자율주행 차량을 위한 음성 명령 인식 시스템

박현묵, 최원석, 최성곤*

충북대학교

qga0809@chungbuk.ac.kr, choiws@cbnu.ac.kr *choisg@cbnu.ac.kr

Voice Command Recognition System for Autonomous Vehicles

Park Hyeon Muk, Choi Won Seok, Choi Seong Gon*

Univ.

요 약

본 논문은 자율주행 차량 환경에서 운전자의 음성 명령을 인식하고 화자 인증을 수행하는 FPGA 기반 음성 인식 시스템을 제안한다. 클라우드 기반 음성인식의 지연 및 보안 문제를 해결하기 위해 Raspberry Pi에서 MFCC(Mel-Frequency Cepstral Coefficients) 특징을 추출하고, Zynq-7020 FPGA에서 DTW(Dynamic Time Warping) 기반 템플릿 매칭을 수행하는 엣지 처리 구조를 설계하였다. 실험 결과, 화자 인증 정확도 95.0%와 명령어 인식 정확도 91.4%를 달성하여 제안 시스템의 유효성을 검증하였다.

Ⅰ. 서 론

자율주행차 기술은 자율/협력 주행, 교통 안전, 인포테인먼트 서비스에 활용하기 위해 활발히 연구되고 있다. 특히, 운전자 인터페이스 기술 중 음성 명령 인터페이스는 운전자가 핸들에서 손을 떼지 않고 차량 기능을 제어할 수 있어 안전성 측면에서 핵심 기술로 주목받고 있으며, 최근에는 네트워크 지연 문제를 해결하기 위한 엣지 AI 기반의 음성인식 시스템에 대한 연구가 활발히 진행되고 있다.[1][2]

차량용 음성인식 시스템 관련 연구로 S. Bae 등은 FPGA에서 MFCC 전처리와 이진화 신경망 기반 키워드 스팟팅 시스템을 구현하여 저전력 음성 명령 인식을 달성하였다.[3] 하지만, 화자 인증 기능이 없어 누구의 음성이든 인식하는 화자 독립 시스템으로 보안성이 부족하다. Y. S. Chong 등은 LSTM과 MFCC를 결합한 2.5μW 키워드 스팟팅 엔진을 제안하여 IoT 환경에서의 초저전력 음성인식을 달성하였다.[4] 하지만, 화자 인증 기능 없이 키워드 인식만 수행하여 보안 응용에는 한계가 있다. T. W. Chen 등은 엣지 디바이스에서 구현 가능한 경량 화자 인증 모델을 제안하여 1.31M 파라미터로 1.76% EER을 달성하였다.[5] 하지만, 음성 명령 인식 기능 없이 화자 인증만 수행하여 차량 제어 시스템으로의 확장이 필요하다. O. Ozturk 등은 엣지 AI 디바이스에서의 음성인식 플랫폼을 비교 분석하여 클라우드 접근이 불가능한 환경에서의 최적화 방안을 제시하였다.[6] 하지만, MCU 기반 구현으로 FPGA 기반의 하드웨어 가속 효과는 분석하지 않았다.

따라서, 본 논문은 자율주행 차량에서의 음성 명령 인식과 운전자 본인 인증을 동시에 수행하는 FPGA 기반 DTW 시스템을 제안한다. 본 논문은 Raspberry Pi에서 MFCC 특징을 추출하고, Zynq-7020 FPGA에서 DTW 템플릿 매칭을 수행하는 구조를 구현하여 화자 인증 및 명령어 인식 정확도를 검증하였다.

Ⅱ. 본론

본 논문에서는 음성 명령 인식과 화자 인증을 동시에 수행하는 FPGA 기반 DTW 시스템의 구조를 설명한다.

