2015 International Conference on Information Technology Systems and Innovation (ICITSI)

Proceedings

November 16 – 19, 2015

Bandung – Bali, Indonesia

IEEE Conference Number #36052

IEEE Catalog Number : CFP1590Y-ART

ISBN : 978-1-4673-6664-9

Organizer:

- School of Electrical Engineering and Informatics – ITB
- Information Networks and System Lab – ITB
- Jurusan Teknologi Informasi - Udayana

Sponsored by:

Proceedings of the

2015 International Conference on Information Technology Systems and Innovation (ICITSI)

IEEE Catalog Number CFP1590Y-ART
ISBN 978-1-4673-6664-9

Additional copies may be ordered from the IEEE Service Center:
IEEE Part Number: CFP1590Y-ART
IEEE Service Center
445 Hoes Lane
P.O. Box 1331
Piscataway, NJ 08855-1331 USA

Telephone (toll-free): 1-800-678-IEEE
Telephone (direct): +1-732-981-0060
Fax: +1-732-981-9667
E-mail: customer-service@ieee.org

General Chair

Suhardi - Institute of Technology Bandung

Steering Committee

- Suhono Harso Supangkat (Chair) – Institute of Technology Bandung
- I Ketut Gede Darma Putra – Udayana University
- Jaka Sembiring - Institute of Technology Bandung
- John Yearwood – Deakin University
- Robin Doss – Deakin University

Organizing Committee

- Kusprasapta Mutijarsa (Chair) – Institute of Technology Bandung
- A. A. Kompiang Oka Sudana – Udayana University
- A. A. Kt. Agung Cahyawan Wiranatha – Udayana University
- Dessy Purnami Singgih Putri – Udayana University
- Dwi Putra Githa – Udayana University
- Gusti Made Arya Sasmita – Udayana University
- I Putu Agus Eka Pratama – Udayana University
- IGB. Baskara - Institute of Technology Bandung
- Kadek Suar Wibawa – Udayana University
- Made Sunia Atmaja – Udayana University
- Ni Kadek Dwi Rusjayanti – Udayana University
- Ni Putu Sutramiani – Udayana University
- Putu Arya Dharmadi – Udayana University
- Putu Wira Buana – Udayana University
- Samuel Andi Kristyan – Institute of Technology Bandung
- Yoanes Bandung – Institute of Technology Bandung

Technical Program Committee

- Armein Z. R. Langi (chair) – Institute of Technology Bandung
- Adit Kurniawan – Institute of Technology Bandung
- Andrian Bayu – Institute of Technology Bandung
- Arry Akhmad Arman – Institute of Technology Bandung
- Ary P. Setijadi – Indonesia IEEE CS chair
- Bambang Riyanto Trilaksono – Institute of Technology Bandung
- Benhard Sitohang – Institute of Technology Bandung
- Budi P. Resosudarmo – Australian National University
- Dinh Phung – Deakin University, Australia
- Endra Joelianto – Institute of Technology Bandung
- Hyoung-Gook Kim – Kwangwoon University, South Korea
- I Ketut Adi Purnawan – Udayana University
- I Made Sukarsa – Udayana University
- I Nyoman Piarsa – Udayana University
- I Putu Agung Bayupati – Udayana University
- Iping Supriana – Institute of Technology Bandung
- Kangbin Yim – Soon Chung Yang University, South Korea

TABLE OF CONTENTS

2015 International Conference on Information Technology Systems and Innovation (ICITSI)
Bandung – Bali, November 16 – 19, 2015
ISBN: 978-1-4673-6664-9

Lip Reading Based on Background Subtraction and Image Projection

Fatchul Arifin
Departement of Electronics Engineering Education, Universitas Negeri Yogyakarta
Yogyakarta, Indonesia
fatchul@uny.ac.id

Aris Nasuha
Department of Electronics Engineering Education, Universitas Negeri Yogyakarta
Yogyakarta, Indonesia
arisnasuha@uny.ac.id

Hardika Dwi Hermawan
Department of Electronics Engineering Education, Universitas Negeri Yogyakarta
Yogyakarta, Indonesia
hardikadwihermawan21@hotmail.com

