



ASSESSMENT OF FACIAL EXPRESSIONS IN PRODUCT APPRECIATION

M.C. Popa*, L.J.M. Rothkrantz†, P. Wiggers‡, C. Shan§

Abstract: In the marketing area, new trends are emerging, as customers are not only interested in the quality of the products or delivered services, but also in a stimulating shopping experience. Creating and influencing customers' experiences has become a valuable differentiation strategy for retailers. Therefore, understanding and assessing the customers' emotional response in relation to products/services represents an important asset. The purpose of this paper consists of investigating whether the customer's facial expressions shown during product appreciation are positive or negative and also which types of emotions are related to product appreciation. We collected a database of emotional facial expressions, by presenting a set of forty product related pictures to a number of test subjects. Next, we analysed the obtained facial expressions, by extracting both geometric and appearance features. Furthermore, we modeled them both in an unsupervised and supervised manner. Clustering techniques proved to be efficient at differentiating between positive and negative facial expressions in 78% of the cases. Next, we performed more refined analysis of the different types of emotions, by employing different classification methods and we achieved 84% accuracy for seven emotional classes and 95% for the positive vs. negative.

Key words: *product emotions, facial expression analysis, geometric features, appearance features, unsupervised learning, supervised learning*

Received: August 8, 2016

DOI: 10.14311/NNW.2017.27.009

Revised and accepted: March 10, 2017

1. Introduction

Nowadays, retailers are looking for new ways of enhancing their customers' satisfaction and differentiating themselves from their competitors, as described in [17]. An important step towards building a reliable and long-term relation with their

*Mirela Carmia Popa – Corresponding author; University of Maastricht, Faculty of Humanities and Sciences, Department of Data Science and Knowledge Engineering, Bouillonstraat 8-10, 6211 LH, Maastricht, the Netherlands, E-mail: Mirela.Popa@maastrichtuniversity.nl

†Leon Rothkrantz; Delft University of Technology, Section of Interactive Intelligence, Department of Intelligence Systems, Mekelweg 4, 2628 CD Delft, The Netherlands, E-mail: L.J.M.Rothkrantz@tudelft.nl

‡Pascal Wiggers; Hogeschool van Amsterdam, Theo Thijssenhuis, Wibautstraat 2-4 1091 GM Amsterdam, E-mail: p.wiggers@hva.nl

§Caifeng Shan; Video and Image Processing Department, Philips Research, HTC 36, 5656 AE Eindhoven, The Netherlands, E-mail: Caifeng.Shan@philips.com

customers regards the analysis of rational and emotional aspects related to purchasing choices and decisions. Understanding customers' shopping behaviour, their relation to products, what catches their attention and what remains unobserved has been of interest for companies since early times [23].

The traditional approach implies using questionnaires or interviews to find out the customers' opinion regarding a set of products. This method has some limitations, as it implies a lot of time and lacks the means of distinguishing between a rational and an emotional answer. Usually people lack the time and energy to respond to an extensive set of questions which could reveal their impressions. Besides, when asked explicitly, people often have no stable answers to what they like and what they don't like, i.e. people cannot always articulate their preferences well [3]. By analysing the non-verbal behaviour of customers a lot of cues can be extracted related to their perception of a product. Body language, gestures, speech tonality and semantics but most of all facial expressions indicate in what kind of product/products the customer is interested.

An approach towards assessing customers' shopping behavior is presented in [18], using the information obtained from different types of cameras. From the video recordings, the authors extract features related to the spatio-temporal behavior of trajectories, the dynamics and the time spent in each region of interest (ROI) in the shop, and regarding the customer-products interaction patterns. An example of the type of images acquired with a fish-eye camera, used in our analysis is displayed in Fig. 1a, the view obtained from a side camera depicting an interaction pattern of a customer with products is presented in Fig. 1b, while for a more refined assessment of the customer reaction to products, we need to use another type of video camera, able to capture the customer's face as shown in Fig. 1c. In a real-

