ISSN 1975-8359(Print) / ISSN 2287-4364(Online)
The Transactions of the Korean Institute of Electrical Engineers Vol. 63, No. 11, pp. 1559~1563, 2014
http://dx.doi.org/10.5370/KIEE.2014.63.11.1559

보안 시스템을 위한 비명 검출 엔진 설계

A Design of a Scream Detecting Engine for Surveillance Systems

서 지 훈* · 이 혜 인* · 이 석 필†
(Ji-Hun Seo · Hye-In Lee · Seok-Pil Lee)

Abstract - Recently, the prevention of crime using CCTV draws special in accordance with the higher crime incidence rate. Therefore security systems like a CCTV with audio capability are developing for giving an instant alarm. This paper proposes a scream detecting engine from various ambient noises in real environment for surveillance systems. The proposed engine detects scream signals among the various ambient noises using the features extracted in time/frequency domain. The experimental result shows the performance of our engine is very promising in comparison with the traditional engines using the model based features like LPC, LPCC and MFCC. The proposed method has a low computational complexity by using FFT and cross correlation coefficients instead of extracting complex features like LPC, LPCC and MFCC. Therefore the proposed engine can be efficient for audio-based surveillance systems with low SNRs in real field.

Key Words : Scream, Ambient noise, Surveillance system, LPC, LPCC, MFCC

1. 서 론

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.

† Corresponding Author : Dept. of Media Software, Sangmyung University, Korea.
E-mail: esprit@smu.ac.kr
* Dept. of Computer Science, Sangmyung University, Korea.
Received : April 10, 2014; Accepted : October 23, 2014

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 대표할 수 있는 오디오 신호로는 비명이 있다. 사람들이 보통 위험하거나 급박한 상황에 처하게 되면 가장 흔하게 표현하는 오디오 신호가 비명이다. 따라서 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 인식하고, 분류하는 연구 중 비정상 상황을 인식하는 방법에 대한 연구를 대상으로 한다. 이를 위해 비명소리와 같은 비정상 상황을 인지하기 위해 비명을 효율적으로 인식할 수 있는 방법에 대해 분석하고, 그것을 이전 연구와 비교해 보고자 한다. 제안하는 방법은 환경잡음과 비명데이터의 주파수 분석을 통해 주파수 영역에서의 특징을 찾는 것이다. 주파수 분석을 위해 직접

녹음을 하여 DB를 구성하였고, 실제 감시 시스템이 설치되는 요소를 고려하여, 2~3미터의 거리를 두고 환경 소리를 녹음 하였고, 감시시스템의 감시 범위를 20M정도로 가정하고 비명 소리를 녹음 하였다. 또한 효율성 검증을 위해 제안하는 방법과 이전 연구에서 사용된 방법과의 검출률과 오인식률(False Alarm Rate)을 비교한다.

본 논문의 구성은 다음과 같다. 2장에서는 본 논문에서 제안하고자 하는 방법에 대해 기술한다. 3장에서는 이전 연구와 제안하고자 하는 방법의 실험 결과를 비교하고, 4장에서는 결과를 통해 결론에 대해 논의 한다 .

2. 제안하는 방법

그림 1은 비명 검출 엔진 구조를 나타낸 순서도 이다. 입력은 비명과 환경잡음이 합성된 데이터이며 Pre-processing을 수행한 뒤 FFT를 이용하는 방법과 Cross correlation을 이용하는 두 가지 방법을 통해 비명을 검출하고 검출되는 시점을 알려주도록 설계하였다.

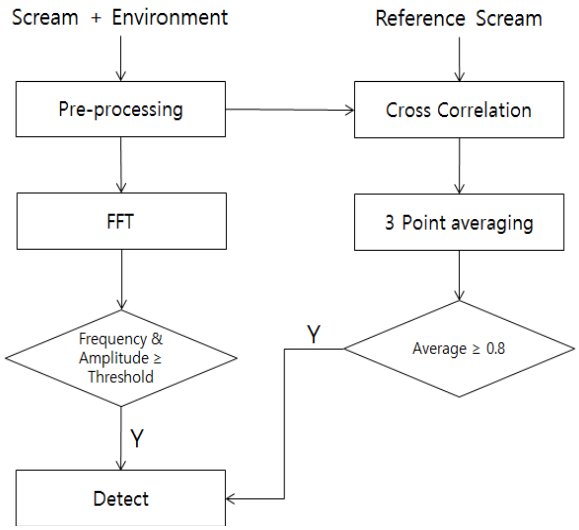

그림 1 비명 검출 엔진 구조
Fig. 1 Architecture of scream detect engine

입력 데이터는 16kHz로 다운 샘플링 된 모노 채널의 데이터인 환경DB와 비명DB를 합성하였다. 주파수 영역에서의 처리를 위해 32ms 길이의 Hamming window를 적용하여 STFT(Short Time Fourier Transform)를 해 준다. 비명의 길이가 짧기 때문에 프레임의 절반을 중첩하여 16ms씩 이동한다.

