Are you Lost? Using Facial Recognition to Detect Customer Emotions in Retail Stores

Valter Borges*, Rui P. Duarte*, Carlos A. Cunha* and David Mota†
*School of Management and Technology
Polytechnic Institute of Viseu, Viseu, Portugal
Email: estgv16626@alunos.estgv.ipv.pt, pduarte@estgv.ipv.pt, cacunha@estgv.ipv.pt
†Bizdirect Competence Center, Viseu, Portugal
Email: david.mota@bizdirect.pt

***Abstract*—The understanding of consumer behavior is a dynamic field, especially relevant to the success of companies and for consumer satisfaction. It is especially important in the situation of intense competition, currently characteristic for the retail store industry, where companies fight for every individual customer. Moreover, companies do not want customers to enter their system and leave without buying products they intended to buy. This has an impact on user satisfaction and retail stores income. In this paper we present a method that targets customer satisfaction in the aforementioned context using a facial recognition system acting at the emotional level of the customer. Our method is based on cumulative negative emotions that are associated to a sadness level, which triggers events for retail store assistants to help customers. Results show that this method is adequate to measure these emotions and is a useful reference for retail store assistant intervention.**

Keywords–Image recognition; sentiment analysis; activity recognition; user satisfaction; retail environments.

I. Introduction

The concept of shopping has been changing during the years. Today shops are not only the place where customers go to buy products but also the place where they spend part of their time. Thereof, retail stores need to adapt to the needs of customers in order to provide them a positive experience. Two perspectives are present: the customer that wants to find and buy a specific product and the retail store that wants to increase sales. Although a relationship can be established between perspectives, in real context scenarios, they have different approaches to achieve a win-win-win solution for the customer–retailer–manufacturer relation. According to Oliver [1], it is more challenging to fidelize an existing customer than to attract new ones. A simple way to lose customers is when they come into a store to buy products they cannot find and leave without buying them. This transforms the process into an unsatisfactory experience for all the players involved, which need to be taken into account.

The application of computer vision techniques in retail dates back to more than two decades [2]. More recently, due to advances in computer vision, machine learning, and data analysis, retail video analytics can provide retailers with much more insightful business intelligence [3][4][5]. Thus it promises much higher business value, far beyond the traditional domain of security, authentication, and loss prevention. Examples include analysis of store traffic, queue data, customer behavior, and purchase decision making, among others. However, it is a real-world scenario, and many technical challenges are present for realistic computer vision techniques: changing and uncontrollable lighting conditions, high-level, complex human and crowd activities, cluttered backgrounds, crowded scenes, occlusion, odd viewing angles, low resolution cameras, limited contrast, and low object discriminability [3].

It is well known that Video Analytics Technology (VAT) mostly focuses on automatic customer detection. However, customer perspective is of most importance since they consume products available in retail stores. One of the potential areas of interest is to determine whether a customer is not finding a specific product. As a consequence, he leaves the store without buying it, which does not relate to a win-win-win situation. If more information about customers is collected using VAT to detect if they are not finding a product and generate triggers to employees informing the occurrence of this, a significant impact in customer satisfaction and retail store sales is foreseen.

Our work focuses on both perspectives and the main contributions are as follows:

- *Emotion Analysis.* To our best knowledge, this is the first scalable attempt to measure negative emotions to determine if a customer is not finding products in a retail store.
- *Real-time notification and intervention.* An integrated platform is developed for the real-time notification of retail stores assistants and intervention with customers.

The remainder of this paper is organized as follows. Section II briefly reviews works in the field of video analytics technology. Section III, identifies the problem to be solved in this work and details our approach, providing a method based on negative emotion analysis. Section IV presents a web interface for the retail store assistant and the method is validated with experimental results carried out in real context scenarios. Section V concludes the paper, providing some hints to future work.

II. Related Work

Our work overlaps with previous research on automatic analysis of human behavior inside retail stores. In this context, several approaches have been studied, as hot zone analysis, automatic activity recognition and sentimental analysis.

