eJAICT
ISSN 2355 1771
JOURNAL OF APPLIED INFORMATION, COMMUNICATION AND TECHNOLOGY
Published by:
Research Center of Technology and Innovation
Academic Research and Community Services
Swiss German University

Multi Modal Gender Recognition for Gender-Based Marketing Using Depth Camera

Adityo Setyonugroho, James Purnama, Maulahikmah Galinium

Faculty of Engineering and Information Technology, Swiss German University

Article Information

Received: 1 June 2015
Accepted: 3 August 2015
Published: 25 October 2015
DOI: 10.33555/ejaict.v2i2.87

Corresponding Author:

Adityo Setyonugroho
Email: adityosn@hotmail.com

ISSN 2355-1771

ABSTRACT

This research is conducted to prove that gender recognition by computer which can be done in real time by using depth camera. Gender recognition can be used on many industries, such as security, marketing, and other sectors. The purpose of this research is to detect gender by using images of user (RGB image) and voice. Furthermore, gender-based marketing is used for the implementation of this system. By using multi modalities, the result is more accurate than only using one factor. Image processing algorithm is used on processing facial image, which is Linear Discriminant Analysis (LDA) algorithm. Furthermore, gender can also be detected by special frequency of each gender speech. Autocorrelation is one of the methods that is able to detect pitch from detected audio. Kinect for Windows v2 was carried out as visual and audio sensor. This research proved that gender can be detected by using those modalities with right algorithm. Several problems are also found during the experiments, such as input data problem, not matching algorithm, and small percentage of accuracy. In conclusion, detecting gender can be done by computer (real time or not) and several ideal conditions must be made to get proper and high accuracy result, such as person distances from camera, lighting, image size.

Keywords: *Autocorrelation, Face Recognition, Gender Recognition, Gender-based Marketing, Kinect for Windows v2, Linear Discriminant Analysis, Pitch Detection*

1. Introduction

Gender is a basic criteria that human have. Humans are separated into two kinds of gender which are male and female. They are different in gesture, speaking, thinking, feeling, style, and needs. For example, male has own kind of clothing styles and female has their own.

The gender identification is important for people to interact and communicate between each other in the correct way. It is very important to the industries and businesses to provide their customer needs based on their gender. They need to classify customer target of their product to do correct marketing or campaign. Nowadays, people or industries separate their customer gender manually using their physical vision. It is so tiring for marketing people to identify the right target by gender from the large amount of customers. Since a product need to be delivered to the right customer, several big companies have already used gender statistical data to promote their product. This marketing is called gender based marketing (Chowney V., 2012).

This research is conducted to implement a gender classification system in order to convert manual work on gender separation into computer work. The system detects the user gender and deliver the product promotion based on detected gender. Therefore, the user should enter the sensor area, and then the system will track and identify the gender. Furthermore the system will give an output, which is a product based on the detected gender. This research is expected to bring a new way to deliver the product promotion in more efficient and modern way. The purpose of this research is developing multimodal gender recognition using facial image and audio. Linear Discriminant Analysis (LDA) algorithm is used for facial image while Autocorellation is used to detect pitch from detected audio.

2. Gender Recognition

The method to differentiate a gender between male and female using several factors is called Gender Recognition. Gender separation is easy for human, but for a machine, it is quite difficult task because several factors must be learnt and trained to be understood well like a human (OpenCV., 2015). There are several situations that need automation of gender detection to help human and computer interaction. For example, on marketing case, product will be delivered well if people can perform a consistent and great detection of each gender needs.

Gender can be classified using several factors, such as face, gesture, gait, and voice. This research use two factors to detect the gender as follow:

2.1. By Using Head or Face

People can naturally differentiate a gender by looking at their faces and heads. Gender recognition by computer vision uses the same aspects as people naturally did. For example, man usually has shorter hair than woman, every man grows moustache, women wear earring, women wear veil, and many other examples.