2.1 DB구성

실험을 위해 27개의 환경DB와 180개의 비명DB를 직접 녹음하여 구성하였다. 또한 제안하는 엔진의 성능 검증 실험을 위해 인터넷에서 20개 녹음된 데이터를 사용하였다.

2.1.1 환경DB

일반적으로 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. 녹음을 할 때 환경잡음과의 거리는 2~3M정도를 두고 하였으며, 이것은 실제 감시 시스템이 설치되는 마이크의 위치를 고려하고자 하였다. 이렇게 녹음된 데이터는 16kHz로 샘플링하고 모노 채널을 사용하였다. 인터넷에서 가져온 환경은 천둥, 번개가 치는 악천후를 녹음한 데이터와 번화가를 녹음된 데이터를 사용하였다.

2.1.2 비명DB

사람들은 보통 고통스럽거나 다급할 때, 그리고 놀랬을 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. 보통 일반적인 감시 시스템의 감시 범위는 20M를 상정한다. 그런데 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 샘플링 했으며 모노 채널을 사용하였다. 인터넷에서 가져온 비명 데이터는 1초 내외의 여성 12명의 비명과 남성 6명의 비명을 사용했다.

2.2 Pre-processing

각 환경마다 특징적으로 나타나는 소리가 다르며, 환경잡음의 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. 이 단계는 이러한 문제를 해결하기 위한 전처리 과정으로 엔진이 일정시간 동안 입력되는 데이터의 환경을 학습하도록 한다. 식 (1)을 이용하여 경계 값을 구한다. AVR은 학습구간에서 구한 평균값이고, w는 실험을 통해서 얻은 가중치 값이다. 이 과정을 통해 엔진은 각 환경에 최적화 된 상대적인 경계 값을 설정한다. 이렇게 설정된 경계 값은 비명 주파수 대역에서 비명 검출의 기준으로 한다.

$$Threshold = AVR * w \tag{1}$$

2.3 비명 특징 추출

본 논문에서는 주파수영역에서의 분석과 연산을 통해 환경 잡음 속에 비명이 섞여 있을 때, 관리자에게 알려주는 것

에 목적을 두고 있다. 구성한 환경DB와 비명DB의 주파수 영역을 분석 하여 비명의 특징을 추출하고, 상호 상관을 통해 비명을 검출할 수 있는 특징을 찾고자 한다.

2.3.1 주파수 영역에서의 특징 추출

그림 2는 골목길의 주파수 영역 그래프, 그림 3은 차도의 주파수 영역 그래프, 그림 4는 번화가의 주파수 영역 그래프 이다. 골목길, 차도, 번화가환경들에서 공통적으로 300Hz 이하에서 큰 에너지를 보이는 것을 확인할 수 있으며, 골목길 환경은 주로 사람들이 지나다니는 발자국 소리, 바람 소리이다. 차도 환경은 주로 차가 지나가는 소리, 경적소리이며 비명과 혼동하기 쉬울 수 있는 경적소리는 2500Hz부근에서 큰 에너지가 발생 한다. 번화가 환경은 주로 사람들의 말소리와 노래 소리이며, 2000Hz부근에서 큰 에너지를 보인다[10].

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 구성한 DB를 통해 얻은 비명 데이터의 주파수를 분석한 결과 비명의 주파수 대역은 625Hz에서 2,031Hz 이다. 그림 5와 그림 6은 각각 직접 구성한 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[8].

