# 가중치 값에 따른 비명검출 성능 연구

*서지훈 **이혜인 ***박주현 ****이석필

*, **, ***, ****상명대학교

*JHSeoPMP@gmail.com

# A study on the scream detecting performance according to weight value

*Seo, Ji-Hun **Lee, Hye-In ***Park, Ju-Hyun ****Lee, Seok-Pil

Sangmyung University

## 요약

본 논문에서는 오디오기반 CCTV에서 비명 구간을 효과적으로 검출하기 위한 가중치 값을 실험을 통해 결정 하고자 한다. 경계값은 학습구간의 평균값에 가중치 값을 곱해주어 계산되며, 이 때 가중치 값에 의해 비명 구간 검출 성능이 결정된다. 따라서 본 논문에서는 가장 좋은 성능을 보이는 가중치 값을 결정하기 위해 가중치 값을 변화시키며 실험을 하였다. 그 결과 $w$값이 3일 때 검출률과 오인식률에서 가장 좋은 성능을 보였다.

## 1. 서론

최근 번화가, 차도주변, 골목길, 공원 등과 같은 공공장소에서의 소매치기, 강도, 성범죄 등 위험 상황 발생에 의해 안전에 대한 문제가 대두 되고 있다. 그 중 국민의 안전을 위협하는 범죄 문제는 발생률이 급증하면서 인력, 장비, 예산 낭비뿐만 아니라 사회적으로 국민의 불안감을 심화시키며 범죄 예방에 대한 경각심을 일깨우고 있다. 그에 따라 해결 방안으로 감시 분야에 대한 중요성이 더욱 강조되고 있다[1]. 현재까지 구축된 방범용 시스템은 특정 지역에 침입이 발생했을 때 센서를 통하여 경비요원이 바로 출동할 수 있도록 하는 무인경비 시스템과 범죄 발생 시 해당 지역에서 녹화된 영상물 수집을 통해 수사에 도움을 주는 블랙박스, 카메라에서 촬영된 화상정보를 이용하여 원하는 지역을 감시할 수 있도록 하는 CCTV 등 영상 정보를 이용하는 데에 집중되어 있다[2]. CCTV는 기능면에서 볼 때 경찰의 부족한 인력과 장비를 보완해주는 중요한 역할을 수행하고 있고 범죄의 예방과 통제의 수단으로 효과적이며, 관리자가 모니터링 중 놓칠 수 있는 영상 정보를 검색, 추적하여 위급한 상황이나 강도, 방범등에서 많은 발전을 이루고 있다. 그러나 영상 정보는 조명이 너무 어둡거나 밝을 경우 인식에 문제가 있으며, 사각지대가 존재 할 수 있기 때문에 비정상적인 상황을 제대로 인지하지 못할 수 있다. 또한 범죄 예방을 위해 들어가는 순찰 인력 문제와 범죄 발생 시 모든 영상을 다 확인해야하는 어려움이 있다.

이러한 문제를 해결하기 위한 방법으로 기존의 영상 데이터뿐만 아니라 오디오 데이터를 함께 사용한 방범용 시스템을 구축한다면 인력 부족 문제를 해결하고 더 효율적으로 범죄를 예방할 수 있다[3]. 비정상 상황을 비명소리로 인식하여 실시간으로 범죄 발생 여부를 알 수 있어 예방적인 면에서 더 효율적이며, 이미 범죄가 발생한 후에도 모든 영상 데이터를 검색할 필요 없이 오디오 데이터로부터 체크 된 시점부터 확인하여 수사 시간을 줄일 수 있다.

오디오기반 CCTV에서 비정상 상황을 인식하는 방법에 대하여 많은 선행연구가 있다. 일반적으로 사용되는 방법으로는 유성음과 무성음을 구분하기 위해 Zero Crossing Rate(ZCR)을 사용하고, 음성부와 비음성부를 구분하기 위해 Linear Prediction Coefficients(LPC), Linear Prediction Cepstral Coefficients(LPCC)을 특징으로 사용하며, 사람의 가청 주파수를 반영하여 특징을 추출하기 위해 Mel Frequency Cepstral Coefficients(MFCC)를 특징 벡터로 사용한다[4][5]. 또한 일반적인 상황과 비정상 상황을 분류하기 위해 확률적 패턴인식인 Gaussian Mixture Model(GMM), 패턴 분류에 우수한 성능을 보이는 Support Vector Machine(SVM)등을 시스템에 적용하고 있다[6][7].

