굴검출), Face Contour Extraction(얼굴윤곽추출), Active Shape Model, Histogram Indexing(히스토그램인덱싱)

1. 서론

시간과 장소의 제약을 받지 않고 24시간 현금 인출을 가능하게 하는 현금 인출기는 주 5일제 근무제

와 더불어 그 사용량이 증가하고 있는 추세이다. 이에 따라 관련 범죄도 증가하고 있어, 이를 예방하거나 사후범인 색출을 용이하게 하기 위한 보안기술이 요구되고 있다. 영상기반의 현금 인출기 보안 시스

※ 교신저자(Corresponding Author) : 고재필, 주소 : 경북

구미시 양호동 1번지(730-701), 전화 : 031)478-7529, FAX

: 031)478-7539, E-mail : nonezero@kumoh.ac.kr

접수일 : 2007년 3월 9일, 완료일 : 2007년 11월 14일

* 금오공과대학교 컴퓨터공학부

(E-mail : agi98@kumoh.ac.kr)

** 정회원, 금오공과대학교 컴퓨터공학부

※ 이 논문은 2005년 정부(교육인적자원부)의 재원으로
한국학술진흥재단의 지원을 받아 수행된 연구임
(KRF-2005-202-D00450).

템은 경제적이며 사용자의 거부감을 줄일 수 있는 방법이다.

국내에서는 1990년대 후반 일부 벤처기업을 중심으로 초보적인 얼굴검출 기술에 기반한 시도가 이루어졌으나 당시로는 상용화 가능 수준에 미치지는 못하였다. 또한, 현금 인출기 응용이 얼굴검출 연구의 대표적인 응용사례로 언급되고는 있으나, 실제로 자체를 위한 연구결과는 찾아보기 어렵다. 초창기에 시도되었던 방법은 정상 얼굴만을 검출할 수 있는 얼굴검출 기법을 단순히 적용하여 검출기의 성공여부로써 정상얼굴과 비정상얼굴을 구분하고자 하였다. 즉, 얼굴이 제대로 검출되지 않는 경우에 선글라스나 마스크 착용을 가정하였기 때문에 마스크와 선글라스 착용을 구분하는 문제가 남아있다. 특히, 마스크나 선글라스를 착용한 얼굴영상은 이들을 착용하지 않은 얼굴영상과는 다르기 때문에 기존의 고성능 얼굴검출 기법[1-8]을 그대로 적용해서는 마스크나 선글라스 착용 얼굴을 검출할 수 없다.

직관적으로 선글라스나 마스크 착용여부를 판별하기 위해서 “일반 얼굴 검출기”, “선글라스 착용 얼굴 검출기”, “마스크 착용 얼굴 검출기”, “선글라스 및 마스크 착용 얼굴 검출기”를 설계하고 이들의 판정결과를 활용할 수 있다. 그러나, 이러한 방법은 회전, 크기, 이동, 조명등에 강인한 각 검출기의 설계가 매우 어렵다는 단점이 있다. 본 논문에서는 선글라스나 마스크 착용 얼굴이라도 얼굴의 유파운 변하지 않는다는 점과 얼굴유파을 성공적으로 검출하면 이로부터 눈 및 입 영역을 쉽게 추정하여 간단한 방법으로 선글라스 및 마스크 착용여부를 판별할 수 있다는 점을 염두 하여 시스템을 설계하였다.

얼굴유파을 검출하는 방법으로, [9]는 얼굴주위의 공간적 에지 성분에 기반한 타원 템플릿을 사용하여 효과적으로 얼굴영역을 추적하였다. 그러나, 동시에 얼굴내부의 컬러정보도 함께 사용하였다. [10]도 얼굴의 모양을 타원으로 모델링 하였는데, 타원추출을 위해 스텝레오 영상을 사용하였으며, 초기 얼굴의 위치를 찾는데 컬러기반 얼굴검출 기법을 도입하였다. 본 논문에서는 얼굴의 모양정보에 만 의존할 수 있는 기법으로 얼굴모양 모델링에 널리 사용되어 왔던 Active Shape Model (ASM)[11]을 사용한다. 다양한 사람을 대상으로 하는 응용인 관계로 개인별 모델링을 해야 하는 Active Appear-

ance Model (AAM) [12,13] 보다 ASM을 적용한다 [14]. ASM의 정합에 기반한 얼굴윤곽 검출 성능은 모델의 초기위치 설정에 매우 민감하므로, 2단계에 걸친 얼굴후보영역 검출을 통해 보다 정확한 초기 위치를 추정할 수 있도록 하였으며, 또한 모델의 초기위치를 다변화하여 반복 정합한 후 가장 정합이 잘된 경우를 선정하기 위한 정합 정확도 측정치를 제시하였다.

본 시스템에서는 얼굴의 외양정보를 활용할 수 없기 때문에, 1차 얼굴후보 영역 검출단계에서는 사용자의 움직임 정보에 주로 의존하면서 형태정보도 함께 활용하여 간단한 방법으로 얼굴의 위치를 대략적으로 파악한다. 2차 얼굴후보 영역 검출에서는 사용자의 움직임이 없거나 미세한 경우에 대처하기 위해 입력영상과 배경영상과의 차이를 활용한다. 이러한 기법의 성능은 배경영상의 적절성에 의존하기 때문에 자동으로 개선 가능한 Robust Principal Component Analysis (RPCA)[15]에 기반한 배경모델을 도입하고 시스템 설계에서 이를 고려하였다. 즉, 배경모델을 개선하기 위한 데이터는 적외선 거리 센싱 결과와 영상으로부터 사용자의 움직임 양을 추정하고 이를 토대로 구분되는 시스템 상태 중 대기상태일 때 주기적으로 획득할 수 있도록 하였다. 그러나, 배경모델을 개선하기 위해서는 사용자의 사용이 전혀 없는 상황에서 획득된 영상들만을 수집해야 하는데, 대기상태에서 획득한 영상 중에는 현금 인출기 앞을 단순히 지나가는 사람이 포함되기도 한다. RPCA는 이러한 문제점을 해결해 준다.