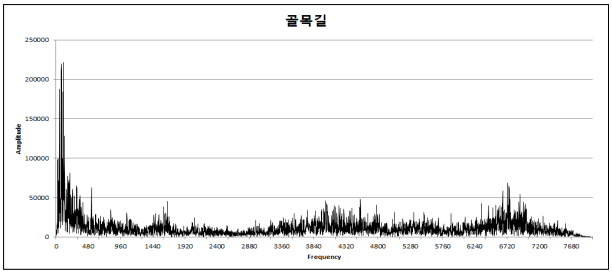

그림 2 골목길 주파수 영역 그래프
Fig. 2 Frequency domain graph of alley

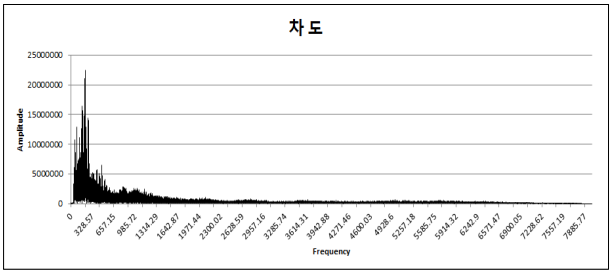

그림 3 차도 주파수 영역 그래프
Fig. 3 Frequency domain graph of driveway

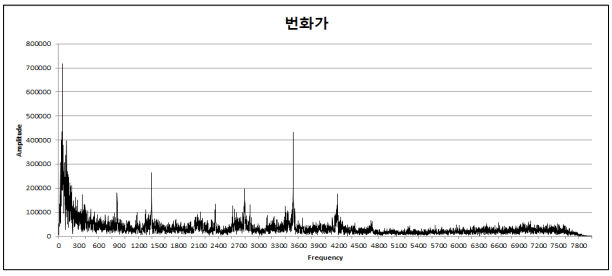

그림 4 번화가 주파수 영역 그래프
Fig. 4 Frequency domain graph of main street

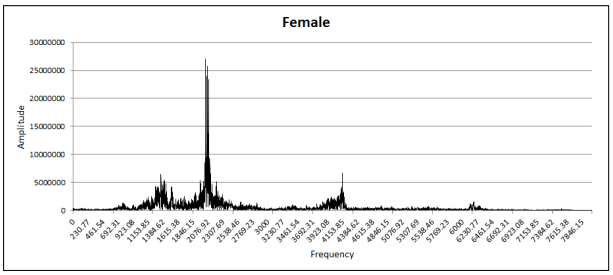

그림 5 여성 비명 주파수 영역 그래프
Fig. 5 Frequency domain graph of female scream

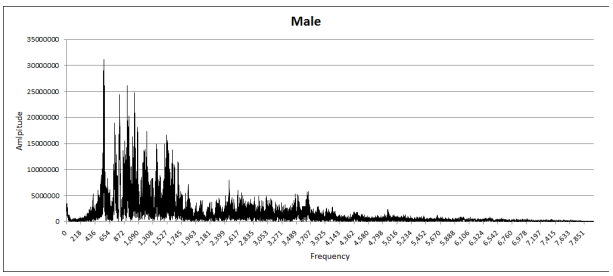

그림 6 남성 비명 주파수 영역 그래프
Fig. 6 Frequency domain graph of male scream

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

2.3.2 Cross correlation

기존의 상호 상관법은 시간영역에서 수행하나, 본 논문에서는 주파수영역의 특징을 좀 더 활용하기 위해서 주파수 영역에서 상호 상관 연산을 수행하였다. 주파수 분석을 위해 FFT연산을 수행하기 때문에 연산속도에는 큰 영향을 미치지 않으며, 주파수 영역으로 상호 상관법을 수행하여도 결과는 같게 나온다[12]. 식(2)은 상호 상관법의 일반적인 식이다.

$$R_{xy}(\tau) = \int_{-\infty}^{\infty} x(t)\, y(t+\tau)\, dt \tag{2}$$

상호 상관법의 결과가 다음 식(3)의 조건을 만족하는 프레임이 일정 개수 이상 연속 되었을 경우 비명으로 판단한다. 경계 값으로 설정된 0.8은 많은 실험을 통해 결정된 값으로, 검출률과 오인식률 측면에서 가장 좋은 성능을 보였다.

$$|R_{xy}(\tau)| > 0.8 \quad (3)$$

2.4 비명 구간 찾기

비명 구간을 정확하게 검출한다면, 처리할 데이터가 줄어들게 되고 효율적인 처리가 가능 해진다[13]. 앞에서 설명한 비명의 특징들을 이용하여 비명의 시작점을 찾고, 그 지점부터는 그 특징의 조건을 만족하지 못하는 프레임이 일정이상 연속되었을 때, 그 지점을 끝점이 된다. 즉, 전처리 과정으로 설정된 경계값을 넘는 프레임이 연속되면 비명 시작이고, 그 경계값을 넘지 못하는 프레임이 연속되면 비명 끝이 된다. 이렇게 하나의 비명 구간을 찾게 된다.

