



A Development Architecture Research of Intelligent Object Recognition and Voice Service for the Visually Impaired

Chul-Jin Kim^{*}, Myung-Soo Park, Min-Hwan Kim

Department of Computer Systems and Engineering, Inha Technical College

ABSTRACT

The development of information technology has made daily life convenient and enjoys cultural benefits through various information technology services. However, due to the limited accessibility of information technology, disabled people do not receive various information technology services. Therefore, in this paper, we propose object recognition and voice service architecture to improve accessibility of information technology service for the disabled. In particular, it enables the visually impaired to receive recognition services and voice services for objects. The visually impaired person recognizes the desired objects using the mobile device, and the recognized object information is provided by voice so that the visually impaired person can recognize the object. In order to recognize the object in the mobile environment, it is possible to recognize the object using the MobileNet model which is a learned model based on the deep learning. The object recognition service works with Amazon's Alexa service, which provides speech recognition and voice conversion services to provide voice services for recognized objects. In this paper, we propose a base architecture that can provide services in three aspects based on object recognition and voice service. The first one is the voice service for the recognized object, the second is the personalized voice service for the recognition object, and finally the knowledge service for the recognition object using the Alexa service.

© 2018 KKITS All rights reserved

KEYWORDS : Object recognition, Voice service, Deep learning, MobileNet, Alexa

ARTICLE INFO: Received 28 June 2018, Revised 23 July 2018, Accepted 10 August 2018.

^{*}Corresponding author is with the Department of
Computer Systems and Engineering, Inha Technical

College, 100 Inha-ro Nam-gu Incheon, 22212, KOREA.
E-mail address: cjkim@inhate.ac.kr

1. 서론

정보기술은 빠르게 발전하고 있으며 우리 일상 생활의 모든 부분과 밀접하게 결합되어 편리함을 제공해 주고 있다. 그러나 이러한 편리한 정보기술을 기반으로 하는 서비스들이 장애인들에게는 불평등하게 제공되고 있다. 많은 정보기술들을 활용하여 장애인의 편리한 정보기술 접근성을 높이려고 노력하고 있으나 아직도 많은 부분에서 미흡한 실정이다. 본 연구에서는 장애인 중에 정보 기술 서비스에 접근하기 가장 어려움을 느끼는 시각 장애인을 위한 서비스를 연구하고자 한다. 시각 장애인에게 모바일 디바이스를 통해 사물을 인식할 수 있도록 하며 인식된 사물에 대한 음성 서비스를 제공한다. 이러한 연구를 통해 불평등한 정보기술 서비스 제공에 대한 편차를 줄이고자 한다.

본 논문의 구성은 2장에서 객체인식 및 음성 서비스 관련 연구를 분석하며, 3장은 시각장애인을 위한 객체 인식과 음성 서비스를 위한 아키텍처와 처리 과정을 제안한다. 4장은 본 연구의 아키텍처와 처리 과정의 적합성을 검증하기 위한 사례 연구를 수행한다. 5장 결론에서는 향후 추가연구 방안을 제시한다.

2. 관련연구

2.1 객체 인식

텐서플로 라이트(TensorFlow Lite)[1]는 모바일 환경에서 머신러닝 서비스를 제공할 수 있는 딥러닝 라이브러리이다. 텐서플로 라이트는 모바일 환경에서 객체 인식을 위해 모바일넷(MobileNet), 인셉션3(Inception v3), 스마트 리플라이(Smart Reply) 모델을 지원한다. 모바일넷은 1천개의 사물을 인식

할 수 있는 학습 모델을 포함하고 있으며[2], 인셉션3는 정확도는 높으나 모바일 환경에서 사용하기 위해서는 높은 사양의 디바이스 자원이 요구된다[3]. 스마트 리플라이는 대화형 채팅 메시지에서 온터치 응답을 제공하기 위한 모델을 제공한다[4]. 본 연구에서는 기본 사양의 모바일 환경에서 객체 인식 및 음성 서비스를 제공하기 위해 모바일넷 모델을 이용하여 객체인식 처리를 수행한다. 모바일넷은 모바일 환경에서 연산될 수 있도록 학습모델의 양을 줄일 수 있도록 제안하고 있다. <그림 1>의 (a)는 기존의 표준 컨볼루션 필터 방식의 학습시 연산 비용을 나타낸 것이며, (b)는 모바일넷에서 제안하고 있는 덤스와이즈 분리 필터 방식과의 연산 비용을 나타낸 것이다. 두 방식을 비교하면 덤스와이즈 분리 필터 방식인 $\frac{1}{N} + \frac{1}{D_k^2}$ 만큼의 연산 비용을 줄일 수 있으며 8~9배의 감소효과를 제공한다[5-7]. 이러한 모바일넷 방식의 객체 인식이 모바일 환경의 자원에 효과적이고 빠르게 처리할 수 있는 장점을 제공한다.

