

자동차 환경의 인포테인먼트 시스템을 위한 음악 검색 알고리즘

Music Search Algorithm for Automotive Infotainment System

김형국*
(Hyoung-Gook Kim)

조재만**
(Jae-Man Kim)

요약

본 논문에서는 자동차 환경의 인포테인먼트 시스템을 위한 음악 검색 알고리즘을 제안한다. 제안된 방법은 음악신호의 로그 스펙트럼기반의 정점을 이용하여 오디오지문을 추출하고, 추출된 음악 핑거프린트에 해시값을 적용하여 클라우드 서버에 저장한다. 클라우드 서버에서는 사용자의 쿼리 음악과 클라우드 서버의 해시 테이블에 저장되어 있는 오디오 지문을 비교함으로써 가장 유사한 음악이 검색된다. 제안된 음악 검색 알고리즘의 성능평가를 위해, 주행 중인 자동차 내부에서 녹음한 잡음에 노출된 다양한 쿼리 음악의 길이에 따른 검색 결과의 정확도를 측정하였고, 해시 테이블의 저장 곡수에 따른 검색 소요 시간을 측정하였다.

Abstract

In this paper, we propose a music search algorithm for automotive infotainment system. The proposed method extracts fingerprints using the high peaks based on log-spectrum of the music signal, and the extracted music fingerprints store in cloud server applying a hash value. In the cloud server, the most similar music is retrieved by comparing the user's query music with the fingerprints stored in hash table of cloud server. To evaluate the performance of the proposed music search algorithm, we measure an accuracy of the retrieved results according to various length of the query music and measure a retrieval time according to the number of stored music database in hash table.

Key words : Music Search Algorithm, Audio Fingerprint, Cloud Server, Hash Table

I. 서론

최근들어 IT 기술의 발전과 융합으로 인해, 운송

수단에 불과했던 자동차가 진화를 거쳐 문화, 생활 공간으로 거듭나게 되면서 보다 편리하고 인간 친화적인 첨단 기능들이 속속 등장하였다. 이러한 추세에 따라 첨단 기능을 가진 차량용 인포테인먼트 (infotainment) 시스템이 각광받고 있다.

* 주저자 : 광운대학교 전파공학과 부교수

** 공저자 및 교신저자 : 광운대학교 전파공학과 석사과정

† 논문접수일 : 2012년 8월 8일

‡ 논문심사일 : 2013년 1월 7일

† 게재확정일 : 2013년 1월 7일

인포테인먼트란, 운전과 길 안내 등 필요한 정보를 뜻하는 인포메이션(information)과, 다양한 인간 친화적인 기능을 말하는 엔터테인먼트의 통합시스템을 말한다. 이러한 인포테인먼트가 적용된 여러 가지 시스템은 과거 차량 내 라디오, TV등을 통해 음악 감상의 기능 만에 국한되어 있었으나, 최근 들어 음악 감상이나 TV시청 중에 흘러나오는 노래의 정보를 획득하고자 하는 음악 검색 기능 및 서비스 까지 확장되고 있다. 이러한 음악 검색 시스템은 현재 Google[1-2], Philips[3], Yahoo[4-5]등에서 PC, 스마트폰 기반으로 개발되어져 왔으며, 최근 글로벌 자동차 업체들이 본격적으로 자동차와 IT와의 결합을 꾀하면서 음악 검색 시스템을 이용한 새로운 서비스를 연구 및 개발 중에 있다. 상기 기술된 방식은 오디오 스펙트럼의 이미지 기반으로 검색하는 방식[6]으로 웨이브렛(wavelet)이나, integral image를 이용하여 adaboost filter를 적용한 32bit descriptor를 오디오지문 정보로 획득하는 방식[7]이며, 본 논문에서 제안하는 방식은 자동차 환경에서도 안정적으로 음악 콘텐츠를 효과적으로 빠르게 검색하기 위한 스펙트럼의 고 에너지 부분인 정점을 검출하여 오디오지문을 추출하는 오디오 스펙트럼 정점 기반의 음악 검색 알고리즘을 제안한다. 제안된 음악 검색 알고리즘은 오디오 지문을 이용하여 방대한 음악 콘텐츠로부터 사용자가 원하는 음악 정보를 빠르고 안정적으로 검색할 수 있는 환경을 제공한다.

