Customer satisfaction measuring based on the most significant facial emotion
Mariem Slim, Rostom Kachouri, Ahmed Atitallah

To cite this version:
Mariem Slim, Rostom Kachouri, Ahmed Atitallah. Customer satisfaction measuring based on the most significant facial emotion. 15th IEEE International Multi-Conference on Systems, Signals & Devices (SSD 2018), Mar 2018, Hammamet, Tunisia. hal-01790317

HAL Id: hal-01790317
https://hal-upec-upem.archives-ouvertes.fr/hal-01790317
Submitted on 11 May 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Customer satisfaction measuring based on the most significant facial emotion

Mariem Slim
University of Sfax, LETI laboratory
ENIS, BP W
3038 Sfax, Tunisia
mariem.slim@esiee.com

Rostom Kachouri
University of Paris-Est,
Gaspard-Monge Computer Science Laboratory
A3SI, ESIEE Paris, CNRS, France
rostom.kachouri@esiee.fr

Ahmed Ben Atitallah
University of Sfax, LETI laboratory
ENIS, BP W
3038 Sfax, Tunisia
ahmed.benatitallah@enetcom.usf.tn

Abstract—Customer satisfaction (CS) measuring has become one of the strategic tools for companies. Many methods exist to measure customer’s satisfaction. However, generally results could be wrong. Otherwise, CS can be deduced from customers emotion state. In this paper, we propose a new end-to-end method for facial emotion detection. For that, six new characteristic features are proposed. We characterize especially the most significant emotions namely “Happy”, “Surprised” and “Neutral”. The proposed method is invariant to camera position. A very challenging datasets as Radboud Faces and Cohn-Kanade (CK+) are used for study and evaluation. Obtained results show that our method reaches a high recognition accuracy and outperforms Action Unit (AU) features based Support Vector Machine (SVM) and K-Nearest Neighbors (KNN).

Index Terms—Facial expression, emotion recognition, most significant emotions, feature extraction, action unit, SVM, KNN.

I. INTRODUCTION

In an increasingly competitive world, negative reviews of customers can have an important impact on a business. Indeed, customer satisfaction is actually measured in many applications such as shops, banks, mobile apps and others. Many methods and metrics are designed. In instance, Customer Satisfaction (CSAT) is asked using emails, phones, face-to-face interview, tablets and touchpads. As well, Net Promoter Score (NPS) is asking about the intention of product or service recommendation and consumer respond by assigning a rating or a number of stars. However, collected information with these methods could be wrong.

In another way, human interaction is recognized essentially through facial expressions which provide a natural and compact way for humans to convey their emotional state to another party. In addition, nowadays, supervision cameras are everywhere. Thus, we can consider that facial expressions are accessible and we can deduce whether the customer is satisfied or not based on his emotional state.

Across different cultures, we found seven universally recognizable emotions [1]: anger, disgust, fear, happy, sad, surprised and neutral. However, in this work, we note that no need to recognize all emotions. In fact, for customer satisfaction, only “Happy”, “Surprised” and “Neutral” are significant. In this paper, we propose an original and performing method for customer satisfaction measuring based on the most significant facial emotion.

The organization of the paper is as follows: In section II, related works are presented. Then, we propose a novel facial emotion recognition in Section III. Experimental results are discussed in Section IV and finally conclusions are drawn in Section V.

II. RELATED WORKS

For the first time, Charles Darwin demonstrated that despite their differences and countries, human beings have innate and common emotion [2]. Historically, there are seven types of human emotions: anger, disgust, fear, happy, sad, surprised and neutral (Fig. 1). Paul Ekman later conducted several experiments and observations on a large number of individuals and confirmed that emotions are universal recognizable expression [2].

![Fig. 1: The basic emotions.](image)

Many approaches are proposed to recognize facial expressions. They can be split into two broad categories: appearance-based approaches and features-based approaches. Appearance-based approaches detect emotion via a model which illustrate the general facial shape and texture ([3], [4]). On the other hand, features-based approaches extract person’s characteristics. Then, extracted features can be combined and classified
Generally, appearance-based approaches are very time-consuming and the human face appearance could be affected due to variations of lighting conditions. Therefore, we consider in this paper features-based approach.

