Towards a Framework for Customer Emotion Detection

Moulay Smail Bouzakraoui
Department of Informatics, Faculty of sciences Ibn Tofail University
Kenitra, Morocco
Bouzakraoui.smail14@gmail.com

Abdelalim SADIQ
Department of Informatics, Faculty of sciences Ibn Tofail University Kenitra, Morocco
Abdelalim.sadiq.MA@ieee.org

Nourddine Enneya
Department of Informatics, Faculty of sciences Ibn Tofail University
Kenitra, Morocco
enneya@gmail.com

Abstract— **Automated analysis of human affective behavior has attracted increasing attention from researchers in psychology *neuroscience, linguistics, computer science and related disciplines. Nowadays,* emotional factors are important as classic functional aspects of customer purchasing behavior. Recent studies have shown that most of our purchasing choices and decisions are the result of a careful analysis of the product advantages and disadvantages and of affective and emotional aspects. Psychological literature recognizes that the emotional conditions are always present and influence every stage of decision-making in purchasing process.**

The main contribution of this paper is the design of a novel framework that provides the essential for customer sentiments analysis and emotion measuring in marketing. We focus our study on automatic facial expression analysis as a tool for measuring emotional customer that can predict their behavior in the decision-making.

Keywords—Big data; customer sentiments; face detection; emotion analysis; consumers' behavior.

I. Introduction

With the growth of technologies services, wireless networks, smartphones, IoT and social networks, etc. the large of diverse forms of data (text, image, video, audio…) are generated. In parallel, the development of storage device, analysis tools and cloud computing as well as new visualization tools for the valuation of these varied and unstructured data in large quantities. This phenomenon is called Big data.

Big data is a popular term used to describe the exponential growth and available data, both structured and unstructured. The more technical definitions, like this from Edd Dumbill, analyst at O'Reilly Media: "Big data is data that exceeds the processing capacity of conventional database systems. The data is too big, moves too fast, or doesn't fit the strictures of your database architectures. To gain value from this data, you must choose an alternative way to process it [1].

The application of big data in enterprises can enhance their production efficiency and competitiveness in many aspects. In particular, on marketing, with correlation analysis of big data, enterprises can more accurately predict the consumer behavior and find new business modes.

Leading marketers are using big data to deliver greater value and relevance to their customer. During product development, enterprise using structured and unstructured data generating from different source such as competitive intelligence and market information, company product and marketing research, press and social media commentary, call center logs, consumer feedback, device data, third-party data, consumer sentiment on retail sites and forums, transactions and other input from sales and enhancement requests.

In this paper, we will focus our study in the impact of face expression to predict the customer sentiments toward a specific product. In section II, we give an overview of the problematic. Next we explain our contribution in which we propose a framework for customer sentiments analysis. In the last section we conclude our work and we open research problems in Big data and predictive analysis in marketing.

II. Problematic

Emotion has become a new trend in marketing, helping enterprise to understand the opinion being expressed on products. Currently the majority of enterprise use conventional marketing methods based on advertising, price, sale points, and satisfaction surveys etc. Those methods were exceeded relative to the cost, time and data reliability. Marketers also gather feedback on attitudes and opinions as they occur without having to invest in lengthy and costly market research activities.

On the other hand, according to human judgment of affect, psychologists have various opinions about the importance of the cues from facial expression, vocal expression and linguistic message. In [18] Mehrabian stated that the facial expression of a message contribute 55% of the overall impression while the vocal part and the semantic contents contribute 38% and 7%, respectively.

This motivates us to propose a framework as an alternative technique capable of triangulating qualitative and quantitative methods through innovative data collection and analysis, allowing initially, to analyze automatically the emotions of Customer's towards a product based on facial expressions. In the following section, we illustrate our contribution in detail.

III. System overview

In this section we propose a new framework allowing to collect a set of video sequences on behaviors of an unlimited number of consumers, and during different periods, to a new product provided by enterprise. Next we extract from these video sequences some representative's images named key frames. These images are those containing only faces. Then we detect basic emotions (Joy, Sadness, Anger, fear, disgust, surprise) expressed by these extracted faces. Finally, we classify detected emotions as expressing a positive, negative or neutral opinion towards a product, and from these result, the enterprise constitutes an opinion on the impact of this product on the customer.

The model that we propose is divided into three phases (figure 2): First step is Data collection from a video camera. Afterwards, data processing (face and emotions detection). Finally, the analysis of these resulting emotions that generate values for the enterprise, which aims to extract business intelligence about how the product is perceived by consumers.