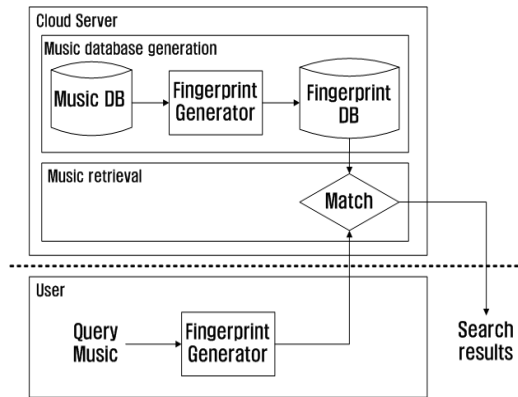
본 논문의 구성은 다음과 같다. 전체 시스템 구조와 세부 알고리즘에 대해 I장에서 설명하고, III장에서는 오디오 검색 알고리즘의 실험결과 및 고찰에 대해서 설명한다. 마지막 IV장에서는 결론을 서술한다.

II. 전체 시스템 구조

본 논문에서 제안하는 음악 검색 알고리즘의 구성도는 <그림 1>과 같다.

음악 검색 알고리즘의 구조는 크게 음악 데이터 베이스 생성부, 음악 검색부를 포함하고 있는 클라우드 서버, 그리고 사용자가 요청하는 쿼리로

부터 오디오 지문을 추출하는 사용자 질의부로 구성된다.



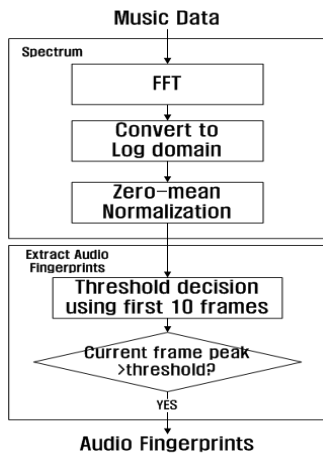
<그림 1> 음악 검색 알고리즘 구성도
<Fig. 1> A block diagram of audio searching algorithm

클라우드 서버 내의 음악 데이터 베이스 생성부는 각 음원 데이터베이스의 오디오 지문을 추출하여 획득하는 과정이며, 추출된 오디오 지문을 해시 테이블에 저장한다. 음악 데이터 베이스를 생성하는 방법은 각각의 음원 데이터 베이스에 색인정보를 부여하고, 입력된 음원 데이터 베이스의 각 음악 신호로부터 획득한 로그스케일 스펙트럼으로부터 오디오 지문을 추출한다. 그리고 추출된 오디오 지문을 해시 테이블에 저장한다. 사용자 질의부에서는 사용자가 검색을 원하는 쿼리 음악으로부터 오디오 지문을 추출하여 클라우드 서버의 음악 검색부에 전송된다. 음악 검색부에서는 사용자의 쿼리 음악의 오디오 지문과 음악 데이터 베이스 생성부에서 음원 데이터 베이스로부터 생성된 해시 테이블과의 비교를 통해 쿼리 음악과 가장 유사한 오디오 지문 정보를 가지고 있는 음악신호의 색인정보를 획득하여 사용자에게 최종 검색된 음악 정보를 제공한다. 예를 들어 사용자가 차량 내부의 라디오, 혹은 TV에서 흘러나오는 노래의 정보를 모르는 경우, 이 노래의 정보를 획득하기 위해서 차량 내 스테레오 시스템의 음악 검색 버튼을 누르게 된다. 이

때, 스테레오 시스템 내부에서 현재 라디오에서 흘러나오는 노래를 쿼리 음악으로 획득하고, 이 쿼리 음악의 오디오 지문을 추출하여 클라우드 서버로 전송한다. 쿼리 음악의 오디오 지문을 전송받은 클라우드 서버는 해시 테이블과의 비교과정을 통해 가장 유사한 음악의 정보를 획득하고, 획득한 음악의 정보를 사용자에게 전송하여 사용자는 클라우드 서버로 전송한 쿼리 음악의 정보를 획득할 수 있다.

1. 음악 데이터 베이스 생성

본 논문에서 제안하는 음악 데이터 베이스 생성부의 세부적인 구조는 <그림 2>와 같다.