***Abstract*— Someone who does not have vocal cords, has no ability to produce voice and speech. This problem is suffered by laryngectomy patients. Over half of all laryngectomy patients worldwide are using electrolarynx for the rehabilitation of their speech ability. Unfortunately, the electrolarynx is relatively expensive, especially for patient from the lower classes. In this research, portable speech aid tool for laryngectomy patient based on lip reading is studied, especially for who use Indonesian language. Two problems in a lip lip reading system, especially if it run in portable device, are lighting and limited resources. Background subtraction and image projection method are studied in this research to overcome these two problems.**

Keywords*— laryngectomy, lip reading, background subtraction, image projection

I. Introduction

Someone who does not have vocal cords, has no ability to produce voice and speech. This problem is suffered by late-stage laryngeal cancer patients. They are usually treated with total laryngectomy, in which larynx, and tissues around it, including vocal cord, should be removed. By doing surgery, a hole in front of the patient's neck, known as stoma, is made. Then, the trachea is attached to this stoma which is used by the patients to breathe. As the vocal cord of the laryngectomy patients have been removed, they will not be able to speech anymore. They lost their ability to speak as they did [1].

Efforts to assist the laryngectomee, person who has no larynx, to speak has much been studied, for example research on silent speech interface [2]. However these research are still in laboratory stage. Other research on automatic lip reading system on cell phone has done, for example in [3], but the result is not encouraging. Moreover, research into this issue for patients using Indonesian language, has not been done or is very limited.

In addition, general research on automatic lip reading for embedded systems is still not produce satisfactory results, one of which is due to existing algorithms usually require large resources, which would be a heavy run on embedded systems. Or need to be always connected to the internet, while in some areas it is not available, or is still relatively expensive.

Another problem on an automatic lip-reading system is lighting. Less light, causing the image obtained by the camera becomes more difficult to be processed, so it requires a more reliable system for minimum lighting conditions. In this research we propose method to use in lip reading using background subtraction as lip segementation and image projection to extract the feature of lip.

II. Segmentation And Feature Extraction

A. Background Subtraction

Background subtraction is a technique in the field of image processing, in which foreground image is extracted for further processing. Generally, region of interest of the image is objects in the foreground of image. After the stage of image preprocessing, the object localization is required, which can use this technique. Background subtraction is a widely used approach to detect moving objects in the video of a static camera. The rationale for this approach is that it detects a moving object from the difference between the frame and the frame of reference [4]. Background subtraction done if the image is a part of the video stream. Background subtraction gives important cues for a variety of applications in computer vision, for example, surveillance or tracking human pose estimation.

Principle of lip reading is observing lip pattern, assuming the lips represent the pronunciation of certain syllables or words. Meanwhile, lip detection is not always easy to do, especially if the lip color does not contrast compared to surrounding areas, or because of the weak illumination. The use of background subtraction method to observe the movement of the lips will be able to solve the above problem, assuming the image included in the frame is just around the lips, no movement other than the lip, and there was no lighting changes during the pronunciation.

The basic principle of background subtraction method is to calculate the difference between the value of the intensity of the pixels in the current frame and pixel values in the previous frame, or can be expressed by equation (1).

$$dI(x,y,t) = I(x,y,t) - I(x,y,t-1) \qquad (1)$$

where I(x,y,t) is intensity value in pixel(x,y) at frame t

B. Image Projection

Image projections are one-dimensional representations of image contents. Horizontal and vertical projection of image is histogram over horizontal and vertical way of grayscale level. Horizontal and vertical projections of image I(u,v) is defined in equation (2) and (3).

$$P_{hor}(v_p) = \sum_{u=0}^{M-1} I(u, v_p) \quad \text{for } 0 < v_p < N \qquad (2)$$

$$P_{ver}(u_p) = \sum_{v=0}^{N-1} I(u_p, v) \quad \text{for } 0 < u_p < M \qquad (3)$$

Each row and each column of image become a bin in the histogram. The count that is stored in a bin is the number of 1-pixels that appear in that row or column.

This method can extract image features quickly and easily. This method has proven successful for Cursive character recognition [5], Amazigh Handwritten Character Recognition [6], and traffic sign recognition for intelligent vehicle [7]

III. Implementation

In order to test our proposed method, we use video recorded from 10 volunteers, 5 men and 5 women. All video are color and focused around subject's mouth. In these video, all volunteers recorded in frontal face. Each of them pronounces 3 simple Indonesian word twice, i.e. "saya", "mau" and "makan", therefore there are 60 video data. Original video size is 640x480 pixels. Video recording time for each volunteer is 1 second for each word or each data consists of 25 frames. Examples of screen shots recording can be seen in Fig.1.