$$D_k \cdot D_k \cdot M \cdot N \cdot D_F \cdot D_F$$

(a) Computational Cost by Standard Convolution Filters

$$D_k \cdot D_k \cdot M \cdot D_F \cdot D_F + M \cdot N \cdot D_F \cdot D_F$$

(b) Computational Cost by Depthwise Separable Filters

D_k : width/height of filters

D_F : width/height of feature maps

M : number of input channels

N : number of output channels

그림 1. 표준 컨볼루션 필터와 덤스와이즈 분리 필터 연산 비교
Figure. 1. Computation's Comparison of Standard Convolution Filter and Depthwise Separable Filter

2.2 음성 서비스

음성 서비스는 음성 인식과 문장을 음성으로 변환하기 위한 TTS(Text To Speech), 음성을 문장으로 변환하기 위한 STT(Speech To Text), 그리고 자연어

인식을 위한 NLU(Natural Language Understanding) 기능이 포함되어 사용자에게 음성 서비스를 제공할 수 있다. 이러한 음성 서비스를 위한 기능들은 애플 시리(Siri)[8], 마이크로소프트 코타나(Cortana)[9], 아마존 알렉사(Alexa)[10]에서 제공하고 있다. 본 연구에서는 아마존의 알렉사의 음성 인식, TTS, NLU를 기반으로 음성 서비스를 제공하며 개인화 음성 서비스를 개발하기 위한 알렉사 스킬(Alexa Skill)[11]을 이용하여 음성 어플리케이션을 개발한다.

2.3 시각장애인 보조를 위한 영상 기반 휴먼 행동 인식 시스템[12]

연구 [12]는 시각장애인을 위한 행동인식 시스템에 대한 연구로서 휴먼 행동인식 서버는 휴먼 검출부, 객체 검출부, 행동 인식부, 객체 인식부로 구성된다. 모바일 디바이스로부터 전송된 영상은 휴먼 검출부와 객체 검출부로 병렬로 입력된다. 휴먼 검출부에서 휴먼의 행동을 검출하며 휴먼 인식부로 전달하여 해동을 인식한다. 객체 검출부에서도 검출된 객체 인식할 수 있도록 객체 인식부로 전달하여 인식한다. 이렇게 휴먼의 행동 인식과 객체 인식을 통해 휴먼의 현재 어떤 객체에서 어떤 행동을 하고 있는지 인식한다. 실험의 결과는 휴먼 행동 인식률이 60.7%이며, 객체 인식률은 48.8%이다. 그리고 휴먼 행동 검출부터 인식, 음성 재생까지 20초가 소요된다. 연구 [12]는 시각장애인을 위한 휴먼 행동 인식을 위한 연구로서는 큰 의미가 있다. 그러나 인식률 및 음성재생 시간에 대한 성능적인 미흡함을 극복해야 할 것이다.

2.4 음성 도우미 기능을 이용한 시각장애인용 커서의 설계 및 구현[13]

연구 [13]은 시각장애인이 웹이나 응용 어플리케이션을 사용할 경우 컴퓨터 화면에서 커서의 위치를 이동하고 선택하는 방법을 음성으로 제공하기 위한 연구이다. 웹이나 응용 어플리케이션의 객체 분리와 확대 기능을 이용한 커서의 설계 및 구현(CSMO, Cursor using Splitting and Magnifying of Objects)[14]에 음성 정보를 추가하여 객체를 선택할 수 있는 CSMO-VH(Voice Helper) 방안을 제안한다. 커서의 위치에 따라 거시적 탐색과 미시적 탐색을 통해 빠르고 정확하게 요구하는 기능을 사용할 수 있도록 한다. 현재 커서가 목표 기능의 화면으로 이동할 경우 먼저 거시적 탐색을 통해 상하좌우 방향을 찾아 이동하고 목표 기능 지역에 근접했을 경우에는 미시적 탐색을 통해 정확하게 기능을 요구할 수 있다.

연구 [13]은 커서와 음성 정보의 결합을 통해 시각장애인에게 응용 어플리케이션을 사용할 수 있도록 하기 위한 명령어 형식의 음성 서비스이며, 본 연구는 시각장애인에게 상호대화 형식의 음성 서비스 및 지식 서비스를 제공한다.