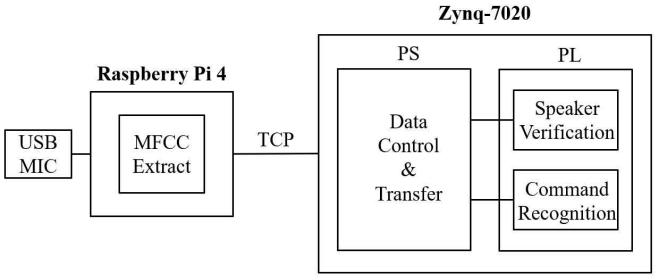

그림 1. 제안하는 FPGA 기반 시스템 구조

그림 1은 제안하는 FPGA 기반 화자 인증 및 음성 명령 인식 시스템의 전체 구조를 보여준다. 시스템은 음성 수집 및 특징 추출부와 DTW 매칭부로 구성되며, 화자 인증과 명령어 인식의 2단계로 동작한다. 음성 수집 및 특징 추출부는 Raspberry Pi 4에서 USB 마이크를 통해 16kHz, 16-bit PCM 음성을 수집하고, MFCC 특징을 추출한다. MFCC 추출은 Pre-emphasis, Framing(25ms, 10ms shift), Hamming Window, 512-point FFT, 40개 Mel Filter Bank, Log, DCT를 거쳐 프레임당 13개의 계수를 생성한다. 추출된 MFCC는 이더넷을 통해 ZedBoard로 전송된다.

DTW 매칭부는 ZedBoard(Zynq-7020)에서 수행된다. 1단계 화자 인증에서 사용자는 웨이크업 워드 "Alice"를 발화하고, 시스템은 등록된 화자의 템플릿과 DTW 매칭을 수행하여 본인 여부를 판단한다. DTW 거리가 임계값 이하이면 인증에 성공하고, 초과하면 비인가 사용자로 거부된다. 2단계 명령어 인식에서 인증된 사용자는 5개의 명령어("가자", "오케이", "아니", "올려", "내려") 중 하나를 발화하고, 시스템은 각 명령어 템플릿과의 DTW 거리를 비교하여 가장 유사한 명령어를 인식한다.

DTW 매칭은 Zynq-7020의 PL에서 Vitis HLS를 사용하여 하드웨어로 구현하였다. PS와 PL 간의 데이터 전송은 AXI 인터페이스를 통해 수행되며, PS는 lwIP 스택을 사용하여 TCP 통신과 UART 출력을 담당한다. 화자 인증 성능 평가를 위해 등록 화자의 음성 20개에 노이즈를 추가하여 100개로 증강하였고, 미등록 화자 테스트를 위해 TTS로 다양한 음성 100

개를 생성하였다. 명령어 인식 성능 평가를 위해 각 명령어당 20개의 음성을 녹음하고 노이즈 증강을 통해 명령어당 100개, 총 500개의 테스트 데이터를 생성하였다.

Ⅲ. 실험 및 결과

본 절에서는 제안 시스템의 화자 인증 및 명령어 인식 성능을 검증한다. 그림 2는 실험 환경을 보여준다.

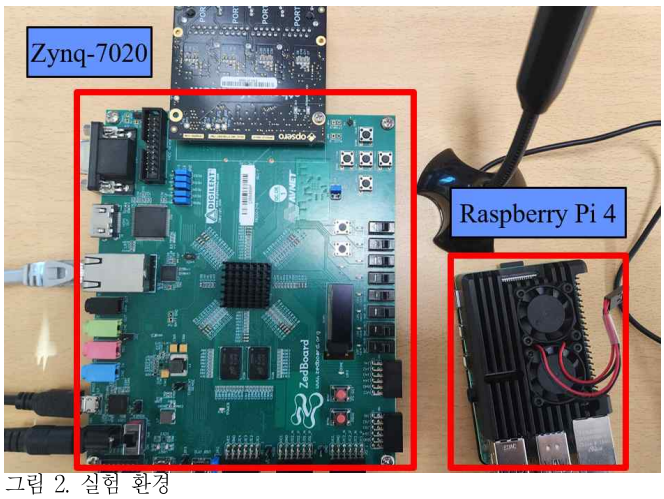

그림 2. 실험 환경

실험 환경은 Raspberry Pi 4에서 MFCC 특징을 추출하고, ZedBoard(Zynq-7020)의 PL에서 DTW 매칭을 수행하는 구조로 구성하였다. Raspberry Pi 4에 USB 마이크를 연결하여 음성을 수집하고, 이더넷을 통해 ZedBoard로 MFCC 데이터를 전송한다. ZedBoard는 UART를 통해 PC에 인식 결과를 출력한다.