An overview of the proposed framework is depicted in figure 3. The framework consists of three layers:

A. *Data collection:*

The first step of our model is the collection of data as a video sequence filmed by a camera set up in a well-defined place in a way to record the maximum possible behaviors of customers who frequent the testing product. For each video sequence, we extract a set of representative image named key frame. In the literature, a lot of research has been done. The widely used key frame extraction techniques are as follows:

- Key frame extraction based on shot activity [3].
- Key frame extraction based on macro-block statistical characteristics of MPEG video stream [4].
- Key frame extraction based on motion analysis [5].

In our case we use the method based on the extraction of the intra-frames (I-frames) of the whole video, frame by frame developed in previous work [6].

B. *Data Processing:*

The processing phase is divided into two steps: Face detection and emotions detection.

1) *Face detection*

The face detection procedure is an essential step for the relevance of our system. Several researches have been developed for this purpose. For this reason a non-exhaustive classification of existing methods has been performed [7], so there are four approaches; Methods based on knowledge, those based on invariant features, those based on the mapping and the methods based on appearance. For our project we chose to use the latter approach is to treat the problem of face detection as a classification problem. In order to determine whether an image belongs to the class of faces or non-faces we use supervised machine learning techniques, such as the AdaBoost technique developed by Viola and Jones [8].This technique allows better use in real time. In this article we used a method based on the work of Viola and Jones which we have developed in such a way as to have the ability to detect faces turned by any angle of rotation in a given image [9]. We have supplied the property group Haar-Like presented by Viola and Jones by others which can extract accurate information on faces taken any pose in a picture.

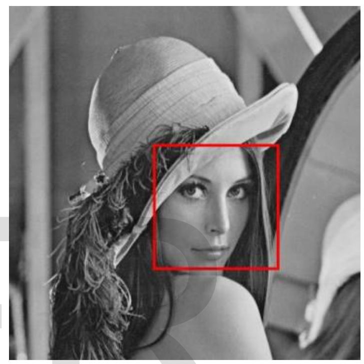

Figure 1: face detection example.

2) *Emotion detection*

Emotion recognition is the ability to identify what people would think and understand the connection between his/her feelings and expressions.

Understanding emotions becomes indispensable for the day-to-day function of humans. Technologies for processing daily activities, including facial expression, speech, and language have expanded the interaction modalities between humans and computer supported communicational artifacts.

In 1970, Ekman et al [10] carried out extensive studies on facial expressions. Their research showed that universal facial expressions provide sufficient clues to detect emotions. They used anger, sadness, surprise, fear, disgust and joy as six basic emotion classes. Such basic affective categories are sufficient to describe most of the emotions exhibited through facial expressions.

