

색상 정보를 이용한 자동 독화 특징 추출

이경호*, 양 룡**, 이상범***

Automatic Speechreading Feature Detection Using Color Information

Kyong Ho Lee *, Ryong Yang **, Sang Burm Rhee ***

요 약

얼굴 특징들을 추출하는 것은 자동 독화나 휴먼컴퓨터 인터페이스, 얼굴 인식, 얼굴 이미지 데이터베이스 관리 등에서 매우 중요하다. 본 논문에서는 영상에 존재하는 다양한 색상 정보를 이용하여 얼굴 영역에서 자동 독화를 위한 특징점이 추출되도록 하였다. 얼굴의 특징들은 휘도와 채도 성분으로 인하여 다양한 색 공간에서 다양한 표현 값을 갖는다. 이를 이용하여 각 표현 값들을 증폭하거나 축소, 대비시킴으로써 얼굴 특징들을 추출되게 하였다. 눈과 코, 안쪽 입의 외곽선, 이의 외곽선을 찾았고 실험하여 좋은 결과를 얻었다.

Abstract

Face feature detection plays an important role in application such as automatic speechreading, human computer interface, face recognition, and face image database management. We proposed an automatic speechreading feature detection algorithm for color image using color information. Face feature pixels is represented for various value because of the luminance and chrominance in various color space. Face features are detected by amplifying, reducing the value and make a comparison between the represented image. The eye and nose position, inner boundary of lips and the outer line of the tooth is detected and show very encouraging result.

▶ Keyword : 자동독화(Automatic Speechreading), 영상처리(Image Processing), 얼굴 특징 추출 (Face Factor Detection)

• 제1저자 : 이경호

• 접수일 : 2008. 11. 4, 심사일 : 2008. 11. 12, 심사완료일 : 2008. 11. 26.

* 한라대학교 정보통신공학부 교수 ** 인하공업전문대학 컴퓨터정보과 교수

*** 단국대학교 컴퓨터학부 교수

I. 서론

최근 음성인식 분야에서 잡음 환경에서 현저하게 떨어지는 인식률을 높이기 위한 보상 방법의 하나로 화상을 분석하여 발화를 이해하려는 자동 독화 연구들이 많이 시도 되고 있다. 영상에서 조음 현상을 나타내는 정보를 추출하여 분석하고 평가하여 음향적 관점에서 추출하지 못하는 정보를 제공하여 음성의 인식률을 높여려는 것이다. 시각적 관점에서 조음 현상을 나타내는 정보를 추출하려면 입력되는 영상에서 발화자를 찾아야 하며, 발화자로부터 조음 현상을 짐작할 수 있는 관심 점들을 찾아 발화를 인식할 수 있는 가치 있는 파라미터를 제공할 수 있어야 한다[1]. 대화 시 전달하고자 하는 의미는 몸의 여러 곳이 표현하지만 많은 발화 정보들은 얼굴에 집중되어 있다. 얼굴 찾기와 관심 점들의 추출은 고도의 영상 처리 기술이 필요하며, 이는 자동 독화에서 제일 먼저 건너야 하는 다리로 많은 연구자들이 자동 독화 성능 저하 요인을 영상처리에서 언급하고 있다.

본 연구에서는 영상에 존재하는 다양한 색공간의 색상 정보를 증폭하거나 축소, 대비시킴으로서 얼굴 특징들을 추출되게 하여 특징점을 추출하였다.

II. 기존 연구

자동 독화 초기 연구자인 Petajan은 Gray 얼굴 영상에 경계 값을 정하여 2진 이미지로 변화시켜 입안은 검게 그리고 밖은 하얗게 만들어 입의 높이와 폭, 넓이, 입의 원주를 계산하는 파라미터를 추출하였다[2,3]. Finn과 Montgomery는 특별한 표식자를 사용하여 입술 사이, 입술과 코 사이, 입술과 턱 사이의 다양한 거리를 추적하였다[4]. Stork도 마커를 이용, 다섯 개의 거리를 측정하고 이를 파라미터로 사용하였다[5]. Yuhas 등은 입주위 20x25 그레이 스케일 영상을 바로 신경망 시스템 파라미터로 제공하였다[6]. Mase 등은 연속된 두 화상에서 입 주위 상하좌우 영상들의 광학적 흐름을 계산하였다. 입을 중심으로 한 사각형에서 여덟 개의 요소의 평균값을 계산하고 입의 열리고 닫히는 평균을 기초로 하여 분류에 사용되는 벡터 값을 파라미터로 사용하였다[7]. Silbee는 조명의 차이를 해결하기 위해 전처리와 정규화를 하고, 입 주위의 영상을 80x80의 벡터 양자화하여 파라미터로 사용하였다[8]. Bregler는 입 주위 13x13 영역을 Fourier 변환을 통해 파라미터로 사용하였고, 나중에는 active

