Real-time Communication Systems powered by Ai for specially Abled

Kedar Potdar[1], Gauri Nagavkar [2]

U.G. Student, Department of Computer Engineering, Watumull Institute of Electronics Engineering and Computer Technology, Mumbai, Maharashtra, India[1]

U.G. Student, Department of Electronics and Telecommunication Engineering, Watumull Institute of Electronics Engineering and Computer Technology, Mumbai, Maharashtra, India[2]

ABSTRACT: This project aims to aid the deaf-mute by creation of a new system that helps convert sign language to text and speech for easier communication with audience. The system consists of a gesture recognizer hand-glove which converts gestures into electrical signals using flex sensors. These electrical signals are then processed using an Arduino microcontroller and a Python-based backend for text-to-speech conversion. The glove includes two modes of operation – phrase fetch mode and letter fetch mode. The phrase fetch mode speaks out words at once, while the letter fetch mode speaks out individual letters. This project forms a base infrastructure which can later be augmented with addition of different Sign Languages and integrating with other hearing impaired aid systems.

KEYWORDS: Gesture recognition, Flex sensor, Arduino, Test to speech, Sign language.

I. INTRODUCTION

Gesture is a non-verbal means of communication. It refers to expressing an idea using position, orientation or movement of a body part.Gesture recognition is the mathematical interpretation of orientation or motion of human body by a computational system. In this project, the words expressed by hand gestures by the speech and hearing impaired are converted into verbal means of communication. The translated output is displayed on a screen and "spoken" on a speaker.

Sign Language is the well-structured code, which uses hand gestures instead of sound to convey meaning, simultaneously combining hand shapes, orientations and movement of the hands. Communicative hand glove is an electronic device that can translate sign language into speech and text in order to make the communication possible between the deaf and/or mute with the general public. This technology has been used in a variety of application areas, which demands accurate interpretation of sign language. In this project, the words/letters conveyed by the disabled person are displayed on a screen and also spoken on a speaker.

The project is divided into two parts: 1) Data acquisition from the flex sensors 2) Processing the acquired data and giving corresponding output on the screen and speaker. Data acquisition is done using Flex sensors mounted on the Hand glove. Next, the analog signals obtained from the flex sensors are converted into digital. The digital signals are processed and compared with the predefined values. If the values match, the corresponding letter is returned.

The paper is organized as follows. The previous projects and papers related to this paper are described in section II. Description of the hardware and software components used in the project is done in section III. Section IV has the system architecture of the project. Section V gives the operation and flowcharts of the main function and the two

INTRODUCTION

Communication is a social process of exchanging information from one entity to another in verbal and non-verbal form. It defines our existence and it is an important instrument that connects people together. It comes naturally as a raw skill embedded in most people at birth and we acquired the ways of communication through cognitive learning. Communication is the basis, which drives the process of development in all the fields (Manohar, 2008) and it is the very core of our civilisation. The ability to communicate allows us to express emotion, feelings, convey our thoughts and ideas as well as to relate our experiences. It plays an important role in the dissemination of information and sharing of knowledge especially in the academic arena. Research has found that human started to learn how to communicate with each other since they are born not only through spoken and written languages but also body gesture, posture, facial expression and eye contacts (Busso, et al., 2004; Cohen, Grag & Huang, 2000).

Communication skill might come as a natural ability in majority of people. However, there are some people inflicted with some form of physical defects which affect their ability to communicate. One of the more severe disabilities is known as "cerebral palsy", a congenital disorder at birth which causes abnormality in their motor system. It affects their muscle movement and coordination, learning and speech abilities. Their malfunctioned motor system causes an uncontrollable and involuntary movement. They are unable to control their oral-facial muscles, thus affects their ability to perform facial expression appropriately.

Many assistive tools or formally termed as Alternative and Augmentative Communication (AAC) has been developed and employed to assist people with impaired communication skills. The term encompasses the whole combination of methods used for communication such as text to speech system, pointing gestures, facial expression and body language. Although these AACs have been widely used to assist the disabled, but it is not potentially effective because most AACs are text

Figure 1. Examples of alternative and augmentative communication (AAC) tools

to speech and touch screen based applications, which are unsuitable for those with severe physical abilities. There are many kinds of AAC stools available in the market which is shown in Figure 1.

