

스테레오 입체음향을 위한 머리 움직임 추정

김현태[†], 이광의^{**}, 박장식^{***}

요 약

입체음향을 구현하는 방법은 두 가지가 있으며, 5.1 채널과 같은 서라운드 시스템(surround system)을 이용하는 방법과 2 채널의 바이노럴 시스템(binaural system)을 이용하는 방법이 있다. 바이노럴 시스템은 사람이 두 귀를 이용하여 음상을 정위하는 원리를 이용하는 방법이다. 일반적으로 라우더 스피커 시스템에서 각 채널사이에 발생하는 크로스토크(crosstalk)는 본래의 입체음향을 재생하기 위해서는 제거되어야만 한다. 이 문제를 해결하기 위해서는 머리의 움직임을 추정할 필요가 있다. 본 논문에서는 청취자의 머리 움직임을 제대로 추정하는 새로운 알고리즘을 제안한다. 제안하는 알고리즘은 얼굴과 눈의 영역 검출을 기본으로 한다. 얼굴 검출은 이미지의 밝기 값을 이용하고 눈 검출은 수학적 형태학(mathematical morphology)을 이용한다. 청취자의 머리가 움직일 때 얼굴 영역과 눈 사이의 경계선의 길이가 변한다. 이 정보를 이용하여 머리 움직임을 추정한다. 컴퓨터 시뮬레이션을 통해 제안하는 알고리즘이 +10 오차 범위 내에서 머리의 움직임을 효율적으로 추정하는 것을 확인하였다.

A Tracking of Head Movement for Stereophonic 3-D Sound

Hyun-Tae Kim[†], Kwang-Eui Lee^{**}, Jang-Sik Park^{***}

ABSTRACT

There are two methods in 3-D sound reproduction: a surround system, like 5.1 channel method and a binaural system using 2-channel method. The binaural system utilizes the sound localization principle of a human using two ears. Generally, a crosstalk between each channel of 2-channel loudspeaker system should be canceled to produce a natural 3-D sound. To solve this problem, it is necessary to trace a head movement. In this paper, we propose a new algorithm to correctly trace the head movement of a listener. The proposed algorithm is based on the detection of face and eye. The face detection uses the intensity of an image and the position of eyes is detected by a mathematical morphology. When the head of the listener moves, length of borderline between face area and eyes may change. We use this information to the tracking of head movement. A computer simulation results show that head movement is effectively estimated within +10 margin of error using the proposed algorithm.

Key words: Face Detection(얼굴 검출), Eye Detection(눈 검출), Mathematical Morphology(수학적 형태학), Head Modeling(머리 모델링), Azimuth Estimation(방위각 추정)

1. 서 론

3 차원 입체음향은 헤드폰이나 라우드 스피커(loud-speaker)를 이용하여 3 차원 공간 상의 임의의

위치에 가상의 음원이 위치하도록 하는 것이다. 가상 공간 속에 있는 청취자는 스피커로부터 소리를 듣고 물리적으로 소리가 발생한 곳이 아닌 임의의 지점에서 소리가 발생한 것으로 느끼게 된다. 이러한 3 차원

※ 교신저자(Corresponding Author) : 김현태, 주소 : 부산광역시 진구 가야동 산24(614-714), 전화 : 051)890-1992, FAX : 051)890-1987, E-mail : htaekim@deu.ac.kr

접수일 : 2005년 3월 4일, 완료일 : 2005년 6월 16일

[†] 정회원, 동의대학교 멀티미디어공학과 조교수

^{**} 종신회원, 동의대학교 멀티미디어공학과 조교수
(E-mail : kelee@deu.ac.kr)

^{***} 정회원, 동의과학대학 디지털정보전과 부교수
(E-mail : jsipark@dit.ac.kr)

입체음향 기술은 가상 현실 시스템에서 중요한 부분을 차지한다.

3차원 입체음향을 구현하는 방법으로는 서라운드 시스템(surround system)을 이용하는 방법과 바이노럴 시스템(binaural system)을 이용하는 방법이 있다. 서라운드 시스템을 이용하는 방법에는 5개의 스피커와 저음을 재생하는 우퍼 스피커를 사용하는 5.1 채널이 널리 사용되고 있다. 5.1 채널로 입체음향 재생을 위해서는 최소 6개의 스피커 시스템이 필요하고 청취자가 스피커에 둘러싸여 중앙에 위치해야만 3차원 효과를 느낄 수 있다.

2개의 채널을 사용하는 바이노럴 사운드 시스템은 일반적으로 헤드폰이나 2개의 라우드 스피커를 사용하여 구현된다. 바이노럴 사운드 시스템에서는 음원과 사람의 두 귀 사이의 음향학적 전달함수인 머리 전달함수(head related transfer function, HRTF)가 매우 중요한 요소로 작용한다. 머리 전달함수를 정확히 부여하면 입체음향 구현이 정확히 이루어진다[1]. 그러나 머리 전달함수가 정확하더라도 사람이 머리를 움직이게 되면 스피커에서 머리까지 음향 전달 경로의 변화로 인하여 입체음향 시스템의 성능이 저하된다.

그리고 왼쪽 스피커에서 출력된 소리가 왼쪽 귀로만 전달되는 것이 아니고 오른쪽 귀로도 전달이 된다. 이러한 현상을 크로스토크(crosstalk)라고 하는데, 입체음향을 재생할 때 음상 정위 효과가 감소되고 명료성도 떨어진다[2]. 크로스토크를 상쇄시키기 위해서 스피커와 사용자 사이에 크로스토크 제거를 설계하여 성능을 향상시켜야 한다.

2개의 라우드 스피커를 이용한 바이노럴 사운드 시스템으로 입체음향을 구현하고자 할 때 청취자의 머리 움직임에 따라 변하는 머리 전달함수와 라우드 스피커와 두 귀 사이의 음향학적 경로변화를 실시간으로 반영해야만 입체음향 구현시의 성능 저하를 방지할 수 있다[3,4].

