

깊이정보 기반 지화 학습 프로그램

이가영, 강동현, 김지온, 신병석
 인하대학교 컴퓨터공학과

e-mail : elain202@gmail.com, hiroomanhihi@gmail.com, 3161508@naver.com, bsshin@inha.ac.kr

Fingerspelling Teaching Program using Depth Information

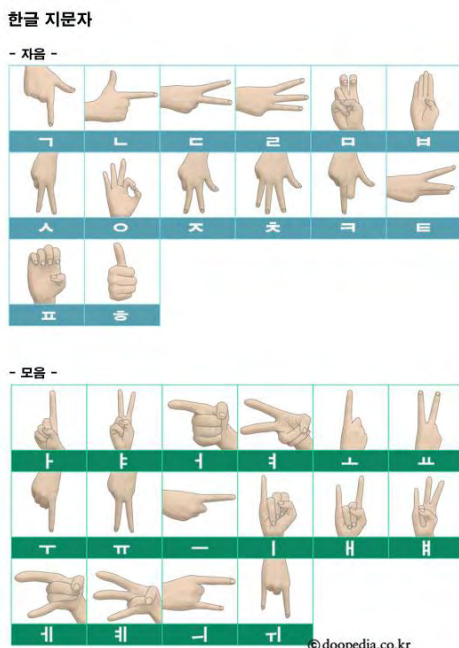
Ka-young Lee, Dong-hyeon Kim, Ji-on Kim, Byeong-Seok Shin
 Dept. of Computer Engineering, In-ha University

요 약

이 논문에서는 깊이 카메라(depth camera)로부터 획득한 깊이 영상을 분석하여 지화를 인식하는 기술을 개발하고 이를 이용한 지화 교육 프로그램의 구현 사례를 소개한다. 먼저 손의 16 개의 특징점을 뽑아내고, 어떤 지화인지 분류한다. 타자 연습 프로그램처럼 단어를 제시하고 그에 맞는 지화를 올바르게 표현하였는지를 검사함으로써 사용자가 지화를 학습할 수 있도록 도와주는 프로그램이다.

1. 서론

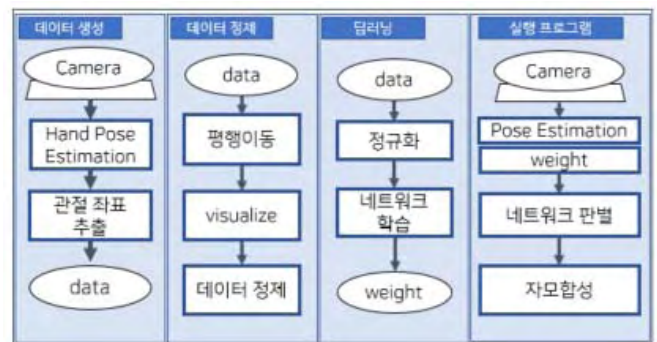
수화(finger language)는 손짓과 몸짓, 표정으로 의미를 전달하는 의사소통 방식으로, 전세계에서 공통적으로 쓰인다. 지화 또는 지문자(fingerspelling)는 수화에 포함되는 개념으로, 문자를 나타내는 손짓이다.[1] 한글 지화는 (그림 1)과 같이 30 종류가 있다. 본 프로그램에서는 손동작 분류의 정확도를 향상시키기 위하여 회전 변환을 사용하였기 때문에, 9 종류의 지화(‘ㄴ’, ‘ㄷ’, ‘ㄹ’, ‘ㅇ’, ‘ㅈ’, ‘ㅊ’, ‘ㅎ’, ‘ㅏ’, ‘ㅑ’)를 분류하는 것으로 한정하였다.



(그림 1)한글 지화

2. 시스템 개요

본 프로그램의 개발은 크게 데이터 생성, 데이터 정제, 딥러닝, 실행 프로그램의 네 부분으로 나눌 수 있다. 데이터 생성 부분에서는 깊이 카메라를 통해 깊이 정보를 받아서 데이터를 생성한다. 데이터 정제 과정에서는 생성된 데이터들 간의 좌표 기준을 맞추고, 잘못된 데이터를 학습하는 것을 막기 위하여 사람이 데이터를 보기 쉽도록 이미지 파일로 시각화를 한다. 이렇게 정제된 데이터는 정규화 단계를 거친 후 딥러닝 네트워크에서 학습이 된다. 실행 프로그램에서는 딥러닝을 통해 얻은 딥러닝 네트워크의 가중치 (weight)를 이용하여 카메라로 들어온 지화 동작을 분류하고, 한글 오토마타를 적용하여 지화를 통해 들어온 한글 자모를 합성하여 보여준다.

