Volume 3, Issue 4, April 2018 ISSN (online): 2456-0006
International Journal of Science Technology Management and Research
Available online at: www.ijstmr.com

Lip Reading Of Digits Using Artificial Intelligence

Shivani Patel
Department of Information Technology
MET's Bhujbal Knowledge City
Institute of Engineering Nashik, India

Smita Jadhav
Department of Information Technology
MET's Bhujbal Knowledge City
Institute of Engineering Nashik, India

Pankaj Pagare
Department of Information Technology
MET's Bhujbal Knowledge City
Institute of Engineering Nashik, India

Saurabh Kasliwal
Department of Information Technology
MET's Bhujbal Knowledge City
Institute of Engineering Nashik, India

Kirti Patil
Assistant Professor,
Department of Information Technology
MET's Bhujbal Knowledge City
Institute of Engineering Nashik, India

Abstract-: ***Laryngectomy is the removal of part or all of the larynx (voice box).They are of two types Partial Laryngectomy it is a smaller cancers of the larynx often can be treated by removing only part of the voice box and in Total Laryngectomy your entire larynx is removed it result in loss of the natural voice and directly affects the basic communication functions in daily life. Reconstructing the basic communication function is an important issue for these patients after total Laryngeal surgery. It is also one of the possibly alternative approaches to reconstructing the basic communication function for these patients after total Laryngeal surgery. Although many human lip-reading recognition methods have been developed to detect lip contour precisely, detecting pronouncing lip contour effectively is still a difficult challenge. In this paper, a lip-reading recognition algorithm is proposed to recognize English digits from zero to nine from the lip contour when speaking. Here, several criteria for detecting the mouth region of interest (ROI) were designed to reduce the error rate of detecting the mouth ROI and lip contour. Moreover, several lip parameters, including the width, height, contour points, area, and the ratio (width/height) of lips, were used to recognize the lip contour and English digits from zero to nine when speaking. The advantages of the proposed method are that it can detect the mouth ROI automatically, reduce the influence of individual differences, such as the individual lip shape or makeup effect, and it also can perform a good performance without per-training. We have also included multiple users giving them more personalizing software for their use. Finally, the performance of lip-reading recognition under different backgrounds and individual differences is tested and the accuracy is maintained.***

Keywords: component Laryngectomy , lip-reading recognition, mouth region of interest, visual-only speech recognition, Digit recognition.

I. INTRODUCTION

A lip-reading recognition algorithm is developed to real-time recognize each English digits when speaking. In this algorithm, the image per-processing and several criteria for detecting the mouth Region Of Interest (ROI) were designed to reduce the influence of the variation of the environmental condition and to avoid the error occurrence results from the detected incomplete contour or the remaining part of the background image. After detecting the mouth ROI, several lip parameters, including the total pixel number of the detected lip contour, and the height, width, area and ratio of the lips, are used as the features of lip-reading recognition. Finally, the correlation between these extracted lip parameters corresponding to each English digit would be calculated to precisely .By using the above techniques, the proposed method can detect the mouth ROI automatically, reduce the influence of individual differences, and it can give good perform without per-training. Total Laryngectomy is a common treatment for patients with advanced laryngeal and hypo pharyngeal cancer, but it is also a result from the loss of the natural voice and directly affects the basic communication functions in daily life. Reconstructing the basic communication function is an important issue for these patients after total Laryngectomy surgery. It is also one of the possibly alternative approaches to reconstructing the basic communication function for these patients after total Laryngectomy

surgery [1][2]. Although many human lip-reading recognition methods have been developed to detect lip contour precisely, detecting pronouncing lip contour effectively is still a difficult challenge. Here, several criteria for detecting the mouth region of interest (ROI) were designed to reduce the error rate of detecting the mouth ROI and lip contour. Moreover, several lip parameters, including the width, height, contour points, area, and the ratio (width/height) of Lip Reading of Digits using Artificial Intelligence lips, are used to recognize the lip contour and English Digits when speaking [3].