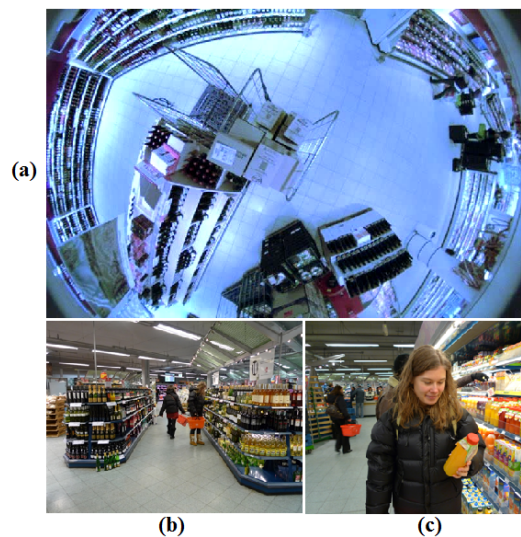


Fig. 1 (a) Surveillance video camera system in a shop. (b) Example of a customer-product interaction. (c) Example of a customer appreciating a product.

life setting, the problem of face detection is challenging, as it needs to cope with different face orientations, changing illumination conditions, occlusions, and huge variation of faces in terms of age, gender, face color, and also other characteristics such as glasses or beards. Still, due to privacy issues, we could not record this type of information, and we limit our analysis in this paper to a pilot study in which participants gave their consent to be recorded.

Our main contribution in this work involves detecting if a customer shows positive or negative facial expressions in appreciating a product. This type of information, whether the customer likes or dislikes a product, can be used to compute statistics about products and exposure to products, and most importantly to assess if the customers buy the product(s) after positive appreciation. Positive appreciated products should be placed in the foreground, while the products which are negatively appreciated should be replaced or presented in another way. Another application of understanding and recognizing customers' emotions to products supports the development of intelligent systems which could react according to the user's preferences in an on-line shopping context leading to an enriched human-computer interaction. Therefore, we present in this paper both types of analysis, a general one, driven by statistical measures and a user-centered one aimed at identifying the user profile and preferences. The methodology followed to obtain emotional data consisted of showing a set of forty product images to a number of twenty-five test subjects. The facial expressions displayed by the subjects while looking at each product image were recorded and used to build a database. Moreover, another contribution of this paper refers to the database acquisition, collecting relevant product images and using an experimental set-up which enabled spontaneous reactions, as participants were not instructed about how they should react. In this paper the techniques used for measuring product related emotions are outlined. We are following two approaches. The first one is an unsupervised one, which aims at discovering if the clusters obtained using different methods are meaningful and group together similar emotions. The second approach exploits the labelled data in a supervised manner and looks for the best feature descriptors and classification method. Consequently, a third contribution of this work encompasses the identification of the types of emotions displayed during product appreciation. Furthermore, an investigation of the best methodology for facial expressions analysis in terms of feature extraction methods and classification algorithms is performed.

The outline of the paper is as follows. In Section 2, the related work is presented. Next, in Section 3, we describe the database acquisition process. The approach followed to develop a facial expression analysis system is discussed in details in Section 4. Experimental results are presented in Section 5 and the paper ends with conclusions and directions for future work.

2. Related work

The human face represents one of the most important non-verbal means of communication, as it can reveal information about the emotional state of a person. Facial analysis has been of interest for cognitive sciences, as psychology and neuroscience, but also for computer scientists and engineers who study the face from the viewpoint of graphics, animation, computer vision, and pattern recognition.

As the focus of this paper regards facial expression analysis we present the main advances in this field. Many studies are focused on recognizing the six basic emotions (anger, disgust, fear, happiness, sadness, and surprise) defined in [8], but real-life situations showed that during human-human communication many more facial expressions can be observed, depicting pure or blended emotional states.