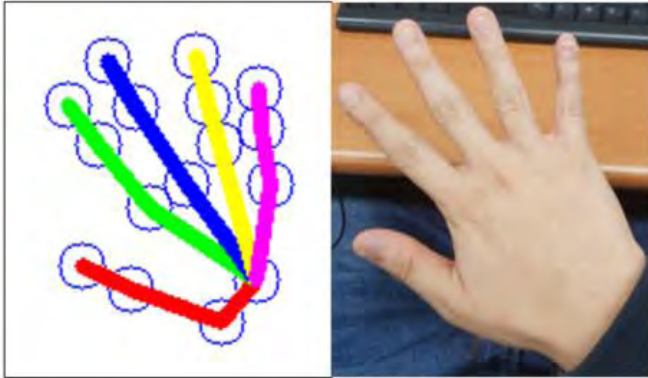


(그림 2)시스템 구성

3. 데이터의 제작과 딥러닝 네트워크 구축

본 연구에서 사용하는 러닝 데이터는 직접 제작하

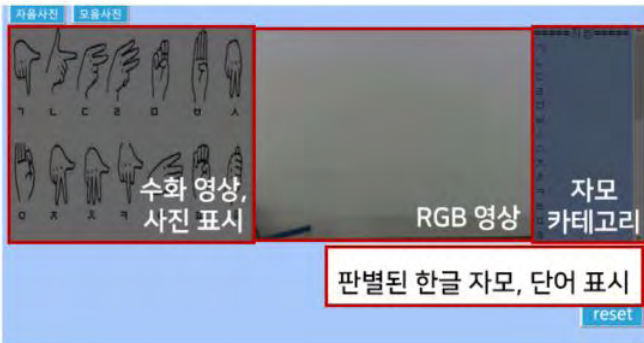
였다. 깊이 카메라를 통해 추정된 16 개의 관절의 픽셀 좌표와 깊이를 얻는다. 이는 16 개의 관절에 대한 공간 좌표이다. 깊이 카메라는 Intel Realsense SR300 을 사용했으며, 관절의 추정은 Intel RealSense SDK 가 지원하는 라이브러리를 사용했다[2]. 평행이동으로 관절 공간좌표의 기준점을 동일하게 했으며, 핸드 스켈레톤 이미지를 만들어서 데이터를 라벨링 하는 작업을 쉽게 하였다.



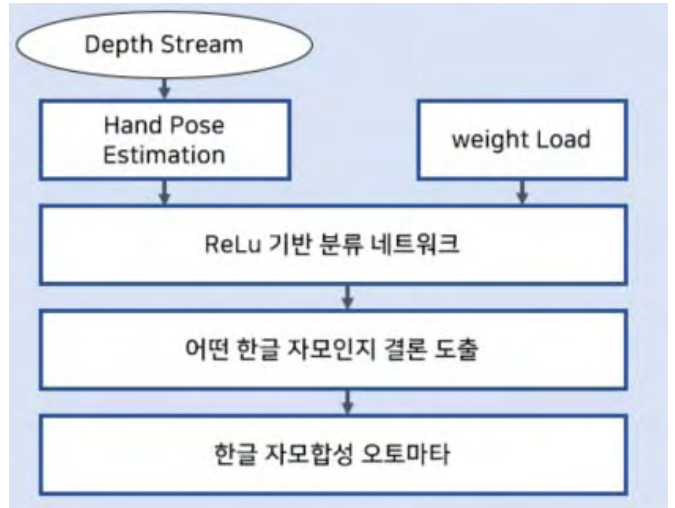
(그림 3)(좌)맵스 기반의 핸드 스켈레톤 이미지 / (우)RGB 영상

16 개의 관절에 대한 공간 좌표값 48 개를 입력으로 받는 딥러닝 네트워크를 학습시킨다. 딥러닝 네트워크는 ReLu 기반의 분류네트워크이다.

4. 실행 프로그램



(그림 4)실행 프로그램의 GUI(Graphical User Interface)



(그림 5)실행 프로그램의 개요도

실행프로그램에서는 이미 학습된 네트워크의 가중치를 받아와서 손동작을 분류한다. GUI 상에서는 사용자가 지화 동작을 따라하는 것을 돕는 영상과 함께, 한글 자모를 표시한다. 한글 자모는 한글 오토마타로 합성된다. 왼쪽에 제시된 표준 지화표를 참조하여 손동작을 하면 영상을 분석하여 자모를 찾아내고 그 자모를 조합하여 제시된 단어를 완성하도록 한다.

본 프로젝트에서는 딥러닝 네트워크에 약 11 만장의 손 사진에 대한 손 관절을 학습 시켰다. 학습 과정에서 99%의 training accuracy 를 보였다.

5. 결론

이 연구에서는 깊이 카메라(depth camera)로 부터 획득한 깊이 영상을 분석하여 지화를 인식하는 기술을 개발하였고 이를 이용한 지화 교육 프로그램의 구현 사례를 소개하였다. 깊이 영상에서 16 개의 특징점을 찾아내고 이를 학습된 네트워크에 입력벡터로 제시한 후 출력된 문자들을 조합해 내는 방식으로 구현되어 있다.

참고문헌

[1] 두산백과(<http://www.doopedia.co.kr>), 수화법
 [2] <https://github.com/IntelRealSense/librealsense>