얼굴모양모델의 정합을 통해 모델의 한 변형으로써 얼굴윤파이 검출되기 때문에, 입력영상과 일치된 변형된 모델로부터 쉽게 눈과 입 영역을 추정할 수 있다. 그러나, 얼굴의 포즈변화와 같은 원인으로 인해 정합 정밀도가 낮은 경우 눈 및 입 영역 추정 오류가 생긴다. 따라서, 눈 및 입 영역에 대한 선글라스 및 마스크 착용판별은 크기, 이동, 회전 등에 강인한 히스토그램인덱싱[16] 기법을 적용한다.

본 논문의 구성은 다음과 같다. 2장에서는 시스템 개요를 설명한다. 3장에서는 얼굴윤파검출과정을, 그리고 4장에서는 선글라스 및 마스크 착용판별 방법을 상세히 설명한다. 5장에서는 얼굴윤파검출과 선글라스 및 마스크 착용 인식결과를 제공한다. 그리고, 마지막 6장에서는 결론을 내린다.

2. 시스템 개요

시스템은 그림 1에 도시한 사용자의 사용 환경을 가정한다. 기기에 설치된 카메라는 지상에서 80센티 정도의 높이에서 60도 위쪽을 향하도록 설치되어 사용자의 상반신과 비교적 배경이 단순한 천정이 촬영 되도록 하였으며 사용자의 바로 뒤에 대기하는 사용자가 촬영될 수 있는 소지를 줄이도록 고려되었다. 또한, 적외선 거리 센서를 장착하여 사용자의 접근을 탐지할 수 있도록 하였다.

그림 2는 제안하는 시스템의 구성을 보여준다. 시스템은 대기상태와 동작상태로 구성된다. 사용자가 시스템에 다가와서 일정 시간 체류하면 시스템은 대기상태에서 동작상태로 넘어간다. 동작상태에서는 2 차례에 걸쳐 얼굴후보영역 검출을 수행하고, 이를 기반으로 얼굴윤곽을 추출한다. 추출된 윤곽정보로부터 눈과 입의 위치를 추정하고, 이로부터 선글라스 및 마스크 착용판별을 실시한다. 대기상태에서는 얼굴검출 단계에서 사용하는 배경모델을 위해 배경영상을 정기적으로 수집한다.

3 얼굴윤곽 추출

3.1 얼굴윤곽 추출 흐름도

사용자 접근탐지에서 얼굴윤곽 추출까지의 세부적인 처리과정은 그림 3에 제시하였다. 사용자의 접근탐지를 위해 차영상과 적외선 거리 센서 정보를 활용한다. 만약 접근이 확인되면 현재 영상을 배경영상을 위한 데이터베이스에 저장한다. 그렇지 않으면 얼굴후보영역 검출 단계로 넘어간다. 얼굴후보 검출은 두 단계로 프로젝션 기법의 1차 후보검출 방법과 배경모델을 이용한 2차 후보검출 단계를 차례대로 적용한다. 얼굴후보영역 검출이 이루어지면 얼굴모양모델을 이용하여 얼굴의 윤곽을 추출한다. 주요 단계들에 대한 상세한 설명은 3.2, 3.3, 3.4.3.5절에서 각각 설명한다.

3.2 사용자 접근탐지

사용자의 접근은 적외선 센서로부터 얻은 사용자와 기기간의 거리와 현 프레임 영상과 이전 프레임 영상의 차영상으로부터 추정하는 움직임 정도로 판단한다. 사용자의 움직임이 느린 경우는 차영상에서

변화의 정도가 작기 때문에 추정한 움직임 영역이 작아지고, 움직임이 빠른 경우는 모션 흐름에 의해 변화의 정도가 크기 때문에 추정한 움직임 영역이 넓어지는 문제점이 있다. 느리거나 빠른 움직임에 상관없이 변화의 정도를 일정하게 유지시켜 주기 위해 다음과 같이 모션양을 정의한다.

$$\sum_{k=1}^N d(f_k^i, f_k^j) / N$$

$$d(f_1, f_2) = \begin{cases} 1, & \text{if } |f_1 - f_2| > 0 \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