Our proposed method consists 4 stages:

1). Grayscaling, i.e. convert video from RGB color space to grayscale

2). resizing the image in each frame, from 640x480 pixels into 32x24 pixels

3). background subtraction to segment the lips from sequential image

4). vertical and horizontal projection, to extract the features of the lip image, that will be used as input of classifier. The results of feature extraction for each frame is (32 + 24) features, or (32 + 24) x 25 features for each data.

5). word recognition using artificial neural network as classifier, i.e. backpropagation with 3 layers. Input layer size depends on the number of features and output layer size is 3. Hidden layer size and number of iteration is varied

Fig. 1. Examples of screen shot recording

to produce the best result. As a comparison, another classifier is used, i.e. SVM (Support Vector Machine).

Other feature extraction methods used for comparison are:

(1) background subtraction followed by a vertical projection, this feature extraction output is (32 x 25) features

(2) background subtraction followed by a horizontal projection, feature extraction output is (24 x 25) features

(3) vertical and horizontal projection (without background subtraction), this feature extraction output is (32 + 24) x 25 features.

IV. Experimental Result

To determine whether the system can recognize words spoken by volunteers, we use 5-fold cross validation. This method of evaluation divide all data into 5 parts, one part, say the first part, is used as the test data, the rest as training data. Then the second part of the data as the test data, the rest for training data, and so on, until all the parts has been used as the test data.

Evaluation is done for each of the feature extraction method, and each uses two classifier, which are artificial neural network (ANN) and Support Vector Machine (SVM). Table I until Table IV show the results of accuracy for each of these methods, which is expressed in a confusion matrix. Rows on the confusion matrix shows the actual classes (true classes), while the columns show predictions. The shaded cells show the test results are correct.

TABLE I. Confusion Matrix For Proposed Method

ANN	Makan	mau	saya
makan	14	1	5
mau	2	13	5
saya	5	2	13

SVM	makan	mau	saya
makan	11	3	6
mau	1	15	4
saya	5	2	13

TABLE II. Confusion Matrix for background Subtraction followed by Vertical Projection

ANN	makan	mau	saya
makan	13	2	5
mau	3	16	1
saya	4	2	14

SVM	makan	mau	saya
makan	2	16	2
mau	15	2	3
saya	10	10	0

TABLE III. Confusion Matrix for background Subtraction followed by Horizontal Projection

ANN	makan	mau	saya
makan	10	5	5
mau	5	11	4
saya	2	6	12

SVM	makan	mau	saya
makan	8	5	7
mau	5	11	4
saya	3	5	12

TABLE IV. CONFUSION MATRIX FOR HORIZONTAL AND VERTICAL PROJECTION (WITHOUT BACKGROUND SUBTRACTION)

ANN	makan	mau	saya
makan	8	6	6
mau	4	15	1
saya	7	2	11

SVM	makan	mau	saya
makan	5	9	6
mau	1	6	13
saya	7	10	3

The test results of each these extraction methods can be measured by calculating the average accuracy from each confusion matrices. However, put attention only in accuracy, in many cases, has not been sufficient, so it is necessary to calculate F1 score, which is a combination of precision and recall. F1 score is expressed by Equation (4).

$$F1 = 2\frac{precision \cdot recall}{precision + recall} \quad (4)$$

$$precision = \frac{\sum true\ positive}{\sum true\ positive + \sum false\ positive} \quad (5)$$

$$recall = \frac{\sum true\ positive}{\sum true\ positive + \sum false\ negative} \quad (6)$$

TABLE V. F1 AND AVERAGE F1 SCORE FOR EACH METHOD

	Proposed method		BS + vert.		BS + hor.		Vert + hor	
	ANN	SVM	ANN	SVM	ANN	SVM	ANN	SVM
saya	0.605	0.605	0.700	0.000	0.585	0.558	0.579	0.143
mau	0.722	0.750	0.800	0.083	0.524	0.537	0.698	0.267
makan	0.683	0.595	0.650	0.085	0.541	0.444	0.410	0.303
average F1	0.669	0.645	0.717	0.056	0.545	0.513	0.562	0.238