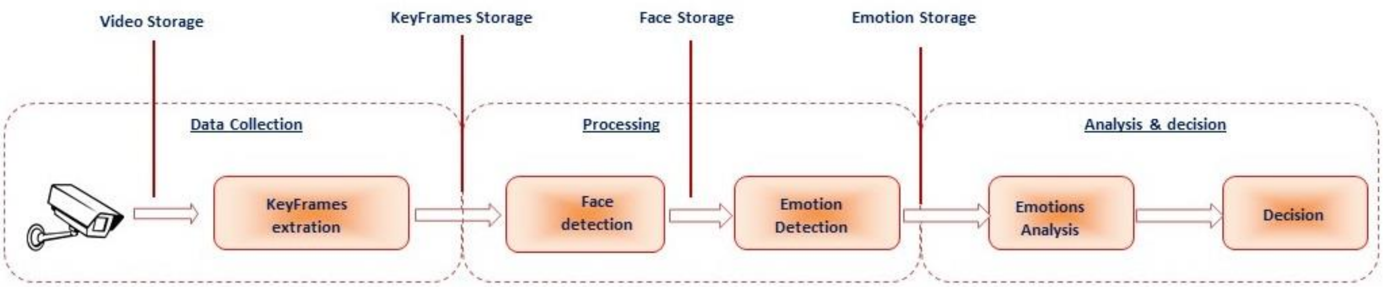

Figure 2: Customer sentiments analysis process

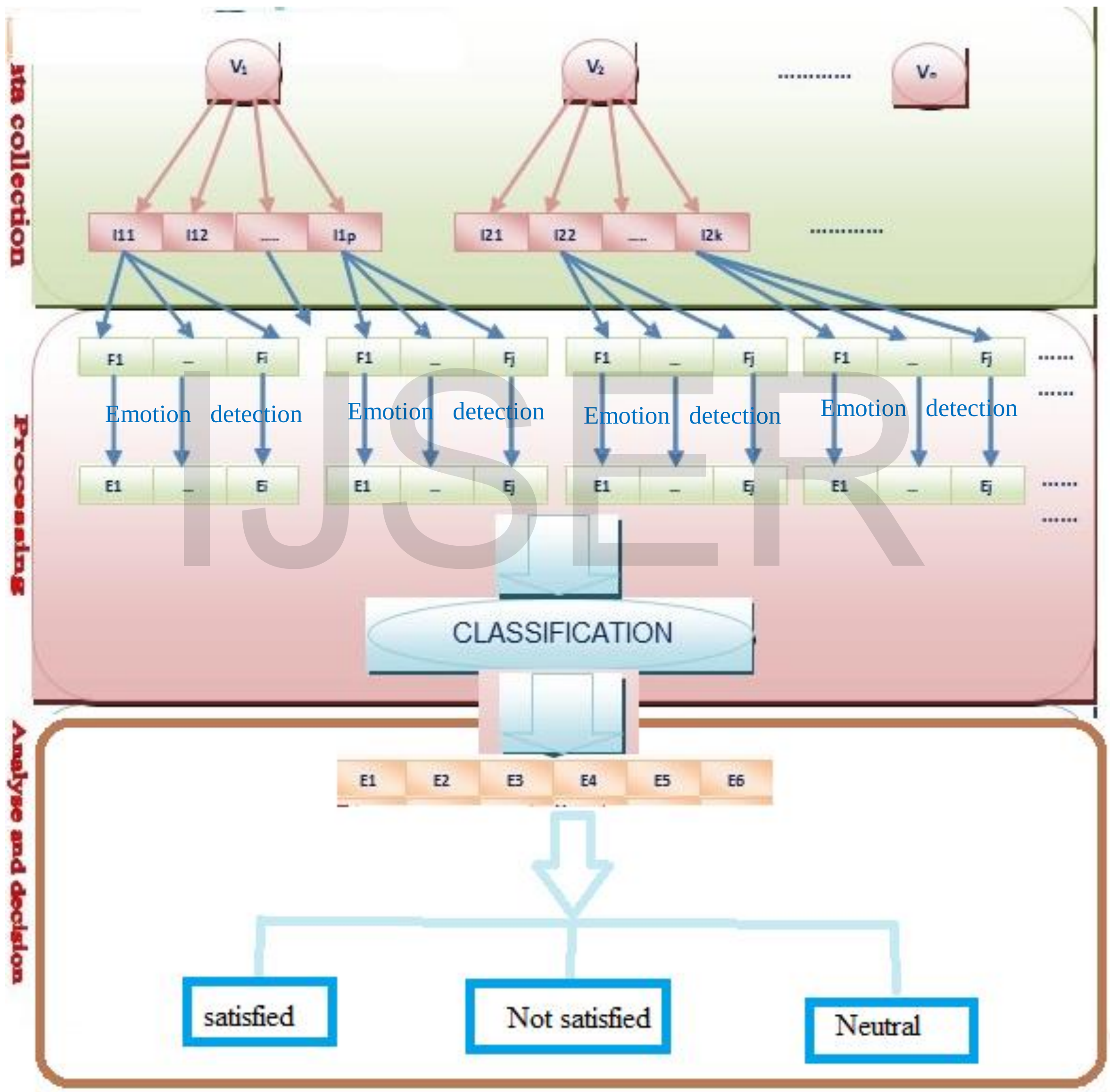

Figure 3: Framework of Customer Emotion Analysis
V: Video, I: Image, F: Face, E: Emotion

a) *Feature extraction faces the LBP algorithm*

From the face detected first, we cropped and resized to a size of 24 × 24 pixels. The preprocessed face image is used to extract features of facial expression using LBP algorithm. Local Binary Pattern (LBP) is a simple yet very efficient texture operator, which labels the pixels of an image by thresholding the neighborhood of each pixel and considers the result as a binary number. Due to its discriminative power and computational simplicity. LBP based feature extraction method is used owing to its excellent light invariance property and low computational complexity [11]. The neighborhood values are thresholded by the center value and the result is treated as a binary number. In this way, it encodes the neighborhood information very efficiently.

We use the follows equations to calculate the LBP codes for each pixel.

$$LBP_{P,R}(x_c, y_c) = \sum_{p=1}^{P} s(g_p - g_c)2^{p-1}$$

Or

$$s(x) = \begin{cases} 1 & if \quad x \geq 0 \\ 0 & if \quad x < 0 \end{cases}$$

P and R represent the number of the block's pixels and the block's radius respectively.