실험은 세 가지 항목으로 구성된다. 첫째, 등록 화자 인증 정확도 측정을 위해 등록 화자의 웨이크업 워드 "Alice" 음성 20개에 노이즈를 추가하여 100개로 증강한 후 인증 성공률을 측정하였다. 둘째, 미등록 화자 거부율 측정을 위해 Edge TTS를 사용하여 9종의 서로 다른 한국어 음성으로 "Alice"를 생성하고, 총 100개의 테스트 데이터에 대해 거부율을 측정하였다. 셋째, 명령어 인식 정확도 측정을 위해 5개의 명령어 각각에 대해 20개의 음성을 녹음하고 노이즈 증강을 통해 명령어당 100개, 총 500개의 테스트 데이터를 생성하여 인식 정확도를 측정하였다.

Test Item	Result
Speaker Verification (Registered)	95.0%
Unregistered Rejection Rate	92.0%
Keyword Recognition Accuracy	91.4%

그림 3. 실험 결과

그림 3은 실험 결과를 보여준다. 등록 화자 인증 정확도는 등록 화자가 웨이크업 워드를 발화했을 때 시스템이 정상적으로 인증한 비율로, 100개의 테스트 데이터 중 95개가 인증되어 95.0%를 달성하였다. 미등록 화자 거부율은 미등록 화자(TTS)가 웨이크업 워드를 발화했을 때 시스템이 정상적으로 거부한 비율로, 100개의 테스트 데이터 중 92개가 거부되어 92.0%를 달성하였다. 명령어 인식 정확도는 발화된 명령어를 시스템이 정확하게 인식한 비율로, 500개의 테스트 데이터 중 457개가 정확히 인식되어 91.4%를 달성하였다. 본 실험 결과는 제안 시스템이 차량 환경에서 화자 인증과 음성 명령 인식을 효과적으로 수행할 수 있음을 보여준다.

Ⅳ. 결론

본 논문에서는 자율주행 차량 환경에서 음성 명령 인식과 화자 인증을 동시에 수행하는 FPGA 기반 DTW 시스템을 제안하였다. 제안 시스템은 Raspberry Pi에서 MFCC 특징을 추출하고, Zynq-7020 FPGA에서 DTW 템플릿 매칭을 수행하는 분산 구조로 설계하였다. 실험 결과, 등록 화자 인증 정확도 95.0%, 미등록 화자 거부율 92.0%, 명령어 인식 정확도 91.4%를 달성하여 제안 시스템의 유효성을 검증하였다.

향후 연구에서는 본 음성 인식 시스템을 트리거로 활용하여 내비게이션, 공조 제어, 인포테인먼트 등 운전자 맞춤형 서비스를 통합 제공하는 차량용 하드웨어 플랫폼으로 확장할 계획이다.

ACKNOWLEDGMENT

이 논문은 2025년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임(No. RS-2020-NR049604).

*교신저자: 최성곤(choisg@cbnu.ac.kr)

참 고 문 헌

[1] G. Bathla, K. Bhadane, R. K. Singh, R. Kumar, R. Aluvalu, R. Krishnamurthi, and S. Basheer, "Autonomous vehicles and intelligent automation: Applications, challenges, and opportunities," Mobile Information Systems, vol. 2022, pp. 1-36, 2022.

[2] A. J. Sulaiman, M. A. Othman, M. H. Misran, and M. A. Meor Said, "Voice Recognition Vehicle Movement System," in Proc. IEEE 12th Int. Conf. on Control System, Computing and Engineering (ICCSCE), pp. 170-175, Oct. 2022.

[3] S. Bae, H. Kim, S. Lee, and Y. Jung, "FPGA Implementation of Keyword Spotting System Using Depthwise Separable Binarized and Ternarized Neural Networks," MDPI Sensors, vol. 23, no. 12, pp. 5701, Jun. 2023.

[4] Y. S. Chong, W. L. Goh, V. P. Nambiar, and A. T. Do, "A 2.5 μW KWS Engine With Pruned LSTM and Embedded MFCC for IoT Applications," IEEE Trans. Circuits Syst. II, vol. 69, no. 3, pp. 1662-1666, Mar. 2022.

[5] T. W. Chen, Y. H. Chen, C. H. Yang, and H. M. Wang, "A Lightweight Speaker Verification Model For Edge Device," in Proc. IEEE Asia Pacific Signal and Information Processing Association (APSIPA ASC), pp. 1-6, Oct. 2023.

[6] O. Ozturk, A. Yilmaz, and M. Karakaya, "Evaluation of Voice Recognition Platforms and Methods for Edge AI Devices," in Proc. IEEE Int. Conf. on Innovations in Intelligent Systems and Applications (INISTA), pp. 1-6, Sep. 2024.