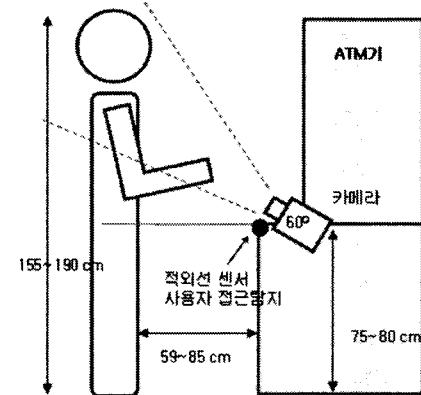


그림 1. 시스템 사용환경

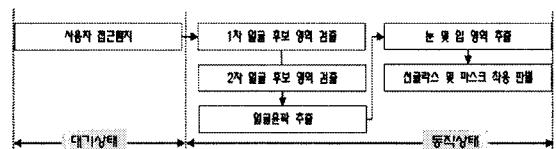


그림 2. 시스템 구성

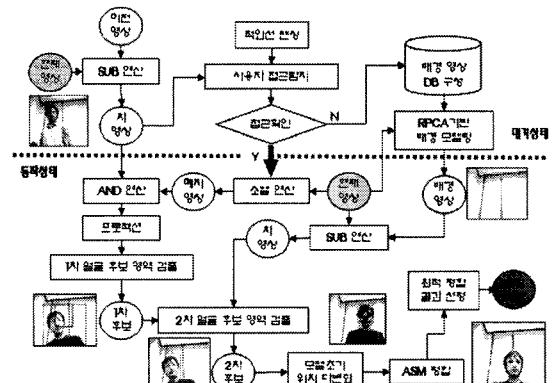


그림 3. 얼굴윤곽 추출 흐름도

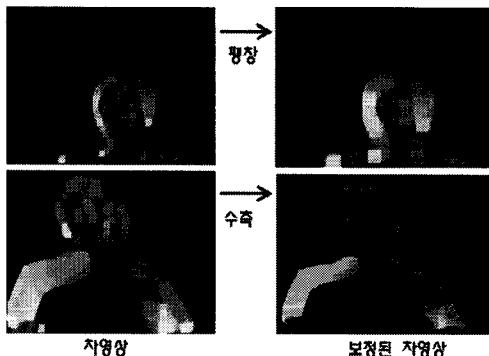


그림 4. 모션양에 기반한 보정된 차영상

여기서, f'_k 는 i 프레임의 k 번째 화소값을, d 는 화소값의 차이가 0이상이면 1을 그렇지 않으면 0을 반환하는 이진함수를 그리고, N 은 영상의 화소수를 나타낸다. 이 값이 하위 임계치 이하 또는 상위 임계치 이상인 경우에 각각 모풀로지[17] 기법의 팽창과 수축을 적용하여 모션양을 보정한다. 그림 4는 이렇게 보정된 차영상을 보여준다.

여기서, 상위 임계치(3.9) 및 하위 임계치(1.56)는 다음과 같이 실험적으로 결정한다. 사용자에게 특정 움직임 속도를 구분하여 요구하는 것이 어렵기 때문에 3단계의 빠르기를 포함한 움직임을 요구하였고 이로부터 임계치를 계산하였다.

- 실험자에게 천천히, 보통, 빠른 움직임이 포함된 움직임을 요구하여 대략 10초 정도로 150프레임 정도의 영상 시퀀스를 획득한다.
- 획득한 시퀀스의 차영상을 구하고 모션양을 구한다. 모션양의 최소값과 최대값 사이를 3등분 하여, 상, 중, 하로 구분한다.
- 10명의 실험자로부터 얻은 각 구간별 모션양 평균을 구하고, 중앙과 상위 그리고 중앙과 하위 평균의 중심을 각각 상위 임계치, 하위 임계치로 결정한다.

시스템은 다음 두 가지 조건을 만족하면 사용자가 기기 사용을 위해 다가왔다고 간주하고 대기상태에서 동작상태로 전환된다.

- 적외선 센싱 기반 거리 측정: 사용자와 ATM과의 간격이 정해진 거리 이내이면서 정해진 시간 동안 유지될 때.
- 시간변화 영상에서의 모션양 측정: 정해진 모션양 이상이 정해진 시간 동안 유지될 때.

본 논문에서 사용한 정해진 간격은 19~45cm, 정해진 시간은 10초, 그리고 정해진 모션양은 0.78로 설정하였다.

3.3 움직임정보 기반 1차 얼굴후보영역 검출

본 시스템에서는 얼굴의 외양정보를 활용할 수 없기 때문에 뛰어난 성능을 보이는 기존의 얼굴검출 기법들을 적용할 수가 없다. 본 시스템에서는 단순히 사용자의 움직임 정보에 주로 의존하면서 형태정보도 함께 활용하여 간단한 투영 기법으로 얼굴의 위치를 대략적으로 파악한다. 그림 5는 1차 얼굴검출 과정을 보여준다.

수직성분의 에지 정보와 차영상의 합을 이진화한 후 수평투영하여 머리 상단을 먼저 알아내고, 수평성분의 에지 정보와 차영상의 합을 이진화한 후 수직 투영하여 얼굴의 너비를 알아낸다. 마지막으로, 얼굴의 높이와 너비의 비율이 1.2라는 정보를 활용하여 얼굴영역을 검출한다.

빠른 움직임이나 느린 움직임이 있는 경우 움직임 영역이 각각 커지고 작아지는 경우, 바로 이전에 찾았던 위치를 그대로 사용한다. 그럼에도 불구하고 검출된 영역이 얼굴의 중심을 벗어나는 경우가 종종 발생한다.

그림 6은 얼굴후보(안쪽 사각형) 및 상반신 영역(외곽 사각형)검출 결과를 보여준다. 그림 6의 (a)는 정확히 얼굴의 중심에 얼굴후보 영역이 찾아진 경우이고, (b)는 얼굴이 포함되어 있으나 중심을 벗어난 경우이고, (c)는 얼굴이 얼굴후보 영역에서 벗어나 있는 경우이다. 얼굴윤곽 추출용으로 ASM을 적용하기 위해서는 (a)와 같은 결과를 도출해야 한다. 세 경우 모두 얼굴이 후보영역의 상하 경계 안쪽에 들어오고 있다.

