대용량 음성인식을 위한 하이브리드 빔 탐색 방법과 가변 플로링 기법을 이용한 고속 디코더 알고리듬 연구

Fast Decoder Algorithm Using Hybrid Beam Search and Variable Flooring for Large Vocabulary Speech Recognition

김 용 민* · 김 진 영* · 김 동 화** · 권 오 일***
Yong Min Kim · Jin Young Kim · Dong Hwa Kim · Oh Il Kwon

ABSTRACT

In this paper, we implement the large variable vocabulary speech recognition system, which is characterized by no additional pre-training process and no limitation of recognized word list. We have designed the system in order to achieve the high recognition rate using the decision tree based state tying algorithm and in order to reduce the processing time using the gaussian selection based variable flooring algorithm, the limitation algorithm of the number of nodes and ENNS algorithm. The gaussian selection based variable flooring algorithm shows that it can reduce the total processing time by more than half of the recognition time, but it brings about the reduction of recognition rate. In other words, there is a trade off between the recognition rate and the processing time. The limitation algorithm of the number of nodes shows the best performance when the number of gaussian mixtures is a three. Both of the off-line and on-line experiments show the same performance. In our experiments, there are some differences of the recognition rate and the average recognition time according to the distinction of genders, speakers, and the number of vocabulary.

Keywords: Variable Vocabulary, Gaussian Selection Based Variable Flooring Algorithm, Hybrid Beam Search, Limitation Algorithm of the Number of Nodes.

1. 서 론

음성 인식에 대한 연구는 동적인 프로그램을 이용하여 입력 음성과 기준이 되는 학습 음성을 비교하여 최소 거리를 갖는 음성을 입력 음성으로 인식하는 DTW(dynamic time warping)[1], 인간의 신경세포 구조와 기능을 모델화하여 사용하는 신경회로망 인식[2], 그리

* 전남대학교 일반대학원 전자공학과, RRC HECS
** 밀양대학교 정보통신공학과
*** (주) 현대오토넷

고 음성을 상태 천이 확률 및 각 상태에서의 출력의 관찰 확률을 갖는 마코프 과정으로 가정한 후 학습을 통하여 확률을 구하고 인식에는 학습된 확률 정보를 이용하는 은닉 마코프 모델(HMM: hidden Markov model) 등을 사용하여 많은 연구들이 행하여져 오고 있다[3]. HMM은 파라미터 모델로서 단어 단위뿐만 아니라 음절, 음소 단위의 모델링도 가능하고 또 높은 인식률을 보임으로써 현재 가장 널리 쓰이는 인식 알고리듬으로써 본 논문에서 구현한 인식 시스템에도 이 HMM에 기반한 HTK(HMM Tool Kit)[4]를 이용하여 학습하였으며 HMM 알고리듬에 기반하여 인식 시스템을 구축하였다. 본 논문에서는 실시간으로 대용량 단어 인식기를 구현하는데 필요한 가변 어휘 음성 인식기를 구현하고 탐색 시간 단축 알고리듬에 대한 연구를 하였다. 대용량 어휘 인식기에서는 탐색 시간이 빠르고, 잡음에 강인해야 되고, 데이터 갱신이 용이해야 하며, 인식률이 높아야 되며, 여러 사용자가 사용할 수 있어야 한다. 이러한 조건들을 해결하기 위해서 본 논문에서는 결정 트리 기반 상태 공유 알고리듬[5]과 가우시안 셀렉션[6]을 구현하고 이를 인식기에 적용 시켜 인식률 향상과 가변어휘 인식, 인식 시간 단축을 가능하게 하였다. 그리고 화자 독립 인식을 할 수 있도록 하여 사용자의 음성에 대한 학습과정이 없어도 인식할 수 있도록 하였다. 본 논문의 구성은 다음과 같다. 2 장에서는 구현된 인식 디코더 및 실험에 사용한 음성 DB에 대한 설명이고, 3 장에서는 평균인식시간을 단축시키는 고속화 알고리듬에 대해서 설명했고, 4 장에서는 인식실험에 대해서 그리고 마지막 5 장에서 결론을 맺었다.

2. 시스템 개요 및 음성 데이터 베이스

2.1 구현된 인식 시스템

본 논문에서 개발한 인식 시스템은 대용량 음성 인식을 위해서 HMM 음향 모델링 기반으로 구축하였고 대용량의 가변어휘 또한 인식시킬 수 있는 인식기이며 전화 다이얼링, 인터넷 웹브라우저, 자동차 항법 장치 등 여러 응용 분야로 사용될 수 있는 시스템이라고 말할 수 있다.

인식 시스템 구동에 대해서 설명을 하면 먼저 메모리를 초기화하고 캘리브레이션 과정을 거치어 사용자가 발성한 음성이 입력장치를 통해서 들어오면 음성검출부(끝점검출부)에서 에너지와 ZCR을 가지고 시작점 및 끝점을 검출 한 다음 전처리부에서 파라미터의 특징을 추출하여 저장한 후 인식부에서 Viterbi 빔 탐색을 이용하여 학습데이터와 확률비교를 하여 가장 큰 확률을 갖는 순서로 단어를 3 개까지 디스플레이하고 음성 입력 대기 상태로 된다. 그리고 WLSET 버튼 제어부는 인식 대상어휘를 변경할 때 사용한다.

이제 단어를 인식시키기 위해 인식기 디코더에서 전처리부와 인식에 필요한 DB 파일을 올리기 위한 메모리를 초기화하는 과정을 거친 후 모든 DB 파일을 읽어들이고 동적 메모리를 할당하고 인식기의 창을 띄운다. 그리고 캘리브레이션 및 음성검출부(끝점검출부)에서 음성신호 입력을 받기 전에 비음성 구간에서 에너지와 ZCR을 계산하는 캘리브레이션 과정을 거치고 음성을 입력받아 시작점 및 끝점 검출을 하고 신호를 디스플레이하고 전처리부를 지나서 인식부로 들어가 최적의 단어를 디스플레이하고 다시 입력음성 대기 상태로 돌아온다.

그런 다음 전처리부에서 mel-cepstrum과 pitch 검출을 비롯하여 여러 가지 특징 파라미터를 추출하여 인식부에 넘기고 인식부에서는 전처리부에서 파라미터 특징을 추출하여 동적 메모리로 할당해 놓고 데이터 수(프레임 수)를 인식부로 넘겨서 가우시안 셀렉션을 하고 Viterbi 빔 탐색을 하여 인식 대상 어휘를 출력하는 과정이고 다시 입력 음성 대기 상태로 돌아간다.

