FOOD PRODUCT ACCEPTANCE AND PREFERENCE PREDECTION THROUGH AUTOMATED FACIAL EXPRESSION ANALYSIS

MEDICIÓN DE ACEPTACIÓN Y PREFERENCIA DE PRODUCTOS ALIMENTICIOS MEDIANTE ANÁLISIS AUTOMATIZADO DE EXPRESIONES FACIALES

Julieta Dominguez Soberanes
Universidad Panamericana Aguascalientes, Mexico
jdominguez@up.edu.mx

Víctor Manuel Alvarez Pato
Universidad Panamericana Aguascalientes, Mexico
valvarez@up.edu.mx

Claudia Nallely Sanchez Gomez
Universidad Panamericana Aguascalientes, Mexico
cnsanchez@up.edu.mx

Jose Sebastian Gutiérrez Calderon
Universidad Panamericana Aguascalientes, Mexico
jsgutierrez@up.edu.mx

David Eduardo Mendoza Perez
Universidad Panamericana Aguascalientes, Mexico
demendoza@up.edu.mx

Ramiro Velazquez Guerrero
Universidad Panamericana Aguascalientes, Mexico
rvelazquez@up.edu.mx

Reception: 22/October/2019 **Acceptation:** 2/December/2019

Abstract

This paper presents a system for determining consumer acceptance and preferences of food products through emotion recognition. Changes in facial expressions of 80 test subjects while tasting five different food samples were captured using the Microsoft Kinect sensor. The expressions were compared to the consumers' sensory evaluations. To determine the facial expressions in every video frame, a neural network was trained and different supervised learning techniques (such as support vector machines, multilayer perceptron and regression trees) were used to predict which of the different tastes could be accepted or rejected. A neuronal net was used, when observing the confusion matrix a percentage of adequate recognition was obtained for the following emotions: neutral (94%), surprise (98%), happiness (99%) and disgust (94%). The industrial application of the proposed system is relevant for the Food Industry Research and Development (R&D) by allowing the sampling of a product by potential consumers and analyzing their emotions before launching into market.

Keywords: consumer's acceptance, sensory analysis, facial expressions, Kinect, machine learning, visualization.

Resumen

Se presenta un sistema para determinar la aceptación y las preferencias de usuarios en productos alimenticios a través del reconocimiento de emociones. Los cambios en expresiones faciales de 80 sujetos mientras probaban cinco distintos productos alimenticios fueron capturados con el sensor Microsoft Kinect. Las expresiones faciales se contrastaron con las evaluaciones sensoriales de los consumidores. Para el reconocimiento de expresiones faciales en cada cuadro del video, se entrenó una red neuronal y se utilizaron diferentes técnicas de aprendizaje supervisado (como máquinas de soporte vectorial, perceptrón multicapa y árboles de regresión) para determinar que sabores podrían ser aceptados y rechazados por el consumidor. Se decidió utilizar la red neuronal, y al observar la matriz de confusión se obtiene un porcentaje de reconocimiento adecuado para las siguientes emociones: neutral 94%, sorpesa 98%, felicidad 99% y disgusto 94%. La aplicación industrial es relevante en el sector de Investigación y Desarrollo de la industria de alimentos.

Palabras Claves: *aceptación de consumidores, análisis sensorial, aprendizaje de máquina, expresiones faciales, Kinect, visualización.*

1. Introduction

Consumers' sensory tests predict market performance and acceptance of new food products [Danner, 2014]. Nowadays, there is a high rate of failure of these products in market, which makes it necessary to implement new methodologies that can guarantee their success. Conventional methods to assess consumer acceptance of a product rely on explicit measurements (e.g. written evaluations with intensity scales), which may be practical and easy to use, but also tend to provide cognitively biased results [Lagast, 2017]. Consumer choices in real life are mostly driven by unconscious mechanisms rather than rational ones [Köster, 2009], and emotions play an important role in the decision-making process [Rozin, 1987], which implies that a proper emotion measurement system might be useful for evaluating new products sales.

