Sign Language Tutor

INF7470
UQAM
Autumn 2018
By
Fares Benslimane
and
Ghaith Dekhili

23 December 2018

1 Introduction

Sign language is the mean of communication of the deaf. It uses the articulation of the hands to transmit a message. In addition to hand gestures, it also uses the orientation of the fingers, arms and body movements, and also the facial expressions. These components (channels) are used together to convey a particular meaning.

According to a recent report from Statistics Canada [5], the number of people who reported speaking a sign language at home was nearly 25,000. In the United States alone, studies show that it is estimated that more than 500,000 people seem to be deaf and therefore use ASL (American Sign Language) [13]. Sign language is an alternative way of communication for people who can not communicate verbally. It is as important as any other spoken language. Unfortunately, most people do not speak or understand sign language. As a result, the deaf community often feels forgotten or even misunderstood most of the time, unable to properly communicate their opinions and ideas.

Intelligent Tutoring Systems (ITS) are usually used for the purpose of tutoring and efficiently delivering knowledge interactively instead of traditional learning methods like books or school. Many have used ITSs for tutoring spoken languages by incorporating Natural Language Processing (NLP) techniques in order to evaluate students' answers by identifying the text-input errors and providing an immediate feedback to them [1]. Like for Spoken languages, ITSs can also be used to effectively teach sign language for both deaf people and also for non-deaf people that want to learn sign language to communicate with their family and deaf friends.

In this paper, we demonstrate the effectiveness of using ITS systems for teaching sign language. Our system only uses the ASL (American Sign Language) alphabet (24 letter excluding Z and J), Nevertheless, it can be extended to teach words or even phrases.

2 Related work

Many researchers currently working on frameworks and systems designed to create educational environments that facilitate the learning of sign language. DeSIGN [18], is an educational application designed for deaf students in grades 5th and 6th. It tends to reinforce both English vocabulary and American Sign Language (ASL) signs using an intelligent tutor.

Since most of these systems are destinated to deaf children, many have incorporated game features to make the learning process more fun and entertaining, like CopyCat. CopyCat [19] is an educational adventure game designed to help deaf children practice their sign language.

Some researchers argue that the majority of countries have already implemented a learning system for sign language. The problem emerges in non-deaf persons ”listeners” that need to learn sign language in order to communicate with their deaf friends/family, since learning methods for sign language are usually designed for deaf people. [7] propose Kinect-Sign, a game with the objective of teaching sign language to non-deaf users.

Sign language incorporates many components to deliver a message. For that reason, sign language recognition can be quite challenging. Almost all of the current studies in sign language recognition depends heavily on the Kinect SDK toolkit [6, 11, 7] or the use of special gadgets like gloves [19] for sign recognition. Thanks to the advancement of recognition and vision techniques, it is possible to interpret the visual signs without using any external devices.

In this paper, we propose a learning environment for sign language that make use of a pixel wise recognition system that automatically detects and evaluates the user's hand gestures. Also, an ergonomic user interface UI application that designed to facilitate the learning of sign language.