별도의 장치를 두어 가변어휘 인식을 하였다. 이곳은 학습에 참여하지 않는 데이터 즉, 가변어휘를 음운변동 과정을 거쳐 트라이폰을 추출하고 트라이폰에 대한 상태 정렬과 천이가 정의된 파일을 얻게 된다. 여기서 적용된 알고리듬이 결정 트리 상태 공유 알고리듬이다. 그래서 입력은 가변어휘 텍스트이고 출력은 상태정렬과 천이가 정의된 파일이다.

그림 1은 본 논문에서 개발한 인식 디코더의 전체 블록 다이어그램이다.

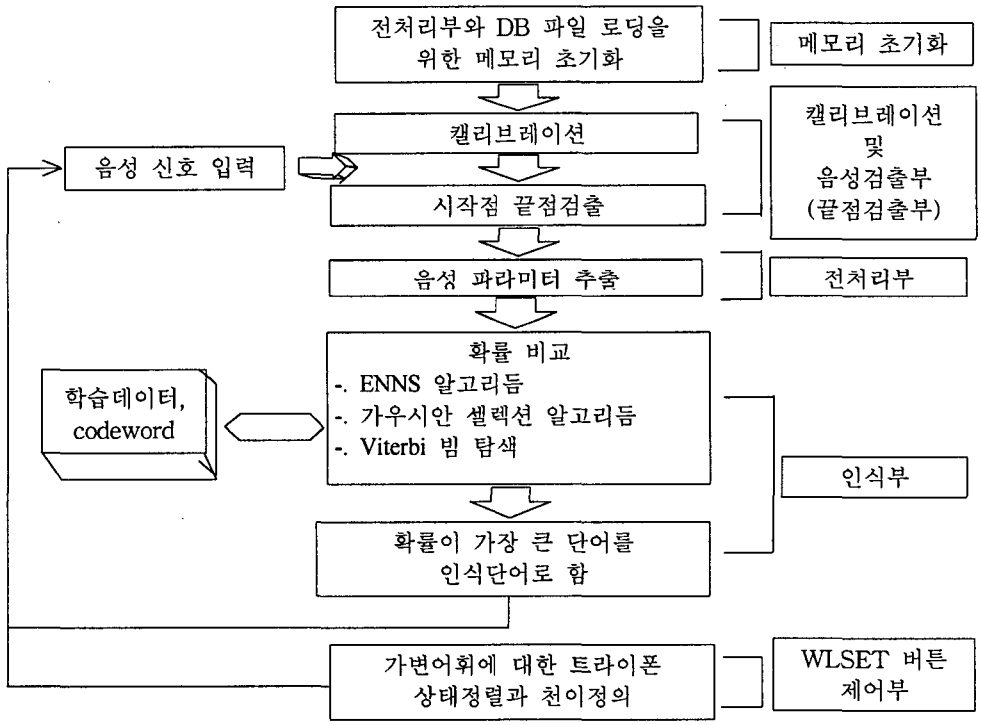

그림 1. 인식 시스템 전체 구성도

2.2 음성 데이터 베이스

본 논문에 사용한 음성 DB는 지역별(서울, 부산, 인천, 대구, 대전, 광주), 성별(남, 여), 연령별(20~50 대) 화자가 지명 또는 상호명 등 총 5,084 단어(1 set)에 대해 발음한 음성 데이터이다. 학습에 사용한 데이터는 지역별로 남자는 총 50 set 중 44 set이고, 여자는 총 30 set 중 24 set이다. 나머지 남·여 각각 6 set씩은 인식에 사용된 데이터이다. 녹음 환경은 잡음이 없는 사무실에서 하였으며 마이크는 콘덴서 마이크를 사용하였다. 음성의 특징 파라미터는 에너지, ZCR, 피치(pitch) 주기, formant, 단구간 spectrum, Filter Bank 출력, LPC 계수(Linear Prediction Coefficient), cepstrum 계수 등 다양하다. 현재 많이 이용되고 있는 특징

파라미터로는 귀의 비 선형적인 특성을 고려해 Mel-scale로 warping시킨 Mel-scale cepstrum 과 이것의 시간적인 변화를 나타내는 delta-cepstrum 등이다. 본 논문에서는 Mel-cepstrum 과 Normalized log energy, delta-cepstrum, delta energy 등을 특징 파라미터로 사용하였다. 이러한 파라미터들에 대한 조건은 표 1에 나타내었다.

표 1. 음성 데이터 및 파라미터 조건

sampling frequency	8 kHz
resolution	16 bits
hamming window	window size: 25 ㎳ shift size: 10 ㎳
filter bank channel number	26
liftering number	22
cepstrum number	12
feature parameter	12 order MFCC+1order energy+12 order delta-cepstrum+1 order energy → 26 order

이러한 음성 DB를 기반으로 HTK 학습을 하였고 다음과 같은 결과를 내었다. 표 2는 트라이폰 단위 학습 결과를 나타내었다.

표 2. HTK 학습 결과 (5,084 단어)

성 별	mixture 수 (단위: 개)	재추정 임계치	총 mixture 수 (단위: 개)	총 상태 수 (단위: 개)
남 자	5	150 50 500	46,955	9,391
	3	150 50 500	28,173	9,391
여 자	5	150 50 500	35,755	7,151
	3	150 50 500	21,453	7,151

표 3은 1set이 1,074 단어인 총 70 set에 남자 음성 DB에 대한 HTK 학습 후 인식결과를 표 4와 인식대상 어휘수에 따른 인식시간을 비교하기 위해 나타낸 것이다.

표 3. HTK 인식 결과 (1,074 단어)

인식률/인식시간 / mixture 수/ viterbi 빔 폭	남 자	
	인식률(%)	개당 인식시간 (초) (CPU 속도)
3 mixtures / 60	93.38	0.8 (PⅢ 600 MHz)

표 4에서는 본 논문에서 사용한 음성 DB에 대한 트라이폰 단위 HTK 인식 결과를 표로

나타내었다. 여자 학습 데이터가 남자 학습 데이터보다 더 적기 때문에 다소 인식률이 남자 데이터에 비해 저조함을 알 수 있다.