Recognition of emotions elicited by food products is still an open problem [Kostyra, 2016], and automatic emotion detection in human faces is a frequent approach to solving it [Lagast, 2017]. There is a claim that facial expressions are universal and mostly unrelated to cultural differences. In addition, they are a result from the activation of a set of facial muscles, psychologist Paul Ekman developed the FACS (Facial Action Coding System) which allows to interpret a person's emotions by examining his facial expressions, described as a combination of isolated activation of facial muscles also known as AUs (Action Units) [Donato, 1999]. For example, a common motion in a face expressing joy is a smile, which results from tension in the zygomaticus major muscle, classified as AU 12 or "lip corner puller" by FACS.

Using Ekman's FACS, a trained expert can manually code nearly any facial expression, deconstructing it into its specific AUs, but this skill demands extensive training and manual coding is time consuming. Nevertheless, the process has been automated with computer algorithms. There is readily available software like Noldus FaceReader and the Computer Expression Recognition Toolbox (CERT) that estimate consumer acceptance of food products [Danner, 2014; Leitch, 2015; Juodeikiene, 2014; Littlewort, 2011].

Previous research associated with sensory analysis shows that facial expressions add insights to the knowledge about perception of basic tastes, mainly when concentration of stimuli was studied. The results show that some of the facial expressions are innate [Wendin, 2011]. Research based on understanding consumer emotions in the food industry is of great interest, since it can be used to differentiate acceptability of food products and understand consumers' emotions. In [Crist, 2018], the authors used an automated facial expression analysis (AFEA) in order to determine if it could predict food acceptability. However, results showed that the AFEA algorithm needed to improve its analysis sensitivity. The purchase intention is correlated to hedonic liking and facial expression [Zi, 2018]. This information was enriched when the subjects measured their expressions according to a scale: 0, 20, 40, 60, 80, and 100%. A study reported [Walsh, 2017] emotional response through AFEA, but added the use of frontal cortex (EEG) and of cardiac electrical activity (ECG). The methodology was based on showing subjects different videos regarding food concerns and analyzing their facial expressions. Other research [He, 2014], studied reactions to pleasant and unpleasant odors. Their results showed that pleasant odors were associated with neutral expressions whereas more intense unpleasant odors were associated with fewer neutral expressions and more expressions of disgust.

The main idea of this research project is to create a novel platform for predicting whether a consumer would like a food product or not (acceptance), as well as the preference for different types of flavors based on his/her facial expressions.