본 논문에서는 비정상 상황을 인식하는 연구 중 비명 발생구간을 검출하기 위한 가중치를 결정하는 방법에 대한 연구를 하였다.

본 논문의 구성은 다음과 같다. 2장에서는 DB구성에 대해 서술하고, 3장에서는 비명 특징 추출, 4장에서는 실험 결과. 5장에서는 결론에 대해 논의 한다.

## 2. DB 구성

실험을 위해 27개의 환경DB와 180개의 비명DB를 직접 녹음하여 구성하였다. 또한, 검증을 위해 인터넷에서 20개의 녹음된 데이터를 가져왔다.

2.1 환경DB

보통 CCTV는 보안이 취약한 지역이나 사람이 밀접한 지역 등에 많이 분포되어 있다. 그래서 일반적으로 CCTV가 설치되는 위치와 시간대를 고려하여 총 27개의 환경 잡음을 녹음하였다. 녹음 장비로는 실제 사용할 방범용 CCTV의 음질을 고려하여 일반적인 휴대용 마이크로폰을 이용하였다. 한적한 골목길, 번화가, 차도에 대해 각각 다른 3가지 장소를 정하고 아침, 점심, 저녁 시간대에 각각 녹음을 수행하였다. 이렇게 녹음된 데이터는 16kHz로 다운 샘플링하고 모노 채널을 사용하였다. 인터넷에서 가져온 환경은 천둥, 번개가 치는 악천후를 녹음한 데이터와 번화가를 녹음된 데이터를 사용하였다.

2.2 비명DB

사람들은 보통 고통스럽거나 다급할 때, 그리고 놀랐을 때 비명을 지르게 된다. 비명은 보통 1초 내외의 짧은 소리이며, 남녀 성별간의 주파수 차이를 보인다[8]. 주변소음이 적은 밀실에서 피실험자와 마이크로폰의 거리를 5M로 두고 20~50대의 남녀 각각 30명에 대하여 3가지의 비명으로 총 180개의 비명을 녹음하였다. 소리의 SPL은 거리가 2배 증가 될 때 마다 약 6dB씩 감소하게 된다[9]. 이러한 성질에 따라 비명 녹음 데이터의 dB을 1/4로 줄여 20M거리의 비명소리로 구성하였다. 이렇게 녹음된 데이터는 16kHz로 다운 샘플링 했으며 모노 채널을 사용하였다. 인터넷에서 가져온 비명 데이터는 1초 내외의 여성 12명의 비명과 남성 6명의 비명을 사용했다.

3. 비명 특징 추출

이전 연구에 따르면 음역대가 낮은 사람의 경우 비명의 주파수 대역은 150Hz에서 533Hz에서 나타나고, 음역대가 높은 사람의 경우 비명의 주파수 대역이 500Hz에서 2,133Hz에서 나타난다[11]. 직접 녹음을 통해 얻은 비명 데이터의 주파수를 분석한 결과 비명의 주파수 대역은 625Hz에서 2,031Hz 이다. <그림 1>과 <그림 2>은 각각 직접 녹음한 DB비명의 남성, 여성의 주파수 영역 그래프를 나타낸다. 여성 비명의 경우 1,000Hz에서 2,000Hz부근 대역에서 특징적인 에너지가 나타나며, 남성의 비명의 경우 500Hz에서 1,500Hz부근 대역에서 특징적인 에너지를 나타낸다[10].