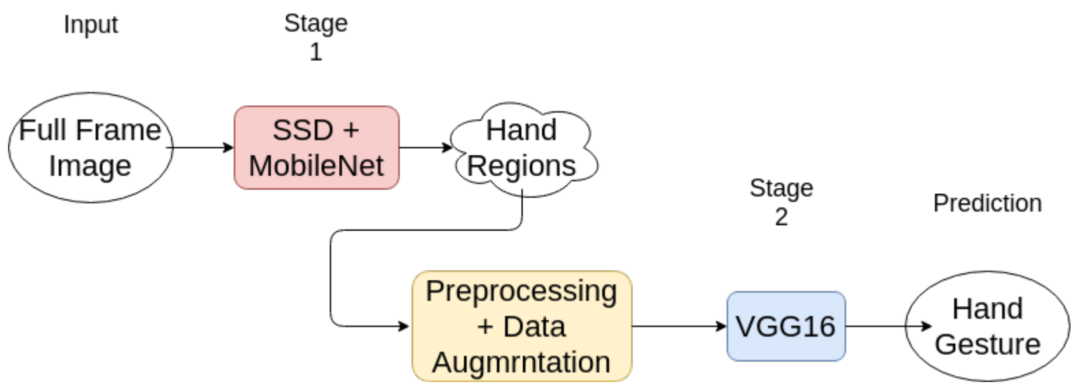

Figure 1: Recognition System architecture

3 System model

3.1 Hand Recognition System

The recognition system consists of two stages. First, the hand detection sub-network. And Second, the gesture recognition sub-network that recognize the hand gesture from the 24 handshape gestures of the ASL alphabet (excluding the letters J and Z). The figure 1 clearly illutrates this. The full frame of the user captured via the webcam is firstly fed into the detection sub-network that will extract the user's hand regions. The extracted images are then fed to the second sub-network to identify the hand gesture for each region.

3.1.1 Hand detection

There exist many deep learning algorithms which are successfully used for object detection tasks. We chose the SSD(Single Shot Detection) [12] with a backbone architecture of Mobilenet [10] for the hand detection task. The reason behind this choice is because it has been shown that SSD w/Mobilenet is the fastest among the other detection models with the best speed/accuracy tradeoff. We used the Pytorch implementation of qfgaohao [16].

For training the detection sub-network, we used the Open Images Dataset V4. In which we only used images that are associated with the 'HUMAN HAND' class. And therefore, we got 22.093 images for training, 1676 for validation and 5328 for testing. We train the model for 200 iterations and we end up with a loss error rate of 3.068.

3.1.2 Gesture recognition

For hand gesture recognition, we used a convolutional neural network based architecture (VGG16 [17]). After detecting the hand regions with the detection sub-network, we predict the hand gesture from the 24 gestures of the alphabet using only the hand's regions of the frame. This will improve the recognition accuracy in which it will discard useless information that doesn't contribute to the sign. Before feeding the gesture recognition sub-network the hand's regions, we preprocess the images by resizing them to 224*224.

For the gesture recognition sub-network, we used the dataset from [15] that comprises 24 static signs (excluding letters j and z because they involve motion). The dataset contains a set of RGB and depth images for each letter in the alphabet. But in this project, we only used the RGB images. We also used the Masey University gesture dataset [3] that contains 2524 images of ASL gestures. The images are in RGB mode, with the hands segmented by color and with a black background.

4 Importance of Human Machine Interface (HMI)

Since the 1960s, the rapid growth of information systems has led to the wide development of research on human computer interaction (HCI) that aims at the designing of human computer interfaces presenting ergonomic properties, such as friendliness, usability, transparency, etc. Various work situations have been covered such as clerical work, computer programming, design, etc [9].

The multidisciplinary and cross-continental roots of HCI combined with its broad scope and multiplicity of paradigms, methods, tools, and application areas have led to its huge diversity [8]. Also, the community is growing, the number and size of scientific and industrial venues is growing, and the body of knowledge is almost exploding, but HCI should always have an early focus on users, tasks, and their contexts [8]. [2] confirms that the primary aims of the researches is still the same wich is to develop an ability to recognize affective state of the user, such an ability is indispensable to have a more human-like nature in human-computer interaction.

5 Description of our interface

In our ptoject we use Kivy[1] as a tool to develop our interface. Kivy is an open source python library for rapid development of applications that make use of innovative user interfaces.

The code for the interface is inspired from a *Kivy Math Tutor* available in this repository[2].