2. Methods

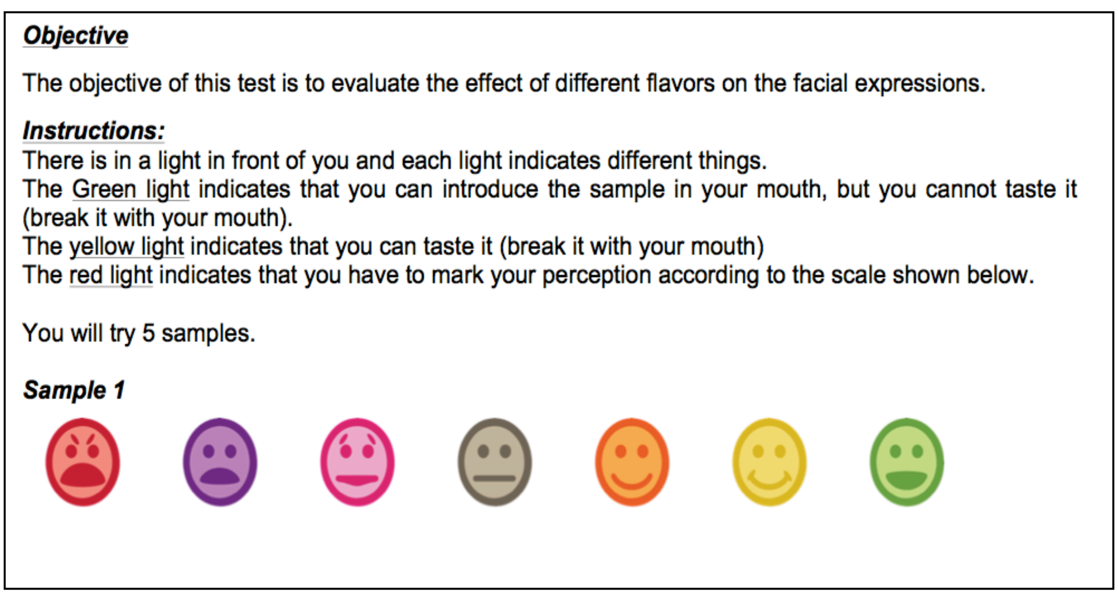
Flavor samples description

The use of different flavors was thought beforehand in order to provide five different sensory sensations. We presented only five samples in order to avoid saturation. From this point of view, three flavors considered as pleasant: mint (Demian, USA), pineapple (Demian, USA) and grape (Demian, USS). Two of them were unpleasant flavors: clam (Bell, 930001, USA) and smoke (Castells, Mexico). After trying each sample, consumers drank water and ate crackers. The flavors were encapsulated (figure 1) in order to keep the volunteers from guessing the sample's taste beforehand and to control the release of these flavors in the precise moment that the facial expression was to be measured. Therefore, when the sample was tasted, the flavor was considered a surprise for the volunteers and the facial expression was measured at that precise moment of the experiment. Later on, the consumers answered a sensory questionnaire. The information was analyzed with the facial expression for each sample evaluated.

Figure 1 Food flavors.

Questionnaire

We designed a questionnaire, which comprises a hedonic scale of seven points for each of the five samples with different flavors were tested (figure 2). These questionnaires are used in sensory science for testing acceptance of different type of food products. The results obtained were compared with the facial expressions.



Objective

The objective of this test is to evaluate the effect of different flavors on the facial expressions.

Instructions:

There is in a light in front of you and each light indicates different things.

The Green light indicates that you can introduce the sample in your mouth, but you cannot taste it (break it with your mouth).

The yellow light indicates that you can taste it (break it with your mouth)

The red light indicates that you have to mark your perception according to the scale shown below.

You will try 5 samples.

Sample 1

Figure 2 Sensory Questionnaire.

Information of the capturing application

We developed an application to store all experiments' data with the following information:

- Name of the volunteer
- Video of every test
- Time and frame in which the volunteer tries each flavor sample

Capture device

The capture device for this research was a Kinect, a system developed by Microsoft which integrates various sensors such as a color camera, an infrared light camera and a depth sensor, making it possible to capture the frontal view and the facial geometry in a single device, and eliminating the need of managing and synchronizing multiple sources of information. For this test, we used the frontal image of the person. However, other images that were gathered and will be analyzed for future publication. The distance in which a subject was placed was 0.5 m to 1 m approximately, but this parameter has no impact in the final results because the camera resolution is adequate for this type of experiments.

Semaphore device

We built a small semaphore device so the staff could let the volunteer know the right moment to try each sample. This allows a better synchronization between the reaction of the user and the records captured by the staff. Figure 3 shows a conceptual design for the experiment.

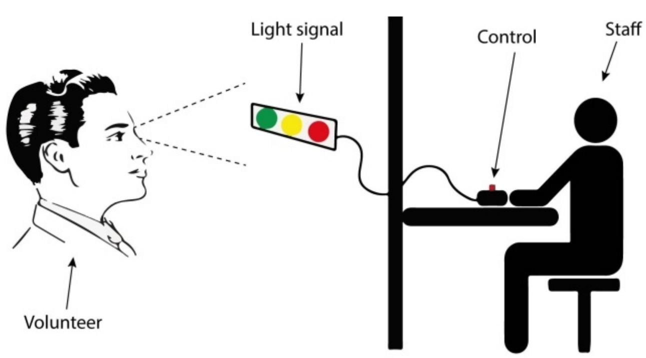


Figure 3 Conceptual design of the experiment.

