보안카메라에서 소리인식 구현

윤태인 · 구하늘 · 김도은 · 장원석 · 권순각 · 권오준
동의대학교 컴퓨터소프트웨어공학과

Implementation of Sound Recognition for Security Camera

Taein Yun · Haneul Ku · Doeun Kim · Wonserk Jang · Soonkak Kwon · Ohjun Kwon
Department of Computer Software Eng., Dongeui University

요 약

소리인식이란 우리 귀에 들리는 모든 소리를 받아 들여 소리의 값과 저장되어 있는 데이터의 값을 비교하여 인식 결과를 도출해내는 과정을 의미한다. 보안 카메라는 현재 다양한 장소에서 설치되어 있어도 여전히 보안의 사각지대는 존재하며, 이를 보완하기 위해서는 여러 방향을 촬영하기 위한 아주 많은 보완 카메라가 설치될 수 밖에 없다. 그렇게 되면 설치비용이 더욱 증가되고, 무수한 카메라는 사람들에게 심적 부담감을 줄 것이다. 본 논문은 보안 카메라에 마이크를 설치하고, 입력되는 소리를 인식하여 발생되는 상황을 판단하는 시스템을 설계하고 구현하기 위한 것이다. 이를 바탕으로 보안 카메라의 사각지대를 소리인식으로 해결할 수 있어서 보완 카메라의 설치 비용을 줄일 수 있다.

키워드
DTW(Dynamic Time Warping) , PCM , Digital Bit Stream

Ⅰ. 서 론

보안 카메라가 설치되어 있어도 여전히 보안 카메라의 사각지대는 존재하며, 이를 보완하기 위해선 여러 방향으로 보완 카메라를 설치할 수 밖에 없다.
그렇게 되면 설치비용이 더욱 증가되고, 무수한 카메라는 사람들에게 심적 부담감으로 다가올 수 있다. 이를 보완하기 위하여 보안 카메라에 소리인식 마이크를 설치하여 보안 카메라의 사각지대를 소리인식으로 해결하는 방법을 구현한다.

Ⅱ. 소리 인식 방법

2-1 소리 인식 과정

(그림2.1)

아날로그 신호가 들어오면 PCM과정을 거쳐 헤더파일이 제거되고 DB데이터와 비교할 수 있는 디지털 비트 스트림이 나오게 된다.

2-2 헤더 파일 제거

소리가 인식되면 먼저 웨이브 파일로 저장되게 된다. HEADER 부분, Information 부분으로 구

분하여 평균소리 데이터와 비교를 하기 위해서 HEADER 부분을 제거하게 된다. 헤더파일 제거는 PCM과정을 이용하여 제거한다.

2-3 PCM 이용

2-3-1 PCM의 필요성

전화로 예를 들어 전화는 음성 파형을 전기적 파형으로 변환하여 상대방에게 정보를 전송하는 것이다. 전화망은 원래 아날로그 망으로 개발되었다. 그러나 아날로그 형태로 정보를 전송하면 전송로를 통과할 때 왜곡되거나 전송 도중에 잡음이 섞여 파형이 흩어져 통신 품질이 좋지 않다. 인간의 목소리는 높고 낮음과 진폭의 크고 작은 요소를 가지고 있는데, 이와 같은 고저 강약을 부호로 바꾸어 전송하면 잡음에 강하고 다중화가 쉬워 경제적인 전송을 할 수 있다. 그렇기 때문에 아날로그 정보를 디지털로 바꾸게 되었다.

2-3-2 PCM을 이용한 변환과정

표본화(Sampling)

아날로그 신호를 일정한 간격으로 샘플링하는 것을 표본화라고 하는데, 『샘플링 주기는 원 신호 최고 주파수의 2배의 빈도로 표본화 하면 원래의 신호는 완전히 복원된다』는 표본화 정리가 PCM의 이론적 근거로 되어 있다. 전화의 경우 최고 주파수가 4kHz(1초간 4,000번 진동)이기 때문에 2배의 8kHz, 즉 매초 8,000회의 빈도로 표본화 하면 좋다는 것이 표본화 정리인 것이다.

즉 8kHz간격 (1÷8Khz=125마이크로 초)마다 보내더라도 좋은 것이다. 이렇게 해서 매초 8,000의 표본치 하나하나가 8bit의 2진 부호화되니까 8,000×8=64,000, 즉 매초 64kbit/s 의 부호 정보가 되는 것이다.

양자화(Quantization)

아날로그 음성 파형은 간단히 수치화하기 힘들다. 어느 수치와 수치 사이에 있을 수 있는 수치가 많기 때문에 이를 4사 5입해서 간단한 수치로 고치는 것을 양자화라고 한다.

