The Intel Realsense Depth-Camera Performance for Real-Time Customer Satisfaction Analysis using Facial Expression Detection

James Purnama, Jason Yapri, Tommy Winarta, Stefanus Oliver, Maulahikmah Galinium
Faculty of Engineering and Information Technology, Swiss German University, Indonesia
james.purnama@sgu.ac.id

Abstract. Customer satisfaction evaluation is an important process in business development. Unfortunately, doing it manually by using survey or by phone will take up the time and resources that can actually be reallocated. In this research, we are going to make an automated customer satisfaction analysis tool by using Intel RealSense. The depth sensor and its emotion recognition API is enable real-time 3-D customer emotion recognition monitoring accurately. Emotions such as anger, disgust, fear, joy, sadness, and surprise that is expressed by the customer is logged. Then, the captured emotions are processed and analyzed by the computer. Finally, the summarized survey presents an interactive data in form of charts that can show the customer service’s performance. As a result, the data that is produced can be used to improve the performance of the customer service representative. In this research, experiments show that while Intel RealSense can capture customer's emotions well most of the time, there are still some cases in which it couldn't present the accurate result due to the limitation in the API. Therefore, further testing need to be done that also takes several factors e.g. head accessories, glasses into account to increase the credibility of the output from the program.

Keywords: Customer Satisfaction, Emotion Recognition, Depth Sensor, Intel Realsense.

1. Introduction
Customer satisfaction plays a major role in service companies. These companies include major banks and insurance companies. Evaluating customers’ response can be used as a reference to improve the current customer service’s performance. For instance, it is much harder to acquire new clients rather than keeping the current ones. 7 out of 10 Americans are expected to be returning customers as long as they are happy with the services provided [1]. Unfortunately, measuring customer satisfaction is still done conventionally, such as the supervisor observed the customer service directly, doing a survey via telephone or questionnaire. This requires more resources and inefficient.

Therefore we creates an analysis program to detect customer emotions from the image captured by Intel RealSense depth camera and produce a detailed report based on customer’s real time emotion. For this research, our current scope is only one person in front of the camera as the customer at a particular time. Furthermore, we distinguish each customer by assuming that whenever the presence of a customer
is not detected for a certain period of time, the next face which will be detected, is a different customer. Our program is limited to computers with USB 3.0 I/O port that runs on Windows 8.1 (64 bit) and equipped with 4th generation Intel processors and 4GB of memory. Other than that, the range of the sensor is also limited about 20-120 cm.

2. Related Works
Reference [2] resulted a system that can detect head movement, gaze direction, and expression of the user for human robot interaction. The researcher designs the experiment by comparing the use of single camera in their project with a stereo camera pair. The data that is received is then processed for further uses. The advantages of using a single camera over a stereo camera for human computer interaction are as follows: larger freedom in camera installation, faster speed in processing, wider measurement range. However the utilization of single camera in this research has shown that the process of face finding is more difficult because the depth cannot be directly obtained. However there is still issue as we still need to calibrate the face to reconstruct the 3D facial model using a stereo camera or 3D scanner as an initial step.

Reference [3] aimed to create a tool that will alert car driver of a possible undesired event, such as the driver is sleepy. Thermal camera is used to detect the driver’s face and facial expression. While reference [4] used webcam to implement real time and robust 3D face tracking framework. Its system can also be integrated with face recognition module. References [5],[6] used Kinect Depth Sensor to recognize the real time facial emotion recognition while reference [7] used Kinect Depth Sensor to detect gaze-tracking based on head pose estimation. Since Intel Realsense is still a new product, there are no many researches using this depth camera. Therefore we used Intel Realsense to conduct the research. The advantage is Intel Realsense camera can detect more dots in the face in short distance, which is suitable for our experiments.

3. Methodology
3.1 Architecture and Activity Diagram
The figure 1 describes the architecture diagram of our system. The depth camera consists of IR Sensor, Color Sensor, IR Laser Projector, and Imaging Processor used for capturing facial expression. The camera is connected to computer through USB port 3.0. Ms. NET Framework is used to connect with the Intel RealSense SDK. This SDK detects the captured facial expression and produce it as the raw data. Then, the detected facial expression is sent into the application, named as CUSALICS. This application has a relation with MySQL Database Server to store the data.
Figure 1: Architecture Diagram

Figure 2 describes the activity diagram of the program. There are six contributors of this process. Three of them are human, i.e. a customer service, a customer, and the supervisor. The other three are hardware (Intel RealSense depth camera) and two softwares (Intel RealSense SDK and Cusalis). The process is started when the customer service serves a customer. When the first time they make conversation, logically, a customer will express whatever he or she is feeling at that time. Then, the Intel RealSense Depth Camera directly captures when the conversation starts. It captures any expressions from the customer, including when the customer does not express or gives a straight face. Then, the captured facial expression is forwarded into the Intel RealSense SDK. It detects coordinate points of the captured facial expression. Then, it gives a result of the detected expression in the form of raw data. After that, the detected expression is passed to CUSALICS application as a data. It arranges the detected expression from the starting time until the ending time of every detected expression. CUSALICS application produces the report of analysis. Finally, the supervisor gives a review about the report result and does an evaluation.

