부 벡터 군집화를 통한 FPGA기반 음성 인식기의 수행 속도 향상

최정욱, 성원용
서울대학교 전기공학부
e-mail : jwchoi@dsp.snu.ac.kr, wysung@snu.ac.kr

Execution Speed Improvement of FPGA-based Speech Recognizer using Sub-vector Clustering

JungWook Choi and Wonyong Sung
School of Electrical Engineering
Seoul National University

Abstract

Reducing the memory bandwidth is very critical for real-time implementation of a large vocabulary speech recognizer. In this paper, we implemented sub-vector clustering algorithm in hardware to alleviate the problem of much extreme memory access. The experimental result shows that this implementation accelerates acoustic likelihood computation twice faster, and increases the speed of overall recognition by 22%.

I. 서론

음성 인식은 가장 자연스러운 인터페이스로써 휴대폰, 네비게이션 등의 휴대용 기기에 널리 사용되고 있다. 휴대용 기기를 위한 대용량 음성 인식기(Large Vcabulary Continuous Speech Recognizer, LVCSR)는 제한된 하드웨어를 이용하여 많은 연산을 실시간으로 수행해야 하므로 효율적인 연산을 위한 최적화 알고리즘의 적용이 필수적이다.

본 논문은 참고논문 [1]의 부 벡터 군집화 알고리즘을 적용 하여 FPGA(Field Programmable Gate Array) 기반 하드웨어 음성 인식기[2]의 연산 처리 성능을 향상시키고자 하였다. 본 논문의 구성은 다음과 같다. 2장에서는 FPGA 기반 음성 인식기의 구조를 간단하게 살펴본다. 3장에서는 부 벡터 군집화 알고리즘의 구현에 대해 알아본다. 4장에서는 부 벡터 군집화로 인한 성능 향상 결과를 제시한다. 마지막으로 5장에서 본 논문의 연구 성과를 정리한다.

이 논문은 지식경제부 출연금으로 ETRI와 시스템반도체산업진흥센터에서 수행한 ITSoC 핵심설계인력양성사업과 교육과학기술부의 재원으로 한국학술진흥재단에서 수행하는 BK21 프로젝트의 지원을 받아 수행된 연구입니다.

II.음성 인식기의 구조

본 논문에서 쓰인 음성 인식기는 문맥 의존적 은닉마르코프모델(context dependent Hidden Markov Model, HMM)을 이용해 입력된 음성에 대한 연속어 음성 인식(continuous speech recognition)을 수행한다.

음성 인식기는 매 10ms마다 30ms 길이의 음성 파형을 입력으로 받는다. 음성 특징 추출 장치(feature extraction unit)는 입력으로 들어온 음성 파형에 대해 MFCC(Mel-Frequency Cepstral Coefficient)를 계산하여 39차원의 특징 벡터(feature vector)를 구한다. 음향적 유사성 계산 장치(acoustic likelihood computation unit)는 음향 모델(acoustic model)의 가우시안 분포(Gaussian density)와 입력으로 받은 특징 벡터 사이의 유사도(likelihood)를 계산한다. 이 유사도는 특정 음향 상태(HMM state)에서 현재 음성 입력이 발생할 확률을 의미한다. 최적 상태열 검색 장치(Viterbi search unit)는 비터비 검색을 이용하여 음향 상태의 유사도와 언어모델로 구성된 인식 네트워크를 검색하고 최적의 문자열을 기록한다. 이 문자열은 역추적 과정을 통해 재구성되어 인식 결과 문장으로서 출력된다.

음향적 유사성 계산의 경우 음향 모델을 구성하는 가우시안 분포들에 대해 단순 유사도 계산을 반복한다. 가우시안 분포들을 메모리로부터 읽어 오는데 필

요한 메모리의 대역폭은 한정되어 있으므로 최적화된 수의 가우시안 분포만을 연산하도록 하는 알고리즘이 필요하다. 본 논문에서는 가우시안 분포를 차원 단위로 나누어 군집화 하는 방법[1]을 하드웨어 적으로 구현하여 음향적 유사성 계산의 수행시간을 줄였다.

Ⅲ. 부 벡터 군집화

참고논문 [1]에서 제시된 부 벡터 군집화는 39차의 가우시안 분포들을 차원별로 묶어서 부 벡터로 만든 후 군집화를 통해 중심 값을 얻는 방법이다. 부 벡터 군집화의 경우 군집화를 하는 단위가 작기 때문에 군집화를 통한 에러가 적게 된다. 즉, 부 벡터 군집화는 메모리로부터 읽어오는 가우시안 분포의 수를 줄임으로써 음향적 유사성의 계산 속도를 높이면서 군집화로 인한 에러를 작게 유지할 수 있는 방법인 것이다.