<그림 2> 오디오 데이터 베이스 생성 블록도
<Fig. 2> Block diagram of audio database generation

음악 데이터 생성부는 음원 데이터 베이스로부터 음원 데이터를 입력받아 이로부터 음원별 오디오 지문을 획득하고 이를 해시 테이블에 저장하는 과정을 의미한다. 음원의 오디오 지문은 잡음 환경과 음성의 왜곡에도 강력한 검색 성능을 갖게 하기 위해 음원 데이터의 프레임별 스펙트럼을 획득하고, 획득한 스펙트럼 중에서 일정 에너지 이상의 에너지 값을 가지는 구간을 이용한다. 이는 음원의 스펙트럼에서 가장 높은 에너지를 가지는 구간은 확인되지 않은 음원이라 할지라도 서로 같은 패턴을

가지게 되므로 스펙트럼에서 가장 높은 에너지를 가지는 부분을 음원의 오디오 지문 정보라 한다.

음원의 오디오 지문을 추출하기 위해 입력된 음원 데이터는 64ms의 길이를 가지는 프레임으로 나누어져 푸리에 변환을 통해 주파수 영역으로 변환한다.

$$X_k = \sum_{n=0}^{N-1} x(n)w(n)e^{-\frac{2\pi i}{N}kn} \quad k=0, \dots, N-1 \quad (1)$$

$$S_{\log} = \log(|X_k|) \quad (2)$$

N은 Window size를 나타내며 본 논문에서는 프레임 사이즈인 64ms이며, w(n)은 Hamming window이다. 식 (1)을 통해 주파수 영역으로 변환된 음원 데이터 중에서 중복된 성분들과 종속된 성분들을 제거하고 입력된 오디오 신호의 고유한 특성을 나타내기 위해 식 (2)를 통해 로그도메인으로 변환하여 현재 프레임에서 일정 주파수 에너지 문턱값 이상인 부분을 오디오 스펙트럼의 정점 정보로 추출한다. 이 때, 오디오 스펙트럼의 정점을 추출하기 위한 문턱값은 음원 데이터의 초기 10개 프레임의 스펙트럼의 각 주파수 인덱스별로 가장 큰 값을 가지는 에너지 값을 정점 추출을 위한 문턱값이라 정의한다. 이는 음원이나 쿼리 음악이 입력되었을 경우, 특히 쿼리 음악은 잡음환경에 노출이 되어 있거나 음원으로부터의 거리를 두고 녹음하는 경우에 어떠한 일정한 문턱값을 적용하게 되면, 정점 검출이 이루어지지 않는 문제점이 발생한다. 이러한 잡음이나 음원으로부터의 거리에 따른 쿼리 음악에 왜곡이 발생하는 경우에도 각각의 쿼리 음악이 녹음된 환경에 맞게 고 에너지 구간인 정점 검출을 할 수 있도록 잡음환경과 녹음된 거리에 적응적으로 문턱값을 결정하기 위함이다. 이때 초기 10개 프레임을 이용하여 문턱값을 결정하는 이유는 다음과 같다. 10개 이상의 프레임으로 문턱값을 결정하는 경우에는 문턱값을 결정하는 10개 이상의 프레임 구간 동안 문턱값이 결정되지 않아 쿼리 음악의 초기구간에는 정점검출이 되지 않는 문제점이 발생한

다. 또한 10개 이하의 프레임을 이용할 경우에 충분하지 않는 길이로부터 문턱값을 계산하여 문턱값이 너무 높게 결정되거나 낮게 결정되는 문제로 인한 정점검출의 오류가 발생하는 문제점이 있다. 따라서 본 논문에서 제안하는 음악검색 알고리즘에서는 처음 10개 프레임을 이용하여 문턱값을 결정한다. 문턱값을 계산하는 과정은 식 (3)과 같다.

$$F_{th}(f) = \max(S_{\log}(l, f)), 1 < l < 10 \quad (3)$$

f 는 주파수 인덱스를, l 은 프레임 번호를 나타내며 F_{th} 는 계산된 정점 검출 문턱값이다. 이렇게 추출된 정점들 중 시간축을 기준으로 가장 인접해있는 2개의 정점을 하나의 pair로 정의하여 해시 테이블에 저장 가능한 형태의 오디오 지문 정보를 획득한다. 오디오 지문은 각 오디오 지문이 위치하는 시간축 정보와, 주파수축 정보인 Start Time(St), Frequency start time(Fst), Frequency End Time(Fet), Delta time(Dt)와 같은 4가지 정보를 가지고 있다.