Customer Service | Customer | Intel RealSense Camera | Intel RealSense SDK | CuSaLics Application | Supervisor / Manager / HRD
--- | --- | --- | --- | --- | ---
Serve the Customer | Facial Expression | Capture Facial Expression | Process the raw data | Produce Result | Review The Report
| | | | Manage the report | | Do Evaluation

Figure 2: Activity Diagram

3.2 Algorithm Flowchart
An array of PXCMEmotion.EmotionData[] holds all of the emotion that are captured from the depth camera. Then, the program takes the most outstanding emotion (with the highest evidence and intensity) by iterating through the array of EmotionData. The array itself has the size of 7 according to the number of emotion that can be captured by the sensor. The iteration number (index) is stored if the most outstanding emotion has been found. As the loop ends, only if the intensity has the value more than 0.4 (to make sure that the emotion that is detected has a high or medium intensity) and an emotion is actually found along the iteration (by making sure that the index of the iteration has a positive value), the index that has been previously acquired is used to identify which emotion is being expressed. Detail algorithm can also be read in Figure 3.

![Algorithm Flowchart](image)

**Figure 3: Algorithm Flowchart**

In the system there are two charts to be used to analyse the result, which are a line chart and a pie chart (can be seen in figure 4 and 5). In line chart, it contains the information of the initial expression
of the customer, the changed expression of the customer, and how long the expression of the customer
occurs. Intel RealSense gives its raw data in the current emotion, current intensity and evidence, and the
timestamp everytime there are any changes in the emotion and intensity. We get all the raw data, then
we represent them on the line chart which can be seen in Figure 4. The X axis represents time, and the
Y axis represents the 7 emotions. For instance:
  0 represents anger
  1 represents contempt
  3 represents disgust
  4 represents joy
  5 represents sadness
  6 represents surprise

For example, if the line represents number 1.56. It means the current customer is on the contempt
emotion, with intensity 0.56. The intensity calculation would be [1.56 – (Base number/1)]. In the pie
chart, it represents the processed data from the line chart, which can be seen in Figure 5. For each
emotion, we measured the sigma which is divided by the total of intensity. Then, divided again by the
spike frequency. The formula is the total of intensity is divided by the spike frequency. Then, after
getting the value, we create the percentage from the value of the other emotions. For example: The
expressions detected are joy and fear.

JOY : 14.957884669303894 / 5 == 2.99157693386 / [Total Value] * 100 == 74.423%
FEAR : 5.140446037054062 / 5 == 1.0280892074108 / [Total Value] * 100 == 25.58%
Total Value : 4.019666141270812

4. Result

4.1 Experiment to Test the Accuracy

The facial expression of people are recorded when they watch two type of videos, which are very
touching and very funny videos. The very touching video would be played prior to the very funny video.
There are 3 men and 3 women as respondents. They are able to express their feeling through their face,
i.e. they can laugh when it’s funny or they feel sad when the condition should be sad. The experiment is
conducted in a good lighting condition room. The respondent sits on the chair in front of the 1st laptop.
The 1st laptop contains two videos and also the webcam. The 2nd laptop stands behind the 1st laptop
with Intel RealSense Depth Camera facing to the respondent. The 2nd laptop has been configured with
Intel RealSense SDK. Furthermore, the results of the recorded video between webcam and Intel
RealSense depth camera are compared to produce the same result. Then, its result is compared with the raw data produced by Intel RealSense SDK to see the accuracy. It means how many seconds the similarity detection between the human facial expressions compared to the detected emotion by the program. The duration of touching video is 181 s while the funny is 74 s.

Based on the experiment result in table 1, we can see that the accuracy are still low for all respondents. One of the reasons is there are many neutral expression shown with different other facial expressions. We realized when watching video or film, we do not give our facial expression sad or joy continuously, sometime we need to focus or serious which is detected as neutral. Therefore, we realized that we need a label for neutral expression. In addition to the accuracy, Respondent F can’t be detected for both sad and joy expression due to the effect of wearing hijab. This still remains challenge for such person.