contour model 화상에서 80개의 파라미터로 묘사되는 입 모양을 추출하여, 입 모양을 맵핑하는 5개의 파라미터를 추출하였다[9]. Chiou는 칼라 영상에서 주성분분석과 active contour model을 통하여 2개 파라미터를 계산하였다[10]. Adjudani 등은 발화자의 머리에 장착시킨 카메라를 이용하여 발화자를 찾는 과정을 생략하였으며, 입력되는 영상 전체를 직접 이용하였다[11]. Meier 등은 통계적 피부색 모델을 이용하여 얼굴 영역을 추적한 다음 입술 영역을 추출하고 입술의 크기를 24x18로 정규화하고 그대로 인식기의 데이터 벡터로 사용하거나 주성분 분석이나 선형 판별 분석을 적용하여 구한 16개의 개수를 파라미터로 사용하였다[12]. 김진병 등은 카메라 입력을 입주위로 제한하고, 입력된 영상은 4영역으로 분할하여 영상 보정을 하고 이진화를 통한 입 찾기와 영상 기반의 이산 코사인 변환과 이산 웨이블릿 변환을 이용, 특징을 추출하였다[13,14]. 이지은 등은 코에서 턱까지만 촬영한 이미지로 그레이레벨에서 선형 명암 마스크를 이용하여 영상 보정을 하고 크기를 줄여 주성분 분석과 은닉 마르코브 모델을 이용하여 자동 독화를 수행하였다[15,16]. 민덕수는 입술로 제한된 범위 내에서 영상에 기반을 둔 방법으로 입술 움직임을 파라미터로 하였다[17]. 민소희 등은 마커를 붙이고, 정면 얼굴과 거울을 이용한 측면 영상을 함께 확보하는 방법으로 정면과 측면을 관측하고 입술의 폭과 높이, 윗입술에서 턱까지의 거리, 코로부터 윗입술까지 거리, 코부터 턱까지의 거리를 파라미터로 취득하였다[18,19]. 신도성 등은 입술 영역으로 제한된 입력으로 다중 선형 회기분석으로 임계값을 설정하는 2진화 임계값으로 취득하였다.[20-23]. 이은숙 등은 피부색 픽셀 분포 히스토그램을 만들어 가로와 세로의 분포를 중심으로 얼굴 범위 추출하고, 입술의 검출은 구조와 색도를 이용하여 검출하며, 입술의 R 성분은 얼굴의 R 성분 보다 크다는 정보를 이용하여 행렬 평균을 내어 입술로 판단하고 블록 매칭 기법을 이용하여 확인하여 특징점을 찾았다[24]. 자동 독화 성능은 발화자를 잘 찾고 조음기관 변화를 정밀히 인식하여 그 변화를 파라미터로 추출하는 것이다[25]. 그러나 많은 연구가 이미 얼굴을 검색한 것을 전제로 하거나 입술만 촬영하여 학습 및 인식실험을 하거나, 입술에 특별한 색칠을 한 후 얼굴을 촬영하여 입술 영역을 검출하거나, 얼굴에 마커를 붙여 영역을 확인하기도 하고, 수작업으로 관심 점들의 변화를 추출하기도 한다.

III. 본론

3.1 전체 과정

특징점 추출 전체 과정은 크게 1) 얼굴 사진을 이용한 사진 분석 작업으로 특징점들의 위치 통계 정보를 파악하며, 다양한 색공간에서 특징점의 표현 상태를 분석하고, 2) 얼굴 검출을 위해 피부색 추출을 위한 휘도기반 sine-wave 조명보정과 영상에서 피부색 화소 추출, 피부색 화소 후처리, 피부색 블록 구성 및 얼굴 후보 영역을 구성하며, 3) 특징점 추출을 위해 눈을 추출하고 이를 바탕으로 입 마스크를 형성하며, 입 마스크 상에서 입 안쪽 경계선 추출하고, 이 외곽선과 코를 특징점으로 추출하여 이를 이용하여 파라미터를 형성한다.

3.2 사전 분석 작업

3.2.1 얼굴 요소별 두 눈 거리 대비 분석

얼굴 구성 요소별 관심 점 위치 파악을 위해 각 구성 요소의 위치들의 통계 분석을 하였다. 180명의 정면 얼굴의 눈 사이 거리를 기준하여 두 눈 중점으로부터 코끝과 윗입술 턱끝의 위치 분포 조사하였다. 이는 얼굴 구성 요소를 찾는 계산량을 줄이고자 하는 의도이다.

표 1. 얼굴 특징점 거리 분석
Table 1. Facial Feature Distance analysis

| 구분 | 코끝 | 윗 입술 | 턱끝 |
|------|-------|--------|--------|
| 최소 | 65.6% | 86.3% | 169.2% |
| 최대 | 99.2% | 126.2% | 226.1% |
| 평균 | 80.0% | 104.2% | 195.8% |
| 표준편차 | 6.4% | 9.2% | 11.6% |
| 범위 | 33.6% | 40.0% | 56.9% |

*. 두 눈간의 거리를 100으로 했을 때 비율임.

3.2.2 색공간 배치를 통한 특징 점 추출용이 환경 파악

칼라 영상 표현에 널리 사용되는 RGB, YCbCr 색공간의 각 단위 성분별로 분리하여 살펴보면 피부색 뿐 아니라 얼굴 구성 요소의 보임의 정도가 성분별로 특색 있게 보임을 알 수 있다. 빛의 영향 때문에 채도 성분을 구분할 수 없기도 하며, 성분별로는 대비되어 보이기도 한다. 특히 Cb Cr 영상에서 눈 요소 성분값이 대비되어 보이며, 채도 성분이 약한 이는 Cr 상태에서 구분하기 힘들다. 그림1의 영상은 R, G, B, Y,

Cb, Cr, R, G, B, Y, Cb, Cr 영상이다.