Facial emotion recognition systems comprised essentially three steps [6] (Fig. 2). The first step is face and landmarks detection. For that, several open source facial behavior analysis systems exist. In instance, CHEHRA\(^1\) [7], TCDCN\(^2\) face alignment tool [8], CLandmark\(^3\) [9], OpenFace\(^4\) [10]... etc. We note that OpenFace is the most recent fully real time and open source system [10].

Feature extraction is the second step of facial emotion recognition system. P. Ekman and W. Friesen developed the Facial Action Coding System (FACS) to describe movements (action units) caused by expansion and contraction of facial muscles [11]. Using FACS, human coders can manually code almost any anatomically possible facial expression [12]. More recently, Hammal et al propose the use of just 5 characteristic distances as features [5]. They supposed that all the necessary information is contained in the castration of the eyes, mouth and eyebrows.

The final step is emotion recognition. Many features based emotion classification methods exist. Two main categories can be identified: supervised and unsupervised classification techniques [13]. Unsupervised learning or clustering don’t require human intervention and don’t need a prior knowledge or predefined categories [14]. In the other side, supervised learning requires an expert labeled database. The process is performed in two phases. During the learning phase, a model is determined from the labeled data. Using the learned model, the test phase called consist on predicting the label of a new non labeled data [15]. We note that, supervised learning is more efficient and used to resolve linear and non-linear classification problems [16]. The most popular supervised machine learning algorithms [17] are: Support vector machine (SVM) [18], k-nearest neighbors (KNN) [19], Convolutional Neural Networks (CNN) [20]. However, generally these methods could require a large number of attempts to move towards the best possible recognition performance [21].

Up to our knowledge, no facial emotion recognition model has yet proposed an affordable and performing system for customer satisfaction. Otherwise, our method deals with emotion recognition from images, and focus on feature extraction and analyze. To measure customer satisfaction, we note that in the rest of this paper, we consider only three significant emotions: “Happy”, “Surprised” and “Neutral”.

III. PROPOSED METHOD

Our goal is to propose a feasible and especially performing method. We use OpenFace for face and landmarks detection (see subsection III-A). Then, we present our extracted features in subsection III-B. To recognize the most significant emotions, we propose in subsection III-C a new approach to classify features. To develop a working model, we use Radboud Faces Database [22]. It is a highly standardized set of pictures. It contains 67 subjects displaying 8 emotional expressions.

A. Face and landmarks detection

OpenFace is an open source real-time facial behavior analysis system [10]. As shown in Fig. 3, it is able to estimate head pose and detect facial landmarks. It uses for that the recently proposed Conditional Local Neural Fields (CLNF) [10]. OpenFace can also recognize facial action unit and estimate eye-gaze. It uses a modern C ++ toolkit called Dlib library [26]. It can be used free of charge in any application because Dlib’s license is free.

Fig. 3: OpenFace: face and landmarks detection.

The provided landmarks are 68 points on the face such as the outline of the mouth, the eyes, the eyebrows, the noise and the facial contour. The facial landmark detection aptitude was evaluated on four sub-dataset Annotated Faces in the Wild (AFW) [27], IBUG [28], LFPW [29], and Helen [30]. It was also compared to other facial landmark detection algorithms. Implementations’ approaches are available online and have been trained to detect the same facial landmarks (or their subsets). The baselines were: Discriminative Response Map Fitting (DRMF) [31], tree based deformable models [32], extended version of Constrained Local Models [33], GaussNewton Deformable Parts Model (GNDPM) [34], and Supervised Descent Method (SDM) [35]. The train and test datasets are different for all of the methods. The obtained results affirmed that OpenFace is state-of-the-art performance.

B. Feature extraction

Our method is based on feature-based approaches. We proposed seven features that we extract from the previously presented face and landmarks. First of all, as shown in Fig. 4, we compute the frame’s diagonal (D) from the detected face. This feature corresponds to the customer’s position from camera. Using the feature D makes our analysis invariant to the position of the face with respect to the camera.