표 4. HTK 인식 결과 (5,084 단어)

성별/인식률 / mixture 수/ viterbi 빔 폭	남 자		여 자	
	인식률(%)	개당 인식시간 (초) (CPU 속도)	인식률(%)	개당 인식시간 (초) (CPU 속도)
3 mixtures / 60	93.19	8 (PIII 600 MHz)	91.92	7.8 (PIII 600 MHz)
5 mixtures / 60	93.35	2.32 (PIII 1.1 GHz)	91.95	2.4 (PIII 1.1 GHz)

3. 고속화 알고리듬

HTK 인식 결과에서 나타난 것과 같이 인식 대상 어휘가 증가하면 증가할수록 인식 속도가 느려짐을 알 수 있었고 cpu 속도 또한 인식 시간에 영향을 준 것으로 나타났다. 이런 이유로 인하여 대용량 음성 인식기에서는 인식률은 변화시키지 않고 탐색 시간만을 단축시키는 알고리듬이 꼭 필요하다. 그래서 본 논문에서는 하이브리드 빔 탐색, 가우시안 셀렉션 기반의 가변 flooring 기법, ENNS 알고리듬[7] 등을 사용하여 인식시간을 단축했다.

3.1 하이브리드 빔 탐색

탐색 공간이 넓으면 넓을수록 탐색의 정확도는 증가하지만 계산량이 너무 많아 비효율적인 탐색이 되기 쉬우므로 본 논문에서는 Viterbi 빔 탐색 방법을 사용하였다. Viterbi 빔 탐색은 매 프레임에서 모든 후보 경로들을 계산하지 않고 확률이 가장 높은 후보들만을 계산한다. 또 프레임이 계속 진행될수록 우도(likelihood) 값들 사이의 차이가 커지므로 일정 임계치 이상 값은 잘라(Pruning)낸다. 이런 탐색 방법도 노드 수나 임계치에 따라 탐색 시간이 차이가 나는 문제점이 있어서 노드 수나 임계치를 가변하는 방법인 하이브리드 빔 탐색 방법을 본 논문에서는 적용했다.

노드 수를 제한하는 방법과 임계치 제한 방법을 조합해서 빔 탐색을 하므로 하이브리드 빔 탐색이라고 하며, 계산해야 할 노드 수를 줄이기 위해서 적당한 임계치 노드수를 정하여 인식시간을 단축시키고자 하는 방법이다. 즉, 목표 노드수가 임계치 노드수보다 크면 Viterbi 빔 임계치를 계산할 때 빔 폭을 조정할 수 있는 빔 폭 임계치만큼 뺀 값을 대입하여 계산한다. 임계치 제한 기법은 빔 폭 임계치를 얼마로 하느냐에 따라서 빔 폭이 결정이 되어 인식률과 평균 인식시간에 영향을 미치므로 빔 폭 임계치를 튜닝해서 가장 좋은 값을 선택하게 된다.

아래에 본 연구에서 적용한 하이브리드 빔 탐색 알고리듬에 대해 나타내었다.

```
Algorithm :
        for i=1 to nData do;
            if ntargetNode > ThreshNumNode then;
                threshold -= DeltaThreshold;
            end;
        end;
알고리듬 :
        프레임 개수만큼 for loop 실행;
            만약 목표 노드 수가 임계치 노드 수보다 크면 viterbi 임계치는
                그 임계치에서 빔 폭 임계치를 뺀 값을 대입;
            종료;
        종료;
```

3.2 기존의 가우시안 셀렉션 알고리듬

대용량 음성인식에서 가장 큰 문제는 인식시에 많은 데이터들을 탐색해서 결과를 나타낼 때까지의 시간이다. 특히 실시간 음성 인식기에서 인식 시간은 매우 중요한 요소이다. 그러므로 대용량 어휘 인식을 실시간으로 구현하기 위해서, 본 논문에서는 가우시안 셀렉션 알고리듬을 사용하여 인식시간을 줄이고자 하였으며, 이 알고리듬은 Bocchieri[8]에 의해서 처음 제안되었다. 가우시안 셀렉션은 학습 중에 음향적인 영역을 벡터 양자화된 영역의 집합으로 나눈다. 즉, 각각의 가우시안들은 하나의 codeword 또는 그 이상의 codeword에 속하게 되고 이들의 집합이 각 codeword에서의 short list이다. 가우시안 셀렉션은 학습 후 나온 가우시안들을 벡터 양자화한 결과를 가지고 시행한다. 그림 2에 벡터 양자화 과정과 codeword들의 집합인 codebook의 생성과정을 나타냈다.