LBP approach is used in which the face image is subdivided to many sub-images and LBP features of each block are calculated. The bin histogram of each sub-image LBP features is determined and hence by concatenation of the block bin histograms (figure2).

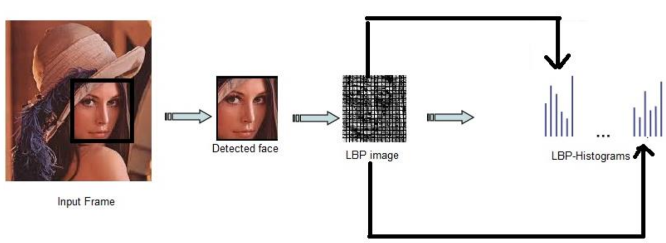

Figure 4: Histogram representing the image of the facial features

b) *Emotion Classification by SVM*

Classifying data has been one of the major parts in machine learning [12]. The idea of Support Vector Machine (SVM) is to create a hyper plane in dimensional feature space and separate the two classes of data with a maximum margin of hyper plane. SVM uses an optimum linear separating hyper plane to separate two sets of data in a feature space, between data sets to indicate which class it belongs to [13]. SVM map a given set of binary labelled training data to a high dimensional feature space and separate the two classes of data with a maximum margin of hyper plane. SVM uses an optimum linear separating hyper plane to separate two sets of data in a feature space. The separating hyper plane is the hyper plane that maximizes the distance between the two parallel hyper planes [14]. This optimum hyper plane is produced by maximizing minimum margin between the two sets. Therefore the resulting hyper plane will only be depended on border training patterns called support vectors. Therefore, Support vectors are the data points that lie closest to the decision surface [15]. Recently several studies have reported that support vector machine (SVM) delivers higher accuracy in terms of data classification compared with other data classification algorithm [16].
Hyper plane acting as the decision surface is defined as:

$$\sum_{i=1}^{N} \alpha_i d_i k(x, x_i) = 0$$

Where

$$K(x, x_i) = \varphi^T(x)\varphi(x_i)$$

Represents the inner product of two vectors induced in the feature space by the input vector x and input pattern x_i pertaining to the ith example. This term is referred to as inner-product kernel [15], [17].
Where

$$W = \sum_{i=1}^{N} \alpha_i d_i \varphi(x_i)$$

$$\varphi(x) = [\varphi_0(x), \varphi_1(x), \ldots\ldots, \varphi_{m1}(x)]^T$$

$\varphi_0(x) = 1$ for all x

The Kernel function is selected as a RBF learning machine.

$$K(x, x_i) = \exp\left(-\frac{\|x - x_i\|^2}{2\sigma^2}\right)$$

The Lagrange multipliers {α_i} for i = 1 to N that maximize the objective function Q(α), denoted by α_0,i is determined.

$$Q(\alpha) = \sum_{i=1}^{N} \alpha_i - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i \alpha_j d_i d_j K(x, x_i)$$

Subject to the following constraints:

$$\sum_{i=1}^{N} \alpha_i d_i = 0$$

$0 \leq \alpha_i \leq C$ for i =1,2,…., N

The linear weight vector w_0 corresponding to the optimum values of the Lagrange multipliers are determined using the following formula:

$$W_0 = \sum_{i=1}^{N} \alpha_{0,i} d_i \varphi(x_i)$$

$\varphi(x_i)$ is the image induced in the feature space due to x_i

We use the multi-class classification; a multi-class classifier must combine all single class classifiers trained for each class with appropriate means.
We directly use the output of the original functions of decision for a multi-class classifier.
To do this, the equation should be changed to a real value function.

To classify the faces, test objects feed in every single class classifier. We'll take the subject in class, which has the largest value of work of decision, namely:

$$output = \arg\max_i (f_i) \qquad i = 1,\ldots,6.$$

To classify six basic emotions namely happiness, sadness, disgust, fear, surprise and anger. Architecture will be as follows:

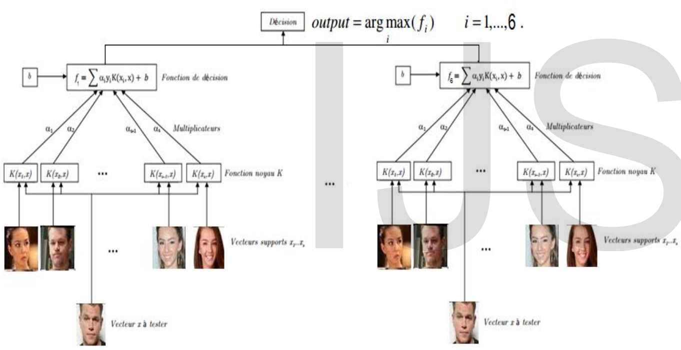

Figure 5: classification algorithm

C. Analysis and Decision:

This phase of analysis and decision allows to have information about the preferences and behaviors of customers for the products.