3. 시각장애인을 위한 객체 인식 및 음성 서비스 아키텍처

본 연구에서는 시각장애인에게 객체 인식을 통해 획득된 정보를 기반으로 지식 서비스를 음성으로 제공하기 위한 아키텍처를 제안한다.

3.1 객체 인식 및 음성 서비스 아키텍처

객체 인식 음성 서비스는 <그림 2>와 같이 모바일 디바이스를 통해 객체를 인식하여 객체의 정보에 대한 음성 서비스를 제공한다. 음성 서비스는 인식 객체에 대한 음성 서비스와 객체 정보에 대한 지식 서비스를 제공한다.

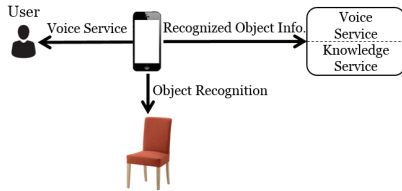


그림 2. 객체 인식 음성 서비스 개념도
Figure. 2. Object Recognition Voice Service Concept

객체 인식 음성 서비스를 제공하기 위한 상세 아키텍처는 <그림 3>과 같다. 시각 장애인은 모바일 디바이스를 통해 객체에 대한 인식을 요청하며 알렉사 기반의 음성 서비스를 통해 인식된 객체에 대한 음성 서비스를 제공한다. 모바일 디바이스 환경에서 객체 인식을 위한 인식 모델로서 모바일넷을 이용하며, 음성 서비스를 위해서는 AVS(Alexa Voice Service)[15]를 이용한다. 모바일 환경에서 음성 서비스는 음성 인식과 음성 제공 서비스이다. 장애인에게 개인화된 음성 서비스를 제공하기 위해서는 음성 어플리케이션 개발 도구인 알렉사 스킬 키트(ASK, Alexa Skill Kit)을 이용하여 개인화된 음성 서비스를 제공한다. 개인화된 음성 서비스에 추가적으로 객체 정보에 대한 지식 서비스를 제공받기 위해서는 AVS를 이용하여 아마존의 지식 서비스를 제공 받을 수 있도록 한다.

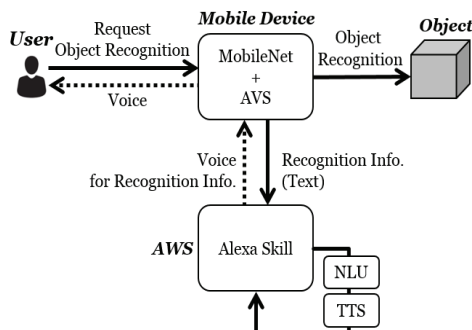


그림 3. 객체 인식 음성 서비스 아키텍처
Figure. 3. Object Recognition Voice Service Architecture

모바일 디바이스는 카메라 센서에 의해 얻은 객체에 대해 모바일넷의 학습된 객체 모델을 통해 인식된다. 모바일넷의 학습모델은 이미지 분류 모델인 인셉션(Inception) 모델을 기반으로 한다. <그림 4>에서와 같이 학습 서버에서 학습 객체에 대해 모바일넷을 기반으로 학습하며 학습된 모델은 모바일 디바이스에서 인식될 수 있는 모델로 변환한다. 모바일 환경의 학습 모델은 JNI(Java Native Interface)의 객체 인식 인터페이스를 통해 객체 인식을 제공한다.

모바일넷 기반의 객체 인식은 모바일 디바이스에서 인식되어 객체에 대한 정보(객체명, 인식률)를 음성 서비스 서버로 전송한다. 음성 서비스 서버에서는 정보를 개인화할 수 있도록 정보를 저장하여 사용자에게 적합한 형태로 서비스한다. <그림 5>와 같이 인식하여 서버에 저장된 객체의 종류를 파악하고 유사한 객체 정보에 대해 개인화된 음성 서비스를 제공할 수 있다.

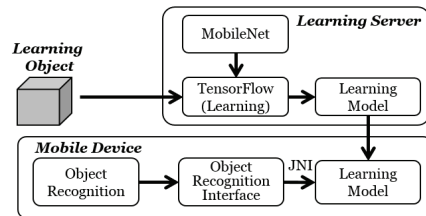


그림 4. 모바일넷 기반의 객체인식 아키텍처
Figure. 4. MobileNet based Object Recognition Architecture

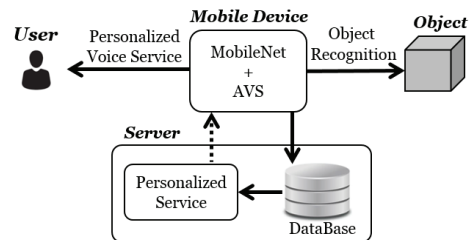


그림 5. 개인화 음성 서비스
Figure. 5. Personalized Voice Service

인식된 객체 정보는 알렉사의 AVS를 이용하여 시각 장애인에게 음성 서비스를 제공한다. 시각 장애인에게 인식 객체에 대한 개인화된 음성 서비스를 제공하기 위해서 음성 서비스 어플리케이션인 알렉사 스킬(Alexa Skill)을 통해 개인화 음성 서비스를 제공한다.