2. 해시 테이블 저장

추출된 오디오 지문을 해시 테이블에 저장하는 구조는 <그림 3>과 같다.

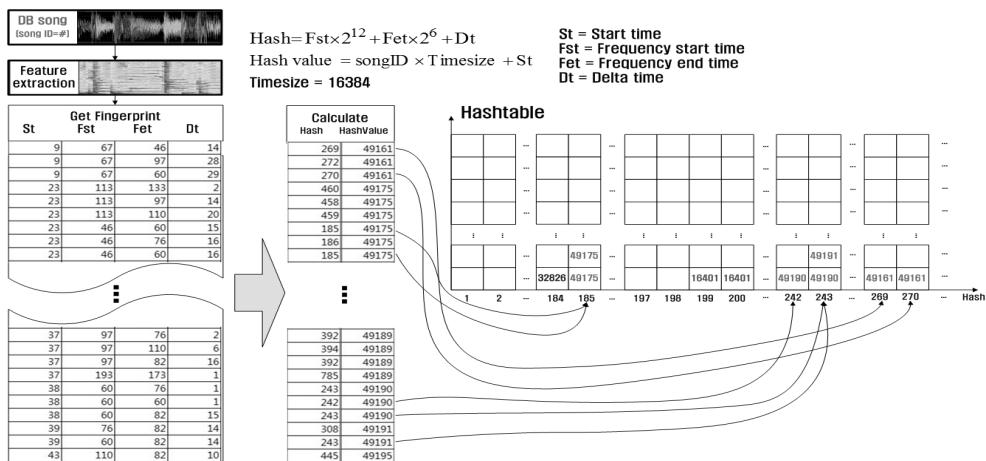
오디오 지문을 해시 테이블에 저장하기 위해서 추출된 오디오 지문의 4가지 위치 정보를 하나의 값으로 변환하는 과정이 요구되며 변환은 식 (4)과 식 (5)를 통해 변환한다.

$$Hash = Fst \times 2^{12} + Fet \times 2^6 + Dt \quad (4)$$

$$Hashvalue = songID \times Time\ size + St \quad (5)$$

여기서 해시($Hash$)는 해시 테이블의 데이터가 저장되는 주소를 의미하며 해시값($Hashvalue$)은 오디오 지문을 하나의 값으로 표현한 고유한 값이다. $Time\ size$ 는 상수이며 St , Fst , Fet , Dt 는 추출된 핑거프린트의 값이다. 이 때, Fet 가 2^6 인 64 이상의 값을 가지는 경우 Fet 에 2^6 을 빼줌으로서 64미만의 값을 가지게 하여 Fet 가 St, Fst, Dt 와 상호 이격이 되는 최종 해시 계산을 수행한다.

식 (4)을 통해 계산된 해시는 <그림 3>의 해시 테이블에서 식 (5)를 통해 계산된 해시값을 저장하는 주소가 되며, 해시테이블 내의 해시 주소의 위치에 해시값을 저장한다. 오디오 지문의 값에 따라서 해시 테이블의 동일한 위치에 서로 다른 오디오 지문이 저장되는 경우도 발생하나, 이미 해시테이블의 주소에 오디오 지문이 저장되어 있는 경우에는 동



<그림 3> 오디오 지문 해시 테이블 저장 구조도
 <Fig. 3> A diagram of audio fingerprint save in hash table

일 주소의 다음 배열에 오디오 지문정보를 저장하여 다른 음원으로부터 동일한 오디오 지문이 계산된 경우에도 해시테이블에 모두 저장한다. 이를 음악검색부에서 주소에 해당하는 모든 저장된 오디오 지문정보를 획득하여 오디오 색인과정을 수행하며 자세한 음악검색 과정은 다음 단락에서 설명한다.