Test

A group of 80 students, professors and employees from Universidad Panamericana (Mexico) agreed to serve as volunteers for the experiments. The test took place in the Sensory Laboratory. Each cabin has its illumination controlled as shown in figure 4 and the distance between the kinect and the subject ranged from 0.5 to 1 m. The volunteer was instructed to try a sample when the green light of the semaphore was turned on. The image/video was taken when the sample was tasted, and later on was processed. The consumer was asked to answer the sensory questionnaire.

Figure 4 Sensory laboratory.

Expression analysis

We processed all captured images through a neural network in order to classify them according to the face expression. The first part of the procedure consisted in extracting several landmarks in an image [Sekhon, 2010; Alvarez, 2018]. This involves two steps. The first, in which we apply an algorithm based on a Histogram of Gradients (HOG) [Kazemi, 2014; Pissaloux, 2013], in this process we could detect if there was a face within the image. If so, the algorithm returns the coordinates of a rectangle containing the detected face. Both coordinates and the original image work as the input for the Kazemi algorithm [King, 2009], which locates the coordinates of 68 landmarks (figure 5) related to specific facial features as mouth and eye corners, nose borders, etc (figure 6).

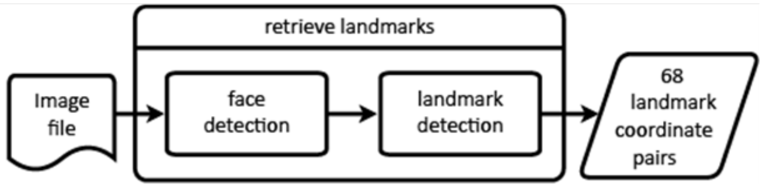


Figure 5 Landmark retrieving process.

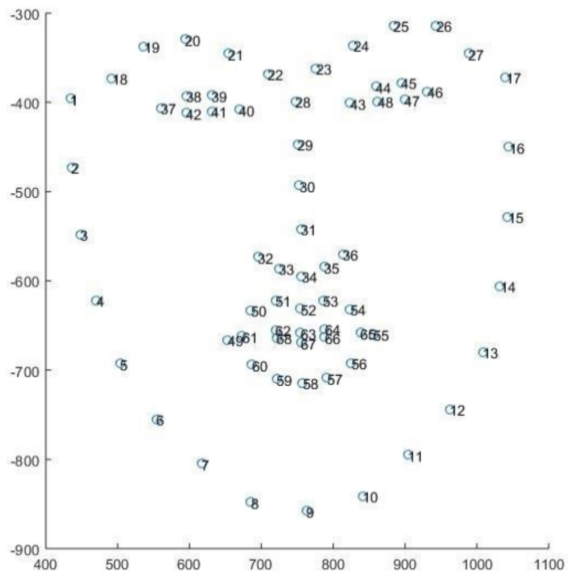


Figure 6 Location of landmarks on a sample face.

Both algorithms were implemented in Dlib [Lucey, 2010], a C++ open source toolkit. We normalized these vectors to diminish the effect of scale and rotation differences among various faces. Afterwards, we employed this process to extract the coordinates for two images of the same person: one showing a neutral expression and another one expressing emotion. The neutral coordinates were subtracted from the other expression's to obtain a set of 68 vectors (deltas) (figure 7).

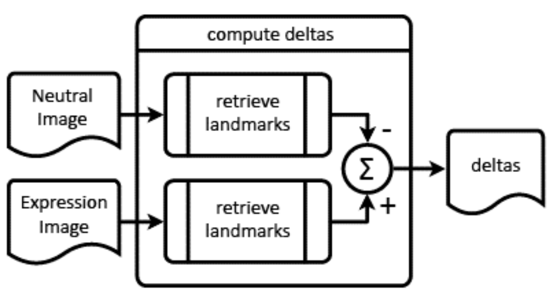


Figure 7 Calculation of deltas.

