

능동 시청각 시스템 설계 및 구현

*정채용, 임성길, 이현수
경희대학교 전자계산공학과

e-mail : jungchw@khu.ac.kr skan72@khu.ac.kr, lee@khu.ac.kr

Design and Implementation of Active Perception System

*Chae-Woong Jung, Sung-Kil Lim, Hyon-Soo Lee
Dept. of Computer Engineering, Kyung Hee University

Abstract

In this paper, we design and implement an active perception system. The proposed active perception system consists of four parts : an active robot, an aural active perception part, a visual perception part and an association part.

I. 서론

능동 인식이란 시각 및 청각적인 인식과정과 정확한 시/청각적 데이터를 관측하기 위한 능동적인 움직임을 가진 시스템을 연구하는 분야이다[1][2]. 능동 시청각 시스템은 주변의 환경을 인식/판단하고 그에 따른 적절한 동작을 해야 하는 로봇에서 필수적인 기술이다.

능동 인식 시스템에 적용되는 시각적, 청각적인 인식 기술은 응답 시간이 매우 짧아야 하며, 특히 청각 처리의 경우는 움직임에 따른 잡음이 발생하기 때문에 잡음에 강한 특성을 요구하고 있다.

본 연구에서는 능동 시/청각 시스템에서 응답 시간이 짧고, 잡음에 강한 알고리즘을 적용하여 능동 인식 시스템을 구현한다.

II. 능동 시청각 시스템 설계

2.1 시스템 개요

능동적인 시/청각 시스템을 구현하기 위하여 본 연구에서는 크게 4가지 부분으로 나누어 연구를 수행하였다. 첫째는 능동 청각 시스템에서 음원의 위치를 판단하는 연구이고, 두 번째는 능동 시각을 이용한 얼굴 인식 및 추적에 대한 연구, 세 번째로는 시/청각 인식 결과로부터 로봇을 지능적으로 제어하는 연구이고, 마지막으로 네 번째는 능동적인 시/청각을 가지는 로봇

의 제작이다. 그림 1은 제안하는 시스템의 블록 구조도이다.

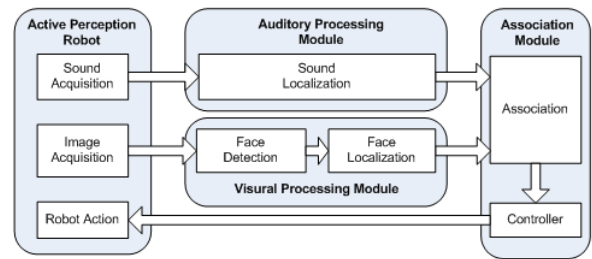


그림 1. 제안하는 능동 인식 시스템의 블록 구조도

2.2 음원의 위치 추적

능동 인식 시스템에서 음원의 위치를 판단하는 알고리즘은 응답시간이 빨라야 하며, 잡음에 강한 특성을 가지고 있어야 한다. 따라서 본 연구에서는 잡음의 영향을 적게 받는 채널들을 선택하여 음원의 위치를 판단하는 알고리즘을 적용하였다.

인간의 음성은 피치(기본주파수)의 정수배에 해당하는 채널에서 높은 진폭을 갖는 배음구조(harmonic structure)를 가지고 있는데, 배음 구조에 해당하는 채널들은 음성의 신호가 강하게 나타나기 때문에 잡음의 영향을 적게 받는다. 본 연구에서는 입력 음성의 피치를 추정하여 그 배음에 해당하는 채널들만 이용하여 음원의 위치를 추정하는 방법을 사용하여 계산량을 줄이고 잡음에 강한 음원 위치 추정 알고리즘을 적용한다. 적용된 음원 위치 추정의 블록 구조도는 그림 2와 같다.

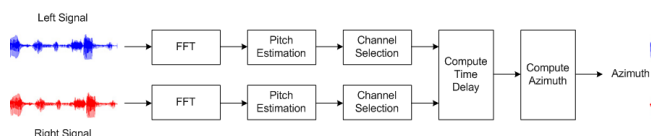


그림 2. 적용된 음원 위치 추정 방법의 구조도

2.3 얼굴 검출

능동 인식 시스템에서 얼굴 검출은 일반적인 정지 영상에서 얼굴을 검출하는 방법과는 다르게 동영상에 대상으로 하기 때문에 빠른 응답시간을 가져야 하며, 대상 얼굴이 일정 거리 안에 존재하기 때문에 측정된 화면과 얼굴의 비율이 비교적 고정되어 있다는 특징을 가지고 있다. 본 연구에서는 [3]에서 제안된 얼굴 검출 알고리즘을 시스템에 맞게 수정하여 적용하였다. 그림3은 적용된 얼굴 검출 알고리즘의 블록 구조도이다.