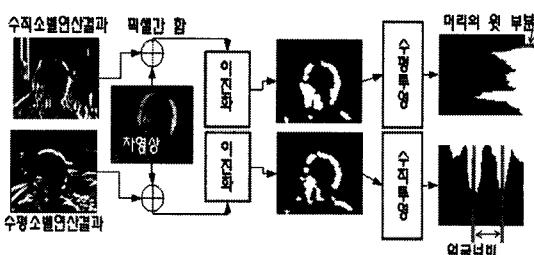


그림 5. 1차 얼굴후보영역 검출 과정

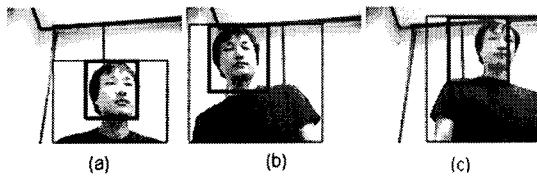


그림 6. 1차 얼굴후보 영역 검출 결과

3.4 RPCA기반 배경모델을 이용한 2차 얼굴후보 영역 검출

움직임이 너무 빠르거나 느린 경우 차영상에 기반한 얼굴영역검출의 정확도를 높이기 위해 입력영상과 배경영상의 차영상을 이용한다. 이 경우 배경영상이 고정되지 않고 지속적으로 생성되어야 성능이 보장된다. 이를 위해, 대기상태에서 (사용자가 현금 인출기 앞에 없을 가능성성이 큰 경우) 지속적으로 획득한 영상을 이용하여 배경모델을 생성하고 이로부터 배경영상을 생성한다. 본 논문에서는 획득한 영상들 중에서 순수하게 배경만 포함되어 있는 영상을 골라낼 필요 없이 대기상태에서 획득한 영상 전체를 그대로 사용하기 위해 RPCA를 도입하였다.

RPCA를 이용하여 얼굴후보 영역을 검출하는 과정은 다음과 같다. 먼저, $\mathbf{F} = [f_1 \ f_2 \ \dots \ f_n]$ 는 $d \times n$ 차원의 행렬이라고 하자. 여기서 f_i 는 하나의 영상을, n 은 학습영상의 수를, d 는 각 영상의 화소 수를 나타낸다. \mathbf{F} 에 대한 k 개의 주성분을 $\mathbf{B} = [\mathbf{b}_1, \dots, \mathbf{b}_k] \in R^{d \times k}$ 라고 하자. 그러면, 입력영상 \mathbf{f} 와 RPCA를 이용하여 얻어낸 복원영상 $\mathbf{BB}^T \mathbf{f}$ 은 그림 7과 같이 배경만으로 구성된다. 이것은 사람이 포함된 10장의 영상을 포함하여 총 100장의 영상을 대상으로 RPCA를 수행하고 분산양이 74%인 상위 12개의 주성분을 사용하였으며 복원한 결과이다. 입력영상에는 사람이 포함되어 있으나 복원된 영상에는 배경만 존재한다.

그림 8의 가장 오른쪽 영상은 입력영상과 복원영상의 차영상을 이진화한 결과이다. 얼굴과 상반신이 배경과 깨끗하게 구분된다.

2차 얼굴후보영역 검출은 다음과 같이 이루어진다. 배경영상을 이용한 차영상에 대해 1차 얼굴영역 검출에서 얻어낸 얼굴의 높이에 해당하는 영역에 한정하여 1차 얼굴영역검출 방법과 동일한 방법으로 얼굴후보영역을 다시 검출한다. 그림 9의 (b) 경우 머리의 상위 절반 가량만 포함된 영역에 한정하여 수평 및 수직 투영이 수행된다. 그림 9의 (c)는 (b)의 일부

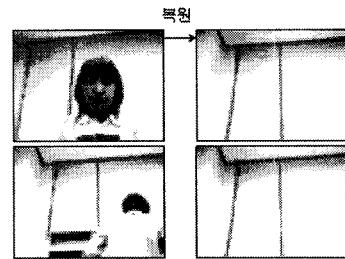


그림 7. PRCA를 이용하여 복원한 영상. 원쪽은 입력영상, 오른쪽은 복원영상



그림 8. 입력영상과 복원영상의 변화영상, 원쪽부터 입력, 복원, 이진화된 차영상



그림 9. 2차 얼굴후보영역 검출 과정

영역에 대해 수평/수직 투영하고 얼굴의 위치를 재추정하여 얻어낸 결과이다.

투영 방식의 한계로 인해 얼굴을 정확하게 찾지는 못한다. 그러나, 얼굴후보 영역의 중심과 얼굴의 중심이 비교적 일치하는 것을 볼 수 있다. 이것은 얼굴 중심 위치에 ASM 모델의 초기위치를 설정하기에는 충분하다는 것을 보여준다.

3.5 ASM을 이용한 얼굴윤곽 추출

마스크나 선글라스 착용을 고려하여 얼굴의 윤곽 위주로 21개의 점만으로 ASM을 모델링 한다. 이로 인해, ASM의 정합성능이 저하될 수 있다. 그림 10은 모델을 구성하는 좌표의 위치를 보여준다. 모델링을 위한 좌표수집은 수작업으로 이루어지며, 이를 위해 그림 11과 같은 프로그램을 작성하였다. 그림 11은 19번 좌표까지 입력한 화면을 보여준다.