II. METHODS

A.LIP-READING RECOGNITION ALGORITHM

The proposed Lip Reading Recognition Algorithm is illustrated in Fig 1. In order to enhance the image contrast for human vision, the image of captured by the webcam would be first processed by extending the original minimum and maximum pixel values to the values of 0 and 255 respectively. Then, the technique of median filtering, which is a nonlinear digital filtering technique and has been widely applied in preserving the contour edges and removing noise, such as salt, pepper noise, speckle noise etc. would be used to reduce noise in the captured image[3]. After detecting the mouth ROI, the information for the lip contour in this region and the bounding box of this region would be used for digit recognition. The used information included the total pixel number of the detected lip contour, and the height, width, area and ratio of the lip bounding box (the mouth ROI)[3]. In order to reduce the influence of the individual difference, the above parameters would be first normalized before digit recognition. Here, the information of the user's face image would be pre-recorded in the beginning of the system program, and is used as a measure of scale for normalization. Here, the information of the user's face image is obtained from five face images that the user closed his/ her mouth. Next, the correlation between the normalized lip contour information vector and each digit feature vector would be obtained by Pearson correlation coefficient.

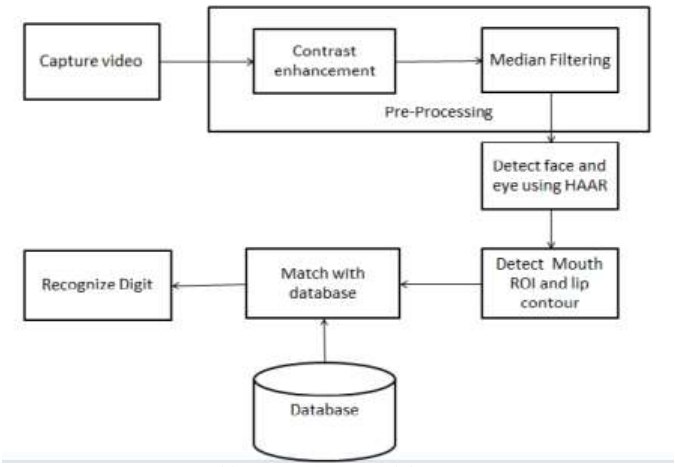

FIGURE 1: Architecture Diagram of the System

B.IMPLEMENTATION OF LIP-READING RECOGNITION SYSTEM

The input image is scanned across location and scale using a scaling factor of 1.1. At each location an independent decision is made regarding the presence of a face. This leads to a very large number of classifier evaluations; approximately 50,000 in a 320x240 image. Following the Ada Boost algorithm [4][5] a set of weak binary classifiers is learned from a training set. Each classifier is a simple function made up of rectangular sums followed by a threshold. In each round of boosting one feature is selected, that with the lowest weighted error. The feature is assigned a weight in the final classifier using the confidence rated Ada Boost procedure. In subsequent rounds incorrectly labeled examples are given a higher weight while correctly labeled examples are given a lower weight. In order to reduce the false positive rate while preserving efficiency, classification is divided into a cascade of classifiers. An input window is passed from one classifier in the cascade to the next as long as each classifier classifies the window as a face. The threshold of each classifier is set to yield a high detection rate. Early classifiers have fewer features while later ones have more so that easy non-face regions are quickly discarded. Each classifier in the cascade is trained on a negative set consisting of the false positives of the previous stages. This allows later stages to focus on the harder examples. In order to train a full cascade to achieve very low false positive rates, a large number of examples are required. After 5 stages the false positive rate is often well below 1%. The image features (see Fig.2) are called Rectangle Features and are reminiscent of HAAR basis functions [6]. Each rectangle feature is binary threshold function constructed from a threshold, and a rectangle filter which is a linear function of the image. The value of a two-rectangle filter is the difference between the sums of the pixels within two rectangular regions. The regions have the same size and shape and are horizontally or vertically adjacent. A three-rectangle filter computes the sum within two outside rectangles subtracted from twice the sum in a center rectangle. Finally a four-rectangle filter computes the difference between diagonal pairs of rectangles. Given that the base resolution of the classifier is 24 by 24 pixels, the exhaustive set of rectangle filters is quite large, over 100,000, which is roughly O(244) (i.e. the number of possible locations times the number of possible sizes). The actual number is smaller since filters must fit within the classification window. Computation of rectangle filters can be accelerated using an intermediate image representation called the integral image. Using this representation any rectangle filter, at any scale or location, can be evaluated in constant time. The form of the final classifier returned by Ada boost is a perceptron a threshold linear combination of features.