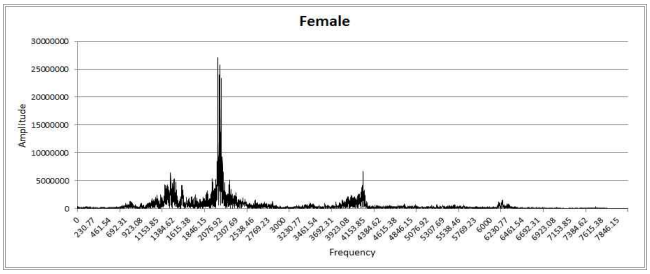


그림 1. 여성 비명 주파수 영역 그래프

Fig 1. Frequency domain graph of female scream

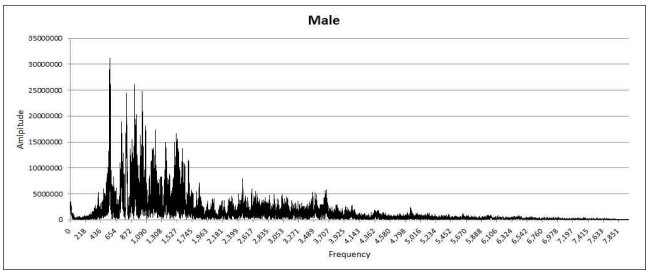


그림 2. 남성 비명 주파수 영역 그래프

Fig 2. Frequency domain graph of male scream

본 논문에서는 분석 결과를 토대로 625Hz과 2,031Hz사이에서 일정이상의 프레임이 연속된다는 비명의 특징을 이용하여 비명을 검출 한다.

각 환경마다 특징적으로 나타나는 소리가 다르며, 환경 잡음의 유형이 다르므로 그에 따른 주파수 대역과 그 에너지가 각각 다르다. 따라서 비명인지 아닌지를 판별하는 경계 값(threshold)을 고정된 값으로 설정하면 오차율(error rate)이 높다는 문제점이 있다. 이러한 문제를 해결하기 위해 일정시간 동안 입력되는 데이터를 학습하도록 한다. 경계값은 식 (1)을 이용하여 구한다. $AVR$은 학습구간에서 구한 평균값이고, $w$는 가중치 값이다. 이 과정을 통해 각 환경에 최적화 된 유동적인 경계 값을 설정한다.

$$threshold = AVR * w \qquad (1)$$

비명의 주파수 영역 특징들 이용하여 설정된 경계 값을 기준으로 경계값을 넘는 프레임이 연속되면 비명의 시작으로 판단하고, 그 경계 값을 넘지 못하는 프레임이 연속되면 비명의 끝점으로 판단된다. 이렇게 하나의 비명 구간을 찾게 된다.

4. 실험 결과

구성한 DB의 비명과 환경잡음을 합성하여 입력데이터를 만들었다. 환경에 따른 성능을 확인하기 위해 10dB의 환경과 4dB의 환경에 가중치 값을 1에서 5까지 변화

시키며 검출률과 오인식률에 대해 실험을 진행 하였다.

골목길이나 차도의 경우 음성과 비음성의 분명한 차이로 인해 좋은 비명 검출률과 낮은 오인식률을 보인다. 반면 사람이 북적거리는 번화가의 경우 비명과 음성간에 차이가 모호하기 때문에 앞의 두 환경에서보다 성능이 좋지 않다. 따라서 환경잡음을 번화가로 설정하여 실험을 진행하였다. <그림 3>은 10dB 번화가의 검출률, 오인식률 그래프이고, <그림 4>는 4dB 번화가의 검출률, 오인식률 그래프 이다. <그림 3>을 보면, 가중치 값이 3이상일때 검출률은 100%가 되고, 오인식률은 0%을 나타낸다. <그림 4>를 보면, 가중치 값이 2와 3일 때 검출률이 93%로 가장 좋다. 그러나 가중치 값이 2일 때 오인식률은 10%이고, 가중치 값이 3일 때 오인식률은 3%이다.


그림 3. 10dB 번화가 검출률, 오인식률 그래프
Fig 3. Accuracy & FAR graph
of 10dB main street

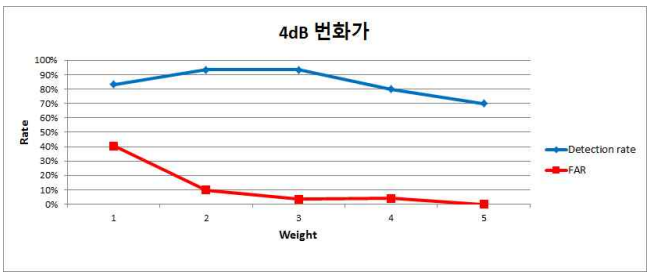


그림 4. 4dB 번화가 검출률, 오인식률 그래프
Fig 4. Accuracy & FAR graph
of 4dB main street

실험결과, 가중치 값이 3일 때 검출률과 오인식률에서 가장 좋은 성능을 보이는 것을 알 수 있다.