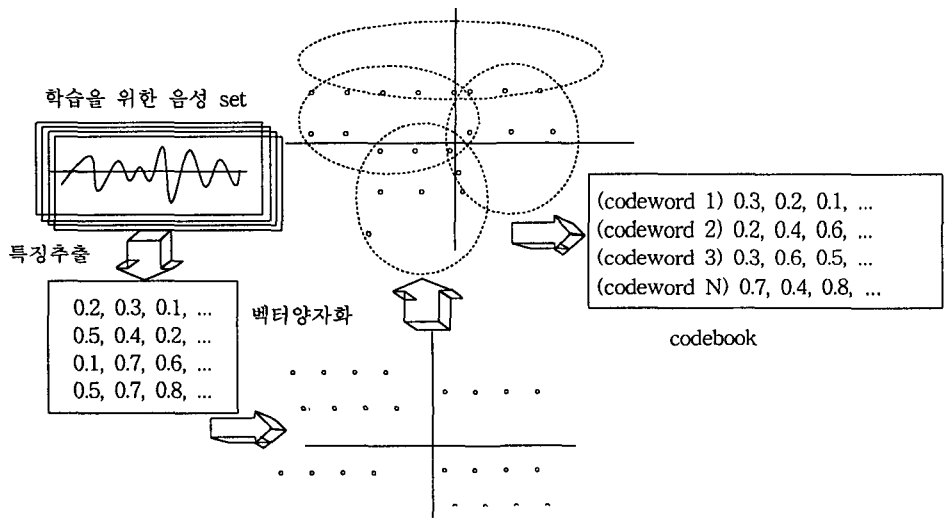

그림 2. 벡터양자화 과정 및 codebook 생성과정

그림 2는 표준 패턴을 만들기 위해서 모아진 음성 set으로부터 특징 추출 과정을 거쳐 벡

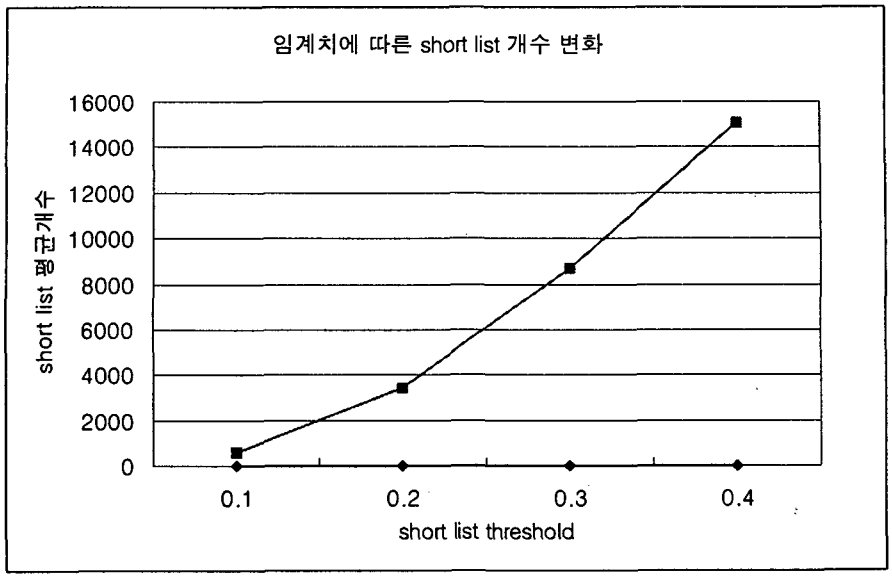

그림 4. 임계치 변화에 따른 short list 평균 개수

3.2.2 제안한 가우시안 셀렉션 기반의 가변 flooring 기법

본 논문에서 제안한 가우시안 셀렉션 기반의 알고리듬인 가변 flooring 기법에 대한 설명이다. HTK 학습 후 가우시안들의 분산이 너무 크게 공유되기 때문에 이 분산의 범위를 좁히기 위해서 아래와 같이 short list의 임계치를 가변하는 방법을 사용하여 확률 계산에 이용했다. 가우시안 셀렉션 사용시 확률 계산을 할 때 관측 벡터와 mixture 사이의 거리는 관측 벡터와 가장 가까운 codeword 1과의 거리, 두 개의 short list 의 codeword 1과 codeword 2 간의 거리, codeword 2와 mixture와의 거리를 모두 합한 것을 근사화한 것으로 계산해서 사용했다. 아래 그림 5는 앞에서 설명한 가변 flooring 기법에 관한 그림이다.

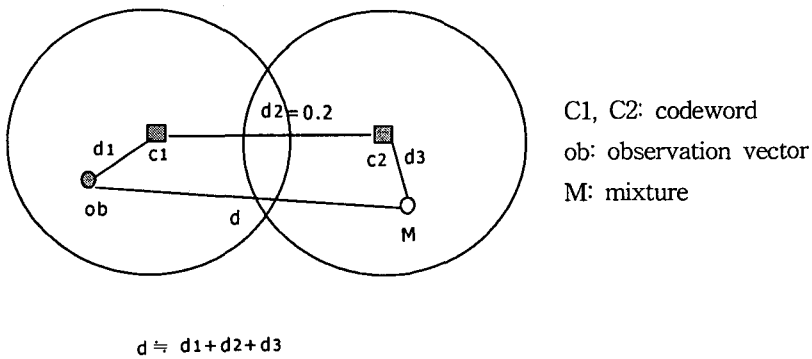

그림 5. 관측 벡터와 가우시안 mixture 사이의 거리계산

광주시 동 이름 100 개를 가지고 가변 flooring 기법에서 관측벡터와 가우시안 mixture사이의 거리 계산 실험을 했을 때 총 평균 계산 수가 134,916 번이었는데 알고리듬이 사용되어

져 계산되었을 때가 평균 98,068 번이었고 바로 계산되어졌을 때가 평균 36,848 번이었다. 그래서 알고리듬을 이용했을 때 가우시안 분산의 범위를 좁힐 수 있었다.

그림 6과 7은 학습 DB 5084 단어 중 표본 1,000 단어에 대한 codeword들 사이의 short list의 임계치의 변화에 따른 평균 인식 시간과 인식률을 그림으로 나타낸 것이다.