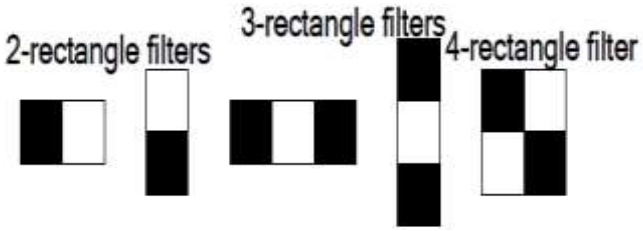

FIGURE 2: HAAR like feature used for face detection

We trained an upright detector using 2000 manually cropped 20x20 pixel faces and 2000 background (non-face) patches. All profile faces were de rotated so that the faces were looking approximately straight right. The resulting cascade has 11 layers of classifiers with the first six classifiers having 9, 9, 3, 7, 10 and 9 features, respectively. We trained only one detector for frontal faces. Therefore we rotate the picture to be detected. The rotation angle is 30 degrees and we make 12 in-plane rotations so that together, the 12 pictures cover the full 360 degrees of possible rotations. We made translations of pixel coordinates for image rotation. Though there are 12 translations, in fact we only need two pair of coordinates, which are (0.866x-0.5y, 0.866y+0.5x) and (0.5x-0.866y, 0.5y+0.866x) ((x, y) is the pixel coordinate before rotation), other translated coordinates are simply the reverse or mirror of the above 3 pair coordinates. The input images are preprocessed using histogram equalization to alleviate luminance. Rotated face can be detected correctly for both color and gray-scale images. It takes less than 0.3 seconds in a Pentium IV 2.8GHz machine to execute the software implementation of our face detection algorithm for a 320x240 image.

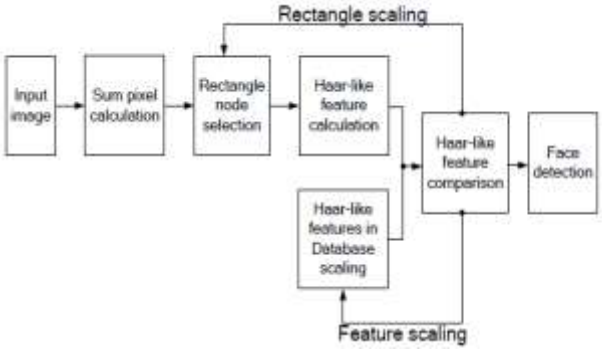

FIGURE 3: Flow diagram of the face detection.

C. EXPERIMENTAL DESIGN

Different with face detection which needs only one training procedure for detection of all faces, each person's face should be trained in the face recognition step. The face size for training is chosen as 30x30 pixels. We use one person's faces under different conditions as positive samples and use other persons' faces as negative samples. In the face recognition step, we only process the detected face region (Fig. 4) of the complete picture. To decrease the false positive rate, the threshold of the final classifier is increased. This unfortunately also reduces the recognition rate. To increase the recognition rate again (now accompanies by a higher false positive rate), classifier layers are removed from the end of the cascade. This is done simultaneously for all of the classification stages of the recognition system. There are three types of environmental conditions and four types of individual differences for tests, including one static background (condition 1), two dynamic backgrounds (fan rotating background: condition 2, and people walking background: condition 3), thick-lip makeup effect (individuality 1), moustache face (individuality 2), and different face angles (30-degree: individuality 3, and 60-degree: individuality 4). Here, a total of ten participants attended this experiment. Each trial contains a least 10 times digit utterances. Before evaluating the performance of the proposed method, several parameters of binary classification test have to be first defined, and they were listed as follows: True positive (*TP*) denotes a specific digit is correctly detected as the specific digit. False positive (*FP*) denotes other digits or silence is incorrectly detected as a specific digit. True negative (*TN*) denotes silence is correctly detected as silence. False negative (*FN*) denotes a specific digit is incorrectly detected as silence. Here, the parameters of sensitivity (also called true positive rate, TPR), precision (also called positive predictive value, PPV) and accuracy (ACC) were used to evaluate the performance of digit recognition.