그림 1. RGB와 YCbCr 색공간의 단일 성분별 영상
Fig 1. Each Component Images of RGB and YCbCr

3.3 얼굴 검출

3.3.1 조명 보정

조명 보정을 위해 널리 사용되는 방법은 Gray-World-Assumption(GWA) 기법과 Reference-White(RW) 기법이 있다. GWA 기법은 휘도 값 평균을 각각의 R, G, B 평균으로 나누어 비례 값을 만들고 비례 값을 R, G, B 별 각각의 화소에 반영하는 것이다. RW 기법은 휘도 성분으로 표현된 Y 영상의 상위 5% 화소의 수를 조사하여 화소 개수가 지정한 경계 값 보다 많다면 조명에 의하여 영향을 받았다고 판단하여 채널 최고값을 R, G, B 각 채널의 평균값으로 나눈 비례 값으로 재조정하는 것이다. GWA은 인간의 시각에는 기여하는 것으로 보이나 본 연구에서 수용한 피부색 판단 함수에 의한 피부색을 추출에는 크게 도움이 되지 않았다. RW 방법은 R, G, B 각 영역에서 어두움에 기여하는 화소를 높은 비율로 재조정하여 어두운 부분을 밝게 하려는 의도를 가지고 있으나, 경계값 설정에 어려움이 있었으며, 성분비가 1보다 큰 값이 나와 전체 영상을 밝게 하는 효과가 있어, 경우에 따라서는 영상을 지나치게 밝게 만드는 결과를 가져오게 된다.

전 영상에 일괄적으로 적용하는 것이 아닌 어두운 곳은 좀 큰 비율로 보정하고 밝은 곳은 작은 비율로 적용하는 차별적 적용이 필요하였고 본 연구에서는 어두운 곳은 좀 큰 값으로 보정하고, 밝은 부분은 좀 적은 값으로 보정하기 위해

sigmoid 함수와 sine 함수로 실험하였다. 빛이 밝기 정보는 Y 휘도 성분을 이용한다. 또한 차별적 변형을 할 때 피부가 아닌 것이 피부색을 갖도록 조절되는 것을 방지할 필요가 있다. 본 논문의 실험에 사용된 sigmoid-wave를 이용한 조명 보정식은 식 (3.1)과 같으며 sine-wave(SW)를 이용한 조명 보정식은 식 (3.2)와 같다.

$$\begin{aligned}
 nR_{(i,j)} &= R_{(i,j)} * (sR * \frac{1}{1 + e^{-(a * R_{(i,j)}/255 - b)}}) \\
 \text{단, } -3 \leq a \leq 0, -1 \leq b \leq 1 \\
 nG_{(i,j)} &= G_{(i,j)} * (sG * \frac{1}{1 + e^{-(a * G_{(i,j)}/255 - b)}}) \\
 \text{단, } -3 \leq a \leq 0, -1 \leq b \leq 1 \\
 nB_{(i,j)} &= B_{(i,j)} * (sB * \frac{1}{1 + e^{-(a * B_{(i,j)}/255 - b)}}) \\
 \text{단, } -3 \leq a \leq 0, -1 \leq b \leq 1
 \end{aligned} \tag{3.1}$$

$$\begin{aligned}
 nR_{(i,j)} &= R_{(i,j)} * (1 + \sin(R_{(i,j)}/m) / p * \pi * 20 / 255) \\
 nG_{(i,j)} &= G_{(i,j)} * (1 + \sin(G_{(i,j)}/m) / p * \pi * 20 / 255) \\
 nB_{(i,j)} &= B_{(i,j)} * (1 + \sin(B_{(i,j)}/m) / p * \pi * 20 / 255) \\
 &* \text{mov는 피부색 평균으로 Y를 중심으로 1.30이며,} \\
 & p \text{ 1.01로 실험하였다.}
 \end{aligned} \tag{3.2}$$

약 50개의 영상을 이용한 실험에서 sigmoid 함수에 의한 변형은 어두운 것을 너무 밝게 하는 경향이 있어 피부색이 아닌 것을 피부색의 영역으로 변환시키는 문제가 있었으며, 여러 가지로 조절해 보았으나 적절한 파라미터의 값을 찾을 수 없었다. 반면에 sine 함수에 의한 변형은 진폭과 주기 조절을 통하여 양극단을 적절히 제어할 수 있어 피부색이 아닌 것을 피부색으로 변환시키는 경우를 많이 제어할 수 있었으며, 조명에 의하여 제외되었던 피부 영역을 많이 회복시키는 결과를 보여주었다. 그림 2에서는 GWA에 의한 보정이 시각적으로 좋게 보이며, RW는 좀 밝게 보이고, SW는 좀 칙칙하게 보이나 그림 3의 피부색 추출 결과에서는 SW 조명 보정 결과가 좋음을 알 수 있다.