개인화된 음성 서비스를 제공하기 위한 문장 표현(Utterance)과 문장의 데이터 구조(Intent Schema) 관계는 <그림 6>과 같다. 음성서비스의 문장 표현을 나타내는 어터런스는 인텐트에 정의된 구조와 관계를 갖는다. 인텐트에 인텐트 명에 정의된 이름과 타입은 음성 서비스에 의해 가변적인 변경될 수 있는 데이터를 의미하며 이러한 데이터는 어터런스에서 정의된 문장에 대체되어 서비스 된다.

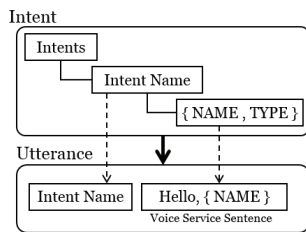


그림 6. 음성서비스 문장과 인텐트와의 관계
Figure. 6. Relationship between Voice Service Sentence and Intent

인식 객체에 대한 음성 서비스와 지식 서비스를 제공하기 위한 아키텍처는 <그림 7>과 같다.

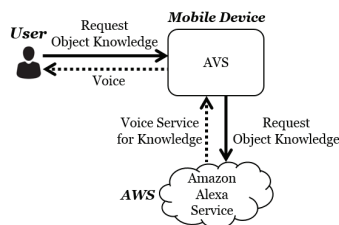


그림 7. 인식 객체에 대한 지식 서비스 아키텍처
Figure. 7. Knowledge Service Architecture for Recognizing Objects

시각 장애인은 인식 객체에 대한 음성 서비스를 통해 어떤 사물인지 인지하며 지식 서비스를 제공받기 위해 모바일 디바이스의 AVS를 통해 객체에 대한 지식 서비스를 요청한다. 요청된 객체의 지식 서비스는 아마존 알렉사 서비스를 통해 제공받을 수 있다.

3.2 객체인식 및 음성서비스 처리 절차

객체 인식 및 음성 서비스 처리 절차는 <그림 8>, <그림 9>와 같다. 객체 인식, 인식 객체에 대한 관리, 인식 객체에 대한 음성 서비스, 그리고 인식 객체에 대한 지식 서비스로 처리된다.

(1) 객체 인식

객체인식은 텐서플로 라이트 기반의 이미지 인식 모델인 모바일넷을 이용하여 객체를 인식한다. 인식된 객체 정보를 인식률이 90.0% 이상 될 경우 인식된 객체에 대해 인식률과 인식 객체명을 제공한다. 모바일 디바이스의 AVS를 통해 인식 객체명에 대한 음성 서비스를 제공한다.

(2) 인식 객체 정보 저장 관리

인식 객체 정보가 의미있는 문장으로 서비스 될 수 있도록 하기 위해 인식 객체의 정보를 서버에 관리한다. 서버에 관리하기 위한 처리를 RESTful API나 아마존의 람다(Lambda) 함수[16]에 정의하며 알렉사 스킬과 연동한다. 알렉사 스킬은 인식 객체 정보를 이용하여 문장을 구성할 수 있으며 개인화된 음성 서비스를 제공할 수 있다.

(3) 인식정보에 대한 음성 서비스

인식 객체 정보에 대해 의미 있는 문장의 음성 서비스를 제공하기 위해 알렉사 스킬을 이용한다. 알렉사 스킬은 의미있는 문장 표현(Utterance)과 객체 정보를 결합하여 음성 서비스를 제공한다. 이러한 음성 서비스는 기존의 저장된 객체 정보를 참조하여 개인화된 음성 서비스를 제공할 수 있다.

(4) 인식 객체에 대한 지식/음성 서비스
 시각장애인은 인식 객체에 대한 상세한 정보의 지식 서비스를 제공 받기 위해 모바일 디바이스의 AVS를 통해 지식 서비스를 요청할 수 있다. 지식

서비스는 아마존의 알렉사 서비스를 이용하며 모바일 디바이스의 AVS와 연동을 통해 제공받을 수 있다.