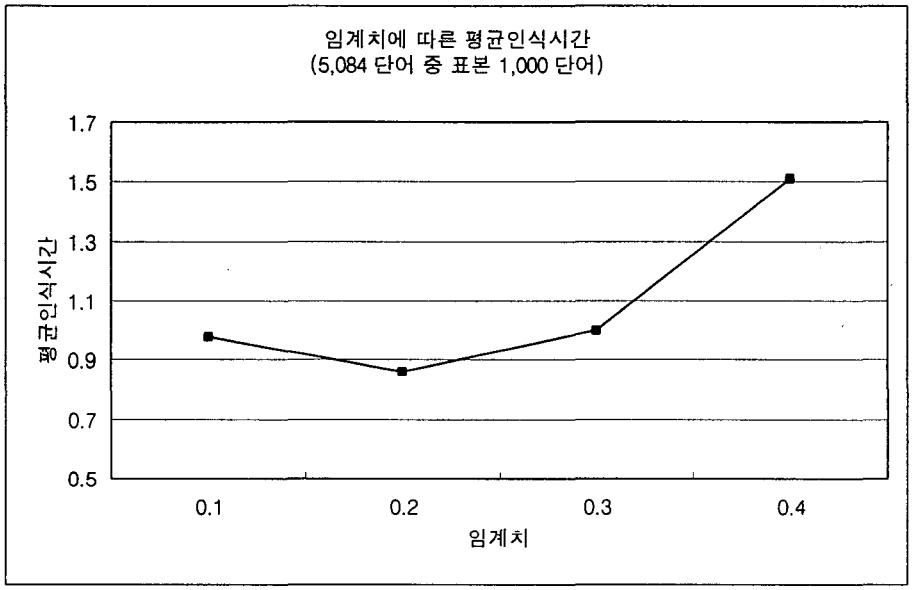

그림 6. short list의 임계치 변화에 따른 평균 인식 시간

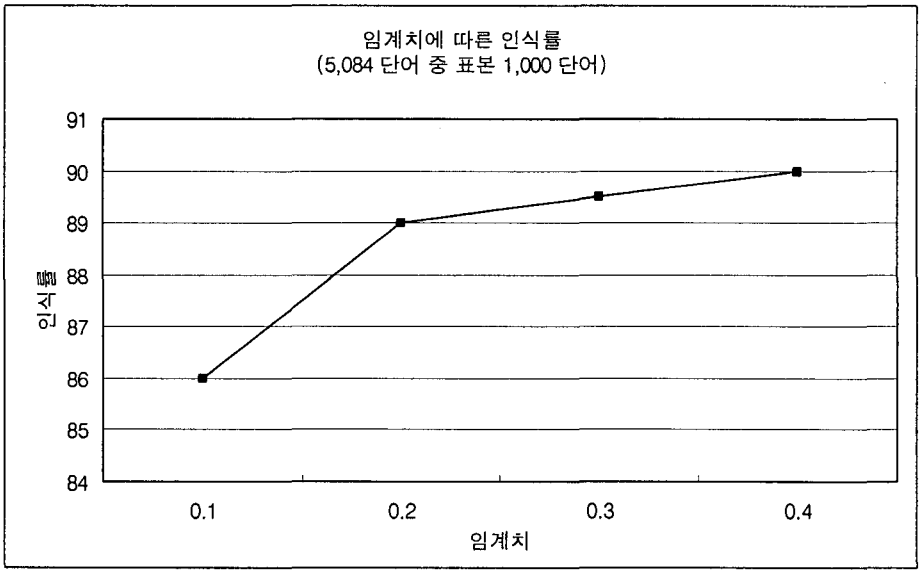

그림 7. short list의 임계치 변화에 따른 인식률

4. 실험 및 결과

본 논문에서 5,000 단어급과 하이브리드 빔 탐색 파라미터 튜닝에 대한 실험은 학습화 된 데이터 5,084 단어 중 1,000 단어에 대해서 조용한 실험실 환경과 펜티엄 컴퓨터 700 MHz급에서 오프라인으로 시행하였다. 그리고 인식에 사용한 데이터는 남자 데이터이며 학습에 참여하지 않은 인식용 데이터 6 set 중 1 set을 사용하였으며 제안한 가우시안 셀렉션 기반의 가변 flooring 기법을 사용했을 때(GS)와 사용하지 않았을 때(NGS), mixture 수, 임계치 노드 수(ThreshNumNode), 빔 폭 임계치(DeltaThreshold)를 변화시키면서 실험을 하였다.

4.1 5,000 단어급 실험

표 5의 specification은 mixture 수는 5 개, Viterbi 빔 임계치는 60, 임계치 노드 수는 5,000 개, 빔 폭 임계치를 20으로 고정시켰고, 제안한 가우시안 셀렉션 기반의 가변 flooring 기법은 1,024 개의 codeword와 codeword들 사이의 short list 임계치는 0.2로 short list 개수를 약 4,000 개를 적용하여 실험을 한 것이다.

표 5. 제안한 가우시안 셀렉션 기반의 가변 flooring 기법 미사용시(NGS)와 사용시(GS)의 비교(총 5,084 단어 중 표본 1,000 단어)

NGS & GS	average time(초)	recognition rate(%)
NGS	3.20	88.7
GS	1.75	88.6

4.2 하이브리드 빔 탐색 파라미터 튜닝 실험

그림 8과 9에 대한 specification은 Viterbi 빔 임계치는 60, 임계치 노드 수를 5,000으로 고정했고, 빔 폭 임계치에 따른 학습 데이터의 mixture 개수가 3인 경우와 5인 경우에 대한 평균 인식시간과 인식률을 나타냈다.