그림 2. 조명 보정 결과
Fig 2. Result of Lighting Compensation

3.3.2 피부색 추출

널리 사용되는 피부색 추출 방법은 RGB 색공간에서 'R > x, G > y, B > z' 와 같은 조건을 통한 추출법, HSI 색공간에서의 H 값 23도~27도 사이 추출 방법이 있다. 이 방법들은 연속적이지 않은 피부색 범위를 연속적인 공간으로 간주함으로써 피부가 아닌 화소도 피부색으로 평가하는 경향이 크다. Hsu는 다양한 인종의 853,571개의 피부색 화소를 추출하여, 추출한 화소를 9개의 색공간 공간에 배치하고 비교 후 기존 색공간에서 피부색 추출의 어려움을 개선하여 피부색 화소가 잘 추출될 수 있는 새로운 색공간으로의 변형하는 함수를 개발하였다. 그 변형 함수는 기존 방법이 색공간에서 연속적이지 않은 피부색 범위를 연속적인 공간으로 간주하는 실수를 많이 피할 수 있었다. 적용한 연산식은 다음과 같다.[26]

$$C'_i(Y) = \begin{cases} (C_i(Y) - \bar{C}_i(Y)) \cdot \frac{W_{C_i}}{W_{C_i}(Y)} + \bar{C}_i(K_h) & \text{if } Y < K_l \text{ or } K_h < Y \\ C_i(Y) & \text{if } Y \in [K_l, K_h] \end{cases} \tag{3.3}$$

$$W_{C_i}(Y) = \begin{cases} WL_{C_i} + \frac{(Y - Y_{\min}) \cdot (W_{C_i} - WL_{C_i})}{K_l - Y_{\min}} & \text{if } Y < K_l \\ WH_{C_i} + \frac{(Y_{\max} - Y) \cdot (W_{C_i} - WH_{C_i})}{Y_{\max} - K_h} & \text{if } K_h < Y \end{cases} \tag{3.4}$$

$$\bar{C}_b(Y) = \begin{cases} 108 + \frac{(K_l - Y) \cdot (118 - 108)}{K_l - Y_{\min}} & \text{if } Y < K_l \\ 108 + \frac{(Y - K_h) \cdot (118 - 108)}{Y_{\min} - K_h} & \text{if } K_h < Y \end{cases} \tag{3.5}$$

$$\bar{C}_r(Y) = \begin{cases} 154 - \frac{(K_l - Y) \cdot (154 - 144)}{K_l - Y_{\min}} & \text{if } Y < K_l \\ 154 + \frac{(Y - K_h) \cdot (154 - 144)}{Y_{\max} - K_h} & \text{if } K_h < Y \end{cases} \tag{3.6}$$

$$\frac{(x - ec_x)^2}{a^2} + \frac{(y - ec_y)^2}{b^2} = 1 \tag{3.7}$$

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} C'_b - C_x \\ C'_r - C_y \end{bmatrix} \tag{3.8}$$

YCbCr 공간에서는 위에서 보여준 그림과 같이 채도 성분 Cb와 Cr은 밝기 성분 Y의 함수로서 간주되어 Cb(Y), Cr(Y)로 되며, 채도 성분은 C'b(Y)와 C'r(Y)로 변형된다. 피부색 모델의 중심은 $\bar{C}_b(Y)$ 와 $\bar{C}_r(Y)$ 로 표현되며, 범위는 $W_{C_b}(Y)$ 와 $W_{C_r}(Y)$ 로 표현된다. C'_i 로 표현된 것은 C_b 와 C_r 을 의미하며, W_{C_b} 는 46.97, WL_{C_b} 는 23, WH_{C_b} 는 14, W_{C_r} 는 38.76, WL_{C_r} 는 20, WH_{C_r} 는 10, K_l 은 125, K_h 는 188 이고, $C_x = 109.38$, $C_y = 152.02$, $\theta = 2.53$,

$ec_x = 1.60, ec_y = 2.41, a = 23.59, b = 14.03$ 이다. 그림 3은 피부색 판단식에 의해 추출된 피부색이다.



그림 3. 피부색 판단 예
Fig 3. Skin Color Detection Examples

3.3.3 피부색 추출 후 후처리

피부색 추출 후 잡음 제거 및 들출 픽셀의 정리 작업이 필요하다. 격리된 피부색 픽셀과 피부색으로 둘러 쌓인 피부색 속에 잡음의 복원은 메디안 필터 방법을 이용하여 복원하고, 들출 픽셀은 4-연결 링크를 이용한 방법으로 제거하였다. 또 수축과 팽창 연산을 이용하여 잡음을 제거하였다. 그림 4의 윗 영상이 수행 전, 아랫 영상이 결과이다.

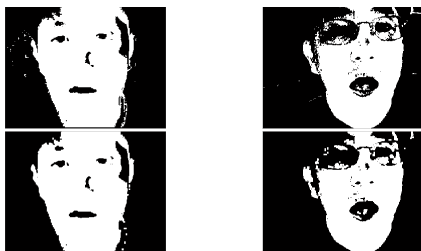


그림 4. 피부색 영역 후처리
Fig 4. Skin Color Regions Postprocess

3.3.4 피부색 블록 구성

자동 독화를 위한 파라미터는 얼굴에 존재하므로 파라미터 추출을 얼굴로 제한하기 위한 범위 설정과정이 본 과정이다. 2차원배열 2진값에서 반복되는 Two Pass Connected Component Labelling(TwoPassCCL) 알고리즘을 이용하여 8-연결 그래프로 구성하였다. 방법은 라벨 배열과 동등 정보 저장 공간을 준비하고 패스1에서는 백그라운드가 아닌 픽셀들을 모두 라벨 공간에 배치한다. 그리고 행 다음 열 순으로 데이터의 각 요소를 반복하되, 백그라운드가 아니어서 라벨링 해야 하면 이웃 요소를 살펴 라벨된 이웃 중 가장 작은 라벨 값을 찾아 붙이고, 이웃 라벨과 동등함을 저장한다. 만일 이웃이 없다면 현 요소에 유일한 라벨을 붙인다. 이런 방법으로 모든 화소를 처리하고 패스2로 들어간다. 패스2에서

동등 정보를 이용하여 연결된 요소에 같은 라벨을 부여 한다. 결과인 피부색 블록이 그림 5이다.