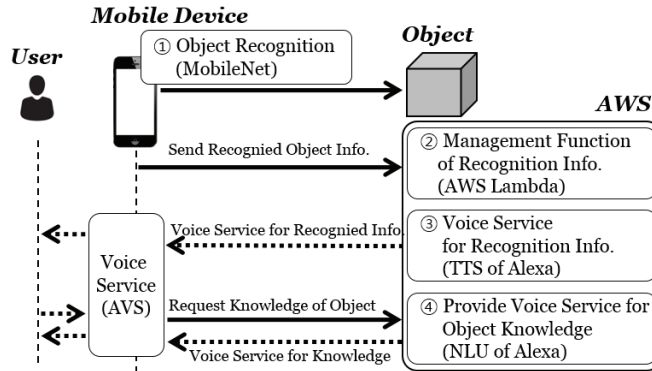


그림 8. 객체인식 및 음성 서비스 처리 구조
 Figure. 8. Object Recognition and Voice Service Architecture

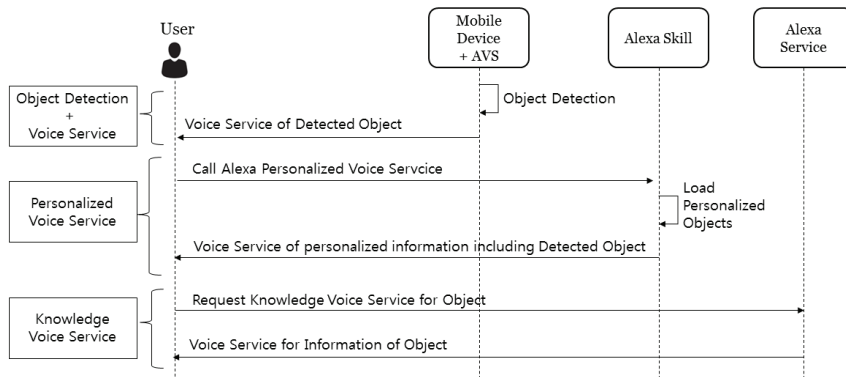


그림 9. 객체인식 및 음성 서비스 처리 절차
 Figure. 9. Object Recognition and Voice Service Process

4. 실험 및 평가

본 연구에서는 객체 인식 및 알렉사 기반의 음성 서비스를 기반으로 객체를 인식하여 음성 서비스 및 지식 서비스를 제공한다.

<그림 10>에서와 같이 객체인식을 위해 디바이스 카메라에 의해 인식된 이미지를 정규화(Normalization)하며, 추론 과정을 위해 모바일넷 모델과 비교 과정을 통해 인식된 객체의 최선의 객체를 찾아낸다.

```
private TensorFlowInferenceInterface inferencelInterface;
...
// Preprocess - Image Nomarization
bitmap.getPixels(intValues, 0, bitmap.getWidth(), 0, 0,
    bitmap.getWidth(), bitmap.getHeight());
...
// Copy Nomarized Image into TensorFlow.
inferencelInterface.feed(inputName, byteValues, 1, inputSize, inputSize, 3);
...
// Inference
inferencelInterface.run(outputNames, logStats);
...
// Detection through comparison with MobileNet Model
final PriorityQueue<Recognition> pq =
    new PriorityQueue<Recognition>(1,
        new Comparator<Recognition>() {
            public int compare(final Recognition lhs, final Recognition rhs) {
                return Float.compare(rhs.getConfidence(), lhs.getConfidence());
            }
        });
```

그림 10. 모바일 디바이스에서 모바일넷 모델 기반의 객체인식
Figure. 10. Object Recognition code based on MobileNet Model

객체 인식은 안드로이드 기반의 모바일 디바이스에서 모바일넷 모델을 통해 인식하며, <그림 11>과 같이 인식된 이미지의 정보(인식 객체명, 인식률)를 서버에 저장한다.



그림 11. 객체 인식 사례
Figure. 11. Object Recognition Case

<그림 11>에서와 같이 인식된 객체의 객체명과 인식률 외에 추가적으로 <그림 12>과 <그림 13>에서와 일정한 시간 간격으로 감지된 객체에 대해 데이터베이스에 저장한다. 저장된 객체 정보는 라벨(label), 정확도(acc), 등록날짜(regdate), 그리고 디바이스id(deviceid)이다. 디바이스 id는 개인화를 위

한 정보로서 서버에 해당 사용자에 의해 관심 갖는 객체에 대한 관심도를 관리하여 사용자에게 노출된 횟수(viewCount)가 높은 객체 정보를 음성으로 제공할 수 있도록 한다.