**Table 1. Accuracy of Watching Touching and Funny Video**

| No. | Respondent | Accurateness for Touching Video | Accurateness for Funny Video |
|-----|------------|---------------------------------|-----------------------------|
| 1   | A          | 101 seconds, 55.8%              | 24 seconds, 32.43%          |
| 2   | B          | 37 s, 20.44%                    | 27 s, 36.49%               |
| 3   | C          | 78 s, 43.09%                    | 12 s, 16.22%               |
| 4   | D          | 52 s, 28.73%                    | 25 s, 33.78%               |
| 5   | E          | 59 s, 32.6%                     | 3 s, 4.05%                 |
| 6   | F          | 0 s, 0%                         | 0 s, 0%                    |
|     | Average    | 54.5 s, 30.11%                 | 15.17 s, 20.5%             |

4.2 Experiment to Test the Satisfaction Compared to the Report

This experiment is conducted to simulate the customer satisfaction scenario. There are 2 scenarios. The first scenario is used for satisfied customer while the second scenario is used for unsatisfied customer. The first scenario is initialized with the happy customer and the customer service is able to keep the joy of the customer. The second scenario contains the customer who initially has been a problem and he/ she is angry to the bank. During the process, the customer becomes more angry since the customer service cannot solve the customer problem. At the end, the customer feel disapponted to the service of the customer service in the bank. The produced report from the first scenario can be seen in Figure 6. The report of first scenario mentions the service is satisfied by showing the dominan of joy facial expression. When the customer gives neutral expression, the program detects his facial expression as disgust or sadness. However, everytime the customer actor acts that he has to show that he is happy, his facial expression is detected as joy. The pie chart shows that the facial expresssions of the respondent 1 are 60.24% disgust, 22.42% joy, 17.35% sadness.

![Figure 6: The Line and Pie Chart Result of Scenario I](image-url)
While the report of second scenario shows the service is not satisfied by showing the dominant of bad facial expression, such as anger, contempt, or sadness. When the customer gives neutral expression, the program detects his facial expression as disgust or sadness. When the respondent acts he is angry, the detected facial expressions becomes fear, surprise, angry, disgust, or sadness itself. The pie chart shows that the facial expressions of the respondent 1, which are 34.48% sadness, 31.18% anger, 19.92% disgust, 10.15% surprise, and 4.27% fear. This result can be read in Figure 7. We realized to reduce facial expressions quantity to only three expressions: joy, sadness, and anger because these three facial expressions are more useful for the evaluation than the other four expressions.

![Figure 7: The Line and Pie Chart Result of Scenario II](image)

5. Conclusion
In conclusion, our program paired with Intel RealSense depth camera has been successfully implemented as a real-time automated customer satisfaction analysis tool. Unfortunately, there are still many data that shows inaccuracies because of the device's and API limitations. Therefore, further testing in real cases is required to make sure that it is suitable for organization needs to assess customer’s satisfaction level. In addition to the device, algorithm improvement is also needed especially for neutral/flat expression.

References
[1] “75 Customer Service Stats and Facts you Can’t Afford to Ignore.” [Online]. Available: http://www.helpscout.net/75-customer-service-facts-quotes-statistics/. [Accessed: 08-Apr-2015].
[2] Y. Matsumoto, N. Sasao, T. Suenaga, and T. Ogasawara, “3D model-based 6-DOF head tracking by a single camera for human-robot interaction,” in Robotics and Automation, 2009. ICRA ’09. IEEE International Conference on, 2009, pp. 3194–3199.
[3] A. Kolli, A. Fasih, F. Al Machot, and K. Kyamakya, “Non-intrusive car driver’s emotion recognition using thermal camera,” in Nonlinear Dynamics and Synchronization (INDS) 16th Int’l Symposium on Theoretical Electrical Engineering (ISTET), 2011 Joint 3rd Int’l Workshop on, 2011, pp. 1–5.
[4] J. Choi, Y. Dumortier, S.-I. Choi, M. B. Ahmad, and G. Medioni, “Real-time 3-D face tracking and modeling from webcam,” in Applications of Computer Vision (WACV), 2012 IEEE Workshop on, 2012, pp. 33–40.
[5] M. Kakarla and G. R. Mohana Reddy, “A real time facial emotion recognition using depth sensor and interfacing with Second Life based Virtual 3D avatar,” in Recent Advances and Innovations in Engineering (ICRAIE), 2014, 2014, pp. 1–7.
[6] S. Polikovsky, Y. Kameda, and Y. Ohta, “Facial micro-expressions recognition using high speed camera and 3D-gradient descriptor,” in Crime Detection and Prevention (ICDP 2009), 3rd International Conference on, 2009, pp. 1–6.

[7] M. Galinium, Yulius, J. Purnama, “Gaze-tracking based on head-pose estimation using depth camera” in International Journal of Applied Engineering Vol.11 No.15, 2016, pp. 8656-8661.