The automatic detection of human emotions is a complex problem that has been applied to several ordinary problems. Techniques addressing this problem spans several types of data sources. Faces' images are one of the most promising sources for data analytics related to the emotion detection problem.

 ISBN: 978-1-61208-754-2

A. Hot Zone Analysis

Hot zone analysis aims to identify the trajectory of customers within a store. Trajectory analysis unveils spots with more activity and reveal where customers spend their time. Human's head position estimation was explored to create the initial estimates for tracking algorithms. Zhao et al. [6] presented a method for the detection and tracking of several humans in video frames. They propose boundary and shape analysis for human detection. On top of that, a 3D walking model predicts motion templates from the captured frames to track humans. This work was later improved by Zao and Ram [7], through the inclusion of a detection technique for human identification using Markov chain Monte Carlo methods. The method was tested in indoor and outdoor high-density scenes. In the outdoor scenes, false positives appear at far ends and dense edges. In the indoor scenes, the subtraction method gives erroneous foreground blobs. For human segmentation in both scenes, 1000 iterations are necessary to segment human objects. Leykinv and Mihran [8] developed a method where the human head coordinates are extracted from video frames to determine the position of customers in a store. These coordinates are further used to track customers in video sequences captured in crowded environments. The low-level extraction of the customers in a frame and the use of camera calibration to locate customer's head and location in the picture allows them to infer their location in the store.

B. Activity Recognition

The Activity recognition is related to the shop behaviour and represents the actions of customers when buying products. Monitoring this behaviour is of most importance to academic as to retail stores. Popa et al. [4] analyzed customer behaviour using background subtraction form images. This approach allowed them to detect customers in the entry point and then track them in the system. Popa et al. [5] improved the method for automatic assessment of customer' appreciation of products. First, they classified customer behaviour by participant observation. Next, they implemented the model for motion detection, trajectory analysis, and face localization and tracking for different customers. Sicre and Nicolas [9] resorted to behaviour models for detection of motion, tracking moving objects, and describing local motion. Results have shown that the approach can correctly classify 73% of the frames, for sequences taken in real environments. Later, Frontoni et al. [10] proposed a method to analyze human behavior in shops in order to increase consumer satisfaction and purchases. In their method, they use vertical red, green and blue depth sensors for people counting and shelf interaction analysis. Their results exhibited areas with both positive and negative interactions with products in shelves. They compared their results with ground truth visually recorded and accuracy varies between 97.2% and 98.5%. Hu et al. [11] investigated the detection of semantic human actions in complex scenes. Their work deal with spatial-temporal ambiguities in frames using bag of instances representing the candidate regions of individual actions. A technique based on the combination of Simulated Annealing and Support Vector Machines has shown better results than standard Support Vector Machines.

C. Sentiment Analysis in Videos

Sentiment analysis is another area of video analytics. This type of problems is strongly connected to the problem addressed in this paper, since it determines the emotional level of the customer. Zadeh ett al. [12] addressed this problem using a multimodal dictionary that exploits jointly words and gestures. The approach has shown better results than straightforward visual and verbal analysis. An alternative approach to methods that adopt bag of words representations and average facial expression intensities is presented by Chen et al. [13]. They propose sentiment prediction using a time-dependent recurrent approach that performs fusion of several modalities (e.g., verbal, acoustic and visual) at every time-step. The implementation of the approach using long short-term memory networks has shown significantly improvements over several other approaches. Wang and Li [14] explored sentiment analysis in social media images. The main challenge of the work lies in the semantic gap between visual features and underlying sentiments. Contextual information is proposed to overcome the semantic gap in prediction of image sentiments. The solution was shown effective when evaluated with two large-scale datasets.

III. Approach

Our approach is based on a machine learning system that runs in background for the intervention of retail store assistants with costumers focusing on the analysis of negative emotions using a facial recognition system. When negative emotions are detected, the retail store assistant is notified for customer intervention.

A. Problem Statement

The study of human behavior in retail stores has been carried out in the last years, and their behavior can be interpreted by analyzing their emotional responses [15] to contexts.