In the first screen the user has the ability to choose between two modes: Learn and Challenge. The Learn mode is made for those who haven't any knowledge about sign language. In this case the user, will be oriented to the Learn mode. Where the alphabet of the ASL sign language is proposed (right side of the screen) as a *gif* (animated image) describing the movements to perform the letter, and its equivalent in English alphabet (left side of the screen). Along side these two, a description of the steps to follow to do the sign is available too. The START button activate the webcam that will evaluate the performed user's gesture of the current letter (using the recognition system shown above). If the sign is correct, a pop-up arises to tell the user that he has done the sign correctly, and it closes after few seconds, if not, the user have the possibility to keep trying to do the sign till a fixed number of trials. If he reaches this number, another pop-up arises to tell him that he failed and asks him if he would like to try again. The same logic will be applied with all the other letters. The functionalities described below are available also for the "challenge" mode. The difference is that in the latter the signs are proposed to the user randomly and don't follow the alphabet order. There is also an *"about this up"* screen that shows the basic information about the application like the used tool, the source code, etc... . There is also music in the background playing in loop that can be muted and adjusted by the user in the settings.

[1] `https://kivy.org/#home`

[2] `https://github.com/gopar/Kivy-Tutor`

6 Intelligent Tutoring System (ITS) architecture

According to Butz [4], the basic architecture of an ITS is composed by a student module, a knowledge module and a tutor module which is also called teaching strategies module. These modules operate interactively and communicate through a central module, which it is often called user interface. This architecture is shown in the figure below.

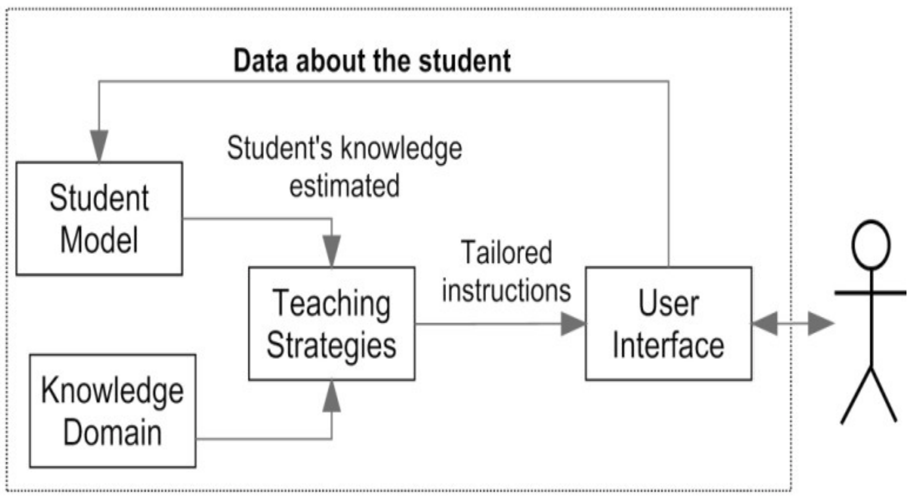

Figure 2: Intelligent Tutoring System architecture [4]

The student module aims to perform the student's cognitive diagnosis and the student's representation for future system feedback[14]. In our application, this model is presented by modes from which the user have to choose.

The knowledge module aims to store the dependent and independent knowledge of the scope[14]. In our application this is represented by the *gifs*, the letters of the alphabet and the description texts.

The tutor module defines and implements a pedagogical teaching strategy, contains the objectives to be achieved and the plans to achieve them. This module selects the exercises, monitors the performance, provides assistance and selects the learning material for the student[14]. In our application, this model is presented by the strategy of proposing signs in different order in the beginner level and the challenge level. The signs are described by three manners: the *gifs* wich describe the animation to be done to learn the sign, its equivalent in english alphabet, and the description provided at the bottom of the screen.

The user interface specifies and provides support to the student's activities and to the methods used to perform these activities. Our interface is easy to use and attractive. Thus, the students quickly learn how to use it, and they can focus all their attention on the process of learning the subject[14].