3. 음악 검색부

음악 검색부는 음악 데이터 베이스 생성부에서 생성된 오디오 지문을 저장한 클라우드 서버의 해시 테이블에서 사용자가 검색을 원하는 쿼리 음악을 비교하여 검색결과를 제공하며 구조는 <그림 4>와 같다. 사용자가 검색을 원하는 쿼리 음악이 입력되면 음악 데이터 생성부의 오디오 지문 생성과정과 마찬가지로 오디오 데이터를 로그 스케일 스펙트럼으로 변환하고, 쿼리 음악의 오디오 지문을 계산한다. 이렇게 계산된 쿼리 음악의 오디오 지문과 클라우드 서버의 해시 테이블에 저장되어 있는 음원 데이터베이스의 오디오 지문 정보와 비교를 하여, 쿼리 음악과 가장 유사한 구간을 검색한다.

이터 생성부에서는 해시 테이블에 저장을 위한 해시값을 계산하나 음악 검색부는 쿼리 음악을 해시 테이블내에서 검색만을 위한 과정이기 때문에 해시 테이블에 저장하는 해시값은 계산하지 않는다.

해시 테이블 내에서 쿼리 음악의 해시와 동일한 주소에 위치하는 해시값은 식 (5)의 과정을 반대로 적용한 식 (6)과 식 (7)을 이용하여 오디오 색인정보(songID, St)를 획득하고 해시와 함께 비교 결과 배열에 저장한다.

$$songID = Hashvalue / Time\ size \quad (6)$$

$$St = Hashvalue - songID \times Time\ size \quad (7)$$

비교 결과 배열에 저장된 오디오 색인정보(songID)중 가장 많은 개수로 획득된 오디오 색인정보를 최종 검색 결과로 사용자에게 제공한다.

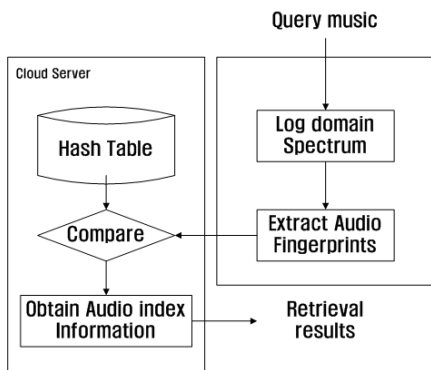
III. 실험 결과 및 고찰

1. 실험 방법

본 논문에서 제안하는 음악 검색 알고리즘의 성능을 측정하기 위해 실험을 다음과 같이 수행하였다.

- 테스트 실험군: 실험을 위해 사용한 음원 데이터 베이스는 가요 2000곡을 사용하였다. 음원 데이터 베이스의 평균 노래 길이는 3분 30초이며, 샘플링 레이트는 16000Hz이며, query 오디오는 5초, 7초, 10초, 15초의 길이를 사용하였다. 자동차 주행 환경에서의 실험 결과의 획득을 위해 클린 환경의 쿼리 음악 외에 실제 자동차의 고속도로와 일반도로 주행 시, 차량의 스피커에서 흘러나오는 노래를 삼성 갤럭시S3, iPhone4S, 삼성 갤럭시노트를 이용하여 녹음한 쿼리 음악의 검색 정확도 측정을 통해서 자동차 주행 환경에서의 성능을 측정하였다.

- 구동 환경: 본 실험은 CPU(Core i5, 2.8GHz), RAM 4GB의 사양을 갖춘 Windows 7 운영체제에서 음악 검색 알고리즘의 검색 정확도를 측정하였다.



<그림 4> 음악 검색부 구조도
<Fig. 4> A block diagram of music searching

쿼리 음악과 유사 구간 검출을 위해서는 쿼리 음악의 오디오 핑거프린트 정보는 음악 데이터 저장부의 핑거프린트 계산과정과 동일하게 식 (1)를 거쳐 해시 테이블의 주소인 해시를 계산한다. 음악 데

2. 실험 결과

음악 검색 성능 측정은 20000곡의 음원 데이터 베이스를 저장한 해시 테이블에서 쿼리 음악의 길이, 노출된 자동차 주행 환경잡음에 따라 녹음된 쿼리 음악 5000개의 검색 정확도를 측정하였다.

또한, 음원 DB를 1000곡, 5000곡, 10000곡, 20000곡으로 늘려가면서 검색에 소요되는 시간 측정을 하였다. <표 1>은 쿼리 음악의 길이에 따른 검색 정확도 결과를 나타낸다.

<표 1> 자동차 주행환경에서 녹음된 쿼리 음악 길이에 따른 검색정확도 실험결과

<Table 1> Accuracy of the experimental results of various length query recorded in the car driving environment.