At the emotional level, one of the problems that currently exist in customer service is trying to understand their state of mind when inside a store. For that purpose, the detection of emotions from customers will be able to increase the quality of service - the more relevant information about the customer, the better the assistance. The measurement of negative or positive emotions can be carried out by several applications that are available in the market. This work aims at the detection of negative emotions in a time window, where sadness is one of the most significant negative emotion to consider. However, other parameters, like anger, disgust, or fear, are relevant for the measurement of negative emotions. Thus, tracking negative emotions in the context of a store is an open problem, which is of most importance to be solved since it serves the automation of customer-assistant situations, resulting in an increase of the speed of attendance, improve customer satisfaction and increase retail stores sales.

B. Machine Learning Implementation

The performance of machine learning models is deeply dependent on the volume of data available for training models. For that reason, the most accurate models are provided by giants of software that have access to large volumes of data for training models capable of accurate detection of emotions in images. Fortunately, these models are widely available through an Internet accessible API like the IBM Watson [16], Face API [17], Kairos [18], and Amazon Rekognition [19].

In this work, we use Face API [17]. It is a cognitive service developed by Microsoft that supplies algorithms to detect, recognize, and analyze human faces in images. Face API features are obtained in two stages: the first is the detection and recognition of face attributes; in the second stage, a JSON file is returned with the fields that contain face attributes.

The detection stage represents the analysis of the existing faces and returns attributes for each face. When a face is detected, the face rectangle attribute is returned, since it contains the pixels to track the face in the image and gets its bounding box. Within this rectangle, other attributes are returned by the API to the JSON file, namely, face Id, face landmarks, age, emotion, gender, and hair. In this paper, except for face landmarks, all the parameters are considered in two contexts. First, for a general characterization of the costumer, age, gender, and hair attributes are used. These attributes will allow the retail store assistant to better identify the customer (note that for security policies, the system cannot store the face of the customer). Next, for the emotion analysis (cf., Section III-C), the emotion attribute is considered which contains a set of different emotions. The parameters returned by the Face API are a basis of knowledge for the implementation of the emotion tracking method presented in this paper.

C. Emotion Analysis

As previously referred, there are several parameters associated to emotions that are returned by facial recognition systems, namely anger (A_p), contempt (C_p), disgust (D_p), fear (F_p), happiness (H_p), neutral (N_p), sadness (S_p) and surprise (Su_p). In the scope of this work, we only consider negative emotions (A_p, D_p, F_p and S_p) that affect the costumer interaction with the system.

The basic idea of our method is presented in Figure 1. When a customer arrives at a shelf, Face API captures his emotions, and a sadness level β is set to zero. This factor updates in the presence of negative emotions, and once a threshold is passed ($\beta > 50\%$), the assistant is asked to go to the customer. Negative emotions manifest in several ways, and one of the most critical parameters is the sadness parameter, $S_p \in [0..1]$ (values near 1 correspond to the total manifestation of sadness). Therefore, every time a frame captures a customer with a high value of sadness, it may indicate a potential customer not finding a product. Other parameters like A_p, D_p or F_p are also present in negative emotions, and their contribution is analyzed.

To determine the weights to consider in each of the negative emotions, an empiric study (presented in Table II) was carried out with users that were asked to express several emotions: S_p, N_p, D_p, H_p and simulate the action of looking for a product and not finding it, referred to as *Simulated*. In the emotion tests considering H_p and N_p, these parameters have high values, representative of the tested emotion. In the tests for forced sadness and simulation, S_p has low values in most cases, which is justified by the fact that the sadness emotion can result in false positives. However, in this case, the presence of other negative emotions is visible, with small values of A_p, D_p and

Figure 1. Problem specification for negative emotion analysis.

F_p. Analyzing the impact of these parameters in the emotion is an essential factor to determine how to infer sadness when S_p should be naturally present and is not.