모바일 디바이스에서 아마존 지식 서비스를 통해 음성 서비스를 제공 받기 위한 코드는 <그림 14>와 같다.

startListening() 함수는 모바일 디바이스로 부터 음성을 입력받기 위한 함수로서 AlexaManager 클래스의 sendAudioRequest() 함수를 통해 알렉사로 음성을 전달한다. 이때 콜백 함수를 지정하여 제공 받은 지식 서비스를 받을 때 호출하는 함수로서 getRequestCallback() 함수를 호출한다.

```
mTask = new TimerTask() {
    @Override
    public void run() {
        runOnUiThread(new Runnable() {
            //Perform the task every 8 seconds.
            @Override
            ...
            getAudioStream(detected_label);
            //updates detected_object
            if(!mapperArrayList.isEmpty())
            {
                for (Mapper mp : mapperArrayList) insert(mp);
                clearList(mapperArrayList);
            }
            ...
            mTimer = new Timer();
            mTimer.schedule(mTask,0,8000);
        }
    }
};

public void insert(Mapper mapper)
{
    // URL setting.
    String url = "http://myserverip:myport/db_insert.jsp?label="+mapper.getLabel()+
        "&acc="+mapper.getAcc()+"&viewcount="+mapper.getView_count()+
        "&deviceid="+mapper.getDeviceid();
    //Perform HttpURLConnection through AsyncTask.
    networkTask = new NetworkTask(url, null);
    networkTask.execute();
}
```

그림 12. 개인화를 위한 객체 정보 저장
Figure. 12. Save Object Information for Personalization

idx	viewcount	label	acc	regdate	deviceid
17	14	mouse	0.94	2018-06-11 19:10	74bfb3cc-4c3a-336a..
18	20	laptop	0.85	2018-06-11 19:14	74bfb3cc-4c3a-336a..

그림 13. 개인화를 위한 저장된 객체 정보
Figure. 13. Saved Object Information for Personalization

```

// Recording Audio
public void startListening() {
    if(recorder == null){
        recorder = new RawAudioRecorder(AUDIO_RATE);
        recorder.start();
        alexaManager.sendAudioRequest(requestBody, getRequestCallback());
    }
}

// Recorded voice transmission
final String url = getEventsUrl();
// Get our access token
TokenManager accessToken =
    TokenManager.getAccessToken(
        ( mAuthorizationManager.getAmazonAuthorizationManager(),
        mContext,
        new TokenManager.TokenCallback() {
            public void onSuccess(final String token) {
                new AsyncTask<Void, Void, AwsResponse>() {
                    protected AwsResponse doInBackground(Void... params) {
                        try {
                            getSpeechSendAudio().sendAudio(url, token,
                                requestBody, new AsyncEventHandler(AlexaManager.this, callback));
                        } catch (IOException e) {
                            e.printStackTrace();
                        }
                    }
                }
            }
        }
    );
}
    
```

그림 14. 모바일 디바이스에서 알렉사 서비스 연동
Figure. 14. Integration of Alexa Service on Mobile Device

인식된 객체에 대한 음성 서비스를 위한 문장의 데이터 구조와 문장 표현은 <그림 15>와 같이 'DetectionIntent' 로 정의하여 문장 표현과 결합하여 음성으로 제공한다. 객체 정보는 사용자에게 노출 횟수(viewCount)가 많은 객체를 선택하여 제공한다.

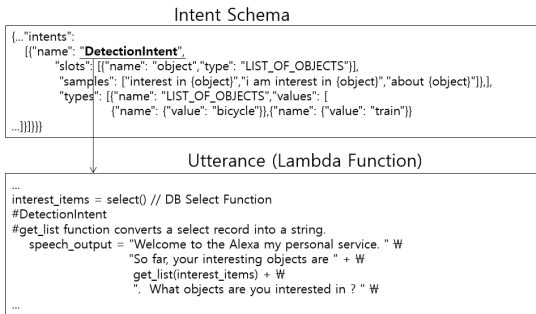


그림 15. 개인화 음성 서비스를 위한 인텐트 구조와 문장표현
Figure. 15. Intent Schema and Utterance for Personalized Voice Service

객체 인식에서 부터 지식 음성 서비스까지의 처리 흐름은 <그림 16>과 같다. 모바일 디바이스에서 객체 인식 후의 음성 서비스는 AVS를 이용하여 단순한 객체 정보(객체명, 인식률) 만을 제공한다. 이후에 사용자에게 의해 개인화 서비스를 요청하기 위해 알렉사 스킬 어플리케이션을 호출하며 서버 저장된 사용자의 인식된 객체 정보 목록을 음성으로

제공하며 사용자는 원하는 객체를 선택할 수 있다.