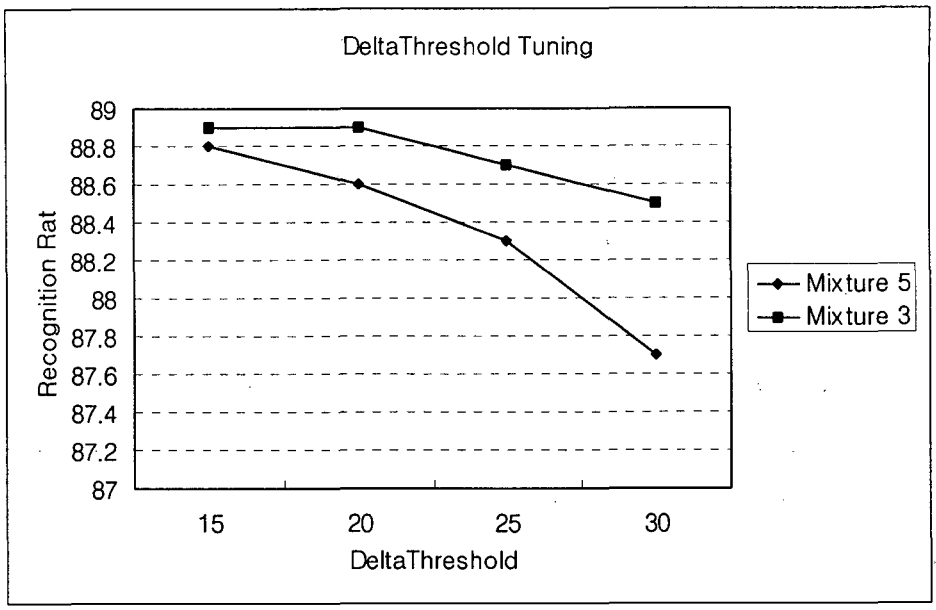

그림 8. 빔 폭 임계치에 따른 인식률 비교

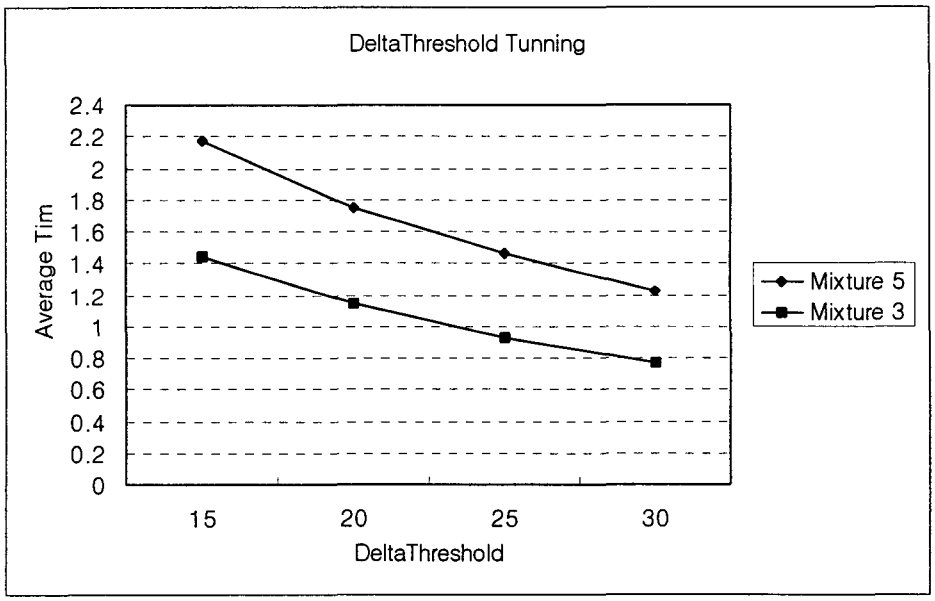

그림 9. 빔 폭 임계치에 따른 평균 인식시간 비교

그림 10과 11에 대한 specification은 Viterbi 빔 임계치는 60, 빔 폭 임계치를 20으로 고정시키었고, 임계치 노드 수에 따른 학습 데이터의 mixture 3개 와 5개에 대한 평균 인식시간과 인식률을 나타냈다.

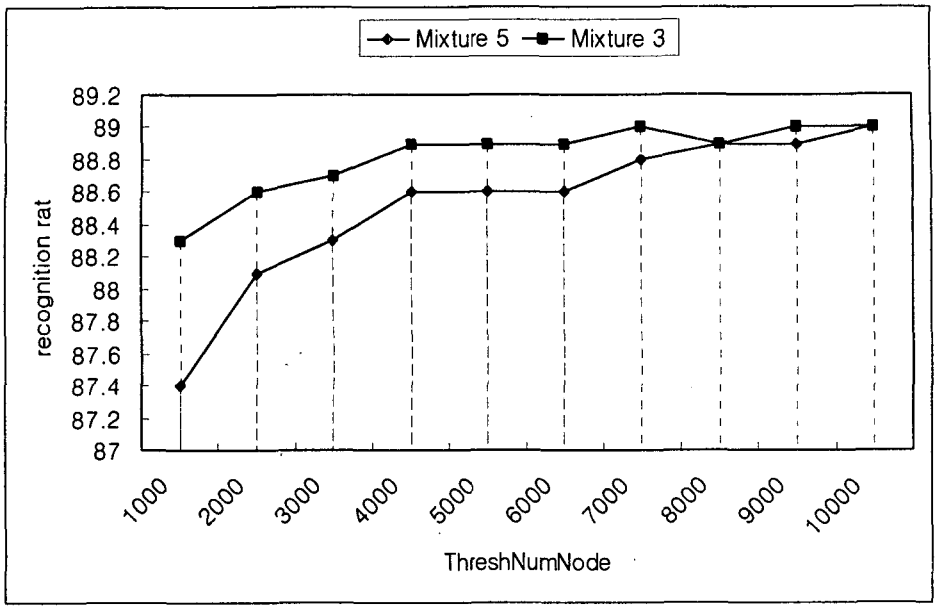

그림 10. 임계치 노드 수에 따른 인식률 비교

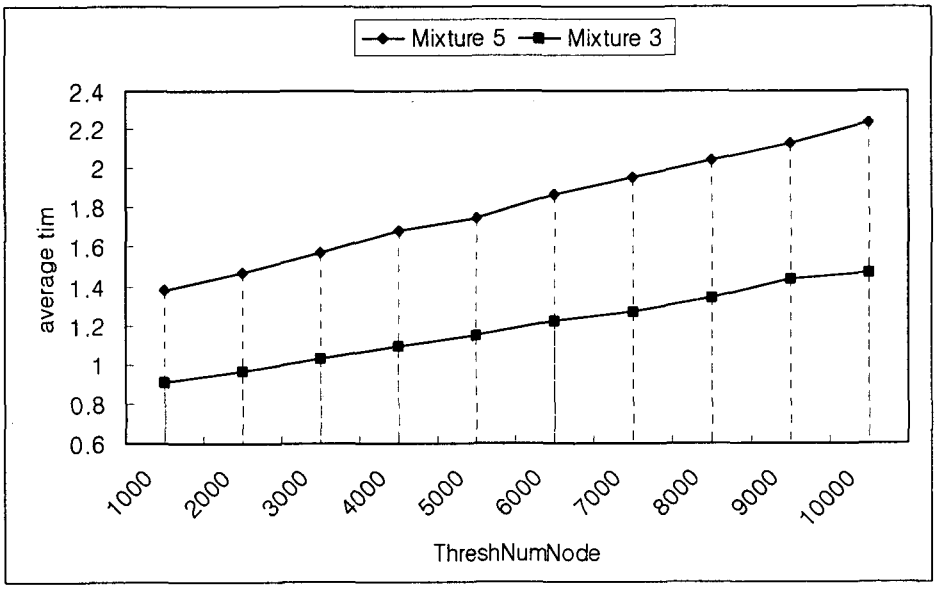

그림 11. 임계치 노드 수에 따른 평균 인식시간 비교

위 실험들에서 mixture 수가 5 개보다 mixture 수가 3 개가 성능이 좋고, 임계치 노드 수가 4,000일 때가 성능이 제일 좋은 걸로 나타났다. 그래서 제일 성능이 좋은 빔 폭 임계치를 찾아내기 위해 파라미터 튜닝에 대한 마지막 튜닝 결과를 그림 12과 13에서 나타내었다.

이에 대한 specification은 Viterbi 빔 임계치를 60, 임계치 노드 수를 4,000으로 고정시키었고, 빔 폭 임계치에 따른 학습 데이터의 mixture 3 개에 대한 평균 인식시간과 인식률을

나타냈다.