We calculated the deltas for two sets of images from the CK+ database, the latter one was used because: 1) it is commonly used in similar studies and 2) because it is well labeled by experts and 3) finally because neutral expressions and another with different facial expressions can be extracted [Lucey, 2010]. From the same database, we extracted labels that classify each expression shown in the images. Using these delta sets as input and the matching labels (anger, contempt, disgust, fear, happiness, sadness, surprise or neutral) as desired outputs. Since we found that neural networks showed the best results on previous work [Alvarez, 2018], we trained a neural network (figure 8). After several tests, the network configuration with the best results proved to be a 10-neuron array for the only hidden layer.

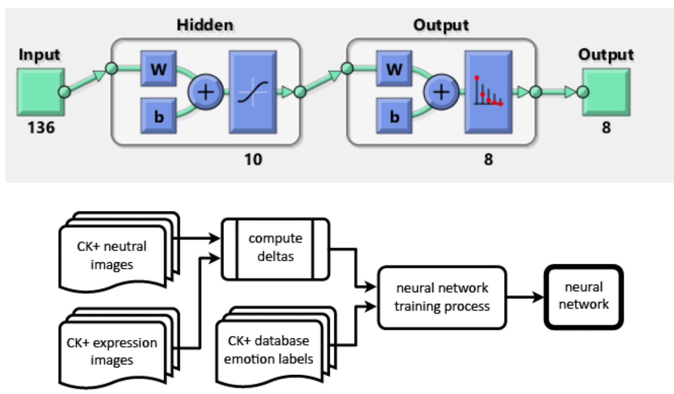


Figure 8 Neural network training in structure

Once trained, we fed (380 images, 70% were used for training, 15% for the validation and the rest for the analysis) the neural network with the delta sets from the images acquired in the experiment (neutral and expressive separately) as shown in figure 9.

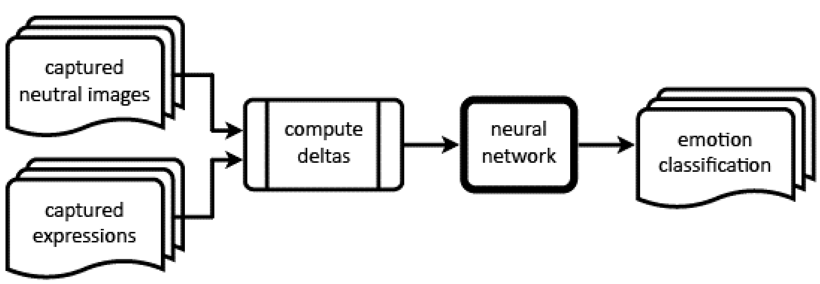


Figure 9 Classification of emotions through a neural network.

We obtained a .csv file for each individual with the following information: video frame number, a number indicating the number of faces found, frame filename and eight columns (one for each classified emotion) with numbers from 0 to 1 which state the probability that the face in the image expresses the matching emotion. In figure 10, we can appreciate an example of the emotion classification in one frame. The trained network's confusion matrix is shown in table 1.

Figure 10 Classification of emotions for each frame

Table 1 Confusion Matrix for the trained Neural Network.

	Anger	**Contempt**	**Disgust**	**Fear**	**Happy**	**Sadness**	**Surprise**	**Neutral**
Anger	**0.71**	0.02	0.13	0.00	0.02	0.04	0.00	0.08
Contempt	0.04	**0.54**	0.00	0.00	0.00	0.06	0.00	0.36
Disgust	0.06	0.00	**0.94**	0.00	0.00	0.00	0.00	0.00
Fear	0.00	0.00	0.00	**0.78**	0.06	0.03	0.07	0.06
Happy	0.00	0.00	0.01	0.00	**0.99**	0.00	0.00	0.00
Sadness	0.09	0.00	0.00	0.00	0.00	**0.62**	0.00	0.28
Surprise	0.00	0.01	0.00	0.00	0.00	0.01	**0.98**	0.00
Neutral	0.03	0.00	0.00	0.00	0.00	0.03	0.00	**0.94**

When observing the confusion matrix a percentage of adequate recognition was obtained for the following emotions: neutral (94%), surprise (98%), happiness (99%) and disgust (94%).