In this context, two types of tests were carried out: first, the evaluation of the impact of each negative emotion and, second, the presence of all negative emotions. In the first test, results obtained ($A_p = 47\%$, $D_p = 16\%$, $F_p = 6\%$ and $S_p = 91\%$), show that negative emotion is present in the tests. However, excluding S_p, the other negative emotions cannot be used individually to complement the sadness test, since they are present in a small number of tests which are not representative of the sample. In the second test, we considered the cumulative presence of all negative emotion parameters ($A_p + D_p + F_p + S_p > tol$) for the same scenario (forced sadness and simulation), as shown in Table I.

TABLE I. TOLERANCE TESTS FOR $A_p + D_p + F_p + S_p > tol$.

	Tolerance (tol)		
	0	0.01	0.02
Negative emotions (%)	97.22%	83.73%	80.16%

Results show that when $tol = 0$, 97.22% of the tests reveal the presence of cumulative negative emotions, which is very representative of the tested scenario. The rate decreases for $tol = 0.01$ and $tol = 0.02$. Therefore, when S_p is not representative in a sadness test, the alternative of considering cumulative negative emotions has success rate of 97.22%. Recall that these criteria are used only to improve the success rate of retail store assistants interventions and are used in two contexts: in the evident presence of sadness (high values of S_p) and in the presence of signs of sadness ($A_p + D_p + F_p + S_p > tol$) for low values of S_p. The resulting method is presented in the algorithm depicted in Figure 2. Let $\mathcal{C} = \{c_j\}$, $j = 1 \ldots M$, where M represents the number of customers that are detected in the system and $\mathcal{F} = \{f_i\}$, $i = 1 \ldots N$, where N represents the number of frames captured in real-time using the Face API for each customer $c_j \in \mathcal{C}$. The algorithm starts by scanning if a customer is detected by the Face API and its faceId is generated. The sadness level of each customer, β_j, is set to zero and frames are captured while the customer is detected in the system. For every captured frame, the Face API returns

 ISBN: 978-1-61208-754-2

TABLE II. USER TESTING IN REAL SCENARIOS: ACTING NORMAL, SIMULATION, FORCE SADNESS, FORCE ANGER AND FORCE HAPPINESS.

Anger (A_p)	Contempt (C_p)	Disgust (D_p)	Fear (F_p)	Happiness (H_p)	Neutral (N_p)	Sadness (S_p)	Surprise (Su_p)	Testing
0	0.001	0	0	0	0.999	0	0	*acting normal*
0.001	0.001	0	0	0	0.985	0.014	0	*simulate scenario*
0	0.002	0	0	0	0.762	0.235	0	*force sadness*
0.004	0.005	0.005	0	0.001	0.962	0.022	0	*sumulate scenario*
0.005	0.002	0.001	0	0.001	0.731	0.261	0	*force sadness*
0	0.002	0	0	0	0.993	0.005	0	*acting normal*
0	0	0	0	1	0	0	0	*force happiness*
0	0.016	0	0	0	0.811	0.172	0	*force sadness*
0.031	0.001	0	0	0	0.967	0.001	0	*simulate scenario*
0.035	0.001	0	0	0	0.966	0.001	0	*force anger*
0	0	0	0	0	0.977	0.023	0	*simulate scenario*
0	0.001	0	0	0	0.905	0.094	0	*simulate scenario*
0	0	0	0	0	0.958	0.041	0	*force sadness*
0	0.089	0.001	0	0	0.58	0.33	0	*force sadness*
0.001	0.027	0	0	0	0.967	0.004	0	*acting normal*
0	0.152	0	0	0.848	0	0	0	*force happiness*
0.172	0.002	0	0	0	0.823	0.003	0	*force sadness*
0.011	0.006	0	0	0	0.962	0.021	0	*simulate scenario*
0.008	0.37	0	0	0	0.621	0.001	0	*force anger*
0.16	0.043	0.001	0	0.001	0.661	0.134	0	*simulate scenario*
0.001	0.025	0	0	0	0.967	0.007	0	*simulate scenario*
0	0.169	0	0	0.009	0.821	0	0	*force sadness*
0.0058	0.011	0	0	0	0.887	0.043	0	*force sadness*
0	0.004	0	0	0.006	0.987	0.004	0	*acting normal*
0	0.001	0	0	0.958	0.04	0.002	0	*force happiness*
0	0	0	0	0	0.857	0.143	0	*force sadness*
0	0	0	0	0	0.84	0.159	0	*simulate scenario*
0.412	0.042	0.09	0.029	0.006	0.57	0.001	0.363	*force anger*
0.001	0.007	0	0	0.001	0.94	0.051	0	*simulate scenario*
0	0.005	0	0	0.038	0.955	0.001	0	*simulate scenario*
0	0.001	0	0	0	0.958	0.041	0	*force sadness*
0	0.001	0	0	0	0.417	0.582	0	*force sadness*
0	0	0	0	0	0.997	0.002	0	*acting normal*
0	0	0	0	1	0	0	0	*force happiness*
0	0	0	0	0	0.965	0.035	0	*force sadness*
0.053	0.004	0	0	0	0.943	0	0	*simulate scenario*
0.127	0.009	0	0	0	0.864	0	0	*force anger*
0	0.0087	0	0	0.036	0.868	0.002	0.007	*simulate scenario*
0	0.001	0	0.001	0.001	0.956	0.033	0.009	*simulate scenario*
0	0.003	0	0	0	0.679	0.318	0	*force sadness*
0	0	0	0	0	0.887	0.113	0	*force sadness*