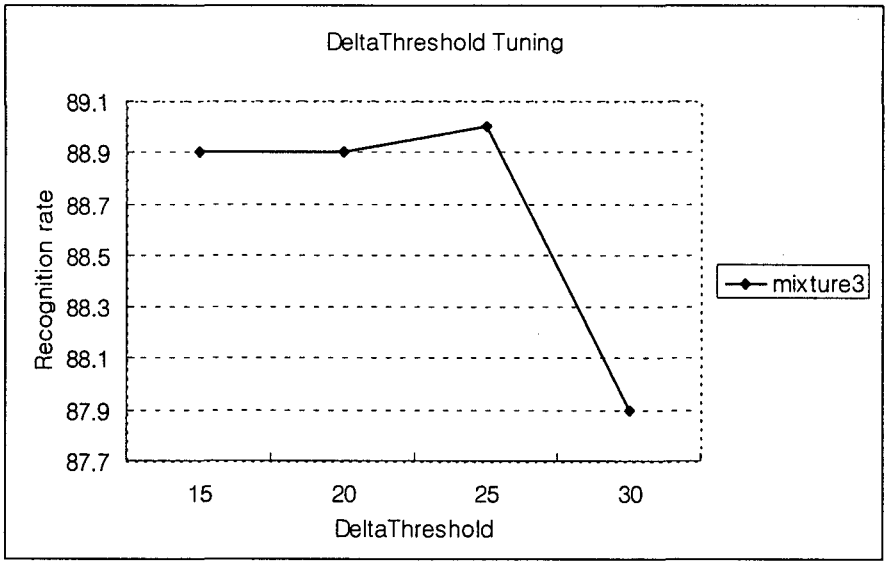

그림 12. 튜닝 데이터를 이용한 빔 폭 임계치에 따른 인식률 비교

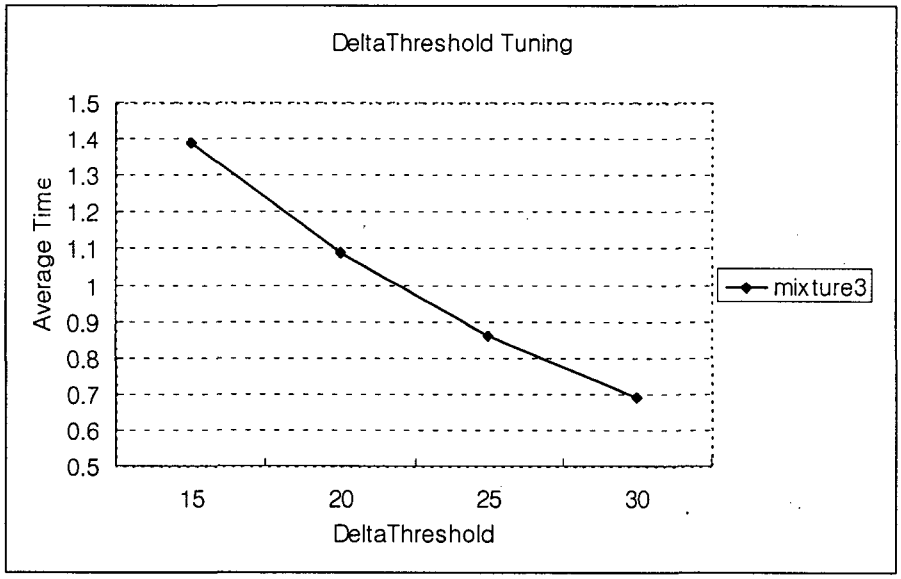

그림 13. 튜닝 데이터를 이용한 빔 폭 임계치에 따른 평균 인식시간 비교

하이브리드 빔 탐색의 파라미터 튜닝에 대한 최종 결과는 다음과 같으며 mixture 수가 3개일 때가 인식시간과 인식률이 보다 더 좋은 결과임을 알 수 있었고, 이 파라미터를 최종 선정했다.

Mixture 5 개 Specification

인식률: 88.6%, 평균 인식시간: 1.75 초

제안한 가우시안 셀렉션 기반의 가변 flooring 기법 사용, Viterbi 빔 임계치: 60.

임계치 노드 수: 5,000, 빔 폭 임계치: 20.

Mixture 3 개 Specification

인식률: 89%, 평균 인식시간: 0.86 초

제안한 가우시안 셀렉션 기반의 가변 flooring 기법 사용, Viterbi 빔 임계치: 60.

임계치 노드 수: 4,000, 빔 폭 임계치: 25.

4.3 20,000 단어급 가변어휘 실험

위 4.2 절에서 선정된 값을 기반으로 가변어휘 실험을 하였다. 본 논문에서 사용한 가변어휘 인식 실험은 두 가지로 나누어서 시행하였다. 첫째, 광주광역시 지명 100 개를 선정하여 자체 구현된 인식 시스템에 포팅 시키고 학습에 참여하지 않은 20 대 화자 5 명(남자 4 명, 여자 1 명)이 실시간 인식 실험을 하였다.

그리고 두 번째로는 한국지명(일명 POI data) 20,000 단어를 선정하여 인식 시스템에 포팅 시키고 표본으로 1,000 단어를 선택해서 학습에 참여하지 않은 20 대 남자 화자 1명이 인식 실험을 하였고 아울러 여러 가지 제약 조건으로 인하여 표본으로 200 단어를 선택해서 학습에 참여하지 않은 20 대 남녀 화자 2 명씩이 실시간 인식 실험을 하였다. 표 6은 첫 번째 실험에 대한 화자별에 따른 제안한 가우시안 셀렉션 기반의 가변 flooring 기법을 사용하지 안 했을 때(NGS)와 사용했을 때(GS)를 단어 하나 당 평균 인식 시간과 인식률로서 비교한 것이다.

표 6. 화자별에 따른 평균인식시간 및 인식률 비교 (100 단어)

NGS & GS / 화자별 분류	NGS		GS	
	average time(초)	recognition rate(%)	average time(초)	recognition rate(%)
화자 1(남자)	0.19	95	0.13	93
화자 2(남자)	0.17	94	0.13	94
화자 3(남자)	0.17	95	0.12	92
화자 4(남자)	0.19	94	0.12	93
화자 5(여자)	0.16	91	0.11	89

그리고 표 7은 표 6의 화자 1(남자)이 20,000 단어 중에서 표본 1,000 단어에 대한 실험 결과이다.

표 7. 2 만 단어급에서 표본 1,000 단어 인식 결과

NGS & GS	average time(초)	recognition rate(%)
NGS	3.31	88.2
GS	2.10	87.4

표 8에서는 마지막으로 20,000 단어 중에서 표본 200 단어에 대한 실험 결과를 나타내었다. 위 실험에서는 부정확한 발음으로 잘못 인식이 된 단어에 대해서는 다시 한번의 실험을 더 하였다. 표에서 보는 봐와 같이 제안한 가우시안 셀렉션 기반의 가변 flooring 기법을 사용하지 안 했을 때(NGS)와 사용했을 때(GS)의 평균 인식 시간과 인식률의 차이를 볼 수 있을 것이다.