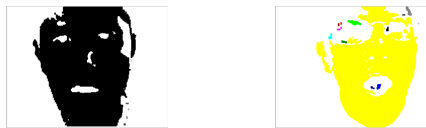


그림 5. TwoPassCCL 수행 방법 및 결과
Fig 5. Method and Result of TwoPassCCL

3.3.5 얼굴 후보 마스크 구성

안전한 얼굴 후보 마스크 구성을 위해 블록 다각형 중 'Graham Scan Algorithm'을 적용하였다. 방법은 가장 낮고 가장 오른쪽 화소를 P0로 선정 후 x축을 기준으로 P0와 각 화소가 이루는 각을 기준으로 정렬하고, P0부터 세 화소로 양끝 화소를 연결한 선을 기준하여 중간 화소가 오른쪽에 위치하면 다각형 구성에 채택하며 왼쪽에 있으면 버림 작업을 수행한다. 그림 6의 윗 영상이 구성된 마스크 영상이며, 아랫 영상이 마스크를 씌워 출력한 것이다.

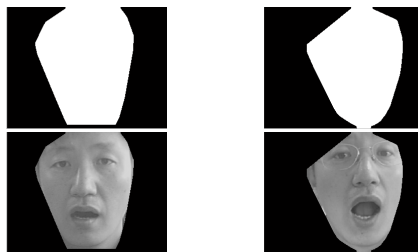


그림 6. 얼굴 마스크와 얼굴 마스크를 씌워 출력한 영상
Fig 6. Face Mask and Face Masked Image

3.4 특징 검출

3.4.1 눈 추출

얼굴 후보 영역 내에 얼굴의 가부를 판단하기 위해 눈 영역을 추출한다. 채도 상의 Cb에서 눈은 비교적 높은 값으로 나타나나 Cr 상에서의 눈은 비교적 낮은 값으로 나타남을 이용하여 대비시켜 찾으며, 휘도 상에서도 눈동자와 흰자위를 이용하여 한번은 밝은 성분이 득세하도록 만들고 한번은 검은 성분이 득세하도록 만든 후 두 영상을 대비시켜 눈을 찾는 후 채도 성분에서 찾은 것과 논리곱 연산을 통하여 검증 확정하는 것이다. 이 방법은 영상의 크기에 의존적이지 않고 매우 견고하게 잘 추출된다[26]. 채도 영역에서의 눈의 추출은 아

래와 같은 식을 이용한다. 여기서 \bar{C}_r 은 C_r 음의 값이다.

$$Eye\ Map\ Chroma = \frac{1}{3} (C_b^2 + \bar{C}_r^2 + C_b/C_r) \quad (3.9)$$

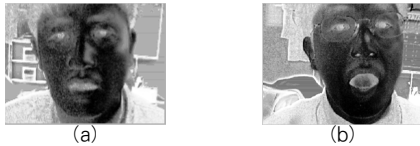


그림 7. EyeMapChroma 결과
Fig 7. Result of EyeMap Chroma

휘도영역을 기반으로 한 눈 영역 검출은 아래 식을 이용하여 '⊕'과 '⊖'은 모폴로지 팽창과 침식연산을 의미한다.

$$Eye\ Map\ Luma = \frac{Y(x,y) \oplus g_\sigma(x,y)}{Y(x,y) \ominus g_\sigma(x,y) + 1} \quad (3.10)$$

$$eyemap - luma1 = Y(x,y) \oplus g_\sigma(x,y) \quad (3.11)$$

$$eyemap - luma2 = Y(x,y) \ominus g_\sigma(x,y) \quad (3.12)$$

식에서 $C_b^2, \bar{C}_r, C_b/C_r$ 은 모두 [0~255]값으로 정규화 하며, 색차의 간격을 넓게 하기 위하여 히스토그램 균등화를 적용한다. 이후 밝기 순 상위 10%를 경계값으로 하여 이진화를 수행한다. 이진화 후 눈의 위치 파악은 피부색 범위 내의 구조상 두 눈의 대칭성 및 위치를 반영하는 평가 함수를 적용하여 눈을 인식하는 방법을 이용하였다.