Algorithm 1: Emotion-based intervention method

Data: $\mathcal{C}$ $\triangleleft$ detected customers $\mathcal{C} = \{c_j\}$
Data: $\mathcal{F}$ $\triangleleft$ API frames $\mathcal{F} = \{f_i\}$
Result: β, $\mathcal{I}$ $\triangleleft$ β = sadness level, $\mathcal{I}$ = Intervention
begin
  foreach $c_j \in \mathcal{C}$ **do**
    $\beta_j \leftarrow 0.0$ $\triangleleft$ set emotion value to zero
    $\mathcal{I} \leftarrow false$ $\triangleleft$ no intervention required
    foreach $f_i \in \mathcal{F}$ **do**
      $A_{p_i} \leftarrow A_p \in f_i$ $\triangleleft$ get anger from f_i
      $F_{p_i} \leftarrow F_p \in f_i$ $\triangleleft$ get fear from f_i
      $S_{p_i} \leftarrow S_p \in f_i$ $\triangleleft$ get sadness from f_i
      $D_{p_i} \leftarrow D_p \in f_i$ $\triangleleft$ get disgust from f_i
      if $(S_{p_i} > 0.5)$ **then**
        $\beta_j \leftarrow \beta + 0.1$ $\triangleleft$ update sadness level
      else if $(A_{p_i} + F_{p_i} + S_{p_i} + D_{p_i} > 0)$ **then**
        $\beta_j \leftarrow \beta + 0.05$ $\triangleleft$ update sadness level
      if $(\beta_j > 0.5)$ **then**
        $\mathcal{I} \leftarrow true$ $\triangleleft$ intervention required
    end
  end
end

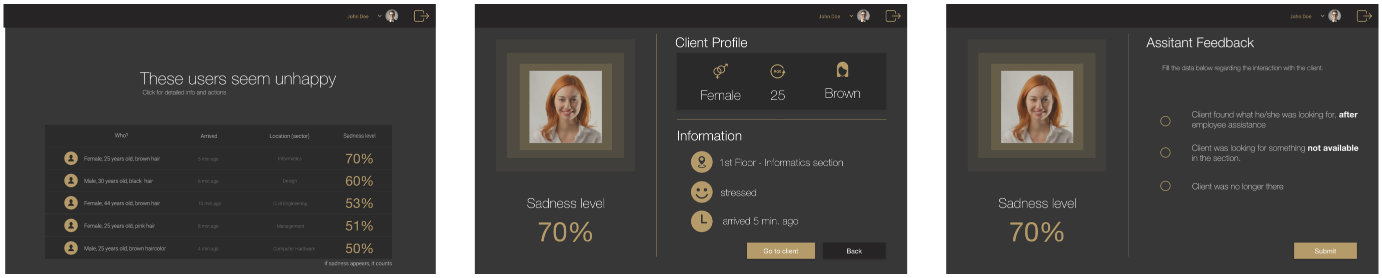
Figure 2. Emotion-based intervention method.

negative emotion values that are stored for processing. Every time the algorithm captures evidence of sadness ($S_{p_i} > 0.5$ or signs of sadness ($A_{p_i} + F_{p_i} + S_{p_i} + D_{p_i} > 0$), the value of β_j is updated in a factor of 0.1 or 0.05, respectively. When the sadness level passes a threshold of 0.5, the assistant is informed that a customer needs intervention.