Another method that can be used on gender recognition is using biometrics that applied on people face or head, such as facial pattern (Matta et al., 2008). There are several algorithms have been established and developed that calculating facial pattern to create gender recognition for better computer vision. Some techniques or algorithms that manipulate scientific value on people face are Local Binary Patterns (LBP) (Huang et al., 2011) (Hyunh et al., 2012), Gradient-LBP (LBP) (Hyunh et al., 2012), 3DLBP (Hyunh et al., 2012), Principal Component Analysis (PCA) (Kumar et al., 2013) (OpenCV., 2015), Linear Discriminant Analysis (LDA) (Saraswathi et al., 2015) (OpenCV., 2015) (OpenCV., 2015), and others.

2.2. By Using Voice

Human voices also has unique frequency for each gender. Male has frequency range between 85 to 155 Hz while female has frequency range between 165 to 255 Hz (Parker, R., 2015). This recognition can be done by using pitch detection algorithm. By capturing the audio or voice using microphone, analysis of voice can be done by using time domain or frequency domain.

3. Related Work

Several experiments have been conducted by many researcher to detect gender by using face and voice. Each modality has unique characteristic to determined human gender.

3.1. Gender by Using Face

By using image processing method and machine learning, Arnulf and Felix was able to detect gender of facial images (Arnulf et al., 2002). They were combining pre-processing methods using Principal Components Analysis (PCA) and Locally Linear Embedding (LLE) with Support Vector Machine (SVM) Classification to work with facial images. Other image processing methods also used by Zhang and Kipman to detect gender using facial images, they are Gabor and Local Binary Pattern (Zhang et al., 2014). Another experiment showed that LDA can be used to detect gender. Philipp Wagner proved that discriminant analysis method is capable to separated discriminant feature of male and female with high percentages of successful detection (Wagner, P., 2011). In our research, LDA is used to detect gender based on facial data that received from depth camera. LDA was chosen based on training and testing result of LDA, PCA, and LBPH at the beginning of this research. PCA and LBPH have some problems or errors when they are used on the experiments. According to Philipp Wagner (Wagner, P., 2011), LDA shows better result in gender recognition using facial features.

3.2. Gender by Using Voice

Parker’s research showed that human gender can be detected by calculating fundamental frequency of human voice (Parker, R., 2015). Fundamental frequency or pitch of each gender stands on different frequency ranges. Male usually has frequency range from 85 to 155 Hz. Meanwhile, woman has 165 until 255 Hz. By using those ranges, Parker performed some experiments to detect gender in real time. Parker’s experiment methodology use Cooley-Tukey FFT algorithm to find frequency spectrum of voice signal. Another algorithm also can be used for detecting gender by using human voice, which is autocorrelation algorithm. Autocorrelation algorithm is also able to detect pitch of human voice (Heath, M., 2011). In this research, autocorrelation algorithm is used to detect human pitch. It is a basic algorithm to detect pitch on normal speech. Cooley-

Tukey FFT is more complex algorithm. It is better on detecting pitch on instrument or singing voice than autocorrelation. However, in this researh autocorrelation is enough to be used due to its simplicity.