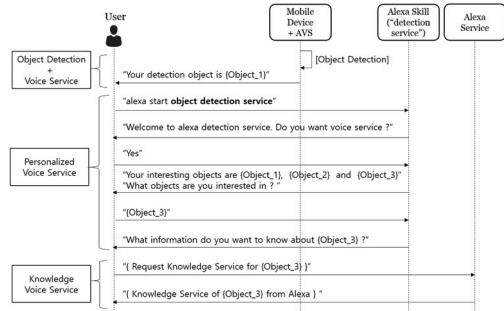


그림 16. 인식 객체에 대한 음성 서비스 흐름
Figure. 16. Voice Service Flow for Recognition Object

선택된 객체에 대한 지식 서비스는 알렉사 서비스를 이용하여 사용자에게 추가적인 정보를 제공할 수 있다. <그림 17>은 개인화 음성 서비스와 지식 서비스에 대한 결과이다.

지금까지 본 연구에서 제안한 아키텍처를 기반으로 한 객체에 대한 인식, 개인화 음성 서비스 그리고 지식 음성 서비스에 대해 사례 연구를 통해 적합성을 검증하였다. 이러한 아키텍처를 통해 시각 장애인의 다양한 웹 서비스 및 모바일 서비스의 접근성을 높일 수 있을 것이다.

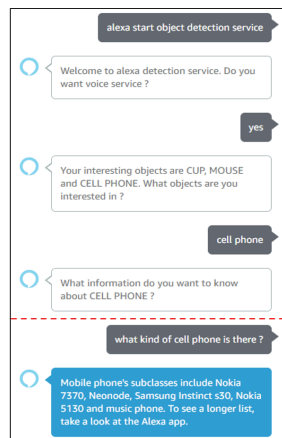


그림 17. 객체 인식 음성 서비스 결과
Figure. 17. Object Recognition Voice Service Results

5. 결 론

본 연구는 객체인식 기술과 음성인식 기술을 활용하여 시각장애인에게 정보기술에 대한 접근성을 향상시키고자 하는 연구로서 객체인식 음성 서비스를 개발하기 위한 아키텍처 및 개발 프로세스를 제안하였다. 제한된 자원의 모바일 환경에서 객체인식을 위해 텐서플로 라이트의 모바일넷 모델을 이용하였으며, 음성 서비스를 위해서 음성인식 기술인 알렉사의 NLU, TTS 등의 기술을 활용하였다. 본 연구의 아키텍처를 통해 제공될 수 있는 시각장애이용 서비스는 3가지로 분류된다. 객체인식 후 인식된 객체 정보에 대한 음성 서비스와 시각장애인에 의해 관심을 갖고 있는 객체 정보를 관리하여 개인화 음성 서비스, 그리고 시각장애인에 의해 관심갖는 객체에 대한 상세한 지식정보 서비스로 제공될 수 있다. 본 연구의 아키텍처 및 개발 프로세스를 통해 다양한 시각장애이용 응용 서비스를 개발하기 위한 기반을 제공할 수 있을 것이다.

향후 연구로는 모바일 환경에서 시각장애인이 인식하는 객체에 대해 노출 횟수에 의한 개인화 음성 서비스를 넘어서 시각장애인의 음성 질문을 관리하여 객체에 대한 관심 타입(구매, 상세정보 확인, 등)에 따른 개인화 음성 서비스를 연구한다.