An important consideration is that our system does not retain personal information of a customer. After detection by a camera, only a faceId is generated to uniquely identify the characteristics of that customer. If he leaves the system, the method still continues to try to track the faceId of the customer for five minutes. After that period, the information of the faceId is removed from the database, but the face attributes are kept. With this, personal information of users is not stored, therefore it does not allow the system to track history of a customer. If that customer is again detected in the system, he will be assigned a new faceId.

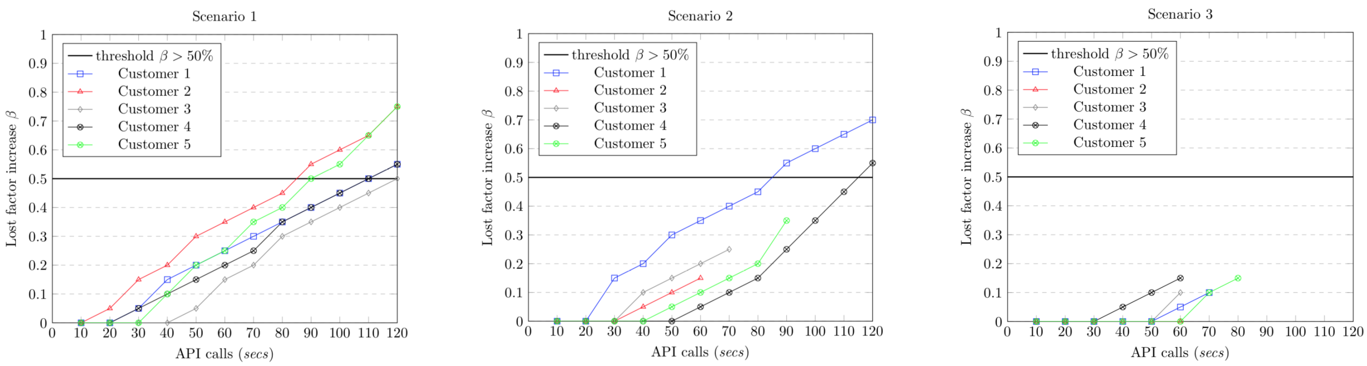
IV. EXPERIMENTAL DESIGN AND RESULTS

The algorithm presented in the previous section runs in background and processes information that can be visualized by the retail store assistant in an web application (see Figure 3).



(a) List of customers for assistant intervention. (b) User info and emotions. (c) Assistant feedback.

Figure 3. Web interface for retail store assistant intervention.



(a) Book is not available in shelves. (b) Book is similar to other products. (c) Book is well identified in shelves.

Figure 4. Results obtained for tests with five customers in the three scenarios.

A. Web Interface

The assistant has access to the notifications management page (see Figure 3a). This page is updated in real time and contains a list of customers that require intervention. Here, some general information of the customers is provided for better identification. When the assistant chooses a customer for intervention, the algorithm running in background stops increasing β for that customer. This value is stored at the moment of intervention, since, otherwise, it would reach 100% for all customers in the time elapsed between the interaction and the time to go to the customers. When the assistant selects a customer, general information is presented (such as hair color, age, gender, location in store, the emotion revealed by the customer and how long the customer is in the system). In addition, the assistant has the possibility of attending the customer and to cancel and return to the call management page as shown in Figure 3b.