In addition to the diagonal (D), as illustrated in Fig. 5, new characteristic features named \(F_1\) to \(F_6\) are defined as...
where:

- \( F_1 = \frac{F_{1L} + F_{1R}}{2} \): the opening of the eyes,
- \( F_2 = \frac{F_{2L} + F_{2R}}{2} \): the distance between the eyebrow and the corner of the eye,
- \( F_3 \): the width of the mouth,
- \( F_4 \): the opening of the mouth,
- \( F_5 = \frac{F_{5L} + F_{5R}}{2} \): the distance between the nose and the tip of the mouth,
- \( F_6 \): the length of the lower inner lip.

**C. Emotion recognition**

Our aim is to characterize each “Happy”, “Surprised” and “Neutral” emotion by a specific hyper-planes which allow to recognize the corresponding emotion depending on the associated characteristic features. The conducted hyper-planes have to be invariant to the frame’s diagonal (D) of each face. Radboud facial database\(^5\) \([22]\) is used. First, we evaluate each extracted features according to the number of images (see Fig. 8(a)-(f)). According to this feature study and based on predicted behavior of each extracted feature for the most significant emotions, we present in the following our proposed hyper-planes for each emotion.

1) **Happy emotion**: We note that in this case, \( F_1 \) have to be small because eyes are little closed. In addition, generally, smiling increases the values of \( F_2 \) and \( F_6 \) and decreases \( F_5 \)’s value. Equation 1 presents our proposed feature combination \( F_h \) to well distinguish the “Happy” emotion:

\[
F_h = F_3 + F_6 - F_5 - F_1
\]  

\(^5\)http://www.socsci.ru.nl:8180/RaFD2/RaFD?p=main

Fig. 6 illustrates the evaluation of the new \( F_h \) representation of the most significant emotions versus the frame’s diagonal D. We can see that the “Happy” emotion curve can be entirely separated from the two other emotions according to the affine hyper-plane \( Y_h \) (equation 2).

\[
Y_h = A_h \times D + B_h
\]  

Where \( A_h = 0.21 \) is the slope and \( B_h = -1 \) is the \( Y_h \)-intercept.

Fig. 6: \( F_h \) representation of the most significant emotions versus the frame’s diagonal D within Radboud Face database.

Finally, the recognition of the “Happy” emotion is ensured through the equation 3:

\[
F_h > Y_h \implies “Happy” \text{ emotion}
\]  

2) **Surprised emotion**: In this case, we note that \( F_2 \) value have to be high because eyebrows are raised. In addition, the mouth is like the letter ”O” so it increases the value of \( F_4 \) and decreases \( F_3 \) value. Thus, the “Surprised” emotion is distinguished by the feature combination \( F_s \) (equation 4):

\[
F_s = 2 \times F_2 + 3 \times (F_4 - F_3)
\]  

The evaluation of the new \( F_s \) representation of the most significant emotions versus the frame’s diagonal D is illustrated in Fig. 7.

Fig. 7: \( F_s \) representation of the most significant emotions versus the frame’s diagonal D within Radboud Face database.

The “Surprised” emotion curve can be entirely separated from the two other emotions (Fig. 7) according to the affine hyper-plane \( Y_s \) (equation 5).

\[
Y_s = A_s \times D + B_s
\]
Fig. 8: Behaviour feature study: (a) $F_1$, (b) $F_2$, (c) $F_3$, (d) $F_4$, (e) $F_5$, (f) $F_6$, curves of the most significant emotions according to 3*67 images from Radboud Face database.

Fig. 9: The three proposed feature representation of the most significant emotions versus the frame’s diagonal D within CK+ database: (a) $F_h$ for “Happy” emotion (b) $F_s$ for “Surprised” emotion and (c) $F_n$ for “Neutral” emotion.
Where $A_s = -0.03$ is the slope and $B_s = -62.2$ is the $Y_s$-intercept.