모델의 초기위치에 특히 민감한 ASM의 정합 성능을 높이기 위해, 얼굴후보 영역을 중심으로 5개의 초기위치를 주어 5번의 독립적인 정합을 수행한다.

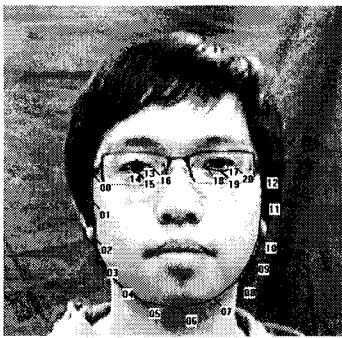


그림 10. 얼굴윤곽 및 눈에 대한 21개의 좌표 위치

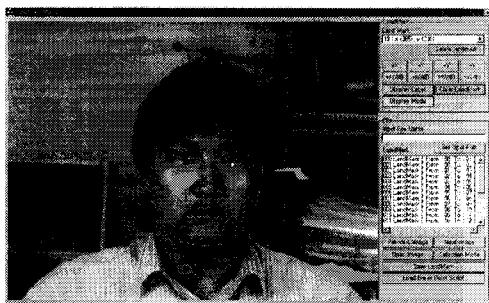


그림 11. 얼굴모양 모델링을 위한 좌표 입력 프로그램 화면

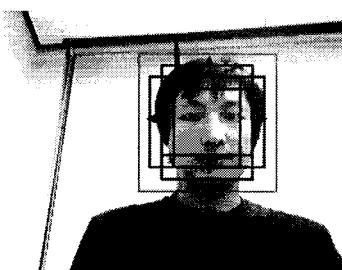


그림 12. 5개의 ASM 모델 초기위치

여기서 5개의 초기위치 설정은 실험을 통해 결정되었다. 그림 12은 2차 얼굴후보 영역을 포함한 5개의 ASM모델 초기위치를 보여준다.

초기위치를 달리하여 5번의 정합 결과 중 최적의 정합 결과를 선정한다. 이때, 최적 정합 판정은 입력 영상과 복원영상의 차영상 g_b 과 수직 애지 영상 g_y , 그리고 정합된 모델의 좌표 s 에 의해 정의하는 거리에 따른다. 그러나, 이때 ASM 모델의 모든 모델 좌표를 사용하지 않고 얼굴의 좌우 측면에 해당하는 좌표만 사용한다. 이를 이외의 눈 및 입에 해당하는 좌표는 선글라스나 마스크 착용에 의해 안정된

차영상을 얻을 수 없기 때문에 제외하였다. 그럼 10의 모델 좌표를 참조하여, 모델의 좌측 및 우측 좌표에 대한 집합을 각각 $s_1 = \{(x_0, y_0), \dots, (x_4, y_4)\}$ 및 $s_2 = \{(x_8, y_8), \dots, (x_{12}, y_{12})\}$ 이라고 하자. 제안하는 거리는 다음과 같다.

$$d_1(s_1, g_b, g_y) + d_2(s_2, g_b, g_y)$$

$$d_i(s_i, g_b, g_y) = \sum_{(x, y) \in s_i} \frac{1 - t(g_b(x, y), g_y(x, y), \theta)}{5} \quad (2)$$

$$t(a, b, \theta) = \begin{cases} 1, & \text{if } (a+b) > \theta \\ 0, & \text{otherwise} \end{cases}$$

여기서, t 는 임계치 θ 에 따라 0 또는 1을 출력하는 이진함수이다. 그리고, 모델이 얼굴의 어느 한쪽 측면에만 잘 정합된 경우를 배제하기 위해 d_1 과 d_2 사이의 거리가 충분히 작은 경우 정합 정확도가 양수가 되도록 다음과 같이 정의한다.

$$\begin{cases} 2 - (d_1 + d_2), & \text{if } |d_1 - d_2| < t, \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

4. 선글라스 및 마스크 착용판별

4.1 얼굴윤곽선 기반 눈 및 입 영역 추출

눈 및 입 영역은 ASM 정합결과로부터 쉽게 추출할 수 있다. 먼저 양쪽 눈을 기준으로 영상을 정면으로 회전시킨다. 즉, 0번 12번 좌표를 잇는 직선을 구하고 수평선과의 각도를 계산하여 그림 12의 오른쪽처럼 얼굴을 수직이 되도록 회전시킨다. 그리고, 그림 13에서와 같이 좌표 0, 12를 이용해 눈 영역을, 좌표 2, 4, 8, 10을 이용해 입 영역을 추출한다.