The intervention level starts when the assistant clicks in the "go to client" button and the page changes so that feedback data can be provided by the assistant which possesses relevant information regarding the intervention with the customer, as shown in Figure 3c. It is important to note that while the assistant is attending the customer, no further changes in the customer emotions are captured. It is intended to capture the emotions that have caused the customer to exceed the emotional threshold and not to register emotion changes while being under intervention.

B. Results

We have tested the approach described in Section III by carrying out a pilot study. Books were placed in shelves with a camera placed to capture emotions. Five customers where asked to find a book from twenty books in three scenarios:

- Scenario 1: The book is not available in the products placed in the shelves.
- Scenario 2: The book is in the shelves, but very similar to other books, making it difficult to be found.
- Scenario 3: The book is available in the shelves and easy to be identified.

Results obtained are presented in Figure 4. For scenario 1 (Figure 4a), costumers reveal signs of cumulative unhappiness, $(A_{p_i} + F_{p_i} + S_{p_i} + D_{p_i} > 0)$, or sadness $(S_p > 50\%)$ as they realize that they are not finding the product. The sadness level threshold is passed for all customers after a few iterations of API calls. The variation of the sadness level cumulative response is due to the fact that, in the API calls, the customer can reveal one of both negative emotions tested. This implies that there can be an increase of 0.05 or 0.1, depending on the most prevalent negative emotion in each call. In this context, the web interface for the assistant is updated with the data related to the new customer that requires intervention (see Figure 4a). In all cases, the assistant reported option 2 in the feedback page (see Figure 4c). In scenario 2, three customers found the product, after some iterations and left the system.

 ISBN: 978-1-61208-754-2

The other two passed the sadness level threshold. In these cases, the assistant reported option 1 in the feedback page. Finally, in scenario 3, all the customers found the product after a few iterations of API calls, never crossing the sadness level threshold without need for assistant intervention.

To provide flexibility to the system, the assistant can decide the moment of the intervention. As previously referred, when the sadness level threshold is passed, the assistant web page is updated with the customer information. However, if he considers that the sadness level is not increasing with time, he can decide not to go to the customer. If the customer continues to reveal cumulative negative emotions, the assistant then makes the decision to assist him. Moreover, if all assistants are occupied, the system continues to increase the sadness level of a customer, until an assistant is available.

V. Conclusion and Future Work

We presented a novel scalable method based on visual recognition of customer emotions when buying products, using Face API. Our method uses a camera to capture the manifestation of negative emotions at two levels: the effective manifestation of sadness and evidence of sadness, in a set of frames. Our evaluation methodology shows that the method presents good results in real scenarios related to the context of the problem. The implementation of an intuitive web interface allows retail shops assistants to carry out interventions with customers if an emotional threshold occurs. This interface will greatly assist retail stores to have an understanding of which customers require intervention and, in a fast way, provide the necessary help. The natural implications are an increase in sales and customer satisfaction.

Future work will follow two directions. An extension of the method to incorporate face landmarks provided by the Face API will add another layer of decision to our algorithm by tracking the motion of the customer's face to detect if he is not finding a product. Moreover, by introducing Artificial Intelligence (AI), we will be able to anticipate the needs of users based on the previous emotional analysis. Finally, the sadness level was obtained empirically. It will be essential to use AI as a means to adjust this parameter.

Acknowledgment

This work is financed by National Funds through the Research Centre in Education, Technology and Heath (CI&DETS) and the Research Centre in Digital Services (CISeD) of the Polytechnic Institute of Viseu.

References

[1] R. L. Oliver, "Whence consumer loyalty?" *Journal of Marketing*, vol. 63, no. 4, pp. 33–44, 1999, ISSN: 00222429.

[2] R. M. Bolle, J. H. Connell, N. Haas, R. Mohan, and G. Taubin, "Veggievision: A produce recognition system," in *Proceedings Third IEEE Workshop on Applications of Computer Vision (WACV'96)*. IEEE, dec 1996, pp. 244–251, ISBN: 0-8186-7620-5.

[3] J. Connell, Q. Fan, P. Gabbur, N. Haas, S. Pankanti, and H. Trinh, "Retail video analytics: an overview and survey," in *Video Surveillance and Transportation Imaging Applications*, vol. 8663. International Society for Optics and Photonics, 2013, p. 86630X.