Regarding the rest of the emotions the percentage was acceptable but not as high, such as sadness (62%), disgust (54%), angry (71%) and fear (78%).

Our net obtains a general accuracy of 81.25%, which is comparable with other studies, which take the same database as reference. For example, [Liu, 2014] report

92.4%, [Jung, 2015] 70.24% and [Zeng, 2018] 94.79%, respectively. To the best of our knowledge, it is important to establish that the database has not been used for sensory purposes.

3. Results

After performing the expression analysis, we obtained the following data:

- The result from the procedure described above, a .csv file for each video and a frame record which contains the frame number and eight values in the range [0,1], expressing the probability to have an expression of anger, contempt, disgust, fear, happy, sadness, surprise and, neutral.
- The volunteer's evaluations for each flavor sample in the range [-3,3].
- The frame number where the volunteer tried each sample.

To analyze the whole information, we created a file containing the volunteer's response to a sample flavor based on his/her written responses. Then, we obtained a frame interval or window size w_S for each flavor sample taking $w_S/4$ frames before and w_S frames after the frame number stored in the application as a reference of the right instant in which the volunteer tried the sample. Afterwards, we calculate a frequency vector that represents the volunteer response for each emotion. The vector has 10 entries, where the first one contains the number of frames where the emotion had a value between 0 and 0.1, the second entry the number of frames where the emotion had a value between 0.1 and 0.2, and so on. This generated a vector of 80 features (10 per expression) that contains the expression frequencies of a volunteer's reaction to a specific flavor sample.

We treated the data as a regression problem for which the inputs are the expression probability histograms and the outputs are the volunteer's acceptance grades, within the range [-3,3]. We later trained a multilayer perceptron model [Santillana, 2011] using the Python library Scikit-learn [Pedregosa, 2011] with default parameters.

We utilized the mean absolute error (MAE) to measure the effectiveness of our approach [Sánchez, 2017]. MAE is defined in equation (1).

$$\mathrm{MAE} = \sum_{i=1}^{n} \frac{\lfloor \hat{y}_i - y_i \rfloor}{n} \tag{1}$$

Where:

MAE : Absolute error

y_i : Volunteer's evaluation of the sample.

$\hat{y}_i$: The predicted value.

As a rule, smaller values indicate better performance. We evaluated our results through cross-validation, which is a model evaluation technique widely used to validate the results of an experiment. It consists of randomly divided the samples into k groups. Then the model is trained and tested k times, the first one the training set is formed by the first k-1 groups and the test set for the last one, the second time the training set is formed by the first group and the last k-1 groups and the test set is formed by the second group, and so on. In this case, the number of groups is k=5. Each time an error is calculated and finally, the error represents the average of the k executions. These represent that the test was performed several times (5 times in total) with different randomly order partitions of 80 volunteers. Table 2 shows the MAE obtained from different window sizes, which means the number of frames taken into account for the measurement. There is little difference among the different window sizes. When evaluating the MAE, which can be from [-3 to 3], we can say that a 1.65 MAE is relatively small.

Table 2 MAE using the histograms of all the expressions.

w_s	20	30	40
MAE	1.65	1.63	1.61

However, 10 features per expression may prove to be too many and some of them may not be adding valuable information to the model. Table 3 shows the MAE using only the features of the histograms related to contempt and disgust, noticeably, all the values are smaller than those in table 2. In this case, the best performance corresponds to the multilayer perceptron with a w_s value of 30 and 40.