3. 실험 결과

논문에서 제안하는 방법의 성능을 기존의 연구의 방법으로 한 것과의 비교 실험을 위하여 이 두 방법을 이용한 프로그램을 같은 컴퓨터에 MFC를 이용하여 구현 하였다. 비명과 환경잡음을 합성하여 입력 데이터를 만들었다. 비교 실험으로는 제안하는 방법과 기존연구들에서 진행했던 LPC, LPCC, LPC+LPCC, MFCC 이렇게 5가지에 대하여, -10dB부터 10dB까지 5dB씩 올리며 5단계에 걸쳐 검출률과 오인식률에 대해 진행 하였다.

주차장법 시행규칙 6조 1항 11호에 따르면, '방범설비는 주차장의 바닥면으로부터 170센티미터의 높이에 있는 사물을 식별할 수 있도록 설치하여야 한다'[14]. 따라서 본 실험을 위해 환경잡음은 2~3M의 거리에서 취득하였고, 감시 범위는 20M로 하였다.

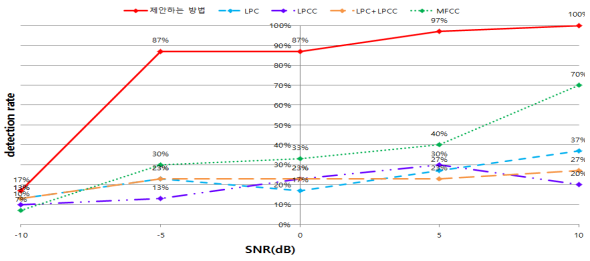

그림 7 SNR에 따른 검출률 비교 실험 결과
Fig. 7 Accuracy comparative experiment result of various SNR

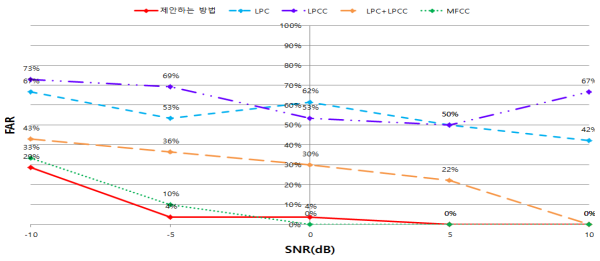

그림 8 SNR에 따른 오인식률 비교 실험 결과
Fig. 8 False Alarm Rate comparative experiment result of various SNR

골목길이나 차도의 경우, 사람소리와 사람소리가 아닌 소리의 분명한 차이로 인해서 다른 방법 모두 좋은 결과를 보이기 때문에, 환경잡음으로는 사람이 북적이는 번화가 소리를 하였다. 입력 데이터는 5가지 방법 모두 같게 하였으며, 검출하는 방법만 다르게 하여 진행 하였다. 그림 7은 SNR에 따른 검출률 그래프 이고, 그림 8은 SNR에 따른 오인식률(FAR) 그래프 이다. 그림 7에서 보는바와 같이 SNR이 -5dB인 환경에서 제안하는 방법의 검출률은 87%인 반면 다른 방법은 30% 이내로 검출률이 낮다. 또한 SNR이 5dB인 환경에서는 제안하는 방법의 검출률은 97%인 반면 다른 방법은 가장 높은 검출률이 40%이다. 그림 8은 SNR이 -5dB인 환경에서 제안하는 방법의 오인식률이 4%인 반면 그 외의 가장 낮은 오인식률이 10%이고, SNR이 5dB인 환경에서는 제안하는 방법은 0%의 오인식률을 가지며 그 외의 방법에서는 최대 50%의 오인식률이 있다는 것을 보여준다. 실험결과 일반적인 환경의 SNR인 10dB일 때 제안하는 방법은 검출률이 100%, 오인식률이 0%로 다른 방법을 적용한 결과와 비교해보았을 때 성능이 우수한 것을 알 수 있다.