본 논문에서는 카메라에 입력된 영상으로부터 청취자의 얼굴 영역과 눈의 위치를 검출하고 머리 움직임에 따른 두 눈과 좌우 얼굴 영역 경계면 사이의 길이 변화를 이용하여 시선의 방위각을 추정함으로써 크로스토크 제거와 머리 전달함수 변화에 대한 정보로 활용한다.

입력된 영상으로부터 머리 위치 검출을 위해 밝기 정보를 바탕으로 얼굴의 영역을 검출한다[5]. 검출된

얼굴 영역에서 수학적 형태학(mathematical morphology) 기반의 연산들로부터 눈의 위치를 검출한다. 얼굴 영역 정보와 눈의 위치 정보를 이용하여 시선의 방위각을 추정한다. 머리의 움직임을 추정하기 위해서는 시선의 방위각(azimuth)과 고도각(elevation)을 구할 필요가 있으나 머리의 좌우 변화가 머리의 상하 변화보다 입체 음향 시스템의 성능 저하에 더 많은 영향을 주기 때문에 본 논문에서는 방위각만을 추정한다.

방위각 추정은 사람의 경우, 머리가 좌우로 움직이면 정면에 있을 때보다 두 눈과 좌우 얼굴 경계면과의 거리비가 현저히 차이가 난다는 사실에 바탕을 둔 방법을 제안한다.

본 논문은 2장에서 스테레오 입체음향 시스템의 기본개념과 스테레오 입체음향에서의 머리 움직임에 관해서 살펴보고 3장에서 제안하는 머리 위치 추정방법을 설명하고 4장에서 컴퓨터 시뮬레이션과 피실험자를 대상으로 실험을 통해 제안하는 머리 움직임 추정 방법으로 시선의 방향을 추정할 수 있음을 보이며 5장에서 본 논문의 결과를 정리하였다.

2. 스테레오 입체음향 시스템

바이노럴 사운드는 사람이 두 귀로 소리의 방향을 인지할 수 있다는 점을 활용하여 양쪽 귀에 들리는 소리를 두 개의 채널을 사용하여 스피커나 헤드폰으로 재현하는 기술이다. 머리 전달함수는 가상음원과 사람의 귀 사이의 음향학적 전달함수로써 입체음향 재생에 중요한 역할을 한다. 3차원 공간 상에서 방향을 인지할 수 있는 것이 바로 음원의 위치에 대한 정보를 담고 있는 머리 전달함수 때문이다.

그림 1은 스테레오 입체음향 시스템을 구현하는 과정을 나타낸다[5]. 그림 1의 (a)에서 보는 것과 같이 실제음원의 위치를 바꿔가면서 실제로 각 위치에 대한 머리 전달함수를 미리 측정하여 데이터베이스를 구축한다[3]. 그림 1의 (b)에서 보는 것과 같이 헤드셋을 이용하여 입체음향을 구현하고자 할 때는 구현하고자하는 가상음원의 위치에 해당하는 머리 전달함수를 이용하여 헤드폰에 소리를 재생시켜주면 입체음향이 재현된다. 라우드 스피커를 이용하는 입체음향 시스템은 그림 1의 (c)에서 보는 것과 같이 헤드셋을 이용하는 경우와 달리 스피커와 두 귀 사이

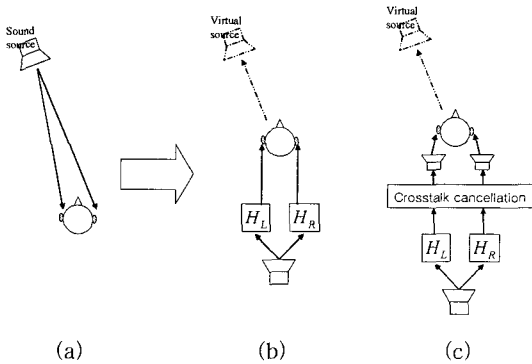


그림 1. 스테레오 입체음향 시스템: (a) 머리 전달 함수의 측정, (b) 헤드셋을 이용한 입체음향 시스템, (c) 스피커를 이용한 입체음향 시스템.

에 음향학적 공간이 존재하게 된다. 즉 왼쪽 스피커에서 출력된 소리가 왼쪽 귀로만 전달되는 것이 아니고 오른쪽 귀로도 전달이 되어 좌우 스피커의 간섭현상이 일어나 소리의 정확한 전달을 방해한다. 이러한 현상을 크로스토크라 하는데, 입체음향 재생 시 음상 정위 효과가 줄어들고 명료성도 떨어지는 요인이 된다. 크로스토크를 제거하기 위하여 스피커와 두 귀 사이의 음향학적 전달함수의 역변환 필터를 구하여 [6] 크로스토크 제거기를 설계하고, 실시간으로 원하는 위치의 가상 음원에 대응하는 머리 전달함수로 필터링하고 크로스토크 제거기를 통과시킴으로써 입체음향을 재생할 수 있다. 그러나 사용자의 머리움직임에 의하여 음향학적 전달함수가 변하게 되면 크로스토크 제거기가 제대로 작동하지 않아 좌/우 두 신호를 정확히 분리해 내지 못해 입체음향 시스템의 성능이 저하된다. 그림 2에 사용자의 머리 움직임에 따른 머리 전달함수의 변화 예를 나타내었다.

본 논문에서는 카메라로 시선이 향하는 방향의 방위각을 추정하여 머리의 위치에 대응하는 크로스토크

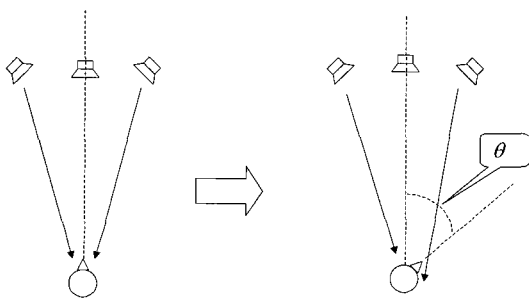


그림 2. 머리전달 함수(HRTF)와 사용자의 머리 움직임

크를 제거하여 효율적으로 3차원 입체음향을 재생하는 방법을 제안한다.