표 8. 화자별에 따른 평균인식시간 및 인식률 비교(20,000 단어 중 표본 200 단어)

NGS & GS / 화자별 분류	NGS		GS	
	average time(초)	recognition rate(%)	average time(초)	recognition rate(%)
화자 1(남자)	3.31	88	1.48	83
화자 2(남자)	2.98	90.5	1.85	88
화자 3(여자)	2.54	87	1.52	84
화자 4(여자)	2.48	85.5	1.45	83.5

5. 결 론

본 논문에서는 대용량 음성인식기의 탐색시간 단축에 초점을 두었다. 대용량 음성인식기를 구현하는데는 인식 대상 어휘와 인식시간 등을 고려하여야 한다. 이러한 문제들을 해결하기 위해서 본 논문에서는 결정 트리 기반 상태 공유 알고리듬을 사용하여서 인식률 향상과 어휘 변환 문제를 해결하였고 제안한 가우시안 셀렉션 기반의 가변 flooring 기법과 노드수 제한 알고리듬을 사용하여서 인식 시간 단축을 시도하였는데 제안한 가우시안 셀렉션 기반의 가변 flooring 기법은 short list의 임계치(Θ) 값에 따라 인식 속도의 변화가 많았으나 속도 증가에 따른 인식률 저하 또한 그에 비례하였다. 또 본 인식 시스템으로 5,000 단어급에서는 학습에 참여한 단어의 오프라인 인식 실험을 시행하였고 20,000 단어급에서는 학습에 참여하지 않은 가변어휘를 여러 화자별로 실시간 온라인 인식 실험을 시행하였다. 오프라인 실험과 온라인 실험에서 인식 결과는 거의 비슷한 양상을 보였다. 그리고 화자별, 제안한 가우시안 셀렉션 기반의 가변 flooring 기법을 사용한 경우와 사용하지 않은 경우, 인식 대상 어휘 수에 따라 인식률과 평균 인식시간의 차이를 보였다. 또한 남자 화자보다는 여자 화자의 인식 결과가 좋지 않음을 볼 수 있었다. 따라서, 본 논문에서 고속화 알고리듬을 사용한 인식 시스템은 인식시간과 인식률이 trade-off 관계임을 알 수 있었다. 본 논문의 학습 음성 데이터가

수집할 때에 발음사전에 기초하여 정확한 발음으로 녹음이 되었는지에 대한 제대로 검증이 되지 않아 인식률 저하를 초래하였다. 더 좋은 인식 성능을 위해서는 좀더 많은 시간을 투자하여 음성 데이터의 확실한 검증 후 재학습을 한다면 본 논문의 인식성능보다 개선되리라 생각된다. 향후에 더욱더 연구해야 할 것은 현재 음성의 시작 부분의 무성음을 제대로 검출하지 못해 인식 성능에 다소 영향을 미치는 경우가 있는데 보다 나은 끝점 검출기의 안정화가 필요하겠다.

또 몇몇 가변 어휘 중 학습 데이터의 트라이폰이 존재하지 않아 인식이 되지 않는 어휘가 생기는 경우가 있는데 이는 몇몇 단어들에 대해서 재학습을 통해 가능한 한 인식이 되지 않는 단어가 생기지 않도록 해야 하겠다.

그리고 탐색시간 단축을 위해 제안한 가우시안 셀렉션 기반의 가변 flooring 기법을 사용했지만 인식대상 어휘가 늘어나면 늘어날수록 탐색 시간이 길어지는 경우를 보완하기 위해 보다 더 견고한 탐색 시간 단축 알고리듬의 연구가 필요하겠다.

참 고 문 헌

[1] Sakoe, Hiroaki. & Seibi Chiba. 1978. "Dynamic Programming Algorithm Optimization for Spoken Word Recognition." *IEEE Trans. On Acoustic Speech and Signal Processing,* vol. 1.

[2] Waibel, A., H. Sawai & K. Shikano. 1989. "Modularity and Scaling in Large Phonemic Neural Networks." *IEEE Trans. On Acoustic Speech and Signal Processing,* Dec.

[3] Rabiner, L. R. & B. H. Juang 1986. "An Introduction to Hidden Markov Models" *IEEE Acoustic Speech and Signal Processing Magazine,* Jan.

[4] Young, Steve., Julian Odell., Dave Ollason., Valtcho Valtchev. & Phil Woodland. 1997. *"The HTK Book (for HTK Version 2.1)",* Entropic Cambridge Research Laboratory Ltd, Cambridge, U.K.

[5] 김동화. 1999. *연속 음성인식을 위한 향상된 결정 트리 기반 상태공유 기법 연구.* 박사학위논문. 부산대학교 전자계산학과.

[6] Knill, K. M., M. J. F. Gales & S. J. Young. 1996. "Use Of Gaussian Selection In Large Vocabulary Continuous Speech Recognition Using HMMs." *IEEE ICSLP'96*

[7] 백성준. 1999. *벡터 부호화를 위한 고속 탐색 알고리듬.* 박사학위논문. 서울대 전기공학부.

[8] Bocchieri, E. 1993. "Vector quantization for efficient computation of continuous density likelihood." *In Proc. ICASSP,* Volume II.

[9] 서봉수. 2001. *가변어휘 음성인식기 구현 및 탐색 시간 단축 알고리듬 비교.* 석사학위논문. 전남대학교 전자공학과.

접수일자: 2001. 10. 28
게재결정: 2001. 12. 6

▲ 김용민
광주광역시 북구 용봉동 300 (우: 500-757)
전남대학교 전자공학과 신호처리실험실 석사과정
Tel: +82-62-530-0472 Fax: +82-62-530-0472
E-mail: ymkim@dsp.chonnam.ac.kr

▲ 김진영
광주광역시 북구 용봉동 300 (우: 500-757)
전남대학교 전자공학과 부교수, RRC HECS
Tel: +82-62-530-1757 Fax: +82-62-530-0472
E-mail: kimjin@dsp.chonnam.ac.kr

▲ 김동화
경남 밀양시 내이동 1025-1 (우: 627-130)
밀양대학교 정보통신공학과 부교수
Tel: +82-55-350-5461 Fax: +82-55-350-5460
E-mail: dhkim@arang.miryang.ac.kr

▲ 권오일
경기도 이천시 부발읍 아미리 산 136-1 (우: 467-860)
(주) 현대오토넷 차장
Tel: +82-31-639-7817 Fax: +82-31-639-7820
E-mail: koi@haco.co.kr