Length of query music	Retrieval accuracy (expressway)	Retrieval accuracy (street)
5초	89.2%	89.3%
7초	95.7%	95.0%
10초	99.9%	99.9%
15초	100%	100%

<표 1>에 따르면 본 논문에서 제안된 방식은 쿼리 음악의 길이가 짧아짐에 따라서 검색 정확도가 감소하는 것을 알 수 있다. 그러나 제안된 방식은 7초 이상의 쿼리 음악의 실험결과에서 95%이상의 우수한 안정된 검색성능 결과를 획득하였다.

<표 2>는 해시 테이블에 저장하는 음원 데이터 베이스의 곡수에 따라 소요되는 검색 시간을 나타낸다.

<표 2> 저장 음원 데이터 베이스 수에 따른 쿼리 음악 검색 시간

<Table 2> Query retrieval time according to the number of stored music DB

Number of stored DB	Retrieval time
1000곡	1200ms
5000곡	1205ms
10000곡	1208ms
20000곡	1212ms

<표 2>는 저장DB 곡의 개수에 따른 검색소요시간을 측정한 결과이다. 측정은 저장 음원 데이터 베이스의 곡의 개수에 따라 각각의 환경에 5000개의 쿼리 음악을 검색하였을 때 검색 소요시간의 평균을 측정하였다. 본 방식은 쿼리 음악으로부터 추출된 오디오 지문을 이용하여 해시 테이블의 주소인 해시를 계산함으로써 쿼리 음악의 검색을 위해 전체 해시 테이블을 검색하지 않고 쿼리 음악의 오디오 지문으로부터 계산된 해시에 의한 검색만을 하여 저장 데이터 베이스의 곡수가 늘어나더라도 검색시간이 크게 증가하지 않는 것을 나타낸다.

IV. 결 론

본 논문에서는 자동차 환경의 인포테인먼트 시스템을 위한 안정적인 오디오 검색 알고리즘을 제안하였다. 제안된 알고리즘은 자동차 환경에서 안정적인 검색 성능으로 사용자에게 오디오 검색 서비스를 제공할 수 있다고 판단된다.

참고문헌

- [1] S. Baluja, "Audio fingerprinting: combining computer vision & data stream processing," *Acoustics, speech and signal processing*, vol. 2, pp.213-216, Apr. 2007.
- [2] Jaap Haitsma, A. Kalker, "A highly robust audio fingerprinting system," *International symposium on music information retrieval (ISMIR 2002)*, pp.107-115, Paris, France, Oct. 2002.
- [3] L. Wang, "An industrial-strength audio search algorithm," *International Symposium on Music Information Retrieval (ISMIR 2003)*, pp.7-13, Baltimore, USA, Oct. 2003.
- [4] J. Bernard, G. Abby, "Searching for multimedia: analysis of audio, video and image web queries," *World Wide Web*, vol. 3, pp.249-254, Dec. 2000.
- [5] Y. Li, Y. Hou, "Search audio data with the wavelet pyramidal algorithm," *Information processing Letters*, vol. 91, pp.49-55, July 2004.

- [6] S.Baluja, and M. Covell, "Waveprint: Efficient wavelet-based audio fingerprinting," *Pattern Recognition*, vol. 41, pp.3467-3480, Nov. 2008.
- [7] B.-J. Jeong, and D.-J. Kim, "Energy and statistical filtering for a robust audio fingerprinting system," *The Korea Contents Association(한국콘텐츠학회 논문지)*, vol. 12 pp.1-9, May 2012.
- [7] B.-J. Jeong, and D.-J. Kim, "Energy and statistical

저자소개



김 형 국 (Kim, Hyoung-Gook)

2007년 3월 ~ 현재 : 광운대학교 전자융합공학과 부교수
2005년 4월 ~ 2007년 2월 : 삼성종합기술원 수석연구원
2002년 8월 ~ 2005년 3월 : 독일 베를린 공과대학교 Assistant Professor
1999년 1월 ~ 2002년 7월 : 독일 SIEMENS/Cortologic AG 책임연구원
e-mail : hkim@kw.ac.kr



조 재 만 (Cho, Jae-Man)

2011년 3월 ~ 현재 : 광운대학교 전자융합공학과 석사과정
2011년 : 광운대학교 전자공학과 공학사
e-mail : coolshut@gmail.com