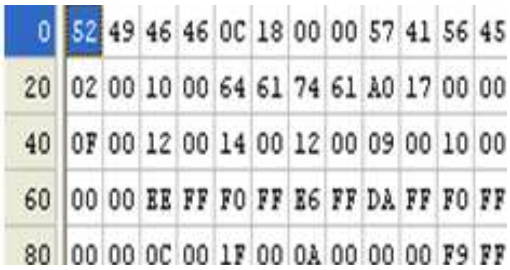
부호화(Encoding)

양자화 값을 2진 디지털 부호로 바꾸는 것을 부호화라고 한다. 즉 LSI(대규모직접회로)가 감지할 수 있거나 없는 펄스 조합으로 표시하는 것이다.

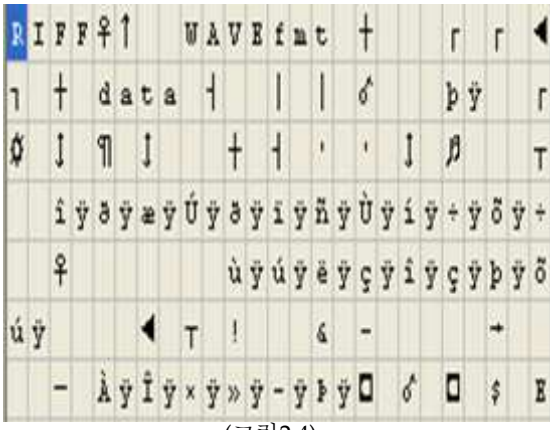
최대 변동폭을 세밀하게 구분하면 할 수록 샘플 값을 정교하게 보낼 수가 있다. 그러나 한 개의 샘플 값을 보내는 데 필요한 [0], [1] 펄스의 수 부호화 비트 수가 늘어나면 그만큼 반도체 부품이 늘어나게 되고, 반대로 듬성듬성 부호화하여 비트 수가 적어지면 품질이 나쁘게 된다.

2-4 디지털 비트열 추출

그림(2.3)은 헤더파일이 제거되기 전의 16진수 정렬이고, 그림(2.4)는 아스키 코드를 나타낸다.



(그림2.3)



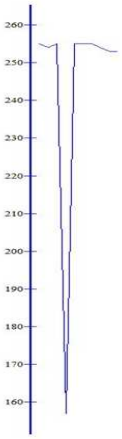
(그림2.4)

Ⅲ. 모의실험

디지털 비트 스트림 파일의 순수한 데이터의 각 자리를 x라 하고, 데이터의 값을 y로 하여 디지털 비트 스트림 파일의 그래프를 만든다. 만들어진 그래프를 사전에 정의해둔 데이터 그

래프와 비교를 하여 일치하는 결과를 보여준다.

추출된 데이터로 그래프 만들기

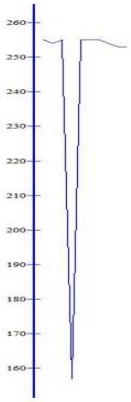


x	y
1	255
2	254
3	255
4	157
5	255
6	255
7	255
8	254
9	253
10	253

Y값이 너무 커서 X좌표가 보이지 않지만 각 X좌표에 의한 Y값들에 대한 그래프가 나온다.

(그림3.1)

정의된 데이터로 만든 그래프



x	y
1	255
2	254
3	255
4	157
5	255
6	255
7	255
8	254
9	253
10	253

미리 정의된 그래프도 인식과정과 마찬가지의 과정을 거쳐 그래프를 추출하였다.

개발자가 미리 인식된 소리와 비교할 수 있는 평균 데이터 그래프를 프로그램에 넣어둔다.

(그림3.2)

음성을 비교하기 위해 음성인식 **DTW 알고리즘 (Dynamic Time Warping)을** 사용한다.
DTW알고리즘은 일단 두열의 각성분에 대한 거리 척도 값을 비용으로 설정한다. 그리고 두열이 이루는 격자상에서 각 열의 시작성분에서 시작하여, 끝성분에 이르기 까지 비용테이블에 최소비용을 순환적으로 택하여 저장하는 점화식을 이용하는 동적계획법으로 매핑함수를 찾아가면서 두열을 비교하는 알고리즘이다.

Ⅳ. 결 론

본 논문은 소리인식을 하여 카메라에 적용하면 국내의 여러 마트와 PC방, 백화점등에 음성인식 보안카메라를 활용할 수 있고, 음식인식만을 따로 기술 이전하여 음성인식 TV나 음성인식 네비게이션등 다양한 방면에서 활용할 수 있다. 기업에서는 음성인식 기술로 IT기기에 익숙하지 않는 유아나 노인 등을 대상으로 음성으로 쉽게 조작하고, 사용할 수 있는 제품을 개발하는데도 도움이 될 것이다.

참고문헌

[1] Han Hag - Yong, “Introduction to Pattern Recognition", 한빛미디어, 2009.

[2] John Coleman, "Introduction to speech processing and natural language processing”, 한국출판사, 2009.