그림 13. ASM 정합 좌표를 이용한 눈 및 입 영역 추출

4.2 히스토그램 인텍싱을 이용한 선글라스 밑 마스크 착용 판별

ASM 정합을 통한 눈, 입 영역 추출 정확도는 ASM 정합 정확도에 의존적이다. 21개의 윤곽중심으로 모델링 한 ASM의 정합 정확도는 비교적 낮기 때문에, 눈, 입 영역이 안정되게 추출되지 않을 수 있다. 따라서, 크기, 위치, 회전에 무관한 특징을 사용하여 선글라스와 마스크 착용 판별을 수행해야 한다. 이를 위해 히스토그램인텍싱 기법[16]을 도입하였다. 히스토그램인텍싱은 두 영상의 유사도를 전역적으로 비교하는 방법으로 각 영상의 히스토그램을 생성한 후 두 히스토그램이 얼마나 유사한지를 검사한다. 이 방법은 영상검색에서 널리 사용되고 있다. 일반적으로 모든 화소값에 대해 히스토그램을 구하는 것이 아니라 화소값의 범위에 따라 몇 개의 구간을 설정하여 히스토그램을 생성한다. 여기서, 각 구간을 빈이라고 한다. 빈의 개수가 너무 많으면 히스토그램이 주어진 영상에 특화되고, 빈의 개수가 너무 적으면 주어진 영상을 제대로 나타내지 못하므로 이는 실험적으로 결정한다. 본 시스템에서는 구간을 여덟 개로 설정하였다. 그림 14은 눈과 입 영역에 대해 선글라스 및 마스크 착용 경우의 히스토그램을 비교하였다. 각각의 히스토그램이 비교적 명확히 구분되는 것을 볼 수 있다. 두 히스토그램의 유사도 판정은 다음과 같이 한다.

$$\sum_{i=1}^N \min(H_a(i), H_b(i))/N \quad (4)$$

여기서, $H(i)$ 는 i 번째 빈의 히스토그램 값이고, $\min()$ 은 작은 값을 출력하는 함수, 그리고 N 은 빈의 개수를 나타낸다.

이때 조명에 따라 히스토그램이 달라지는데 이를 해결하기 위해 히스토그램 평활화와 같은 전처리 기법을 사용할 수는 없다. 왜냐하면, 추출된 눈 및 입역이 작고 유사한 화소값으로 구성되어 있어 히스토그램 평활화를 수행하면 오히려 잡영을 강조하는 결과를 낳기 때문이다. 이를 해결하기 위해 가장 큰 값을 가지는 빈을 기준으로, 입력영상의 히스토그램을 비교하고자 하는 참조 히스토그램 방향으로 이동시켜 비교하는 방법을 적용하였다. 그림 15에 이를 도식화하였다. 어두운 눈 영역에 대한 히스토그램을 평균 눈과 선글라스의 히스토그램을 고려하여 변환한 후

비교한 결과를 보면 이 히스토그램이 평균 선글라스 히스토그램 보다 평균 눈 히스토그램과 더욱 유사해지는 것을 확인할 수 있다.

5. 실험

5.1 얼굴윤곽추출

얼굴윤곽을 추출하기 위한 ASM 모델 학습은 연구실에서 획득한 10명으로부터 5가지 포즈에 대해 10장씩 수집하여 총 500장의 320×240 해상도의 영상에 대해 이루어졌다. 그림 16은 5가지 포즈 (정면, 상, 하, 좌, 우)에 해당하는 일부 영상을 보여준다.

얼굴영역 검출은 100장의 연속프레임을 대상으로 실험해 본 결과 1차 얼굴검출에서 84프레임에 대해 정확하게 검출하였고, 배경모델을 사용하는 2차 검출방법에서는 100프레임 모두에서 정확하게 얼굴을 검출하였다. 5개의 ASM 모델 초기위치 설정 방법은 96프레임에서 성공하였다.

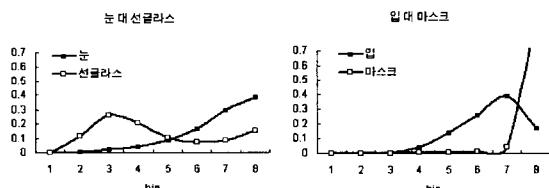


그림 14. 눈과 선글라스, 입과 마스크 영역에 대한 히스토그램 비교

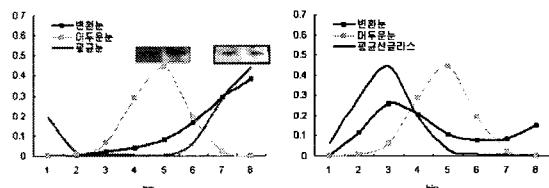


그림 15. 어두운 눈에 대한 히스토그램을 이동시켜 평균 눈(왼쪽) 및 선글라스(오른쪽)의 히스토그램과 비교한 결과



그림 16. 5가지 포즈에 대한 일부 영상

그림 17은 ASM 모델의 초기위치 설정 방법에 따른 인식률을 실험하였다. 실험결과 이동 방향 변화보다 이동 화소의 크기에 더 영향을 많이 받음을 알 수 있다. 실험한 4가지 중에서는 4방향으로 10화소씩 이동하면서 초기 모델의 크기를 크게 잡은 A 경우에 가장 낮은 실패율을 보였다. 이러한 결과는 2단계에 걸친 얼굴후보영역 검출결과가 비교적 정확함을 의미한다. 즉, 모델의 초기위치가 얼굴의 중심부에 위치해 있는데 이동 화소의 크기를 크게 하면 중심에서 벗어날 가능성이 커지기 때문이다.

그림 18은 5방향의 얼굴 포즈변화와 선글라스 및 마스크 착용에 따른 ASM 정합에 대한 선수실험 결과를 보여준다.