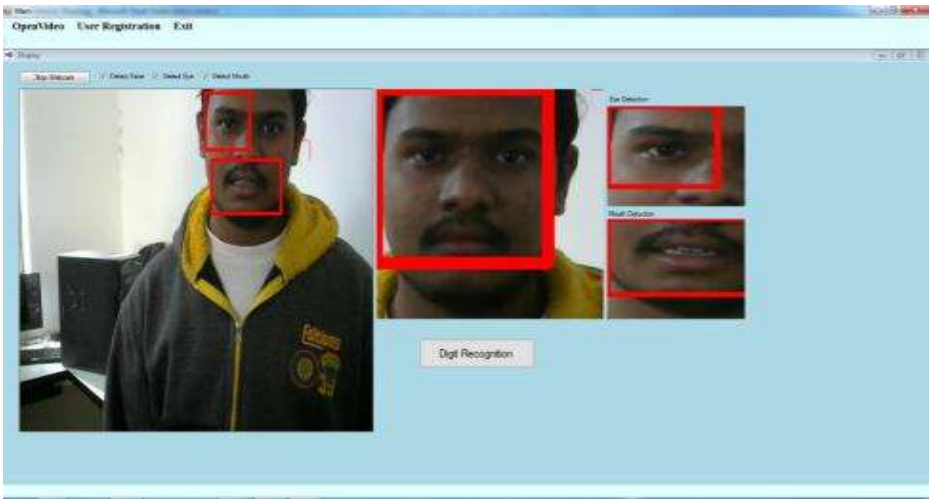

FIGURE 4: Snapshot of proposed system

III. DISCUSSION

Moreover, the above methods can just detect the individual lip shape or check the opening or closing state of the mouth, and they can precisely recognize each English digit when speaking. For the proposed system, the total initialization time and the total tracking time on the computer with Pentium 2.94-GHz CPU were about 0.5 seconds and 0.27 seconds respectively. Different from other human lip recognition methods which can just detect the individual lip shape or check the opening or closing state of the mouth, the proposed method can recognize each English digit when speaking. Moreover, the pre-training is also required for the proposed system, and this also improve the convenience of use.

CONCLUSION

A Lip recognition algorithm is developed to extract the features of lip contour and real-time recognize each English Digit when speaking. Here, the criteria settings were developed to improve the stability of detecting the mouth ROI and lip contour. Five lip parameters extracted from one frame in the video sequence, including the width, height, contour points, area and the ratio (width/ height) of lips, were directly used to recognize English Digit when speaking. The mouth ROI can be detected automatically. Different from other human lip recognition algorithms, not only the lip contour but also English Digit can be recognized when speaking. The pre-training and the database for modeling individual lip contour were also required, and this greatly improves the convenience of use in various kinds of application. From the experimental results, the accuracy of the proposed algorithm on lip-reading recognition is good and is insensitive to the varying background. This also improves its practicability and might contain the potential of applying in the lip-reading recognition under car driving in the future. However, reducing the brightness of the environmental background still obviously affected the performance of lip-reading recognition, and this is a problem to solve in the future.

REFERENCE

1. P. da Silva, T. Feliciano, S. V. Freitas, S. Esteves, and C. A. E. Sousa,``Quality of life in patients submitted to total laryngectomy," *J. Voice*,vol. 29, no. 3, pp. 382388, May 2015.
2. N. Agrawal and D. Goldenberg, ``Primary and salvage total laryngectomy," *Otolaryngol. Clin. North Amer.*, vol. 41, no. 4, pp. 771780, Aug. 2008.
3. BOR-SHING LIN,YU-HSIEN YAO, CHING-FENG LIU, CHING-FENG LIEN, and BOR-SHYH LIN,``Development of Novel Lip- Reading Recognition Algorithm,"vol. 5, January 9, 2017.
4. R. Schapire and Y. Singer. Improving boosting algorithms using confidence-rated predictions, 1999.
   C. Papageorgiou, M. Oren, and T. Poggio. A general framework for object detection. In International Conference on Computer Vision, 1998.
5. Lienhart, R. and Maydt, J. An extended set of HAAR-like features for rapid object detection. IEEE ICIP 2002, Vol. 1, pp. 900-903, Sep. 2002.