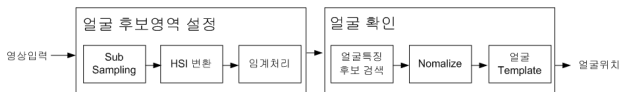


그림3. 적용된 얼굴 검출 방법의 구조도

2.4 시각 및 청각 연합 및 제어

청각에 의한 방향 결정은 음성신호가 시간적으로 연속적이지 않고 잡음의 영향을 받아 정확하지 않은 경우가 발생하게 된다. 또한 청각에 의한 방향 결정의 해상도가 낮다는 단점을 가지고 있다. 반면 시각에 의한 방향 결정은 대상이 시간적으로 연속적으로 존재하여, 방향 결정의 해상도가 높은 특징을 가지고 있지만, 시야가 좁다는 단점을 가지고 있다.

본 연구에서는 청각에 의한 방향을 측정하여 로봇의 초기 방향을 설정한 후에, 얼굴을 검출하여 정확한 로봇의 방향을 결정하는 방법을 사용하여 시각 및 청각적인 정보를 연합한다. 일정 시간동안 대상으로부터 음성 신호가 검출되지 않거나, 새로운 음성 신호가 검출된 경우에 다시 로봇의 방향 설정을 반복한다.

2.5 능동 로봇 제작

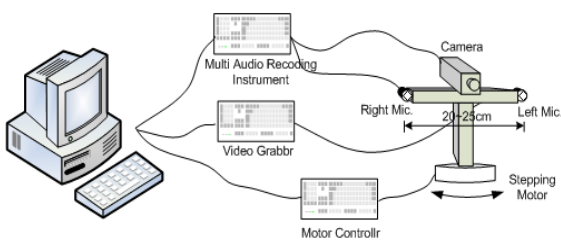


그림4. 구현된 시스템의 하드웨어 구성도

본 연구에서 제작한 능동 로봇의 하드웨어 구성도는 그림4와 같다. 시각 입력 장치로써 한 대의 카메라를 사용하며, 청각 입력 장치로써 두 대의 마이크를 사용하였다. 인식 과정 처리는 일반적인 펜티엄4 PC를 사용한다. 능동 시각의 경우 카메라의 가로방향 회전(pan), 상하방향 움직임(tilt), 줌(zoom)의 자유도 3인 시스템이 일반적이지만, 본 연구에서는 시각과 청각의 결합에 의한 지능적인 제어 시스템의 구현이 목표이므

로, 능동 시각과 능동 청각의 최소 움직임인 가로방향 회전만 수행하는 시스템을 구현하였다. 로봇 제작에 사용된 시스템의 사양은 표1과 같고 제작한 로봇 및 구현된 프로그램은 그림 5와 같다.

표 1. 구현 시스템 사양

| | 장비 | 인터페이스 |
|------------|---------------------------|----------------|
| Microphone | AKG C417 (x2) | WaMi Rack 192L |
| Camera | MPC-C10 (x1) | USB |
| Motor | RC Servo Moter (x2) | SMC Pro |
| PC | Pentium4 1.9GHz 512MB RAM | - |

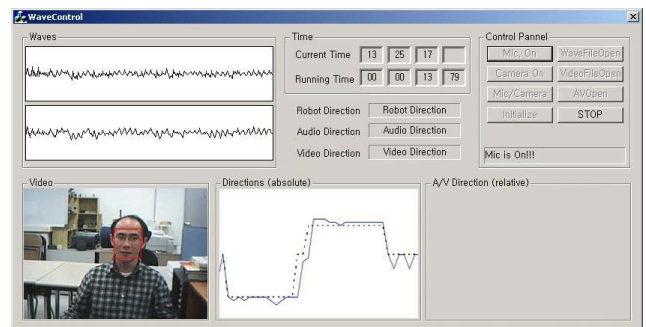
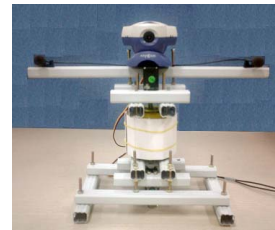


그림 5. 구현된 능동 시청각 시스템 및 프로그램 인터페이스

V. 결론 및 향후 연구 방향

본 연구에서는 능동 시/청각 인식 및 제어 부분으로 구성된 능동 인식 시스템을 제안하고 능동 인식 로봇을 시제작하여 제안한 시스템을 적용하였다. 향후에는 여러 화자가 발화하는 환경에서 화자 인식 및 얼굴 인식 알고리즘을 적용하여 보다 지능적으로 동작하는 능동 시청각 시스템에 대한 연구가 필요하다.

참고문헌

- [1] Y. Aloimonos, I. Weiss, A. Bandyopadhyay, "Active vision", International Journal of Computer Vision, 1987
- [2] K. Nakadai, T. Matsui, H.G. Okuno, H. Kitano, "Active audition system and humanoid exterior design", Proc. of IEEE International Conference on Intelligent Robots and Systems, pp. 1453-1461, 2000
- [3] Ikeda, O., "Segmentation of faces in video footage using HSV color for face detection and image retrieval", Image Processing, 2003. Vol. 3, pp. 913-916, 2003