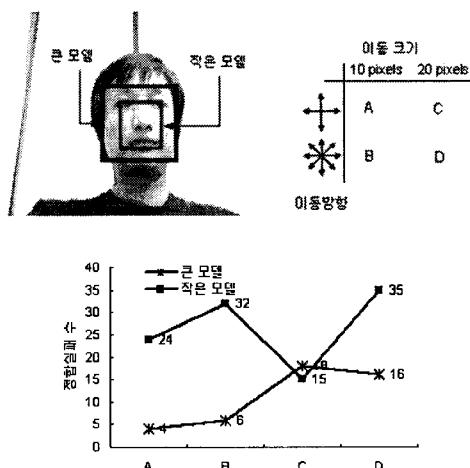


그림 17. ASM 모델 초기위치 크기, 이동방향, 이동크기에 따른 정합성등

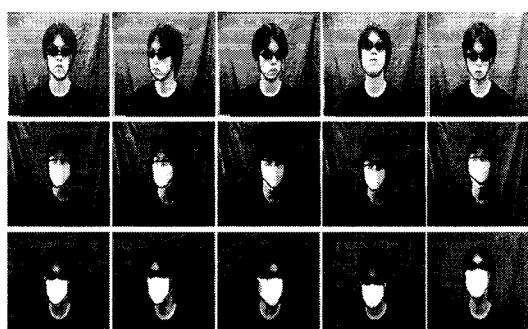


그림 18. ASM 정합결과. 맨 윗줄은 선글라스만 착용한 경우. 가운데는 마스크만 착용한 경우. 아래는 선글라스 및 마스크를 착용한 경우. 왼쪽 줄부터 차례대로 정면, 왼쪽, 오른쪽, 위, 아래 방향을 향하는 얼굴영상

5.2 선글라스 및 마스크 착용판별

선글라스 및 마스크 착용 판별을 위한 학습 데이터는 ASM을 위한 학습 데이터에 대해 모델을 정합시키고 이로부터 눈과 입, 선글라스와 마스크 영역 영상을 수집하였다. 데이터베이스는 10명에 대하여 눈, 입, 선글라스, 마스크 영상을 각각 5장씩 선정하여 총 200장의 영상으로 구성하였다. 실험에서는 조명의 밝기 차이가 있는 경우에도 제안하는 히스토그램 인텍싱이 적용 가능한지를 검사하였다. 그림 19은 각 영상의 일부를 보여준다.

실험은 데이터의 양이 적은 경우에 유용한 방식인 Leave-One-Out[18] 방식으로 49장을 학습하고 1장을 테스트하는 것을 50번 수행하였다. 표 1은 히스토그램 빈의 개수에 따라 잘 못 판정한 경우의 수와 전체 에러율을 보여준다.

빈의 개수가 8개일 때 선글라스 착용을 제대로 판정하지 못한 경우 3건만 제외하곤 에러가 없어 가장 낮은 에러율을 보였다. 빈의 개수가 4일 때는 눈과 입에서 모두 에러가 발생하였는데, 이는 빈이 8개 이하의 경우는 전체 변화를 충분히 반영하지 못한다는 것을 의미한다. 빈이 8개 초과하는 경우 증가된 에러는 주어진 데이터에 특화되었다는 것을 의미한다. 특히, 선글라스에 대한 오류가 급격하게 증가하였는데, 이러한 결과는 그림 19의 선글라스 영상으로부터 추측할 수 있다. 선글라스 특성상 빛 반사, 눈 비침, 마스크 영역의 겹침 등으로 인해 선글라스 영상들의 변화 정도가 크기 때문에, 빈의 수를 늘릴수록 이러한 개별적 특성이 반영되었기 때문으로 추정할 수 있다.

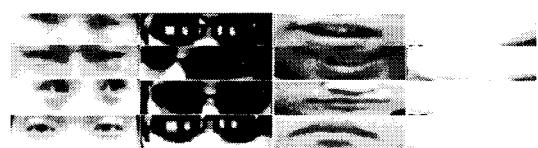


그림 19. 눈, 선글라스, 입, 마스크 영상

표 1. 장신구 착용 판별 성능

빈의 개수	눈	선글라스	입	마스크	전체 에러율
4	2/50	0/50	8/50	0/50	0.20
8	0/50	3/50	0/50	0/50	0.06
16	1/50	5/50	0/50	0/50	0.12
32	0/50	19/50	0/50	0/50	0.38

표 2. 제약조건을 적용한 경우의 장신구 착용 판별 성능

빈의 개수	눈	선글라스	입	마스크	전체 에러율
4	2/50	1/50	2/50	0/50	0.10
8	0/50	0/50	0/50	0/50	0.00
16	1/50	4/50	0/50	0/50	0.10
32	0/50	19/50	0/50	1/50	0.40

조명변화에 대처하기 위해 히스토그램 평활화를 적용하지 않고, 히스토그램 비교단계에서 가장 큰 값을 가지는 빈을 기준으로, 입력영상의 히스토그램을 비교하고자 하는 참조 히스토그램 방향으로 이동시켜 비교하는 방법을 적용하였다. 그러나, 빈의 개수가 작은 경우 빈의 이동이 조명변화에 의한 변화보다 더 큰 변화를 유발하는 경우가 있어, 두 히스토그램에서 최대값을 가지는 빈의 차이가 1인 경우는 이를 적용하지 않도록 제약하였다. 표 2는 이러한 제약을 적용한 경우의 성능을 보여준다.

실험 데이터 집합이 작아 신뢰도는 낮지만 빈의 개수가 적은 4와 8의 경우, 전체 에러율 측면에서 각각 10%와 6%의 성능 향상을 보여주었다.