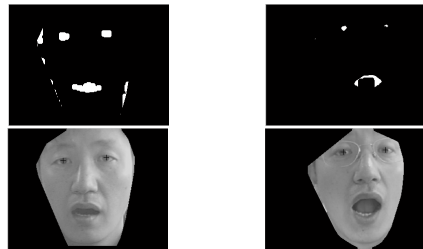


그림 8. 눈 찾기
Fig 8. Eyes Detection

3.4.2 입 마스크 추출

얼굴 마스크 설정 이유와 동일하게 오직 입으로부터의 정보를 추출하기 위하여 입 마스크를 구성한다. 우선 이미 찾은 눈의 정보와 특징점 통계 정보를 이용하여 입이 있을 만한 위치로 얼굴 마스크를 제한한다. 이 범위 내에서 스케일 스페이스 침식 연산 후 얼굴 마스크로 제한된 영상에서 추출된 최고 픽셀 값의 1/3 이상인 픽셀과 C_r 얼굴 마스크로 제한된 영상에서 추출된 최고 픽셀 값의 1/2 이상 이진화 영상을 논리합 연산을 한 후 모폴로지 연산을 통해 마스크를 구성한다. 이는 대략적인 입의 위치를 찾는 것이다.



그림 9. 입 마스크 형성
Fig 9. Construction of Mouth Mask

3.4.3 입 추출

C_r 영역 영상을 보면 입 내부의 영역은 그림자 진 무채색부분과 입술에 해당하는 붉은색 개통의 영역 경계가 확실히 보인다. 따라서 입 추출 연산은 통계정보를 이용 입 부분만을 Canny필터로 외곽선을 추출하고 단일화소와 보푸라기 잡음 화소를 제거하며, 한 칸 떨어진 선들은 연결하고, 수축과 팽창 연산을 통해 외각 선을 정리한 후 발화하는 입은 구멍이므로 구멍 채움 연산을 수행한다. 이후 앞에서 한 팽창 연산보다 하나 큰 값으로 스케일 구조를 통한 모폴로지 침식 연산을 하여 라인 이미지를 모두 제거하고, 한 픽셀 정도의 값을 팽창 연산을 통하여 복원해 준다. 그리고 남은 영상에서 외곽선을 추출한다.

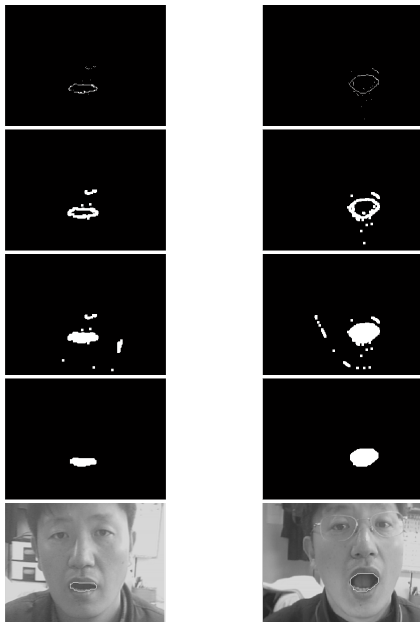


그림 10. 입 찾기
Fig 10. Mouth Detection

3.4.4 이 추출

이 추출은 입 영역 내에서 채도 성분은 죽이고 휘도 성분이 높은 것은 더욱 높이는 방법을 이용한다. 입의 영역 안에서 Y 성분을 이용하여 제공한 후 [0~255]로 정규화 한 후 Canny 필터 0.1 값으로 외각선을 구한 후 혀로 인한 경계선 부분을 제거하기 위하여 혀의 채도 성분을 이용한 마스크를 통하여 제거한다. 잘못된 혀의 경계선 제거 마스크 생성은 Cr 성분을 이용하여 10% 이상으로 구성하였다.



그림 11. 이의 외각선 찾기
Fig 11. Outline of teeth Detection

3.5.5 코 추출

찾은 눈과 입 상단의 범위 영역내에서 Canny 필터 외각선을 이용하여 추출한 화소수 중 3픽셀 두께 합 최고 지점으로 두 눈을 이용하여 그 눈 얼굴 중심선 교차점을 코의 위치로 추출하였다.

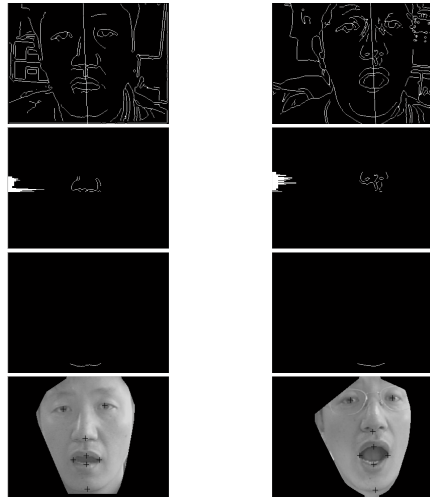


그림 12. 특징점 추출 결과
Fig 12. Results of Feature Detection

3.5.6 파라미터 추출

동 작업을 통하여 추출할 수 있는 파라미터는 입의 면적, 입의 외각선 형태, 입의 폭과 높이의 거리, 이 보임 길이, 코와 입술 상하단과의 거리 등이다.