References

- [1] TensorFlow Lite, <https://developers.googleblog.com/2017/11/announcing-tensorflow-lite.html>, retrieved Apr. 2018.
- [2] MobileNets: Open-Source Models for Efficient On-Device Vision, <https://research.googleblog.com/2017/06/mobilenets-open-source-models-for.html>, retrieved April 2018.
- [3] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, *Rethinking the inception architecture for computer vision*, Rethinking the Inception Architecture for Computer Vision. CVPR, 2016.
- [4] On-Device Conversational Modeling with TensorFlow Lite, <https://research.googleblog.com/2017/11/on-device-conversational-modeling-with.html>, retrieved Apr. 2018.
- [5] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, *MobileNets: Efficient convolutional neural networks for mobile vision applications*, arXiv:1704.04861v1, 2017.
- [6] J. Jin, A. Dundar, and E. Culurciello, *Flattened convolutional neural networks for feedforward acceleration*, arXiv preprint arXiv:1412.5474, 2014.
- [7] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, *Rethinking the inception architecture for computer vision*, arXiv preprint arXiv:1512.00567, 2015.
- [8] Apple Siri, <https://www.apple.com/kr/ios/siri>, retrieved Jan. 2018.
- [9] MicroSoft Cortana, <https://developer.microsoft.com/en-us/cortana>, retrieved Jan. 2018.
- [10] Amazon Alexa, <https://developer.amazon.com/alexa>, retrieved Jan. 2018.
- [11] Alexa Skills Kit, <https://developer.amazon.com/alexa-skills-kit>, retrieved Jan. 2018.
- [12] B. C. Ko, M. C. Hwang, and J. Y. Nam, *Image based human action recognition system to support the blind*, Journal of KIISE, Vol. 42, No. 1, pp. 138-143, Jan. 2015.
- [13] J. W. Lee, J. S. Park, and J. G. Shon, *Design and implementation of a cursor using voice helper for persons with visual impairment*, Journal of Korean Institute of

Information Technology, Vol. 14, No. 4, pp. 181-186, Apr. 2016.

- [14] J. W. Lee, J. S. Park, H. G. Kim, and J. G. Shon, *Design and implementation of a cursor using splitting and magnifying objects for persons with low vision*, Journal of KIIT, Vol. 13, No. 8, pp. 133-138, Aug. 2015.
- [15] Alexa Voice Service, <https://developer.amazon.com/alexa-voice-service>, retrived March 2018.
- [16] AWS Lambda, <https://aws.amazon.com/ko/lambda>, retrived Mar. 2018.

시각 장애인을 위한 지능형 객체인식 및 음성 서비스 개발 아키텍처 연구

김철진¹, 박명수², 김민환²

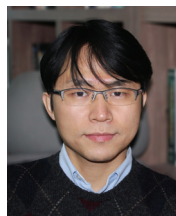
¹인하공업전문대학 컴퓨터시스템과 교수

²인하공업전문대학 컴퓨터시스템과 학부생

요 약

정보기술의 발전으로 일상생활은 편리해 졌으며 다양한 정보기술 서비스를 통해 문화적 혜택을 누리고 있다. 그러나 장애인들은 정보기술의 접근성 한계로 인해 다양한 정보기술 서비스를 제공 받지 못하고 있다. 이에, 본 논문에서는 장애인에게 정보기술 서비스의 접근성을 높이기 위한 객체인식 및 음성 서비스 아키텍처를 제안하고자 한다. 특히 시각장애인이 사물(객체)의 인식 서비스와 음성 서비스를 제공받을 수 있도록 한다. 시각장애인은 모바일 디바이스를 이용하여 원하는 사물을 인식하고, 인식된 객체 정보는 음성으로 서비스하여 시각장애인이 어떤 사물인지를 인식할 수 있도록 한다. 모바일 환경에서 객체를 인식할 수 있도록 딥러닝 기반의 학습된 모델인 모바일넷 모델을 이용하여 객체를 인식할 수 있도록 한다. 객체인식 서비스는 인식된 객체에 대해 음성 서비스를 제공할 수 있도록 음성인식과 음성변환 서비스를 제공하는 아마존의 알렉사 서비스와 연동한다. 본 논문에서는 이러한 객체 인식과 음성 서비스를 기반으로 3가

지 측면에서 서비스를 제공할 수 있는 기반 아키텍처를 제안한다. 첫째는 인식된 객체에 대한 음성 서비스이며, 둘째는 인식 객체에 대한 개인화 음성 서비스, 마지막으로 알렉사 서비스를 이용하여 인식 객체에 대한 지식 서비스를 제공한다.



Chul Jin Kim received both M.S. and Ph. D. degree in Computer Science from the Soongsil University in 1998 and 2004. He worked as a senior researcher in SAMSUNG Electronics from 2004 to 2009. He has been a professor in the Department of Computer Systems and Engineering at Inha Technical College since 2009. He research focus is software engineering, object-oriented & component technique, software architecture, customization, and deep learning.

E-mail address: cjkim@inhatc.ac.kr



Myung Soo Park received a associate degree from Inha Technical College in 2018 and is currently studying bachelor's course in Dept. of Computer Engineering from Inha Technical College. He research focus is deep learning (object detection) and mobile service.

E-mail address: lyine0924@gmail.com



Min Hwan Kim received a associate degree from Inha Technical College in 2018 and is currently studying bachelor's course in Dept. of Computer Engineering from Inha Technical College. He research focus is deep learning (voice recognition) and embedd software.

E-mail address: yurino3@naver.com