Learning smiling expression for Human-robot Interaction: A Malaysia Case Study

Winal Zikril Zulkifli[1], Syamimi Shamsuddin[1], Lim Thiam Hwee[2], Ahamad Zaki Mohamed Noor[1]

[1]Faculty of Manufacturing Engineering, Universiti Teknikal Malaysia Melaka, Melaka, Malaysia
[2]Psychosocial Department, SOCSO Tun Razak Rehabilitation Center, Melaka, Malaysia

*Corresponding e-mail: syamimi@utem.edu.my

Keywords: supervised learning; OpenCV; smile detection; human-robot interaction

ABSTRACT – In human-robot interaction (HRI), outcome assessment is crucial. This paper presents a tool constructed for HRI analysis based on machine learning. Data preparation for the machine learning involved a small program whose function is to first process a dataset which contains a number of face images. Then, the dataset is fed through a machine learning algorithm. Finally using live detection from a webcam, the live input for the program is used to predict whether or not the person is smiling during HRI. The constructed dataset in this study was successfully able to detect the human smile.

1. INTRODUCTION

Human-robot interaction (HRI) is an emerging field of study committed to design, understand and examine the framework of robotics for human use [1]. In the field of HRI, recognizing the interaction between human and robot requires a tool to analyse the outcome. Machine learning can be a tool for analysing and it is divided into three main categories which are supervised learning, unsupervised learning and reinforce learning. To classify an image, supervised learning is to choose to employ this task. Support Vector Machine (SVM) is used as a classifier for datasets with cascade. One of the analyses that can be used is by recognizing a human's mood while interacting with the robot. A smile can be categorized as a positive emotion and image processing is one of the applications to identify the changes in the human face.

One of the ingredient to construct programming for image processing is a dataset. Dataset is the source of data to be feed in image processing algorithm [2]. The often problem found with researcher was to acquire dataset for predicting and categorize [3]. The problem can be found from the dataset is from the dataset provided is not like on specific topic researcher working on, the database is not kept current and Application Programming Interface (API) are either expensive. One the solution to this problem train the new dataset and convolute it into existing dataset. This research aims to build a dataset that has high accuracy to detect a person in smiling as positive emotion especially in Malaysia. The new dataset introduces sample image from Malaysian.

2. METHODOLOGY

When we enter into the virtual world, we literally surrounded by data, since it is fundamentally building a block of everything we see. Data is raw information, its representation of human and machine toward observation of the world [4]. Before feeding datasets into machine learning module, a dataset must be prepared. This research codes basically identify the person either it is smiling or not, by training on the dataset.

It is important to get the exact data. For machine learning, the kind of dataset require is the one that able to predict a column from the other column in a dataset. The criteria for machine learning dataset form the data isn't messy and clean data, there is interesting target column to make a prediction and the other variables have some explanatory power for the target column. There is known public access to data from [5] [6]. In [5] the data come out form data science community that hosts machine learning competitions. The contribution data came up from a variety of externally-contributed data sets on the storage. Usually, a website that offers dataset in the form of API. In the API from, a dataset is easier to use. But if it doesn't, the dataset can be constructed from the library by web scaping using python. It is done by scaping HTML data and convert into a file such as XLS, CSV and SQL.

The second step to prepare a dataset is by the process the data. A set of data function was written in python program to extract the data. A prepared dataset for smiling person come in the form of the standard image size of 20x20 pixel while preserving aspect ratio. This data came out in the format of CSV. The prepared data in CSV format have to convert to XML format to be used in the smile recognition program. Once the data is formatted to XML, data imputation is performed by parametric machine learning to ensure all the instances have complete attributes to be useful for the supervised learning. The data have to be iterated by each row to check the value either it is empty or not. The model works best by selecting only specific features are to be used.