Table 3 MAE using the histograms of contempt and disgust.

w_s	20	30	40
MAE	1.56	**1.51**	**1.51**

The smaller MAE error was 1.51, which is small considering the used scale of seven points, from -3 to 3, and the complexity of this problem.

4. Discussion

The methodology proposed in this paper has given good results for sensory evaluation of food products in consumer analysis. The practical application of this study is fit for objectively measuring consumer acceptance or preference and market analysis for new food products based on Ekman categories, however, it would be interesting to determine whether the like or dislike expressions are expressed properly through this categories. Our neural network easily confuses sadness with contempt and a neutral expression. This will require finer tuning in the future, or a different classification scheme. Nevertheless, these emotions are not as relevant to our line of study as disgust, for example, most of the similar studies use commercial solutions whereas this study can be fined tuned to suit particular problems, because we have a better control of the algorithm.

As future work, it could be useful to find a way to pre-process the data, segmenting those volunteers that are very expressive from the least expressive ones, in order to attain clearer measurements. More precise correlations between sensory science and facial expressions are required.

5. Conclusions

This work could benefit from a feature selection technique to identify the exact expressions related to food preference/acceptance, these would lead to a better understanding of the acceptance of the product.

Future studies could be centered in obtaining data from different sources besides images, for example: temperature, galvanic skin response, heart rate, etc. Another interesting project would be to measure the emotional impact of food products while taking the environment into account i.e. restaurant mood, decoration, etc. This might be achieved through the use of a VR system.

6. Bibliography and References

[1] Álvarez, V.M., Velázquez, R., Gutiérrez, S. & Enríquez-Zarate, J. A method for facial emotion recognition based on interest points, 2018 International Conference on Research in Intelligent and Computing in Engineering (RICE), San Salvador, 2018, pp. 1-4.

[2] Alvarez, V.M., Sánchez, C.N., Gutiérrez, S., Domínguez-Soberanes, J. & Velázquez, R. Facial emotion recognition: a comparison of different landmark-based classifiers, 2018 International Conference on Research in Intelligent and Computing in Engineering (RICE), San Salvador, 2018, pp. 1-4.

[3] Crist, C.A., Duncan, S.E., Arnade, E.A., Leitch, L.A., O'Keefe, S.F. & Gallagher, D.L. Automated facial expression analysis for emotional responsivity using an aqueous bitter model, Food Quality and Preference, Vol. 68, 349-359, 2018.

[4] Danner, L., Haindl, S. Joechl, M. & Duerrschmid, K., Facial expressions and autonomous nervous system responses elicited by tasting different juices, Food Research International, Vol. 64, 81-90, 2014.

[5] Donato, G., Bartlett, M.S., Hager, J.C., Ekman, P. & Sejnowski, T.J., Classifying facial actions, IEEE Trans. Pattern Anal. Mach. Intell., vol. 21, no. 10, 974–989, 1999.

[6] He, W., Boesveldt, S., de Graaf, C. & de Wijk, R. A. The relation between continuous and discrete emotional responses to food odors with facial expressions and non-verbal reports, Food Quality and Preference, Vol.48, 130-137, 2016.

[7] Juodeikiene, G., Basinskiene, L., Vidmantiene, D., Klupsaite, D. & Bartkiene, E. The use of face reading technology to predict consumer acceptance of confectionery products, 9th Baltic Conference on Food Science and Technology, Latvia, 276-279, 2014.

[8] King, D.E., “Dlib-ml: A Machine Learning Toolkit, Journal of Machine Learning Research, Vol. 10, 1755-1758, 2009.

[9] Jung, H., Lee, S., Yim, J., Park, S., & Kim, J. Joint fine-tuning in deep neural networks for facial expression recognition. In Proceedings of the IEEE international conference on computer vision (pp. 2983-2991), 2015.