## 5. 결 론

본 논문에서는 비명을 검출하는 특징으로 주파수영역을 이용하였고, 주파수 영역에서 비명의 특징을 이용할 때 가장 좋은 성능을 보일 수 있는 가중치 값을 실험을 통해 결정하였다. 실험을 위해 DB를 구성하였고, 주파수 특징을 분석하여 비명이 특정 대역에서 큰 에너지를 갖는다는 것을 알아내었다. 환경 잡음 속에서 비명구간을 검출하기 위해 경계값을 계산하는데 있어 가중치에 따른 검출률과 오인식률의 성능을 실험하였다.

실험 결과, 10dB 번화가 환경에서 가중치 값이 3이상일 때 검출률이 100%이고, 오인식률이 0%임을 확인 할 수 있었다. 4dB 번화가 환경에서 가중치 값이 2와 3일 때 검출률이 93%로 가장 좋았으나, 가중치 값이 2일 때 오인식률이 10%인 반면 가중치 값이 3일 때는 3%이므로 더 성능이 좋은 것을 확인 할 수 있다.

10dB 번화가 실험과 4dB 번화가 실험을 종합해 보면, 10dB 번화가는 가중치 값이 3이상일 때, 모두 같은 결과를 보이는 것을 확인할 수 있으나, 4dB 번화가 실험에서는 가중치 값이 3이상으로 올라가면 검출률은 낮아지고, 오인식률은 비슷한 수준이므로 가중치 값이 3일 때가 10dB 환경과 4dB 환경에서 가장 좋은 성능을 보임을 알 수 있다.

따라서, 비명을 주파수 영역에서 검출 하고자 할 때 가중치 값을 3으로 할 때 가장 좋은 검출 결과를 얻을 수 있다. 추후 연구로 비명 검출하는 방법을 주파수 영역에서 시간 영역으로 옮겨 실시간 검출하는 시스템에 적용될 수 있도록 하는 연구가 필요할 것으로 판단된다.

## 참 고 문 헌

[1] Aki Harma, Martin F. McKinney, and Janto Skowronek, “Automatic surveillance of the acoustic activity in our living environment,” in IEEE International Conference on Multimedia and Expo, Amsterdam, July 2005.

[2] I. Haritaoglu, D. Harwood, and L. Davis, “W4: real-time surveillance of people and their activities,” IEEE Transactions on Pattern Analysis and Machine Intelligence, pp. 809-830, 2000.

[3] C. Clavel, T. Ehrette, and G. Richard, “Events Detection for an Audio-Based Surveillance System,” Multimedia and Expo, 2005. ICME 2005. IEEE International Conference on, pp. 1306 - 1309, 2005.

[4] J. Rouas, J. Louradour, and S. Ambellouis, “Audio Events Detection in Public Transport Vehicle,” Proc. of the 9th International IEEE Conference on Intelligent Transportation Systems, 2006.

[5] M. Pleva, E. Vozáriková, S. Ondáš, J. Juhár, A. Čižmár, “Automatic detection of audio events indicating threats”, IEEE International Conference on Multimedia Communications, Services and Security, Krakow 6.-7.5.2010, AGH Krakow, pp. 198-201

[6] P. Atrey, N. Maddage, and M. Kankanhalli, “Audio Based Event Detection for Multimedia Surveillance,” IEEE International Conference on Acoustics, Speech, and Signal Processing, 2006, 2006.

[7] G. Valenzise, L. Gerosa, M. Tagliasacchi, F.Antonacci, A. Sarti, "Scream and Gunshot Detection and Localization for Audio-Surveillance Systems," IEEE International Conference

on Advanced Video and Signal Based Surveillance (AVSS 2007), pp. 21-26, 2007.

[8] S. M. Lee, S. W. Byun, S. C. Li, K. Y. Kim, I. G. Chung, S. P. Lee, "Screaming data analysis for security system with audio capability" 2013년도 한국방송공학회 추계 학술대회, 2013.11, 85-87 (3 pages)

[9] http://en.wikipedia.org/wiki/Inverse-square_law

[10] J. H. Park, H. I. Lee, J. H. Seo, G. Y. Kim, I. G. Chung, S. P. Lee "Environment noise analysis for Security system with Audio capability" 2013년도 한국방송공학회 추계 학술대회, 2013.11, 81-84 (4 pages)