The recognition of the “Surprised” emotion is ensured through the equation 6:

$$F_s > Y_s \implies \text{“Surprised” emotion} \quad (6)$$

3) Neutral emotion: To differentiate “Neutral” emotion, $F_2$ and $F_6$ have to be smaller compared to their respectively “Surprised” and “Happy” values. In addition, we note that the closed mouth in this case decreases $F_4$ value. Then, we consider that feature combination $F_n$ (equation 7) allows to well distinguish the “Neutral” emotion:

$$F_n = F_4 + F_2 - 1.5 \times F_6 \quad (7)$$

Fig. 10 illustrates the evaluation of the new $F_n$ representation of the most significant emotions versus the frame’s diagonal $D$. We can see that the “Neutral” emotion curve can be entirely separated from the two other emotions according to the two affine hyper-planes $Y_{n1}$ and $Y_{n2}$ (equations 8 and 9).

$$Y_{n1} = A_{n1} \times D + B_{n1} \quad (8)$$

$$Y_{n2} = A_{n2} \times D + B_{n2} \quad (9)$$

Where $A_{n1} = -0.02$, $B_{n1} = -20$, $A_{n2} = -0.18$ and $B_{n2} = 0.5$.

The recognition of the “Neutral” emotion is finally ensured through the equation 10.

$$Y_{n1} > F_n > Y_{n2} \implies \text{“Neutral” emotion} \quad (10)$$

IV. EXPERIMENTAL RESULTS

Many facial expression databases (e.g. The Cohn-Kanade database [23], Radboud Faces Database [22], the MMI-Facial Expression Database [24], and the JAFFE database [25]) exist. In our experiments, we use a well known and popular facial expression dataset. Indeed, the extended Cohn-Kanade database (CK+)\(^6\) [23] is used in this work to evaluate the performance of our method. This database have a large number of images with expressions’ labels. The CK+ database contains 327 image sequences for “Happy”, “Surprised” and “Neutral” emotions. To develop our method, we use Microsoft Visual Studio 2015 to employ OpenFace and MATLAB R2017a for drawing curves and illustrate results.

Fig. 11 and 12 shows the performance of our method on both Radboud Faces and CK+ databases. We note that our reached accuracy is higher then 94% for all the three used most significant emotions.

Fig. 9(a)-(c) confirm on CK+ database the previously conducted study using Radboud Faces Database. That validates our original feature presentation $F_h$, $F_s$ and $F_n$ and proposed affine hyper-planes $Y_h$, $Y_s$, $Y_{n1}$ and $Y_{n2}$.

For comparison reasons, we employ two machine learning methods SVM [18] and KNN [19]. Facial Action Coding System (FACS) [11] are used as features with these two classifiers. For fair comparison, we use Radboud dataset for training and CK+ dataset for prediction. Obtained results illustrated in Fig. 13(a)-(b) and presented in table I prove that our proposed method outperform SVM and KNN classifiers.

\(^6\)http://www.consortium.ri.cmu.edu/ckagree/
Table I shows the obtained average accuracy (Av Acc) of the three compared methods. We can see that our method reach 97.02% and outperforms SVM (84.04%) and KNN (95.32%) classifiers.

<table>
<thead>
<tr>
<th>Method</th>
<th>Happy</th>
<th>Surprised</th>
<th>Neutral</th>
</tr>
</thead>
<tbody>
<tr>
<td>Our method</td>
<td>98.59%</td>
<td>94.83%</td>
<td>98.12%</td>
</tr>
<tr>
<td>KNN + AU</td>
<td>97.05%</td>
<td>92.34%</td>
<td>94.84%</td>
</tr>
<tr>
<td>SVM + AU</td>
<td>88.73%</td>
<td>76.33%</td>
<td>86.83%</td>
</tr>
</tbody>
</table>

TABLE I: Our proposed method, SVM, KNN recognition rate comparison within CK+ database

V. CONCLUSION

We propose in this paper a new and performing method for customer satisfaction measuring based on the most significant facial emotion. Six new characteristic features are used and original feature representation and hyper-planes emotion separation are proposed. The proposed method is invariant to camera position. A very challenging datasets as Radboud Faces and Cohn-Kanade (CK+) are used for study and evaluation. Obtained results show that our method reaches a high recognition accuracy and outperforms Action Unit (AU) features based Support Vector Machine (SVM) and K-Nearest Neighbors (KNN).

REFERENCES