From the limitation of the existing tools reviewed (Novita, 2006; Macsolvers, 2009; Standup, 2006; Universiteit van Amsterdam, 2008; Crestwood, 2009; ScienceDaily, 2008), there is still a pressing need for more effective and efficient tools to alleviate this problem. One the possible methods is to implement a facial expression recognition system to predict or determine the emotional state of a disabled person through his expression projected on his face. The implementation of such method can be made possible through biometrics information systems. According to Gregory and Simon (2008), biometrics information system can be employed as a means to detect and classify the physiological aspect of a person in real time. Franco and Treves (2001) further support the notion that facial expression can be used for human computer interaction and usability enhancement.

Based on the problem statements deliberated above, we propose an improved real time behaviour monitoring application for the disabled by employing real time biometric information i.e. the facial expression recognition system. The aim to create a model that is capable of detecting user's emotion without engaging any physical action from the users. To increase the usability and interactivity of the tool, the emotion detected by the system will be sent to the care-taker's mobile phone in the form of SMS.

FACIAL EXPRESSION

Facial expressions recognition is an ability to recognize people by their facial characteristic and differentiate it with one another. Human is born with the ability to recognize other people easily by identifying their facial features such as shape, appearance, skin texture and skin complexion. Other than that, humans also have the ability to express, interpret and differentiate facial expressions. The regular recurring ones are happiness, anger, disgust, fear, surprise and sad (Ekman & Friesen, 1978). The six facial emotions stated above are important and play a major role in expressing emotion as well as recognising facial expression (Busso, et al, 2004).

In real life, inter personal human interaction are performed not only using speech or spoken language, but also non verbal cues for example hand gesture, body gesture, facial expression and tone of the voice. All these cues are sometimes being used for expressing feeling and give feedback (Busso, et al, 2004; Cohen, et al., 2000). We can see how human interact with each other using non-verbal cues everyday. For example a child cries in front of his mother because he is not happy or dissatisfied with something. Other people might interpret it differently thinking that the child might be in pain.

Facial expression interaction is relevant mainly for community social life, teacher and student interaction, credibility in difference contexts, medicine and so on. Besides, facial expression recognition is useful for designing new interactive devices which offers the possibility of new ways for human computer interaction - HCI (Franco & Treves, 2001). Cohen, et al. (2000) conducted survey on their users and noticed that they have been through traditionally HCI consists of the keyboard, mouse, joystick, trackballs, data gloves and touch screen monitors. The interaction can be improved and enhanced by introducing facial expression recognition that requires no direct contact from the user.

Facial Expression Recognition System (FER) has been a topic for research since Ekman and Friesen (1978) who pioneered this research and worked from the psychology perspective. In the past 20 years, many researchers have tried to adopt their idea and make improvement, innovation and modification on facial expression recognition by introducing different techniques, mainly concentrated on the improvement in term of accuracy, efficiency, mobility, and speed (Kotsia & Pitas, 2007). With all the enhancements on techniques for facial detection and recognition, the development of the facial expression recognition has also improved (Zhan & Zhou, 2007). The most active researches in computer vision and pattern recognition is face recognition in forensic identification, access control, user interface design (Wang, Plataniotis & Venetsanopoulos, 2005), emotion analysis, interactive video, indexing and retrieval of image and video database, image understanding and synthetic face animation (Zhan & Zhou, 2007).

In real world, humans are able to read complex communication where the synthesis of verbal and non-verbal communication is used to express feelings and opinions. Human can interpret and generate major facial expressions but a computer is not built with any facial recognition ability unless through the use of some software. It is even more complicated for the computer to interpret irregular facial expression, especially from those suffering from cerebral palsy. Due to their disorder, they do not have the ability to reflect their emotions like a normal typical person. Thus, a more natural and naive method has to be employed for the system to work by a manual labelling of the image captured with the emotion of the user.

FACIAL EXPRESSION RECOGNITION

The same concept of inter-human interaction can be applied for human-computer interaction in facial expression recognition. A computer uses microphone and camera to "see" and "hear" human expressions and learns to recognize it (Cohen, et al., 2000). The human face is captured by the camera attached to a computer and the captured images is stored for processing with some methods to recognize the identity and emotion of a user based on the features achieved. With automated facial expression recognition, the subjects do not need to operate the computer device or performing any actions in order to get a task done. Besides, facial expression recognition method has been recommended by many researchers to be a good technique to interpret a person's emotion if compared with other recognition methods such as speech (Zhan & Zhou, 2007; Pantic & Patras, 2006; Lau, 2009).