3. 제안하는 머리 위치 추정 방법

3.1 머리위치 추정방법

머리 위치 추정은 신체의 일부분들이 어느 위치에 존재하고 어느 방향으로 이동할지를 결정할 수 있는 사용자 인터페이스를 조성하기 위한 중요하고 근본적인 단계이다[7]. 1990년대 초반부터 전 세계의 컴퓨터 비전 전문가들이 얼굴의 크기와 위치를 자동으로 찾기 위한 많은 연구들이 진행되고 있다[8-10]. 사람의 성별, 나이와 얼굴의 모양, 표정 등의 얼굴 자체의 내부적 요인과 다른 물체에 의하여 일부분이 가려지거나 머리카락, 수염, 안경, 화장, 카메라의 해상도, 노이즈, 초점, 조명에 의한 그림자, 복잡한 배경, 칼라의 왜곡, 포즈 등에 의한 외부적인 요인들 때문에 정확하게 추정하기 어렵다[11-13]. 영상처리, 컴퓨터 비전, 패턴 인식 등의 분야에서 얼굴을 자동으로 검출하는 연구가 지속적으로 진행되어 최근에는 처리 시간이 빠르면서 정확한 얼굴 검출을 위한 실시간 방법들이 연구 되고 있다[7,14,16].

3.2 제안하는 머리위치 추정방법

본 논문에서 제안하는 머리의 위치 검출 절차는 그림 3과 같다. 카메라나 캠코더로부터 입력되어진 사진이나 영상들의 밝기정보를 이용하여 사람의 얼굴 영역을 검출하고 검출된 영역을 바탕으로 수학적 형태학의 연산을 조합하여 눈의 위치를 검출한다. 검출된 얼굴의 영역과 눈의 위치를 분석하여 최종적으로 시선이 향하고 있는 방향을 결정한다.

3.2.1 얼굴 영역 검출 알고리즘

그림 4는 입력영상으로부터의 얼굴 영역을 검출하기 위한 알고리즘을 나타내고 있다. 본 논문에선 비교적 계산량이 적은 단색 배경이미지 또는 고정 배경이미지를 이용하는 통제된 배경이미지에서의 얼굴 검출 알고리즘을 사용하며, 주어진 영상의 밝기 정보를 바탕으로 얼굴의 세로 영역 임계값을 추정한 뒤 얼굴의 사전 정보를 이용하여 가로 영역을 임계값을 결정한다. 구한 임계값을 바탕으로 얼굴 영역을 검출한다.

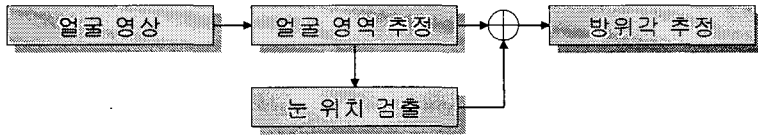


그림 3. 제안하는 머리 위치 검출 과정

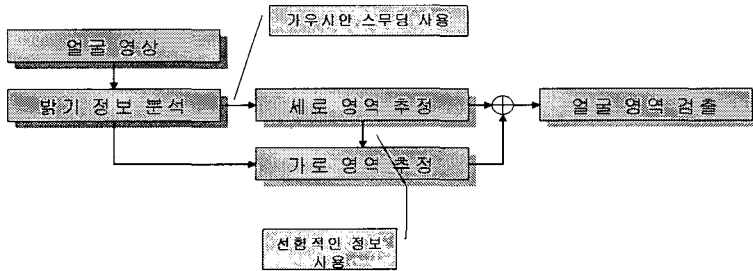


그림 4. 제안하는 얼굴 영역 검출 알고리즘

주어진 이미지의 세로 영역을 추정하기 위해서 우선 얼굴 영역의 세로 방향 평균 밝기 값의 변화폭을 줄이기 위해서 평활(smoothing)화를 시킨다. 그리고 평활화 된 값의 위아래를 선형적인 정보를 이용 각각 10%씩 자른 값을 사용한다. 그림 5에서 보듯이 보통 정지 영상의 경우 사람의 얼굴은 중앙에 위치하게 되므로 원이미지를 사용하는 경우에 비해 계산량이 줄어드는 장점이 있다.

이렇게 10%씩 자른 값을 10 픽셀 단위에서 최대값을 구하고 그 중에서 2개의 최대 값을 순서대로 구한다. 그리고 위에서부터 첫 번째 최대값 사이의 최소 값을 세로 영역 윗부분으로 설정하고 2번째 최대값에서부터 아랫부분까지의 최소값을 얼굴의 아랫부분으로 설정하여 얼굴의 세로 영역을 구할 수 있다.

구한 세로 영역을 바탕으로 이미 알고 있는 얼굴의 세로 대 가로 비로써 얼굴의 가로 영역을 추정한다. 그림 6은 얼굴의 평균 가로 세로의 비를 나타낸 것이다. 얼굴의 가로와 세로의 비는 사람마다 다르겠지만 보통 1 : 1.3이다. 이 비를 이용하여 우리가 구하고자 하는 얼굴의 가로 영역을 구한다.

먼저 가로 방향의 평균 밝기값의 중심값의 좌표를 기준으로 세로 영역 얼굴 길이의 1/2에 대해 1/1.3 배 곱한 길이를 가로 영역의 왼쪽과 오른쪽 끝점 좌표로 설정한다. 즉

$$F_{left} = V_c - |d_2 - d_1| * \frac{1}{1.3} * 0.5 \quad (1)$$

$$F_{right} = V_c + |d_2 - d_1| * \frac{1}{1.3} * 0.5 \quad (2)$$

(1)과 (2)에서 V_0 는 가로 방향 평균 밝기값의 중심

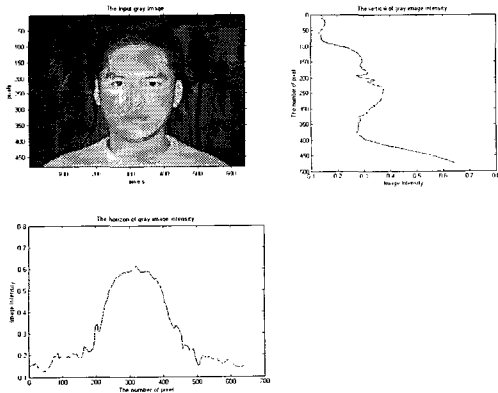


그림 5. 가로 및 세로 방향의 평균 밝기 값

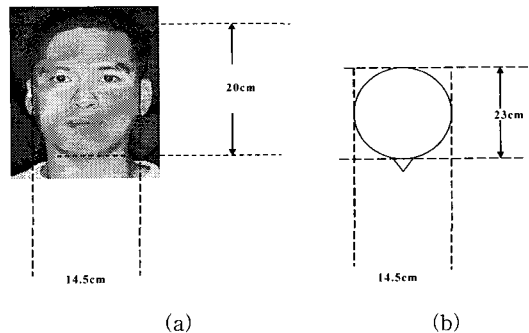


그림 6. 얼굴의 평균 가로 세로 비: (a) 정면. (b) 위.