TABLE VI. ACCURACY OF EACH METHOD

Proposed method		BS + vert.		BS + hor.		Vert + hor	
ANN	SVM	ANN	SVM	ANN	SVM	ANN	SVM
0.67	0.65	0.71	0.07	0.55	0.52	0.57	0.23

By observing Fig. 2, Table V, and Table VI, it is known that the accuracy of the test is always offset by the value of F1, which means that the accuracy of the test is not in doubt. As of the classifier used, ANN showed better accuracy than SVM, for all the feature extraction methods. In terms of feature extraction methods, the best assay results obtained from the Background subtraction method followed by the vertical projection, which uses ANN classifier, i.e. 71.7% accuracy. However, this method with SVM classifier, produces poor accuracy, which is only 6.7%. SVM parameters had been varied, but the results were not much different.

The second best feature extraction methods are the proposed method, the method of background subtraction followed by vertical and horizontal projection, which generates 67% accuracy. This method using SVM classifier also results remain fairly good accuracy, namely 65%.

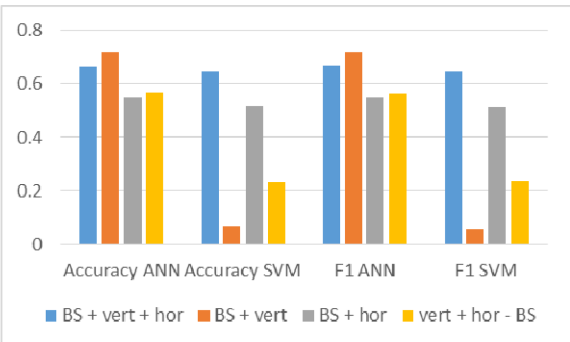

Fig. 2.

In terms of the number of features produced, method of background subtraction followed by the vertical projection, produce fewer features than the proposed method. Fewer number of features will make this method lighter when run on portable devices. However, this method still need to be evaluated, considering its accuracy is so low for SVM classifier.

V. CONCLUSION

This paper presented a method for lip reading based on background subtraction and image projection. The result shows that our proposed method using ANN classifier achieves 67% accuracy. A comparison method, i.e. using background subtraction followed by vertical projection and ANN classifier, yields better accuracy, namely 71%. However the last method need further evaluation, because it produces so low accuracy for SVM classifer.

ACKNOWLEDGMENT

We would like to express our gratitude to Ministry of Research and Higher Education of Republic Indonesia which has given the financial support to this research, on research grant of Hibah Bersaing in 2015.

REFERENCES

[1] Fatchul Arifin, Tri Arief Sardjono, and Mauridhi Hery Purnomo, "The Relationship Between Electromyography Signal Of Neck Muscle And Human Voice Signal For Controlling Loudness Of Electrolarynx," *Biomed. Eng. Appl. Basis Commun.* **26**, 1450054 (2014)

[2] B. Denby, T. Schultz, K. Honda, T. Hueber, J.M. Gilbert, and J.S. Brumberg, "Silent speech interfaces," Speech Communication 52 (2010) 270–287

[3] Young-Un Kim, Sun-Kyung Kang, and Sung-Tae Jung, "Design and implementation of a lip reading system in smart phone environment," *Information Reuse & Integration, 2009. IRI '09. IEEE International Conference on* , vol., no., pp.101,104, 10-12 Aug. 2009

[4] Benezeth, Y.; Jodoin, P.-M.; Emile, B.; Laurent, H.; Rosenberger, C., "Review and evaluation of commonly-implemented background subtraction algorithms," in *Pattern Recognition, 2008. ICPR 2008. 19th International Conference on* , vol., no., pp.1-4, 8-11 Dec. 2008

[5] Rodriguez, Roberto J; Antonio Carlos Gay Thomé, "Cursive character recognition – a character segmentation method using projection profile-based technique", url: http://www.nce.ufrj.br/labic/downloads/isa2000.pdf

[6] Saady, et al. "Amazigh Handwritten Character Recognition based on Horizontal and Vertical Centerline of Character", International Journal of Advanced Science and Technology, Vol. 33, August, 2011

[7] Lorsakul, Auranuch; Jackrit Suthakorn, "Traffic Sign Recognition for Intelligent Vehicle/Driver Assistance System Using Neural Network on OpenCV", Mahidol University. Faculty of Engineering. Center for Biomedical and Robotics Technology (BART LAB), 2007

Link bukti terindex scopus:

https://www.scopus.com/authid/detail.uri?authorId=56163418400