IV. 실험결과

입력 영상은 PC용 카메라로 촬영한 것이다. 피부색 검출 후 피부색의 한 블록의 크기를 10000 픽셀 크기로 필터하여 얼굴의 크기가 대략 150x150 정도 이하일 경우 잘 검출되지 않으나 이 보다 큰 영상의 얼굴을 찾는 데는 문제가 없었다. 그러나 독화를 바탕으로 하지 않는다면 블록의 크기 조절로 해결할 수 있었다. 독화를 전제로 한 실험에 사용된 영상 크기는 640x480, 320x240였다. 얼굴의 방향은 가능한 정면이며, 조명도 특별히 조작하지 않은 상태로 역광과 심하게 그림자 지는 상태만 조심하였다. 한국어 8 단모음을 발화하는 20명의 160장 영상과 1명의 80장 총 240장의 영상을 이용하여 검출 실험을 하였다. 얼굴의 크기는 눈과 눈 사이의 거리가 대략 100픽셀이 되도록 하였다. 얼굴 마스크 구성에 실패한 영상은 없었으며 눈 찾기 실패가 3장으로 두 눈을 감은 경우가 1장 한 쪽 눈을 찡그린 경우가 2장으로 눈 찾기는 큰 문제가 없었다. 특히 안경 착용 유무에는 크게 좌우되지 않았으며, 눈의 감음 정도가 눈의 인식을 좌우하는 것으로 보인다. 입찾기는 18장이 실패를 하였는데 주로 모폴로지 연산을 통

한 잡음 제거 실패로 인한 입 경계 추출 실패이다. 입 구조상 발화시 심하게 그림자 지고 골곡이 생기는 경우도 왜곡되어 추출된다. 코의 위치 추출은 눈과 입을 찾은 영상의 경우 모두 찾고 있으나 얼굴 각도에 따라 몇 픽셀의 위치 변화가 있다. 본 연구의 효과는 표 2에서 보는 바와 같이 배경의 다양성에도 문제없이 얼굴을 찾고 있으며, 휘도에 의존하는 조명 보정으로 피부색 추출에 좋은 효과를 보이고, 안경 유무를 가리지 않고 다양한 화자에도 잘 적용하며, 발화자의 얼굴 기술 어짐에 문제가 되지 않으며, 독화 특징점 추출도 입에 제한하지 않고 코와 눈까지 추출하고 있다.

표 2. 독화 특징점 추출 방법 비교
Fig 2. Speechreading Feature Extraction Approaches

| 연구 | MB | LC | MS | SP | IS | 비고 |
|----------------------|----|----|----|----|------|-----------|
| (13) (14) | UC | C | C | UM | M | 입주위만 영상입력 |
| (15) (16) | UC | C | C | UC | M | 입주위만 영상입력 |
| (17) (22) (23) | UC | C | C | C | M | 입주위만 영상입력 |
| (24) | C | UC | UM | UM | M | 단일배경 단일조명 |
| This | C | C | C | C | M, N | 다양배경 다양조명 |

MB : 배경의 다양성 처리 / LC : 조명 보정 기능. /
MS : 화자의 다양성 처리. / SP : 발화자 기술어짐 처리
기능. / IS : 특징점 추출. / PM : 파라미터 추출. /
UC : 구현 않됨. / UM : 구현 내용 없음. / C : 구현.
/ M : 입 / N : 코



그림 13. 특징점 추출 결과 예
Fig 13. Feature Detection Examples

V. 결론

본 논문에서는 칼라 영상이 갖는 다양한 색상 정보를 이용하여 얼굴에 있는 특징 점들이 검출되도록 하였다. 특히 다양한 색공간의 다양한 표현 방법들이 얼굴 구성 요소가 성분별로 특색 있게 보이게 함을 이용하면 적절한 채도 성분이나 휘도 성분을 증폭하거나 축소시키고 대비시킴으로서 구성 요소들을 특징 점으로 추출할 수 있음을 보였다. 특히 본 논문에서 제시한 방법은 주어진 영상에서 일정 크기 이상의 얼굴 영상이면 영상 크기에 상관없이 얼굴 찾기부터 시작하여 얼굴의 구성 요소들을 찾도록 하였다. 기존 자동 독화 연구에서 얼굴을 찾은 것으로 전체 하거나 입을 찾은 것으로 하는 가정의 상태들을 해결하고자 하였고, 약간의 오차는 있으나 기계적인 관점에서 발화하는 눈의 위치와 입의 외곽선, 이의 외곽선, 코의 위치를 찾게 하여, 이들로부터 입의 형태, 입의 면적과 폭, 높이, 코와의 거리, 이의 보임 비율 또는 분포를 파라미터로 사용할 수 있도록 하였다.

본 연구는 자동 독화 뿐 아니라 휴먼컴퓨터 인터페이스, 얼굴 인식, 얼굴 영상 데이터베이스 관리 등에 응용할 수 있다. 현재 비교적 좋은 추출을 하고 있으나 다양한 조명 환경에서도 경계 추출의 정확성과 연산량을 줄이려는 연구가 필요하다.

참고문헌

- [1] Bailly, G., Vatikiotis-Bateson, E., & Perrier, P., Issues in Visual and Audio-Visual Speech Processing, MIT Press, 2004.
- [2] Petajan, E. D., "Automatic lipreading to enhance speech recognition," Ph.D. Dissertation, University of Illinois at Urbana-Champaign, 1984.
- [3] Petajan, E. D., "Automatic lipreading to enhance speech recognition," Proceedings of the IEEE Communication Society Global Telecommunications Conference, November 26-29, Atlanta, Georgia, 1984.
- [4] Finn, E. K. & Montgomery A.A. "Automatic optically based recognition of speech," Pattern Recognition Letters, 8(3):159-164, 1988.