값 좌표이고 d_1 과 d_2 는 얼굴의 세로영역에서 임계값 1과 임계값 2를, F_{left} 는 얼굴 가로 영역의 왼쪽 끝부분 좌표를 나타내고, F_{right} 은 얼굴 가로 영역의 오른쪽 끝부분 좌표를 나타낸다.

3.2.2 눈 위치 검출 알고리즘

수학적 형태학은 영상분석, 영상분리 등의 다양한 분야에 활용되고 있는 영상처리 방법으로써 영상의 형태를 기반으로 팽창(dilation), 침식(erosion), complementary 연산을 기반으로 다양한 패턴인식 및 영상처리를 할 수 있다[10,17].

입력 이미지에 대하여 처리하는 픽셀의 주변영역의 픽셀들과 비교하여 출력값을 결정한다. 주변영역의 형태와 크기는 입력 이미지의 특성에 맞게 임의로 결정할 수 있다.

수학적 형태학의 기본 연산은 팽창과 침식이며, 팽창은 이미지에서 객체의 경계를 확장시키고 침식은 객체의 경계를 수축시킨다. 팽창과 침식의 조합으로 다양한 영상처리를 할 수 있다. 본 논문에서는 그림 7에 나타난 절차에 따라 배경으로부터 얼굴을 분리한다. 입력된 계조 영상(gray image)에 대하여 경계를 검출한다. 경계 검출을 위하여 Sobel 경계 필터를 이용하여 경계를 검출하였다. 경계 검출된 이미지에 대하여 가로, 세로 방향으로 선 형태소(structuring)로 팽창을 적용함으로써 두 줄로 된 경계선의 내부를 채워서 하나의 굵은 선으로 변환한다.

그리고 객체를 분리하기 위하여 팽창된 선의 한

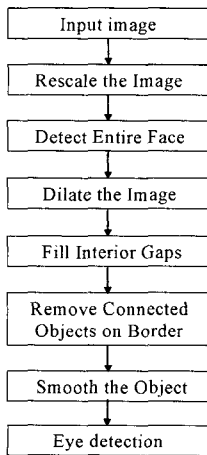


그림 7. 수학적 형태학을 이용한 눈의 위치 검출 알고리즘

줄의 가는 선, 즉 골격선으로 변환한다. (3) 및 (4)와 같이, Lantuejou가 제안한 침식과 열림(opening) 연산을 활용하여 골격선을 구하는 방법을 적용하였다.

$$S(A) = \prod_{k=0}^K S_k(A) \tag{3}$$

여기서, $S(A)$ 는 A 에 대한 골격선 처리 결과를 나타낸다.

$$S_k(A) = \prod_{k=0}^K \{(A \otimes kB) - [(A \otimes kB) \circ B]\} \tag{4}$$

A 는 팽창된 이미지를 나타내며 B 는 형태소를 의미한다. \circ 는 열림 연산을 의미한다. $A \otimes kB$ 는 아래와 같이 연속 침식으로 나타낸다.

$$(A \otimes kB) = ((\dots (A \otimes B) \otimes B) \otimes \dots) \otimes B \tag{5}$$

열림 연산은 일반적으로 영상의 윤곽을 부드럽게 하는 연산으로 (6)과 같이 표현된다.

$$A \circ B = (A \otimes B) \oplus B \tag{6}$$

골격선 처리가 된 이미지에 대하여 골격선이 폐곡선으로 되어 있는 영역을 채운다. 영역을 채우는 알고리즘은 (7)와 같이 팽창, complementation, 그리고 intersection 연산으로 이루어진다.

$$X_k = (X_{k-1} \oplus B) \cap A^c, \quad k=1,2,3 \tag{7}$$

그리고 일반적으로 분리시키는 객체는 경계와 연결되어 있지 않기 때문에 경계에 연결된 객체는 지운다. 경계와 연결된 영역을 지운 다음 erosion 연산을 통하여 작은 독립된 영역을 지우고 경계선을 추출한다. 경계선의 추출은 (8)과 같이 이진영상에 대하여 침식을 먼저 실행하고 원 이진영상과 침식된 영상간의 차를 구함으로써 경계를 얻을 수 있다.

$$\beta(A) = A - (A \otimes B) \tag{8}$$

수학적 형태학은 형태소의 선경과 크기에 따라서 영상처리에 영향을 많이 준다. 다양한 영상에 알고리즘을 적용하여 크기와 형태를 결정하는 것이 필요하다.

그림 8은 수학적 형태학의 연산을 이용하여 눈의 위치검출 예를 나타낸다.