4. Research Methodology

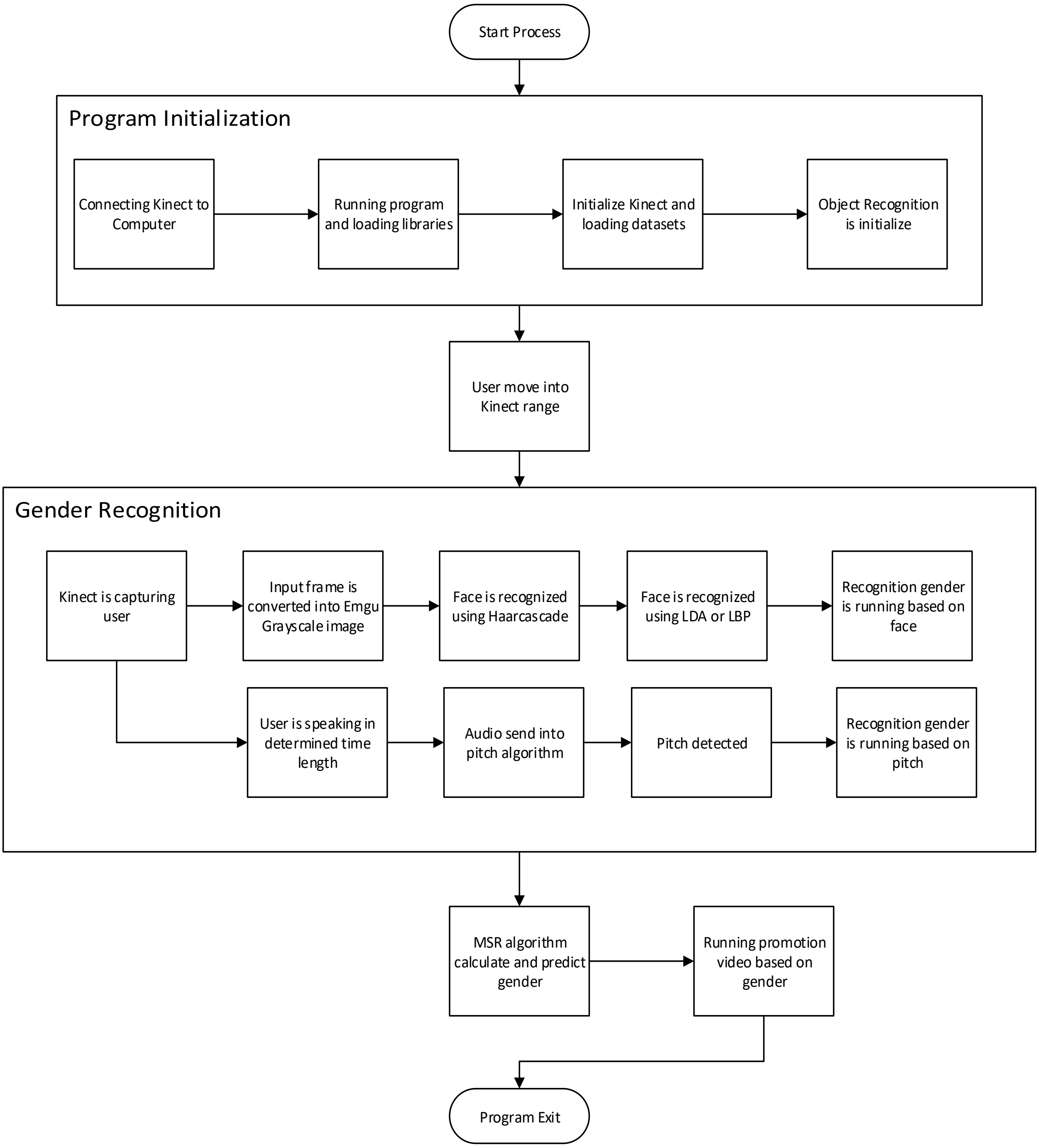

Fig. 1. Full System Workflow

This research uses two modalities to detect gender, including face and voice. In this research, each modality has different methods to recognize the gender as showed in Fig 1. First method of this research is using face to work with. EURECOM datasets of facial images is used as training set of image processing algorithm. PCA, LDA, and LBP from OpenCV library (OpenCV., 2015) (OpenCV., 2015) were compared with same training set to find out which one is better to use in this research. After some training and testing, LDA was elected.

After learning and testing gender recognition by using face is finished, the research is continued to another modality which is voice. To detect gender based on human voice, pitch detection algorithm must be used to determine frequency domain. Autocorrelation pitch detection algorithm is carried out to detect pitch of human voice.

After two modalities were developed, an algorithm must be constructed to calculate gender prediction based on those modalities result. Most Shown Result (MSR) algorithm is developed to calculate the final result for gender prediction based on most shown result from two modalities. This algorithm calculation steps are listed as follow:

1. Each modality generates predictions in determined time length.
2. Find most appeared prediction for each modality.
3. Compare both modality result with condition: If both results are the same, then the calculation is terminated and result is obtained. If both results are not the same, then most appears between both modalities is calculated. Then, the larger result will be the answer of the gender.

MSR algorithm is proposed in this research to carry out good prediction based on two modalities result.