[10] Kazemi, V. & Sullivan, J., One millisecond face alignment with an ensemble of regression trees, IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, 1867-1874, 2014.

[11] Köster, E. P. Diversity in the determinants of food choice: A psychological perspective, Food Quality and Preference, Vol. 20, No. 2, 70-82, 2009.

[12] Kostyra, E., Rambuszek, M., Waszkiewicz-Robak, B., Laskowski, W., Blicharski, T. & Poławska, E. Consumer facial expression in relation to smoked ham with the use of face reading technology. The methodological aspects and informative value of research results, Meat Sci., vol. 119, 22–31, 2016.

[13] Lagast,S., Gellynck, X., Schouteten, J.J., De Herdt, V. & De Steur, H., Consumers' emotions elicited by food: A systematic review of explicit and implicit methods, Trends in Food Science & Technology, Vol. 69, Part A, 172-189, 2017.

[14] Leitch, K.A., Duncan, S.E., O'Keefe, S., Rudd, R. & Gallagher, D.L., Characterizing consumer emotional response to sweeteners using an emotion terminology questionnaire and facial expression analysis, Food Research International, Vol. 76, Part 2, 283-292, 2015.

[15] Liu, M., Li, S. Shan, S. Wang, R. & Chen, X. Deeply learning deformable facial action parts model for dynamic expression analysis. In ACCV, 2014, pages 1749–1756. IEEE, 2014.

[16] Littlewort, G. et al., "The computer expression recognition toolbox (CERT), 2011 IEEE International Conference on Automatic Face and Gesture Recognition and Workshops, 298–305, 2011.

[17] Lucey, P., Cohn, J.F., Kanade,T, Saragih, J., Ambadar, Z. and Matthews, I., The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression, IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Francisco, CA, USA, 94-101, 2010.

[18] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O. & Vanderplas, J. Scikit-learn: Machine learning in Python. Journal of machine learning research, 2011.

[19] Pissaloux, E., Maybank, S. & Velázquez, R. On Image Matching and Feature Tracking for Embedded Systems: A State of the Art. In: A. Chatterjee, H. Nobahari & P. Siarry (Eds.), Advances in Heuristic Signal Processing and Applications, Springer-Verlag (Berlin-Heidelberg), 2013, pp. 357-380.

[20] Rozin, P. & Fallon, A.E. A perspective on disgust, Psychological Review, vol. 94, No. 1, 23-41, 1987.

[21] Sánchez, C., Rivera, M. & Velázquez, R. Robust multiband image segmentation method based on user clues, 2017 IEEE 37th Central America and Panama Convention (CONCAPAN XXXVII), Managua, 2017, pp. 1-6.

[22] Santillana, A., Delgado-Mata, C., & Velázquez, R., Training a single-layer perceptron for an approximate edge detection on a digital image, 2011 International Conference on Technologies and Applications of Artificial Intelligence, Taiwan, pp. 189-193, 2011.

[23] Sekhon, B.S. Food nanotechnology –an overview, Nanotechnology, science and applications, Vol. 3, 1-15, 2010.

[24] Walsh, A.M., Duncan, S.E., Bell, M.A., O'Keefe, S.F. & Gallagher, D. L. Integrating implicit and explicit emotional assessment of food quality and safety concerns, Food Quality and Preference, Vol. 56, 212-224, 2017.

[25] Wendin, K., Allesen-Holm, B.H. & Bredie, W. L., Do facial reactions add new dimensions to measuring sensory responses to basic tastes?. Food quality and preference, Vol. 22, No. 4, 346-354, 2011.

[26] Zeng, N., Zhang, H., Song, B., Liu, W., Li, Y., & Dobaie, A. M. Facial expression recognition via learning deep sparse autoencoders. Neurocomputing, 273, 643-649, 2018

[27] Zhi, R., Wan, R., Zhang, D. & Li, W., Correlation between hedonic liking and facial expression measurement using dynamic affective response representation, Food Research International, Vol. 108, 237-245, 2018.