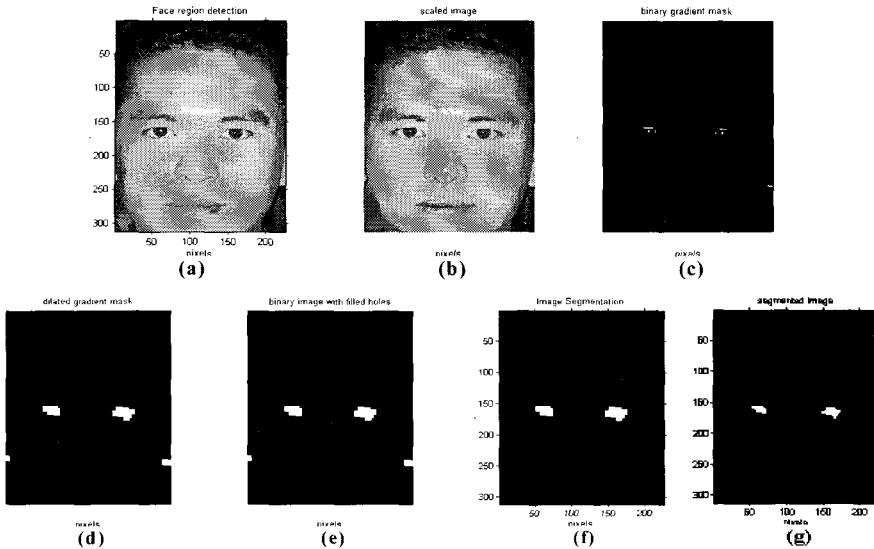


그림 8. 눈 위치 검출 예: (a) 추출된 얼굴영역, (b) 스케일 조정, (c) 이진 기울기 마스크, (d) 팽창, (e) 내부 채우기, (f) 영상 분할, (g) 열림과 침식 후 눈 검출.

그림 8의 (b)에서는 전처리 과정을 거치는데, 전처리 과정에서는 밝기 정도를 조절하고 입력된 영상을 0과 1 사이의 범위로 맵핑한다.

얼굴 영역이 검출된 이미지 내에 있는 여러 사물들(눈, 코, 입 등) 중에서 사물(눈)을 추출하려는 것을 분할(segmentation)이라 한다. 분할하고자 하는 사물은 주변 환경에 비해 밝기 정도가 상당히 다르므로 그림 8의 (c)는 밝기 변화를 이용하여 사물(눈)을 찾을 수 있다.

밝기 변화는 영상의 미분값을 측정하면 간단히 구할 수 있지만, 본 논문에선 이미지의 미분값을 측정하는 방법으로 여러 방법 중 기울기(gradient) 연산 중 하나인 Sobel 필터를 사용한다.

이진 기울기 마스크(binary gradient mask)는 주어진 영상에서 밝기 정도가 높은 선들의 집합이다. 이러한 선들이 사물의 윤곽만을 나타내는 것이 아니다. 주어진 이미지와 비교했을 때 기울기 마스크 내에 있는 사물의 경계선들 사이에 틈(gap)이 존재함을 알 수 있다. 그림 8의 (d)는 이러한 틈들을 없애기 위해서 수직 및 수평 구성요소를 사용하여 이진 기울기 마스크를 팽창시켜 경계선 사이를 채운 결과이다.

팽창된 기울기 마스크의 윤곽선이 그 전 단계보단 많은 향상을 보이나, 아직까지도 폐곡선 내에 많은 구멍(hole)들이 존재한다. 그림 8의 (e)는 폐곡선으로 구성된 내부를 채운 결과이다.

눈의 위치를 정확하게 검출하기 위해 눈을 제외한 사물들을 제거해야 한다. 그림 8의 (f)에서는 눈의 경계 면과 연결되어있는 다른 사물들을 검출하고자 하는 목표가 아니라고 판단하고 경계에 연결되어 있는 다른 사물들을 제거한다.

그림 8의 (g)는 열림과 침식으로 경계면을 부드럽게 하여 최종적으로 검출된 눈의 위치이다.

3.2.3 제안하는 방위각 추정

얼굴 영역에서 가로 영역의 경계값들과 좌, 우의 눈의 위치의 차에 비를 이용하여 사람의 시선방향 즉, 머리 위치를 검출한다.

그림 9는 사람 머리의 단면도이며 (a)는 사람의 얼굴이 정면을 향하는 것이고 (b)는 사람의 얼굴이 중심에서 θ° 만큼의 각도로 머리를 왼쪽으로 회전한 경우이다.

그림 10은 방위각 추정을 위해서 제안하는 모델이다. 머리 모양을 원으로 가정하고 원의 반지름은 그림 9의 (b)에서 장축의 길이라고 가정한다. 앞에서 언급한 것과 같이 얼굴 가로 영역의 왼쪽 좌표와 왼쪽 눈과의 거리(x_{left})와 얼굴 가로 영역의 오른쪽 경계면과 오른쪽 눈과의 거리(x_{right})와의 비를 이용하여 방위각을 추정한다.

거리의 비와 방위각과의 관계를 알아보기 위해 사람이 그림 11의 경우처럼 얼굴을 왼쪽으로 돌렸다고

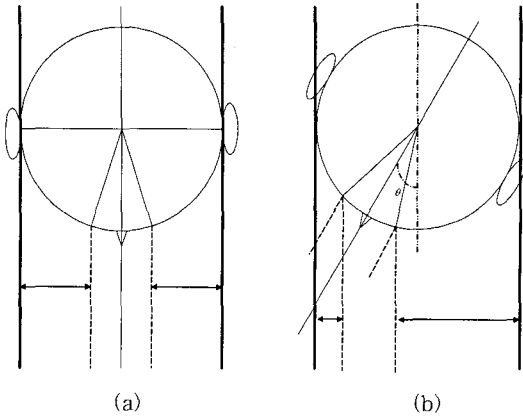


그림 9. 시선의 방향: (a) 정면을 향하고 있는 경우, (b) 중심에서 θ° 만큼 회전한 경우.

