

YCbCr 농도 대비를 이용한 입술특징 추출

김우성*, 민경원**, 고한석*
고려대학교 전자컴퓨터 공학과*
전자부품 연구원 디지털미디어 연구센터**

Lip Feature Extraction using Contrast of YCbCr

Woosung Kim*, Kyungwon Min**, Hanseok Ko*

Department of Electronics and Computer Engineering, Korea University*

Digital Media Research Center, Korea Electronics Technology Institute**

E-mail : *wskim@ispl.korea.ac.kr, **minkw@keti.re.kr, *hsko@korea.ac.kr

Abstract

Since audio speech recognition is affected by noise in real environment, visual speech recognition is used to support speech recognition. For the visual speech recognition, this paper suggests the extraction of lip-feature using two types of image segmentation and reduced ASM. Input images are transformed to YCbCr based images and lips are segmented using the contrast of Y/Cb/Cr between lip and face. Subsequently, lip-shape model trained by PCA is placed on segmented lip region and then lip features are extracted using ASM.

I. 서론

실생활에서 존재하는 잡음은 음성인식에 많은 제한을 준다. 따라서 잡음으로 인한 음성인식을 보조하기 위한 수단으로 영상을 이용한 인식연구가 활성화되었다 [1]. 영상으로 음성을 인식하기 위해서는 화자 입술의 움직임을 정확하게 추적하는 것이 중요하다.

본 논문에서는 입술 특징을 추출하는 방법을 제안하였다. 먼저 카메라에서 입력된 RGB 기반의 정면 얼굴 영상을 YCbCr 칼라 공간으로 변환 후, 얼굴에서 피부와 입술의 Y, Cb 및 Cr의 농도가 다른 점을 이용하여 얼굴에서 입술영역을 검출한다. 검출된 입술영역에서는 PCA (Principal Component Analysis)를 이용한 입술모델을 적용해서 ASM(Active Shape Model)의 수렴연산을 이용하여 입술의 특징을 추출한다[2]. 따라서 사람의 피부색에

* 본 연구는 산업 자원부 차세대신기술개발사업의 다중 생체기반 실감형 디바이스 개발과제의 지원을 받아 수행되었음.

무관하게 입술 검출이 가능하고, 초기 입술모델을 정확하게 위치시켜 ASM의 반복연산을 줄일 수 있다.

II. 본론

YCbCr 칼라 공간에서 입술은 Cr 과 Cb 성분이 얼굴의 피부색보다 우세하다. 그리고 입술은 Cr 성분값이 Cb 보다 크다. 또한 Y 성분은 입술 윗부분에서 높고, 좌우 양끝점 및 윗입술 아래 부분은 작다. 이러한 특성을 사용해서 Cr 과 Cb 성분을 이용하여 만든 영상과, Cr/Cb 색채성분과 더불어 Y 까지 이용한 영상을 융합하면 “그림 1” 에서와 같이 얼굴에서 입술영역을 검출할 수가 있다. “그림 1”의 “영상 A”는 색채 성분만을 이용하여 만든 것으로 얼굴 인식에서 사용하는 MouthMap 영상 생성 기술을 이용하였다[4].

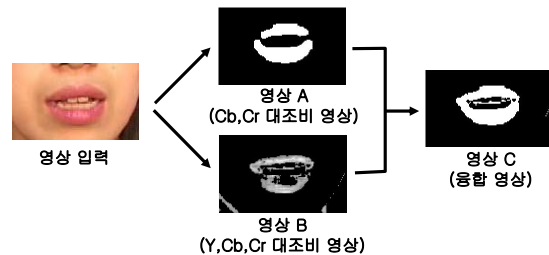


그림 1. Segmentation 과정

MouthMap은 정규화된 Cr2 과 Cr/Cb 성분을 적용하지만, 여기서는 정규화 없이 사용하여 Cr2 과 Cr/Cb의 대비를 크게 하였다. 이렇게 획득한 영상은 다시 n 값을 조절한 임계값을 사용하여 이진영상으로 변환한다. “영상 B”는 얼굴의 피부와 입술에서 각 색채성분과 Y 성분이 갖는 정보의 차이를 이용한 것이다. 입술을

제외한 얼굴영역에서는 Y 성분값이 색채 성분값보다 크다. 따라서 Cr 및 Cb 성분값에서 Y 성분값을 빼면 색채 성분값이 높은 입술만 부분적으로 남는다. 정규화 후 식 (1)와 같이 연산을 수행하면 입술을 제외한 부분이 제거된다.

$$\text{“영상 B”} = ((Cr - Y)' - (Cb - Y)')^2 \quad (1)$$

“영상 A”는 조명의 영향을 덜 받는 아랫입술과 윗입술의 중앙부분을 정확하게 표현한다. 반면에 Y 성분을 활용한 “영상 B”는 윗입술의 모서리 부분과 입술의 양끝점을 잘 표현한다. 따라서 각각의 특색 있는 두 영상을 융합하여 이진화하면 “영상 C”와 같이 입술모양을 검출할 수 있다.