5. Design of Experiment

There are several experiments conducted in this research, which are:

a. LDA Algorithm with Internet RGB Image Input (Not Real Time)

Images from internet are used in this experiment with several conditions. For example, female with glasses, male with long hair, male with hat, and many other conditions. The images are transformed into 256 x 256 pixels scale. Then, the system is run to read each picture and predicts the gender.

b. LDA Algorithm with Half RGB Datasets of EURECOM Input (Not Real Time)

Half of datasets is used as training data. Then, another half of datasets is used as sample pictures to detect gender. Then, program is implemented to read each picture and predict the gender from it.

c. Autocorrelation Pitch Detection with Internet Wav File (Not Real Time)

Wav files that contain human speech are downloaded and separated by folder based on gender. Then, each wav file is loaded into program to be analyzed by autocorrelation pitch detection. Then, the pitch is used to detect gender based on each gender frequency range.

d. LDA Algorithm, Autocorrelation Pitch Detection, and Combination of LDA and Autocorrelation with MSR Algorithm

LDA algorithm, Autocorrelation pitch detection, and combination of LDA and autocorrelation with MSR algorithm experiments are done at the same time. The system detects face frame and audio within determined time length.

6. Result and Discussion

a. LDA Algorithm with Internet RGB Image Input (Not Real Time)

Total testing in the first experiment (section 5.A) are 18 times with different gender and situation. The total correct prediction is 13 and wrong prediction is 5. Total percentage of the first experiment is 72 percent. Result of this experiment can be read on table I.

Based on the current evaluation, the taken picture should have the same format and condition with the trained data. In this research, testing images must be converted into 256 x 256 pixels format. Since several images from Internet have different size, they could be not good converted into the algorithm format. Therefore several failed prediction result has problem with the conversion step.

b. LDA Algorithm with Half RGB Datasets Of EURECOM Input (Not Real Time)

Testing accuracy for EURECOM male images testing were 75 percent. Meanwhile, EURECOM female images testing were shown smaller accuracy than the male images testing as figured in table II. The accuracy of EURECOM female picture testing were 49 percent.

From the second experiment (section 5.B) result, this research found that female is harder to be detected than male faces. It is caused by the total training set of female images is too small if just using half of EURECOM datasets of female faces.

Table I. Result with Internet Images

Gender	Testing Image Condition	Prediction Result Correct or Wrong
Male	Natural	Correct
	Glasses	Correct
	Long Hair	Correct
	Bald	Correct
	Sun Glasses	Correct
	Beard	Correct
	Hat	Wrong
	Bandana	Wrong
	Headphone	Wrong
	Wig	Correct
Female	Long Hair	Correct
	Short Hair	Correct
	Glasses	Correct
	Sun Glasses	Wrong
	Hat	Correct
	Bandana	Wrong
	Veil	Correct
	Beard	Correct

Table II. Result of Experiments

	Result			Accuracy rate
	Total Image	Predicted as male	Predicted as female	
RESULT OF FACE IMAGES WITH LDA				
Male faces	304	228	76	75%
Female face	111	57	54	49%
TOTAL ACCURACY RATE				49%
RESULT OF FACE IMAGES WITH LDA IN REAL TIME				
Male faces	37	32	5	86%
Female face	9	5	4	44%
TOTAL ACCURACY RATE				82%
RESULT OF AUTOCORRELATION PITCH DETECTION IN REAL				
Male voice	17	15	2	88%
Female voice	9	0	9	100%
TOTAL ACCURACY RATE				92%
RESULT OF MSR ALGORITHM IN REAL TIME				
Male	17	17	0	100%
Female	9	5	4	44%
TOTAL ACCURACY RATE				80%

c. Autocorrelation Pitch Detection with Internet Wav File (Not Real Time)

In the third experiment (section 5.C), this research were using wav files of human speech that are downloaded from internet. Wav files were tested one by one with pitch detection algorithm using the same sample rate settings. Sample rate were set in 16.000 Hz. Percentages of third experiment success were 40 percent. From the third experiment, the wav files have different setting, such as different of sample rate while it was recorded.

d. LDA, Autocorrelation Pitch Detection, and Combination of LDA and Autocorrelation with MSR Algorithms

For the fourth experiment (section 5.D), LDA algorithm in real time condition were having 82 percentage of success. Detail of LDA algorithm result using real time stream from Kinect (Microsoft., 2015) is showed in table II.