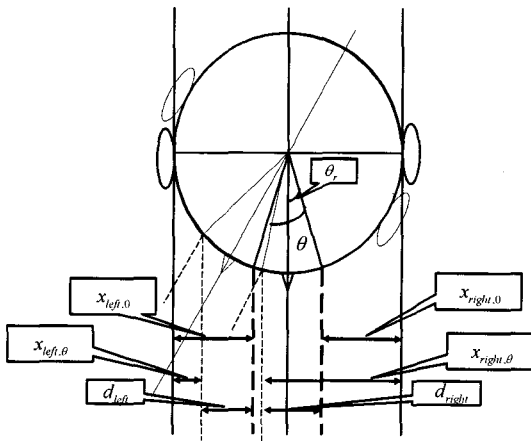


그림 10. 방위각 추정을 위해 제안하는 머리 모델링

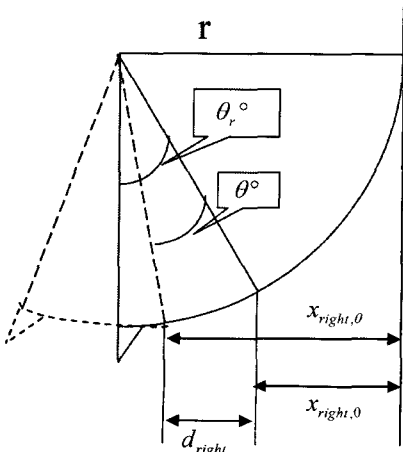


그림 11. 머리의 오른쪽 부분 모델링

가정하면 사람 얼굴의 오른쪽 부분에서의 거리의 변화는

$$x_{right,\theta} = x_{right,0} + d_{right} \tag{9}$$

이다. $x_{right,\theta}$ 는 사람의 얼굴이 θ° 만큼 회전한 경우 오른쪽 눈과 얼굴의 가로 영역의 경계면과의 거리를 나타내고, $x_{right,0}$ 는 사람의 얼굴이 정면을 향하고 있을 때 오른쪽 눈과 얼굴의 가로 영역의 경계면과의 거리를 나타낸다.

$$x_{right,0} = r - r\sin\theta_r \tag{10}$$

(9)에서 $x_{right,0}$ 은 (10)로 나타 낼 수 있는데, 그림 12에서 보면 r 은 사람의 얼굴을 원이라고 가정하여 구한 얼굴의 반지름이고 θ_r° 은 눈과 코사이의 각을 나타낸다.

또한 $x_{right,\theta}$ 와 $x_{right,0}$ 의 변위 차이인 d_{right} 는

$$d_{right} = r\sin\theta_r - r\sin(\theta_r - \theta) \tag{11}$$

이다.

(10)와 (11)을 (9)에 넣으면

$$x_{right,\theta} = r - r\sin(\theta_r - \theta) \tag{12}$$

얻을 수 있다.

그림 12는 사람의 왼쪽 부분을 나타낸 것으로 거리의 변화는

$$x_{left,\theta} = x_{left,0} - d_{left} \tag{13}$$

이고, 여기서 $x_{left,\theta}$ 는 사람의 얼굴이 θ° 만큼 회전한 경우 왼쪽 눈과 얼굴의 가로 영역의 경계면과의 거리를 나타내고, $x_{left,0}$ 는 사람의 얼굴이 정면을 향하고 있을 때의 왼쪽 눈과 얼굴의 가로 영역의 경계면과의 거리를 나타낸다.

$$x_{left,0} = r - r\sin\theta_r \tag{14}$$

(13)에서의 $x_{left,0}$ 은 (14)로 나타낼 수 있는데, 오른쪽과 같이 그림 12에서 보면 r 은 사람의 얼굴을 원이라고 가정하여 구한 얼굴의 반지름이고 θ_r° 은 눈과 코사이의 각을 나타낸다.

또한 $x_{left,\theta}$ 와 $x_{left,0}$ 의 변위 차이인 d_{left} 는

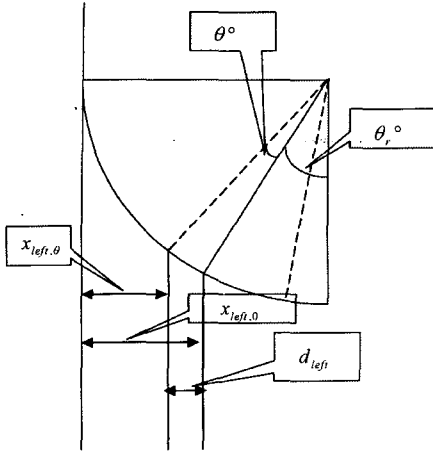


그림 12. 머리의 왼쪽 부분 모델링

$$d_{left} = r \sin(\theta_r + \theta) - r \sin \theta_r \tag{15}$$

이다.

(15)와 (14)을 (13)에 넣으면

$$x_{left, \theta} = r - r \sin(\theta_r + \theta) \tag{16}$$

얻을 수 있다. (16)과 (12)를 이용하여 x_{left} 와 x_{right} 의 비를 구하면

$$1 - \sin(\theta_r + \theta) : 1 - \sin(\theta_r - \theta) \tag{17}$$

로 나타낼 수 있다.

그리고, 사용자가 오른쪽으로 θ 만큼 머리를 회전하게 되면 머리 회전과 반대로 가상음원이 $-\theta$ 에 있다고 감지하게 된다. 즉 머리를 회전하게 되면 가상음원은 사용자의 회전 값 θ 와 크기는 같고 위상이 반대인 $-\theta$ 의 위치에 존재하게 된다. $-\theta$ 값의 머리 전달함수로 보상하면 머리 위치 변화에 의한 성능 저하를 보상할 수 있다. 따라서 x_{left} 와 x_{right} 의 차를 이용하여 좌우를 구한다. $x_{left} - x_{right} > 0$ 일 때는 얼굴이 왼쪽으로 회전한 것이므로 $+\theta$ 값의 머리 전달함수를 출력하고 $x_{left} - x_{right} < 0$ 일 때는 얼굴이 오른쪽으로 회전한 것이므로 $-\theta$ 값의 머리 전달함수를 출력한다. 본 논문에서는 정면을 중심으로 오른쪽 방향을 +로, 왼쪽 방향을 -로 둔다.

4. 실험 결과 및 검토

제안하는 방위각 추정방법의 성능을 평가하기 위

하여 시뮬레이션과 실제 환경에서 방위각을 추정하였다. 방위각 추정 실험을 위하여 얼굴 방향이 0° , 10° , 20° , 30° , 45° 만큼 회전하였을 때의 정지 영상을 사용하였다. 또한 실제 환경에서의 적용을 알아보기 위해서, 정면의 얼굴 영상과 30° 와 -30° 만큼 회전하였을 때의 얼굴 영상을 사용하여, 16 kHz로 샘플링 (sampling) 된 여자의 음성신호를 적용하였다.