- [5] Stork, D. G., & Hennecke, M.E., *Speechreading by Humans and Machines*, Berlin, Germany: Springer, 1996.
- [6] Yuhas, B. P., Goldstein, M.H. & Sejnowski, T.J., "Integration of acoustic and visual speech signals using neural networks," *IEEE Communications Magazine*, 27:65-71, 1989
- [7] Mase, K. & Pentland, A., "Automatic lipreading by computer," *Trans. Inst. Elec. Info. and Com. Eng.*, J73-D-II(6):796-803, 1990.
- [8] Silsbee, P. L., "Computer Lipreading for Improved Accuracy in Automatic Speech Recognition," Ph.D. dissertation, The University of Texas at Austin, 1993.
- [9] Bregler, C., Omohundro, S. M. & Konig, Y., "A Hybrid Approach to Bimodal Speech Recognition," in 28th Annual Asilomar Conference on Signals, Systems, and Computers, 1:556-560. 1994.
- [10] Chiou, G. I. & Hwang, J. N. "A neural network based stochastic active contour model (NNS-SNAKE) for contour finding of distinct features," *IEEE Trans. on Image Processing*, 4(19):1192-1195, 1995.
- [11] Adjoudani, A. et al., "A multimedia platform for audio-visual speech processing," *Proc. European Conference on Speech Communication and Technology*, Rhodes, Greece, 1671-1674, 1997.
- [12] Meier, U., Stiefelwagen, R., Yang, J., Waibel, A., "Towards Unrestricted Lipreading," *International Journal of Pattern Recognition and Artificial Intelligence*, 14(5):571-785, 2000.
- [13] 김진범, 김진영, "이미지 변환과 HMM에 기반한 자동 립리딩", *대한전자공학회 추계학술대회 논문집*, 22(2):585-588, 1999.
- [14] 김진범, 김진영, "입술 대칭성에 기반한 효율적인 립리딩 방법", *전자공학회논문지*, 37(5):455-464, 2000.
- [15] 이지은, 김진영, 이주현, "시간영역 이미지 필터링에 의한 립리딩 성능 향상", *한국음향학회 학술발표대회논문집*, 20(2):45-48, 2001.
- [16] 이지은, 시간영역 이미지 필터링에 의한 립리딩 성능 향상, *전남대학교대학원 석사학위논문*, 2002.
- [17] 민덕수, "동적 환경에서 립리딩 성능저하 요인 분석 및 인식성능 향상에 관한 연구", *전남대학교대학원 석사학위논문*, 2002.
- [18] 민소희, 김진영, 최승호, "입술 정보를 이용한 음성 특징 파라미터 추정 및 음성 인식 성능 향상", *대한음성학회지*, 44:83-92, 2002.
- [19] 김진영, 민소희, 최승호, "음성인식에서 입술 파라미터 열화에 따른 견인성 연구", *음성과학*, 10(2):27-33, 2003.
- [20] 신도성, "입술영상접기와 프레임간 필터링을 이용한 립리딩 성능 개선", *전남대학교대학원 박사학위논문*, 2004.
- [21] 김진영, 신도성, "상태공유 HMM을 이용한 서브워드 단위 기반 립리딩", *음성과학*, 8(3):123-131, 2001.
- [22] 신도성, 김진영, 최승호, "시간영역 필터를 이용한 립리딩 성능향상에 관한 연구", *한국음향학회*, 22(5):375-382, 2003.
- [23] 신도성, 김진영, 이주현, "동적 환경에서의 립리딩 인식 성능저하 요인분석에 대한 연구", *한국음향학회*, 21(5):471-477, 2002.
- [24] 이은숙, "PC 카메라를 이용한 실시간 립리딩 시스템 설계 및 구현", *원광대학교대학원 석사학위논문*, 2003.
- [25] Bailly, G., Vatikiotis-Bateson, E., and Perrier, P., *Visual and Audio-Visual Speech Processing*, MIT Press, 2004.
- [26] Hsu, R. L, Abdel-Mottaleb, M., Anil K. J., "Face Detection in Color Images," *IEEE Trans. on Pattern Analysis*, 24(5):696-706, 2002.

저자 소개



이경호

1987: 인하공업전문대학.
1991: 한국방송통신대학교.
1994: 한국과학기술원 공학석사.
2008: 단국대학교 공학박사.
1996 - 현재: 한라대학교 정보통신
공학부 교수

관심분야: 패턴인식, HCI, 디지털 신
호처리, 컴퓨터 기술 응용



양 룡

1972: 한국항공대학교 공학사.
1980: 동아대학교 공학석사.
1990: 단국대학교 공학박사
1979 - 현재: 인하공업전문대학 컴
퓨터정보과 교수

관심분야: 병렬처리컴퓨터, HCI, 디
지탈콘텐츠, 컴퓨터윤리,



이상범

1974 : 연세대학교 공학사
1978 : 서울대학교 공학석사
1986 : 연세대학교 공학박사
2000 : 미국 SanJose대학 컴퓨터
공학과 객원교수

2005 : 한국정보처리학회 회장
1979-현재 : 단국대학교 전자컴퓨터
공학과 교수

1997-현재 : 단국대학교 멀티미디어
산업기술연구소 소장

관심분야 : 컴퓨터구조, 패턴인식,
디지털 신호처리