Meanwhile, Autocorrelation pitch detection were having 92 percent in real time condition using Kinect microphone array. Detail of autocorrelation pitch detection result using real time audio stream from Kinect can be seen in table II.

In this experiment, MSR algorithm is used to detect gender based on face and voice gender recognition result. MSR algorithm were having 80 percent of success detection. MSR algorithm result can be seen on table II.

From the last experiment, the accuracy of female detection by using veil is always wrong in real time. This error detection might be occurred because the training set of female images has no images of female with veil. And also, this research has figured out about the error detection might be occurred because the training dataset were not included with Indonesia people, while the volunteers are from Indonesia.

7. Conclusion

From the experiment, gender recognition using LDA were not highly accurate when training set were not diverse enough to use with many kind of race. This factor affects most the result of experiment. While, Autocorrelation reached 92 percent in real time pitch detection. Means, gender recognition using audio modalities has highly recommended to use. For final recognition result, MSR algorithm must be enhanced and tuned well in the future. MSR algorithm still has several drawbacks for making fair and accurate prediction based on multi modal factors. Based on the real time experiment result, the prediction based on face is still dominant on MSR algorithm calcution. This domination occurs when face modality in determined length of time has more prediction than audio input prediction. That domination must be fixed in the future development of the MSR algorithm.

References

Arnulf, A.G.B. Felix, A.W. (2002). *Gender Classification of Human Faces*.

Chowney V. (2012). *Gender-based marketing: infographic. [Online]*. Cited 2015 May 13. Available from Hyperlink: https://econsultancy.com/blog/9194-gender-based-marketing-infographic/.

Heath, M. (2011). *Autotune.NET. [Online]*. Cited 5 5 2015. Available from Hyperlink: https://channel9.msdn.com/coding4fun/articles/AutotuneNET.

Huang, D. Caifeng, S. Ardabilian, M. Yunhong, W. Liming, C. (2011). *Local Binary Patterns and Its Application to Facial Image Analysis: A Survey. Systems, Man, and Cybernetics, Part C*. Applications and Reviews: IEEE Transactions on Biometrics Compendium.

Hyunh, T. Min, R. Dugelay, J. (2012). *An efficient LBP-based descriptor for facial depth images applied to gender recognition using RGB-D face data*. Workshops: In Computer Vision-ACCV. eurocom.fr.

Kumar, A. Gov. Eng. Coll. AI, Rawat K, Gupta, D. (2013). *An advance approach Of PCA for gender recognition.* International Conference: Information Communication and Embedded Systems (ICICES).

Matta, F. Saeed, U. Mallauran, C. Dugelay, J.L. (2008). *Facial gender recognition using multiple sources of visual information*. Multimedia Signal Processing: IEEE 10th Workshop.

Microsoft. (2015). *Kinect feature. [Online]*. Cited 2015 5 5. Available from Hyperlink: http://www.microsoft.com/en-us/kinectforwindows/meetkinect/features.aspx.

OpenCV. (2015). *Gender Classification with OpenCV. [Online]*. Cited 2015 5 5. Available from Hyperlink: http://docs.opencv.org/trunk/modules/contrib/doc/facerec/tutorial/facerec_gender_classification.html.

OpenCV. (2015). *Face Recognition with OpenCV. [Online]*. Cited 2015 5 5. Available from Hyperlink: http://docs.opencv.org/modules/contrib/doc/facerec/facerec_tutorial.html.

Parker, R. (2015). *http://www.radfordparker.com/papers/. [Online]*. Cited 2015 5 5. Available from Hyperlink: http://www.radfordparker.com/papers.

Saraswathi, M. Muthiah, K. (2015). *Improving Gender Recognition Rate in Face Recognition System Based On Linear Discriminant Analysis Technique*. International Journal: ijarcsms.com.

Wagner, P. (2011). *Principal Component Analysis and Linear Discriminant Analysis with GNU Octave. [Online]*. Cited 2015 5 13. Available from Hyperlink: http://www.bytefish.de/blog/pca_lda_with_gnu_octave.

Zhang, Z. Kipman, A.A.A. (2014 March 18). *Inventors; multi-modal gender recognition including depth data*. United States of America: Patent US 8,675,981 B2.