부 벡터 군집화 알고리즘은 다음과 같다. 먼저, 전체 N개의 가우시안 분포들로 구성된 본 부호록(original codebook)을 일정한 차수로 묶는다. 이때 한 묶음 속의 부호(codeword)를 부 벡터라 부르며, 부 벡터의 차수는 전체 차수(D=39)를 묶음의 수(K)로 나눈 값이 된다. 그 후, 각 묶음마다 N개의 부 벡터를 M개의 그룹으로 군집화(clustering)하고 각 그룹마다 중심값(centroid)을 결정한다. 이렇게 생성된 값들을 군집화된 부호록(clustered codebook)이라고 부른다. 마지막으로 각 음향 상태 마다 군집화 된 부호록의 중심값으로 색인(index)을 연결한다.

부 벡터 군집화 알고리즘의 하드웨어 구조는 다음과 같다. 총 M*K개의 가우시안 분포로 이루어진 군집화된 부호록은 외부 메모리에 저장되어있다. 입력 특징 벡터가 들어오면 군집화 된 부호록을 이용해 부 벡터의 음향적 유사성을 계산하여 내부 메모리에 저장한다. 그 후, 총 N*K개의 색인을 외부 메모리로부터 읽어와 모든 음향 상태에 대한 음향적 유사성을 계산한다. 이 경우 메모리 접근의 총 횟수(M*K+N*K)는 부 벡터 군집화 알고리즘을 적용하지 않았을 때의 메모리 접근 횟수(N*D)보다 적다. 즉, 부 벡터를 나누는 정도(K)와 군집화 정도(M)가 작을수록 필요한 메모리 대역폭이 감소하는 것이다. 본 논문에서는 K=39, M=256, N=62896(=7862*8)를 사용하였다.

Ⅳ. 실험 결과

4.1 실험 환경

본 논문에서 사용된 음성 인식기[2]는 자이링스(Xilinx)사의 ML402 검증 보드에서 구현된 것이다. 음성 인식기의 동작 주파수는 100MHz이다.

음성 인식에 사용된 음향 모델은 화자 독립적인(speaker independent) 월 스트리트 저널 1 전집을 데이터를 이용하여 오픈소스 음성 인식 툴킷인 HTK로 훈련(trained)한 것이다. 음향 모델은 7,862개의 음향 상태로 구성되어 있으며 각 상태는 8개의 가우시안 분포로 이루어져 있다. 바이그램(bi-gram) 언어모델을 사용하였으며, 월 스트리트 저널 5000단어 연속어 음성 인식 태스크(task)로 음성 인식을 테스트 하였다.

4.2 연산 수행 시간

음성 인식의 단계별로 걸린 수행 시간을 측정하여 표.1에 나타냈다. 부 벡터 군집화 알고리즘으로 K=39, M=256을 사용한 결과 음향 유사성 계산이 약 2.06배 빨라졌다. 전체 수행속도 실시간에 대비 1.35배 에서 1.73배로 약 22% 빨라졌다.

	부 벡터 군집화 비적용 (Mcycle/sec)	부 벡터 군집화 적용 (Mcycle/sec)
부호록 계산	0	0.06
음향 유사성 계산	**29.9**	**14.5**
최적 상태열 검색	44.2	44.2
수행 속도향상	**1.35**	**1.73**

표.1. 부 벡터 군집화 알고리즘의 적용 유무에 따른 음성 인식 수행시간의 변화.

4.3 음성 인식 오류 비율

부 벡터 군집화 알고리즘을 적용 유/무에 따른 음성 인식 오류 비율은 각각 8.69와 8.74%였다. 이 사실로 미루어볼 때 부 벡터 군집화 알고리즘은 군집화로 인한 에러가 실제 음성 인식 오류에 영향을 크게 미치지 않음을 알 수 있다.

Ⅴ. 결론

본 논문에서는 FPGA기반 음성 인식기의 연산 속도를 향상시키고자 부 벡터 클러스터링 알고리즘을 하드웨어로 구현하였다. 그 결과 음향 유사성 계산 단계에서 약 2배의 속도 향상을 거두었으며, 전체 음성 인식의 속도도 실시간에 비해 1.73배 빠르게 수행할 수 있었다.

참고문헌

[1] M. Ravishankar 외, “Sub-vector clustering to improve memory and speed performance of acoustic likelihood computation”, Proceedings of Eurospeech, pp. 151-154, 1997.

[2] Y. Choi 외, “FPGA-based implementation of a real-time 5000-word continuous speech recognizer”, Proceedings of 16th European Signal Processing Conference, 2008.