The last step for preparing a dataset is transforming the data. One the smile is detect in facial regions the data transform is by decomposition the data. The features that are too complex like the other features on the face are decompose beside the lips only. By excluding the unimportant features, the model can be performed more accurate. The selected features and class label are transformed into vectors. Vectors are numerical representation of features. This vector is feed

into neural network directly. Once the data has been processed, it is feed into graph. When the model is run it is separated into smiling and not smiling

ROI specify where the features is specified. Through OpenCV programming, the pulled images are resizing as 20x20 pixel. The larger size of image would make it exponentially longer for trainer to run. Small size is enough for the decent output result. The resize position is necessary the increase teaching efficiency. And those images also converted into grey scale images. The specify images both positive and negative are converted into .txt file as the background information. The quantity images of positive and negative images are at ratio 2:1. The ratio of 2:1 was chosen because classifier need to prior the positive images more than negative images.

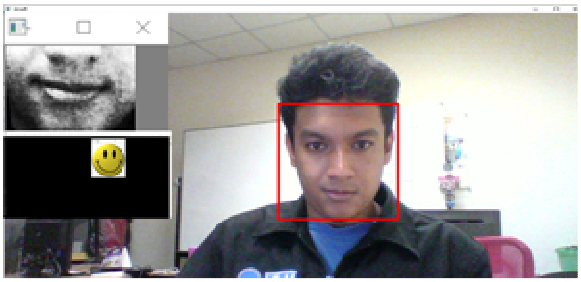

Figure 1 Testing Result on OpenCV

The stages for classifying was chosen at the quantity of 100 stages. This high stage takes almost half a day with average computing power. Minimum suggested running RAM are about 2GB is recommended for this training. When a dataset is completed we use it as Haarcascade classifier in the form of xml file. Finally, the completed data are feed into OpenCV programming as shown in Figure 1. When dataset applied, the program capable to detect a human in smiling condition. Successful results of smile detection are shown in Figure 2. This result showed that this program capable to detect different types of smiles. Given these points, the dataset that has been building is considered success.

Figure 2 Sample Results on Smile Detection

3. RESULTS AND DISCUSSION

Figure 3 shows a some of the collective datasets with a sample of 300 humans in smile condition. The instances where the person is smiling is considered 'positive images, which are labelled as '1'. Other instances where people are not smiling are labelled as '0'. 900 negative images are prepared for sample other than people with the smiling condition. The accuracy match with Gabor-feature based support vector machine. This approach provides 81% accuracy.

Figure 3 Malaysian Dataset for Human Smiling

4. CONCLUSION

In this paper, the dataset that was built is capable to learn smiling expression for HRI that provide the same image characteristic with the existing database to enable other researcher use them easily. In the future research, this data will be going to use to analyses rehabilitation patients with depression in Malaysia. In short, image processing application used in HRI still in early stage.

ACKNOWLEDGEMENT

The authors gratefully acknowledge the Ministry of Higher Education Malay Asia, Universiti Teknikal Malaysia Melaka and Tun Razak Rehabilitation Center Melaka for their support. This project is funded under the Fundamental Research Grants Scheme (FRGS) [FRGS/1/2016/SKK06/FKP-AMC/ F00321].

REFERENCES

[1] S. Shamsuddin, H. Yussof, S. Mohamed, F. A. Hanapiah, and H. A. Ainudin, "Telerehabilitation Service with a Robot for Autism Intervention," in *Procedia Computer Science*, 2015, vol. 76, pp. 349–354.

[2] M. Egmont-Petersen, D. de Ridder, and H. Handels, "Image processing with neural networks—a review," *Pattern Recognit.*, vol. 35, no. 10, pp. 2279–2301, 2002.

[3] J. Wu, H. I. Christensen, and J. M. Rehg, "Visual place categorization: Problem, dataset, and algorithm," in *Intelligent Robots and Systems, 2009. IROS 2009. IEEE/RSJ International Conference on*, 2009, pp. 4763–4770.

[4] H. He and E. A. Garcia, "Learning from imbalanced data," *IEEE Trans. Knowl. Data Eng.*, vol. 21, no. 9, pp. 1263–1284, 2009.

[5] Kaggle, "Data Science Bowl 2017 | Kaggle," *Kaggle*, 2017. [Online]. Available: https://www.kaggle.com/c/data-science-bowl-2017.

[6] Quandl, "Quandl," 2018. [Online]. Available: https://www.quandl.com/. [Accessed: 26-Feb-2018].