각각의 실험에는 제한된 배경인 단색 배경에서의 그레이 스케일 (gray scale) 이미지를 사용하였다. 또한, 카메라와 사람 간의 거리는 1.4 m이고 640×480 의 이미지를 사용하였다. 그림 13 및 그림 14에 방위

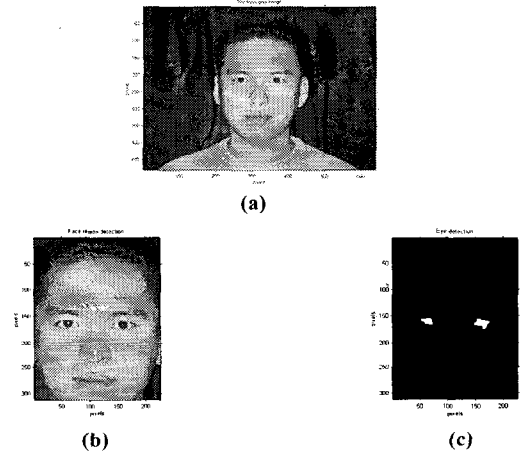


그림 13. 0° 회전 했을 때의 검출 결과: (a) 원 이미지, (b) 얼굴 검출 이미지, (c) 눈 검출 이미지.

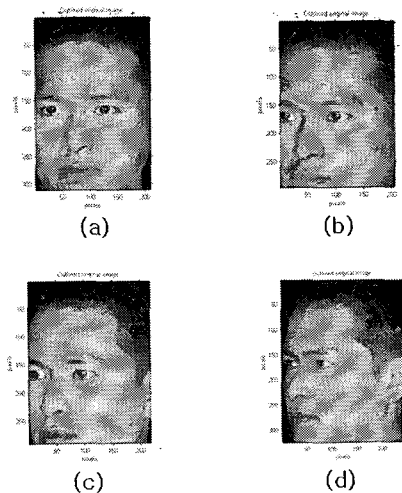


그림 14. $10^\circ \sim 45^\circ$ 회전하였을 때 검출된 눈의 위치: (a) 10° , (b) 20° , (c) 30° , (d) 45° .

각에 따른 눈의 위치를 검출한 이미지를 나타내었다.

이 결과를 바탕으로 실제적으로 구한 얼굴 가로 영역의 왼쪽 좌표와 왼쪽 눈과의 거리와 얼굴 가로 영역의 오른쪽 경계면과 오른쪽 눈과의 거리 비와 제안하는 방위각 추정법을 이용하여 구한 거리 비를 비교하면 아래의 표 1, 그림 15와 같다. 눈과 코 사이의 각은 12.7°로 설정하였다. 0°와 45°에서만 비슷한 값을 가지고 다른 실제값은 추정된 방위값과 비교하였을 때, 10°정도의 오차가 발생함을 알 수 있다. 오차의 발생원인은 머리 모델링에 의한 방위각 추정 단계에서 눈의 위치를 정확하게 추정하지 못하였기 때문에 발생한 것으로 판단된다.

그림 16과 같이 피실험자들을 0°, 30°, -30°방향으로 향하게 하고 카메라로 각 피실험자에 대하여 방위각을 추정하고 스피커를 통하여 가상 음원이 30°와 -30°에 위치하도록 머리 전달함수를 통과시키고 크로스토크 제거 처리한 후 소리를 재생하여 음원의 위치를 판별하는 실험을 10명에 대하여 실시하였다.

그림 17은 음상의 위치를 추정한 결과로써 30°, -30°에 대하여 다소 오차는 있으나, 일반적인 머리 전달함수에 의한 음상 정위와 차이가 없음을 알 수 있다.

표 1. 실제 값과 방위각 추정법을 사용한 결과

	0°	10°	20°	30°	45°
Real data	1.11	2.33	4.04	6.76	8.88
Estimation data	1	1.55	2.45	4.03	8.65

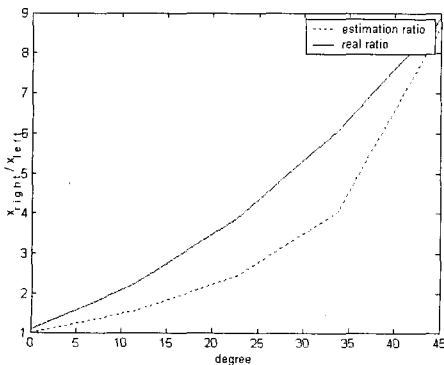


그림 15. 머리의 실제 회전각과 방위각 추정 결과

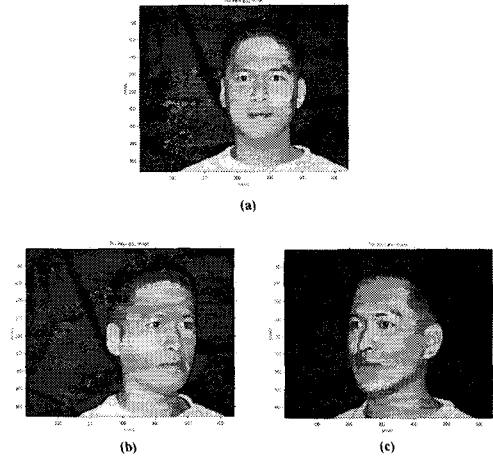


그림 16. 실제 환경에서의 이미지: (a) 머리가 0°회전 했을 때의 이미지, (b) 머리가 -30°회전 했을 때의 이미지, (c) 머리가 30°회전 했을 때의 이미지.