6. 요약 및 향후 연구

본 논문에서는 현금 인출기 사용자의 선글라스나 마스크 착용여부를 판별하기 위한 시스템을 설명하였다. 현금 인출기 응용은 얼굴검출 연구의 대표적 응용사례이나 얼굴의 외양정보에 기반하는 기존의 고성능 얼굴검출 기법들을 적용할 수는 없다. 제안하는 시스템은 얼굴모양모델에 기반한 얼굴윤곽 검출에 주안점을 두었다. 모델의 초기위치에 민감한 얼굴 모양 모델의 정합성능을 확보하기 위하여 2단계에 걸친 얼굴후보 검출과정을 제시하였으며, 다중 정합 후 이들의 정합 정도를 평가할 수 있는 측정치를 제시하였다. 안정적인 얼굴후보 검출을 위해 움직임 정보와 배경정보를 함께 활용하였다. 배경모델을 갱신하기 위해 시스템을 대기상태와 동작상태로 구분하고 대기상태에서 배경모델을 위한 영상을 주기적으로 취득할 수 있도록 하였다. 마스크나 선글라스 착용 판별은 얼굴모양모델의 정합 정확도가 떨어지는 경우에도 동작할 수 있도록 히스토그램인텍싱을 도입하였다. 이러한 시스템 설계를 통해 실험실에서 획득한 영상에 대해 선글라스와 마스크 착용을 구분해

낼 수 있었다. 그러나, 향후 실험실 환경이 아닌 실제 환경에서 취득한 데이터에 대한 실험을 통해 파악된 문제점을 토대로 성능 개선이 필요하며, 특히 손 등에 의해 가려진 얼굴에 대한 추가 연구가 필요하다.

참 고 문 헌

- [1] M. Yang, D. Kriegman, and N. Ahuja, "Detecting Faces in Images: A Survey," *IEEE Trans. PAMI*, Vol. 24, No. 1, pp. 34–58, 2002.
- [2] K. Sung and T. Poggio, "Example-Based Learning for View-Based Human Face Detection," *IEEE Trans. PAMI*, Vol. 20, No. 1, pp. 39–51, 1998.
- [3] H. Rowley, S. Baluja, and T. Kanade, "Neural Network-Based Face Detection," *IEEE Trans. PAMI*, Vol. 20. No. 1, pp. 23–38, 1998.
- [4] C. Liu, "A Bayesian Discriminating Features Method for Face Detection," *IEEE Trans. PAMI*, Vol. 25, No. 6, pp. 725–740, 2003.
- [5] H. Schneiderman and T. Kanade, "Probabilistic Modeling of Local Appearance and Spatial Relationships for Object Recognition," *IEEE Conf. on CVPR*, pp. 45–51, 1998.
- [6] B. Heisele, T. Serre, M. Pontil, and T. Poggio, "Component-Based Face Detection," *Proc. of the IEEE Conf. on CVPR*, pp. 657–662, 2001.
- [7] R. Hsu, M. Abdel, and A. Jain, "Face detection in color images," *IEEE Trans. PAMI*, Vol. 24, No. 5, pp. 696–706, 2002.
- [8] R. Viola and M. Jones, "Robust Real-Time Face Detection," *Int'l Journal of Computer Vision*, Vol. 57, No. 2, pp. 137–154, 2004.
- [9] S. Birchfield, "Elliptical head tracking using intensity gradients and color histograms," *Proc. of IEEE Int'l Conf. on Computer Vision and Pattern Recognition*, pp. 232–237, 1998.
- [10] H. Nanda and K. Fujimura, "A robust elliptical head tracker," *Proc. of IEEE Int'l Conf. on Automatic Face and Gesture Recognition*, pp. 469–474, 2004.
- [11] T. Cootes, D. Cooper, C. Taylor, and J.

- Graham, "Active Shape Models-Their Training and Application," *Computer Vision and Image Understanding*, Vol. 61, No. 1, pp. 38-59, 1995.
- [12] T. Cootes, G. Edwards, and C. Taylor, "Active Appearance Models," *IEEE Trans. PAMI*, Vol. 23, No. 6, pp. 681-685, 2001.
- [13] I. Matthews and S. Baker, "Active Appearance Models Revisited," *Int'l Journal of Computer Vision*, Vol. 60, No. 2, pp. 135-164, 2004.
- [14] R. Gross, I. Matthews, and S. Baker, "Generic vs. Person Specific Active Appearance Models," *Image and Vision Computing*, Vol. 23, No. 11, pp. 1080-1093, 2005.
- [15] F. Torre and M. J. Black, "Robust Principal Component Analysis for Computer Vision," *Int'l Conf. on Computer Vision*, pp. 362-349, 2001.
- [16] B. Funt and G. Finlayson, "Color Constant Color Indexing," *IEEE Trans. PAMI*, Vol. 17, No. 5, pp. 522-529, 1995.
- [17] R. Gonzalez and R. Woods, *Digital Image Processing*, Prentice Hall, 2002.
- [18] T. Mitchell, *Machine Learning*, McGraw Hill, 1997.



임 동 악

2005년 금오공과대학교 컴퓨터
공학과 (학사)
2007년 금오공과대학교 컴퓨터
공학과 (석사)
관심분야 : 얼굴검출, 영상처리,
패턴인식



고 재 필

2006년 연세대학교 전산학과
(학사)
2008년 연세대학교 컴퓨터과학
과 (석사)
2004년 연세대학교 컴퓨터과학
과 (박사)
2004년~현재 금오공과대학교 컴퓨터
공학과 조교수
관심분야 : 객체인식, 패턴인식, 머신러닝