[4] M. Popa, L. Rothkrantz, Z. Yang, P. Wiggers, R. Braspenning, and C. Shan, "Analysis of shopping behavior based on surveillance system," in *2010 IEEE International Conference on Systems, Man and Cybernetics*. IEEE, oct 2010, pp. 2512–2519, ISBN: 978-1-4244-6588-0.

[5] M. C. Popa, L. Rothkrantz, C. Shan, T. Gritti, and P. Wiggers, "Semantic assessment of shopping behavior using trajectories, shopping related actions, and context information," *Pattern Recognition Letters*, vol. 34, no. 7, pp. 809–819, may 2013, ISSN: 0167-8655.

[6] T. Zhao, R. Nevatia, and F. Lv, "Segmentation and tracking of multiple humans in complex situations," in *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2001)*, vol. 2. IEEE, dec 2001, pp. II–II, ISBN: 0-7695-1272-0.

[7] T. Zhao and R. Nevatia, "Stochastic human segmentation from a static camera," in *Workshop on Motion and Video Computing. Proceedings.* IEEE, Dec 2002, pp. 9–14, ISBN: 0-7695-1860-5.

[8] A. Leykin and M. Tuceryan, "A vision system for automated customer tracking for marketing analysis: Low level feature extraction," in *Human Activity Recognition and Modelling Workshop*, vol. 3. Citeseer, 2005, pp. 6–13.

[9] R. Sicre and H. Nicolas, "Human behaviour analysis and event recognition at a point of sale," in *2010 Fourth Pacific-Rim Symposium on Image and Video Technology*. IEEE, Nov 2010, pp. 127–132, ISBN: 978-1-4244-8890-2.

[10] E. Frontoni, P. Raspa, A. Mancini, P. Zingaretti, and V. Placidi, "Customers' activity recognition in intelligent retail environments," in *New Trends in Image Analysis and Processing (ICIAP 2013)*. Springer Berlin Heidelberg, 2013, pp. 509–516, ISBN: 978-3-642-41190-8.

[11] Y. Hu, L. Cao, F. Lv, S. Yan, Y. Gong, and T. S. Huang, "Action detection in complex scenes with spatial and temporal ambiguities," in *2009 IEEE 12th International Conference on Computer Vision*. IEEE, 2009, pp. 128–135, ISBN: 978-1-4244-4420-5.

[12] A. Zadeh, R. Zellers, E. Pincus, and L.-P. Morency, "Multimodal sentiment intensity analysis in videos: Facial gestures and verbal messages," *IEEE Intelligent Systems*, vol. 31, no. 6, pp. 82–88, Nov 2016, ISSN: 1941-1294.

[13] M. Chen, S. Wang, P. P. Liang, T. Baltrušaitis, A. Zadeh, and L.-P. Morency, "Multimodal sentiment analysis with word-level fusion and reinforcement learning," in *Proceedings of the 19th ACM International Conference on Multimodal Interaction (ICMI '17)*. New York, NY, USA: ACM, 2017, pp. 163–171, ISBN: 978-1-4503-5543-8.

[14] Y. Wang and B. Li, "Sentiment analysis for social media images," in *2015 IEEE International Conference on Data Mining Workshop (ICDMW)*. IEEE, 2015, pp. 1584–1591, ISBN: 978-1-4673-8493-3.

[15] R. Ekman, *What the face reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS)*. Oxford University Press, USA, 1997.

[16] "Ibm watson - visual recognition," URL: https://www.ibm.com/watson/services/visual-recognition/ [accessed: 2019-09-29].

[17] "Microsoft cognitive services: Face api," 2019, URL: https://azure.microsoft.com/en-us/services/cognitive-services/face/ [accessed: 2019-09-29].

[18] "Kairos apis and sdks," 2019, URL: https://www.kairos.com/ [accessed: 2019-09-29].

[19] "Amazon rekognition - video and image," URL: https://aws.amazon.com/rekognition [accessed: 2019-09-29].