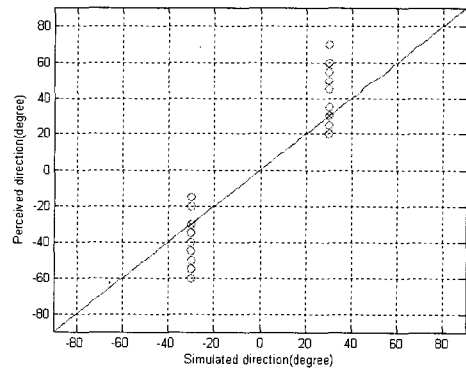


그림 17. 음상 정위에 대한 실험 결과

5. 결론

머리 전달함수를 이용한 바이노럴 방식의 입체음향을 재생하는데 있어 실질적인 입체음향 효과를 얻기 위해 크로스토크는 반드시 제거되어야 한다. 특히 청취자의 머리가 움직이면 머리 방위각이 변하고 머리 전달함수가 달라져 크로스토크를 정확히 제거할 수 없다. 본 논문에서는 2 채널 입체음향 재생에 필요한 크로스토크 제거를 위하여 청취자의 시선 방향을 실시간으로 추정하는 방법을 제안한다. 제안하는 방법은 카메라로 입력된 얼굴의 밝기 정보를 이용하여 얼굴 영역을 검출하고 수학적형태학의 연산들을 이용하여 눈의 위치를 검출하고 얼굴 영역에서의 눈 위치를 이용하여 시선의 방위각을 추정한다.

시뮬레이션 결과 제안하는 방법으로 방위각을 추정하는데 다소 오차가 발생하였지만 일반적인 음상 정위의 오차 범위를 벗어나지는 않아 방위각 추정에 효율적인 것을 확인하였다. 보다 정확한 시선의 방향을 추정하기 위해서는 얼굴 영역과 눈의 위치 추정을 보다 정확하게 하는 것이 필요하다.

참 고 문 헌

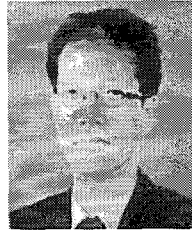
- [1] Yasuyo Yasuda and Wai C. Chu, "All-Pole Modeling of Head-Related Transfer Functions: Objective and Subjective Evaluation Results," *Proc. Int. Congress on Acoustics, ICA2004*, Vol. V, pp. 3745-3748, 2004.
- [2] Kraemer, A., "Two speakers are better than 5.1 [surround sound]," *IEEE Spectrum*, Vol. 38, Issue 5, May 2001.
- [3] B. Gardner and K. Martin, "HRTF measurements of a KERMA dummy-head microphone," *MIT Media Lab. Perceptual Computing-Technical Reports 280*, pp. 1-7, 1994.
- [4] S. S. Oh and K. M. Kim, "A robust cross-talk canceller using 2 loud speakers." *Proc. Int. Conf. Acoust. Speech Signal Proc.*, Vol. 5, pp. 2689-2692, Salt Lake City, Utah, May 2001.
- [5] V. R. Algazi, R. O. Duda, and D. M. Thompson, "Motion-Tracked Binaural Sound," *J. Aud. Eng. Soc.*, Vol. 52, No. 11, pp. 1142-1156, Nov. 2004.
- [6] J. B. Melick, V. R. Algazi, R. O. Duda, and Thompson, D. M., "Customization for personalized rendering of motion-tracked binaural sound," *Paper 6225, 117th Convention of the Audio Engineering Society*, San Francisco, CA, Oct. 2004.
- [7] Liang Tao and H. K. Kwan, "Automatic localization of human eyes in complex background," *Circuits and Systems, 2002. ISCAS 2002. IEEE International Symposium on*, pp. 669-672. Vol. V. 2002.
- [8] Jae-Hyub Park, Kyu-Sik Park, Syng-Yup Ohn, and Chun-Kuk Kim, "Face Detection for Intelligent Video Conference System," *KIPS, 2001*, Vol 8-B., 2001.
- [9] Shaogang Gong, Stephen J McKenna, and Alexandra Psarrou, *Dynamic Vision*, Imperial College press, 2000.
- [10] Rafael. C. Gonzalez and Richard E. Wood, *Digital Image Processing*, Addison Wesley, 1992.
- [11] Y. Suzuki, F. Asano, and T. Sone, "On the simulation of a transfer function of an acoustic system(Part1)," *J. Acoust. Soc. Jpn.*, 44, pp. 936-942, 1988.
- [12] Y. Suzuki, F. Asano, and T. Sone, "On the simulation of a transfer function of an acoustic system (Part 2)," *J. Acoust. Soc. Jpn.*, 45, pp. 44-50, 1989.
- [13] Watanabe, Y., Tokuno, H., Hamada, H., and Nelson, P: A, "Subjective investigation of the new transaural system (stereo dipole system)," *ACTIVE 97*, pp. 919-926, 1997.
- [14] Ming-Hsuan Yang, David Kriegman, and Narendra Ahuja, "Detecting Face in Images: A Survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 2001.
- [15] Changhoon Hyun, "Real-time Face Detection And Tracking System Using Web Camera," Hanyang univ, 2001.
- [16] Shaick, B. and Yaroslavsky, L., "Accelerating face detection by means of image segmentation," *Video/Image Processing and Multimedia Communications, 2003. 4th EURASIP Conference*, Vol. 1, pp. 411-416, July 2003.
- [17] R. Jain, "Extraction of motion information from peripheral processes." *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 3. pp. 489-503, 1981.



김 현 태

1989년 2월 부산대학교 전자공학과(학사)
1995년 2월 부산대학교 전자공학과(석사)
2000년 2월 부산대학교 전자공학과(박사)
2002년 3월~현재 동의대학교 멀티미디어공학과 조교수

관심분야: 음성 및 음향신호처리, 멀티미디어신호처리, 입체음향



박 장 식

1992년 2월 부산대학교 전자공학과(학사)
1994년 2월 부산대학교 전자공학과(석사)
1999년 2월 부산대학교 전자공학과(박사)
1997년 3월~현재 동의과학대학 디지털정보전자과 부교수

관심분야: 음성 및 음향신호처리, 멀티미디어통신, 입체음향



이 광 의

1990년 서강대학교 전자계산학과(공학사)
1992년 서강대학교 전자계산학과(공학석사)
1997년 서강대학교 전자계산학과(공학박사)
1997년 12월~2001년 2월 한국 전자통신연구원 선임연구원

2001년 3월~현재 동의대학교 멀티미디어공학과 조교수
관심분야: 계산이론, 가상현실, 상황인지, 유비쿼터스 컴퓨팅