4. 결 론

본 연구에서는 CCTV에 적용시킬 수 있는 실시간 비명 검출 엔진을 구현하였다. 엔진은 실시간으로 검출이 가능하며 검출된 시점을 사용자에게 제공한다. 엔진을 구현하기 위해 사람의 비명소리와 환경 소리를 직접 녹음하여 DB를 구축하였고, 각각의 주파수 분석을 수행하였다. 주파수 분석을 수행한 결과 비명이 특정 주파수 대역에서 큰 에너지 값을 갖는다는 특징을 알아내었고, 이 특징을 이용하여 비명 검출 알고리즘을 설계했다. 엔진은 모든 환경에 적용될 수 있도록 주변 잡음을 일정시간동안 학습하는 전처리 과정을 통해 비명으로 판단하는 경계 값을 설정하며, 이 값으로 비명의 시작점과 끝점을 결정하여 비명 구간을 찾게 된다. 또한 상호상관계수를 이용하여 계수 값이 실험을 통해 얻은 값 0.8 이상일 경우 비명으로 판정하여 오인식률을 낮추기 위해 보조적인 수단으로 사용하였다.

본 연구에서 제안하고자 하는 방법과 엔진의 성능 검증을 위해 선행 연구에서 많이 사용되는 LPC, LPCC, MFCC를 이용하여 비명을 검출한 결과와 비교 실험을 수행한 결과 검출률과 오인식률에서 제안하는 방법이 성능이 더 좋다는 것을 확인할 수 있다. 실험 결과 SNR이 10dB인 환경에서는 검출률이 100%이며, SNR이 5dB인 환경에서는 97%의 검출률을 보였다. 오인식률은 SNR이 5dB 이상인 환경에서는 0%를 보였다.

구현된 엔진은 간단하게 특정 주파수 대역에서의 에너지 값과 상호상관 계수값을 통해 비명을 검출하여 이전의 복잡한 방법의 검출 엔진보다 효율적이라고 판단된다. 또한, 다른 방법보다 SNR의 변화에 민감하지 않기 때문에 오디오기반 CCTV에 적용된다면 범죄예방에 좋은 효과를 보일 수 있을 것으로 보인다. 추후 연구를 통해 비명과 환경잡음을 분리하여 분리된 비명소리가 범죄의 예방 및 증거로 사용될 수 있도록 발전시키는 연구가 필요하다고 판단된다.

감사의 글

본 연구는 2014년도 상명대학교 교내연구비를 지원받아 수행하였음.

References

[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”, trans. KOSBE Conference, 85-87(3 pages), Nov 2013

[9] Wikipedia contributors, Inverse-sqaure Law [Online]. Available: http://en.wikipedia.org/w/index.php?title=Inverse_square_law&oldid = 16039900

[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”, KOSBE Conference, 81-84 (4 pages), Nov 2013.

[11] C. F. Chan, Eric W.M. Yu, “AN ABNORMAL SOUND DETECTION AND CLASSIFICATION SYSTEM FOR SURVEILLANCE APPLICATIONS”, 18th European Signal Processing Conference, EURASIP, 2010 ISSN 2076-1465, 2010

[12] Wangrae Jo, Jongkuk Kim, Myungjin Bae, “A Study on Pitch Detection in Time-Frequency Hybrid Domain”, Springer-Verlag, Lecture Notes in Computer Science, Vol.-LNCS3406, pp.437-440, February 2005.

[13] T. J. Lee, H. J. Kwon, H. G. Choi, Y. S. Shin, J. G. Kim, “A Study on Endpoint Detection Method in Noise Environment”, Trans. IEEK vol. 10, no. 1, pp. 257-260, 1997.

[14] Ministry of Government Legislation, ENFORCEMENT DECREE OF THE PARKING LOT ACT [Online]. Available: http://www.law.go.kr/LSW/lsInfoP.do?lsiSeq=157062#0000

저 자 소 개

서 지 훈(Ji-Hun Seo)

2014년 상명대학교 디지털미디어학과 이학사

2014년～현재 상명대학교 컴퓨터과학과 석사과정

<주관심분야> 오디오 신호처리, 패턴인식

이 혜 인(Hye-In Lee)

2014년 상명대학교 디지털미디어학과 이학사

2014년～현재 상명대학교 컴퓨터과학과 석사과정

<주관심분야> 오디오 신호처리, 패턴인식

이 석 필(Seok-Pil Lee)

1990년 연세대학교 전기공학과 공학사

1992년 연세대학교 전기공학과 공학석사

1997년 연세대학교 전기공학과 공학박사

1997년～2002년 대우전자 영상연구소 선임연구원

2002년～2012년 KETI 디지털미디어연구센터 센터장

2010년～2011년 미국 Georgia Tech. 방문 연구원

2012년～현재 상명대학교 디지털미디어학과 교수

<주관심분야> 멀티미디어 검색, 방송통신시스템, 인공지능