검출된 입술영상에서 입술특징을 추출하기 위하여 ASM 을 사용한다. ASM 에 적용할 평균입술 모델은 PCA 를 이용하여 16 개의 특징점을 가지는 열린입술과 10 개의 특징점을 가지는 닫힌입술 모델을 만든다.

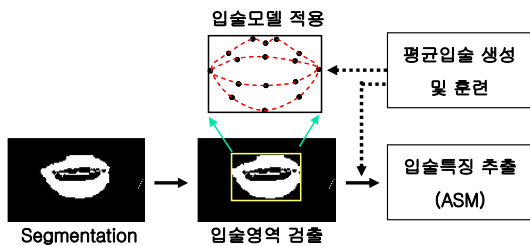


그림 2. Lip-Feature Extraction

일반적인 ASM 은 평균모델의 초기 위치를 입력영상의 중앙에 위치시킨 후 각 특징점이 최대 에너지값에 수렴할 때까지 반복적인 연산을 수행한다. 하지만 본 논문에서는 입술영역이 앞의 과정에서 이미 검출되었으므로, 평균입술 모델을 검출된 이진 입술 영상 위에 정렬시킨다. 따라서 입력영상의 중앙에서 탐색 및 연산을 수행하는 일반적인 ASM 방법에 비해 적은 반복 연산량으로도 원하는 효과를 얻을 수가 있다.

III. 실험 결과

본 논문에서 제안된 알고리즘의 성능을 판단하기 위하여 입술의 움직임을 “그림 3”에 표현된 특징 벡터의 크기(size)로 비교하였다.

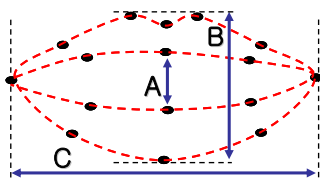


그림 3. 특징벡터(A, B, C)

“표 1”은 수작업으로 직접 구한 특징벡터(A,B,C)와 제안된 알고리즘에 의한 특징벡터 (A',B',C')의 관계를 상관관계수 “ρ” (범위: -1 ≤ ρ ≤ 1)로 표현하였다. 상관관계 “ρ”가 1에 가까우므로 제안된 알고리즘이 입술의 움직임의 표현에 적합함을 알 수가 있다.

표 1. 수동 및 알고리즘 상관관계

상관계수	ρ _{AA'}	ρ _{BB'}	ρ _{CC'}	평균
남자 31명	0.83	0.83	0.83	0.83
여자 34명	0.82	0.82	0.82	0.82
평균	0.82	0.82	0.82	0.82

“표 2”는 ASM 에서의 특징추출과정 중 최대 반복 연산횟수에 따른 MATLAB 알고리즘 처리시간이다. 최대 반복 연산 횟수의 변화에 관계 없이 처리시간의 차가 작은 이유는, 적은 반복 연산으로도 입술모델에 수렴이 가능함을 보여주는 것이다. 따라서 ASM 의 반복연산 횟수를 작게 제한하는 것이 가능하므로, 전체 연산 소요시간을 줄일 수가 있다.

표 2. ASM 반복 연산과 처리 시간

최대 반복 연산	2	5	10	100
65 단어 처리시간	135 sec	143 sec	144 sec	148 sec

IV. 결론

본 논문에서는 YCbCr 색상 표현 방식을 이용하였다. 얼굴에서 입술과 피부색이 가지는 색채와 휘도의 농도 차이를 이용하여, 서로 다른 특징의 이진영상을 융합함으로써 입술을 얼굴로부터 구별하였다. 또한 입술영역이 얼굴로부터 분리된 이진영상을 ASM 을 적용한 단계에 활용하므로 평균입술 모델의 초기위치 설정과 정렬이 용이하여 반복연산의 횟수를 줄일 수가 있다.

참고문헌

- [1] T. Chen, “Audiovisual Speech Processing”, IEEE Signal Processing Magazine, volume 18 Issue: 1, 9-21, 2001.
- [2] T.F.Cootes, C.J.Taylor, D.H.Cooper, J.Garham, “Active Shape Models; Their Training and Application”, Computer Vision and Image Understanding, 61(1) 38-59, 1995.
- [3] A. Caplier, “Lip Detection and Tracking”, IEEE Image Analysis and Processing, 8-13, 2001.
- [4] R. Hsu, M. A. Mottaleb, “Face Detection in Color Images”, IEEE Pattern Analysis and Machine Intelligence, volume 24, 696-706, May 2002.