There are myriads of attempts and research done to produce computer algorithms that serve as models to automate human face recognition function. One of the widely applied systems is known as Facial Action Coding System (FACS). Facial Action Coding System (FACS) was initiated by Ekman and Friesen (1978). It involves the analysis of facial muscle anatomy. It detects the changes to the facial muscle, the contraction and relaxation of a group of muscles to produce certain facial expression. Due to the method employed, it is not suitable to be used by a disabled person namely "Cerebral Palsy" candidates as they have difficulty controlling their muscles and their facial expressions are non-typical. Apart from that, Cohen, Sebe, Garg, Chen, and Huang (2003) stated that FACS processes were very time consuming and tedious.

Another technique is the Skin Color Model that extracts the skin colour from any region of the image as a matching based to detect the face. It will not be accurate if it is taken from other region like legs, arms, and neck appear in the image or video (Singh et. al., 2003). It is also easily affected by the lighting condition and camera settings.

There are other numerous methods developed over the years with their own technique and accuracy. We have chosen to integrate the Viola and Jones Face detection algorithm alongside template matching technique into our image processing engine.

Viola-Jones algorithm is one of the most widely used techniques for face detection and template matching is also the proven effective algorithm for facial expression classification and matching. Viola-Jones algorithm was used by Barlett, Littlewort, Fasel, and Movellan (2003) in their research to perform real time face detection and facial expression recognition process. Shan, Gong, and McOwan (2005) employed template matching techniques to perform person-independent facial expression recognition. They obtained recognition accuracy of 79.1% for 7-class recognition and 84.5% recognition accuracy for 6-class recognition. Refer to the Appendix for a summary of techniques and algorithms used in automatic facial expression recognition system.

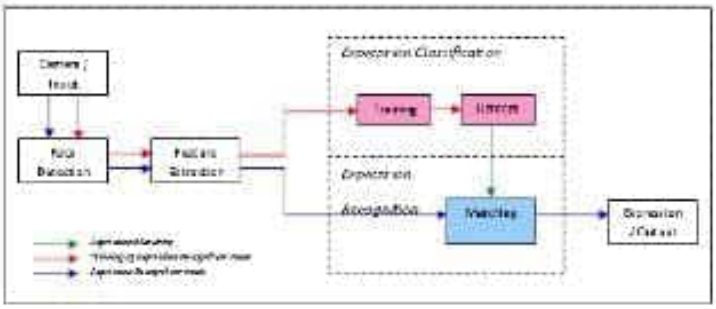
We adopted Viola-Jones Face Detection algorithm (2004) and used the default Haar cascade for human face searching because this algorithm is considered the most common and useful algorithm in the field of facial expression recognition research (Cho et al, 2009; Brubaker et al, 2008; Zhan et al, 2006; Datcu and Rothkrantz, 2007; Bartlett, et al, 2003).

As for template matching, it had been a conventional method for object detection and pattern recognition especially facial features at the early stage of face recognition research (Chai, et al, 2009). The advantage of template matching for our proposed prototype is that it is simple, easy to implement, and does not take so much time and memory. We tried on other algorithm such as Support Vector Machine (SVM) as well as Principal Component Analysis (PCA) as both techniques is widely used and recommended by researchers (Dubuisson, et al., 2002; Chen and Huang, 2002; Bartlett, et al, 2003; Littlewort, et al., 2007; Ashraf et al, 2009) but we found that the expression training using these techniques is processor intensive and is consuming memory.

PROTOTYPE MODELLING

Generally, automated facial expression recognition system involves three major steps or stages. There are the face detection stage, feature extraction stage as well as expression recognition and classification stage (Zhang, Lyons, Schuster, & Akamatsu, 1998; Chibelushi & Bourel, 2002; Zhan et al, 2006). Figure 2 shows two major processes in an automated facial expression recognition system. The blue arrows indicate the process of expression recognition while red arrows indicate the training process of expression recognition. When performing expression recognition, a camera is used to capture the subject's face in video sequence and these video sequences or

Figure 2. Facial expression recognition system



the similar to the expression recognition process but the extracted features is used as training data for classifying subject's expression.

By implementing automated facial expression recognition system, we proposed a real time behavior monitoring for physical and communication disabled which is shown in Figure 3. The automated facial expression recognition system is mounted on the disabled wheelchair for expression recognition purposes. The system will detect the disabled face, extract the face and perform emotion or expression recognition. Critical or preset abnormal expression recognized will trigger notification module via Short Messaging System (SMS) gateway to notify the parents, teachers and the care taker for the disabled.

Figure 3. Real time behavior monitoring system