According to the classification of the six emotions, we use a statistical system to determine the number of customers who are satisfied (happy, surprise), those who are not satisfied (disgust, anger, sadness), and which are neutral (fear, neutral).

Development teams can quickly determine the needs of consumers with analyses of this information. Marketing teams can evaluate the feelings of consumers and create models to predict sales effectiveness, also optimize the effectiveness of advertising campaigns.

IV. Conclusion and Perspectives

This position paper has provided a framework of customer emotion analysis. In this framework, a facial expression classification algorithm is proposed which uses Haar classifier for face detection purpose, Local Binary Patterns (LBP) histogram of different block sizes of a face image as feature vector. In addition, classifies various facial expressions using Support Vector Machine (SVM).

As a result, emotions about facial expressions are much richer for prediction and decision making of customers. Proactive marketing or product design strategies can be developed to enhance the business operations and the competitive power of the corresponding firms.

This work provides the basis for conducting our research. Outlook, we will implement our framework in an environment parallel that supported the processing of large data sets in a distributed computing environment, and include others components like speech, sentiments analysis from social media in order to design a multimodal framework and exploit big data analysis to improve prediction and decision making to the customer.

References

[1] What is big data? An introduction to the big data landscape. (2015). http://radar.oreilly.com/2012/01/what-is-big-data.html

[2] http://searchcloudcomputing.techtarget.com/definition/Hadoop.

[3] P. Gresle, T. S. Huang, “Gisting of video documents: a key frames selection algorithm using relative activity measure,” The 2nd International Conference On Visual Information System, 1997.

[4] J. Calic, E. Izquierdo, “Efficient key-frame extraction and video analysis information technology: coding and computing,” international symposium on information technology, pp. 28-33, 2002.

[5] W. Wolf, “Key frame selection by motion analysis,” Proc. IEEE Int. Conf. Acoust., Speech Signal Proc., vol. 2, pp. 1228–1231, 1996.M. Young, The Technical Writer’s Handbook. Mill Valley, CA: University Science, 1989.

[6] Youness TABII, Abdelalim SADIQ "Shot Boundary Detection in Videos Sequences Using Motion Activities" Advances in Multimedia - An International Journal (AMIJ) Volume -5 Issue–1 pp. 1-7 .December 2014.

[7] Cha Zhang and Zhengyou Zhang. "A Survey of Recent Advances in Face Detection". Technical Report, Microsoft Research MSR-TR-2010-66, June 2010.

[8] Viola P. and Jones, M., "Rapid Object Detection Using a Boosted Cascade of Simple Features", IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Vol. 1, pp. 511-518, 2001.

[9] M. OUALLA, A. SADIQ, S. MBARKI, "A New Approach to Represent rotated Haar-Like Features for objects detection", Journal of Theoretical and Applied Information Technology, July 2015. Vol. 77 No.3.

[10] Ekman, P. Universal Facial Expressions of Emotions. California Mental Health Research Digest, 8,151-158 (1970).

[11] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation Invariant texture classification with local binary patterns," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, no. 7, pp. 971-987, 2002.

[12] Kim, Sangwook, et al. "Deep learning of support vector machines with class probability output networks." Neural Networks (2014)F. Abdat, C.Maaoui et A.Pruski : Gradient based method for static facial features localization. International Conference Visualization, Imaging and Image Processig , VIIP07, 2007.

[13] Srivastava, Durgesh K., and Lekha Bhambhu. "Data classification using support vector machine." (2010).

[14] Wong, Wai-Tak, and Sheng-Hsun Hsu. "Application of SVM and ANN for image retrieval." European Journal of Operational Research 173.3 (2006): 938-950.

[15] Foody, Giles M., and Ajay Mathur. "The use of small training sets containing mixed pixels for accurate hard image classification: Training on mixed spectral responses for classification by a SVM." Remote Sensing of Environment 103.2 (2006): 179-189

[16] Lima, Ana Carolina ES, and Leandro Nunes De Castro. "A multi-label, semi-supervised classification approach applied to personality prediction in social media." Neural Networks 58 (2014);122-130;

[17] Ng, Jeffrey, and Shaogang Gong. "Composite support vector machines for detection of faces across views and pose estimation." Image and Vision Computing 20.5 (2002): 359-368.