7 Conclusion

In our work we propose a graphical easy-to-use interface for learning sign language. In our interface we take into consideration some strategies that support the student's learning pace and multiple techniques to guide the student in his learning such as the use of *gifs*, pictures and texts. We use a recognition system to evaluate the user's gesture and give useful feedback.

References

[1] Mustafa Al Emran and Khaled Shaalan. A survey of intelligent language tutoring systems. In *Advances in Computing, Communications and Informatics (ICACCI, 2014 International Conference on*, pages 393–399. IEEE, 2014.

[2] Halis Altun. Human computer interaction and its impact on e-learning in engineering education. 12 2018.

[3] ALC Barczak, NH Reyes, M Abastillas, A Piccio, and T Susnjak. A new 2d static hand gesture colour image dataset for asl gestures. 2011.

[4] C. J. Butz, S. Hua, and R. B. Maguire. A web-based intelligent tutoring system for computer programming. In *IEEE/WIC/ACM International Conference on Web Intelligence (WI'04)*, pages 159–165, 2004.

[5] Statistique canada. Caractéristiques linguistiques des canadiens, 2018.

[6] Paula Escudeiro, Nuno Escudeiro, Rosa Reis, Fernando Maciel Barbosa, José Bidarra, Ana Bela Baltazar, Pedro Rodrigues, Jorge Lopes, and Marcelo Norberto. Serious game on sign language. In *Proceedings of the XV International Conference on Human Computer Interaction*, page 97. ACM, 2014.

[7] João Gameiro, Tiago Cardoso, and Yves Rybarczyk. Kinect-sign: teaching sign language to "listeners" through a game. In *International Summer Workshop on Multimodal Interfaces*, pages 141–159. Springer, 2013.

[8] Tom Gross. Human-computer interaction education and diversity. In *Human-Computer Interaction. Theories, Methods, and Tools*, pages 187–198. Springer International Publishing, 2014.

[9] Jean-Michel Hoc. From human – machine interaction to human – machine cooperation. *Ergonomics*, 43(7):833–843, 2000.

[10] Andrew G Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*, 2017.

[11] Hatice Kose, Neziha Akalin, Rabia Yorganci, Bekir S Ertugrul, Hasan Kivrak, Semih Kavak, Ahmet Ozkul, Cemal Gurpinar, Pinar Uluer, and Gökhan Ince. isign: an architecture for humanoid assisted sign language tutoring. In *Intelligent Assistive Robots*, pages 157–184. Springer, 2015.

[12] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C Berg. Ssd: Single shot multibox detector. In *European conference on computer vision*, pages 21–37. Springer, 2016.

[13] Ross E Mitchell, Travas A Young, BELLAMIE BACHELDA, and Michael A Karchmer. How many people use asl in the united states? why estimates need updating. *Sign Language Studies*, 6(3):306–335, 2006.

[14] María Lucila Morales-Rodríguez, Apolinar Ramírez-Saldivar, Arturo Hernández-Ramírez, Julia Patricia Sánchez-Solís, and José Antonio Martínez Flores. Architecture for an intelligent tutoring system that considers learning styles. *Research in Computing Science*, 47:37–47, 2012.

[15] Nicolas Pugeault and Richard Bowden. Spelling it out: Real-time asl fingerspelling recognition. In *Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on*, pages 1114–1119. IEEE, 2011.

[16] qfgaohao. pytorsh-ssd, 2018.

[17] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[18] Ling Xu, Vinithra Varadharajan, Joyce Maravich, Rahul Tongia, and Jack Mostow. Design: an intelligent tutor to teach american sign language. In *Workshop on Speech and Language Technology in Education*, 2007.

[19] Zahoor Zafrulla, Helene Brashear, Peter Presti, Harley Hamilton, and Thad Starner. Copycat: an american sign language game for deaf children. In *Automatic Face & Gesture Recognition and Workshops (FG*

2011), 2011 IEEE International Conference on, pages 647–647. IEEE, 2011.