The palette of approaches varies not only in the selection of emotion triggering stimuli and the choice of emotional facial expressions database, but also in the feature extraction techniques and in the classification methods. There are several emotional databases available, such as Cohn-Kanade Facial Expressions Database, or the MMI database [15], but also 3D facial expressions databases such as BU-3DFE or UMP-3DFE [9]. Still, they are limited to the six basic emotions, which are less common in the product appreciation domain where the interest resides in distinguishing between positive and negative emotions. There is a focus on the six basic emotions because they are assumed to be universal. In product appreciation customers can express their emotional state in many ways, but still we assume positive/negative facial expressions are universal. Regarding the representation model of emotions there are two main streams: discrete emotion states and dimensional continuous emotion space. The well-known valence and arousal continuous emotional model [19] seems more appropriate for assessing customers' appreciation as it enables facial expressions ranking based on the pleasantness (valence) of the stimuli on one axis and the activation (arousal) on the other one. Approaches towards analysing user relation with products were made by [13]. They use gaze estimation to discover which type of products draw customers' attention. A deeper understanding of the emotions elicited by products was done by [6] who proposes 21 product emotional classes. This study is very useful for designers as it shows which product features influence the user's emotional state. We used a selection of the emotional classes proposed by them to label facial expressions in our study. Facial expressions analysis proved to be useful also in other fields, for inferring player experiences automatically as presented by [21], for assisting the emotional development of children with autism spectrum disorders, at identifying generalized patterns of pain or for deception detection [14]. Concerning visual information representation used to distinguish facial movements we can observe two main streams: local and holistic approaches. The local methods analyse only a set of points or special regions of the face. One technique efficient at detecting facial landmarks, besides extracting, aligning, and scaling faces from images is the Active Appearance Model (AAM) [4]. Another local feature extraction technique is represented by Facial Animation Parameters (FAPs) extraction [16]. Several facial emotion recognition systems have employed the use of geometrical features, which measure the displacement of special points on the face between the initial and the current frame or applied to 3D data by [12]. Holistic methods process the whole face and from the widely used ones we can mention optical flow estimation [24]. A number of studies applied Gabor wavelet analysis, Principal Component Analysis (PCA), and also small-scale appearance features using Local Binary Patterns (LBP) [20].

Even though the amount of research on emotion recognition using facial expression analysis is impressive, most of the studies focus on the six basic emotions due to the universality of the emotions or the availability of labelled databases. Still, our current interest is driven towards investigating emotions elicited by products and we present in the next section our proposed methodology.

3. Data description

Our objective consisted of investigating which emotions are induced by products and how to recognize them from facial expressions. The methodology followed in our study had several steps and had to satisfy several constraints. We wanted to elicit spontaneous emotions and in the same time to be able to record them. A simple, yet efficient solution was to use a set of product pictures displayed on a laptop equipped with an integrated web cam. The experimental setup is depicted in Fig. 2a. The test subjects were asked to express their opinion regarding a set of product images. Afterwards, they were asked to rate their appreciation on two five-points scales, representing the valence and arousal SAM (Self-Assessment Manikin [2]). An example of the developed graphical interface used in the experiments is provided in Fig. 2b. The experiment was done in two phases. First, subjects were shown the product pictures and their reaction was recorded. In a second step, we asked them to rate the products in terms of pleasure (from disliking a product to really liking it) and arousal (from indifference to excitement). Deciding which set of product pictures to use, represented another research question. The International Affective Picture System (IAPS) [11] contains a set of pictures well distributed over the valence and arousal dimensions and has been used in studies about inducing emotions, but it was not a suitable alternative for our experiment, as we needed product related pictures. Therefore, we used several sources (e.g. marketing magazines and companies websites) to gather an appropriate set of product pictures which could span a wide range of emotions, while examples of the pictures used in the experiment are shown in Fig. 3.

The final image set contained forty product pictures, from several categories (e.g. cars, bicycles, mobile phones, clothes, shoes, bags, jewellery, watches, beds, and food) of interest for both men and women. Participants in the experiment were 25 individuals (20–60 years old, 10 women and 15 men) having a computer science background. The session duration was about 20–25 minutes/test. We analyzed the subjects' appreciation scores and obtained a correlation of 0.52 on the valence self-report scores (negative, neutral, and positive) and of 0.38 for the arousal scores (calm, neutral, and aroused). A reason why the valence confidence was higher than the arousal one was due to the subjects' difficulty to rate their arousal level.



Fig. 2 (a) Experimental setup. (b) Graphical Interface used in the experiments.



Fig. 3 Selection of product images.

If the pleasantness of a product seemed intuitive for all participants, the scoring of their arousal was less natural and more time consuming for most of them. The products for which most of the test subjects gave a similar score were represented in the valence and arousal space (see Fig. 4), in order to provide a view of their overall appreciation. A representation of the distribution across all subjects of the given product scores on the valence and arousal scales is presented in Fig. 5. The

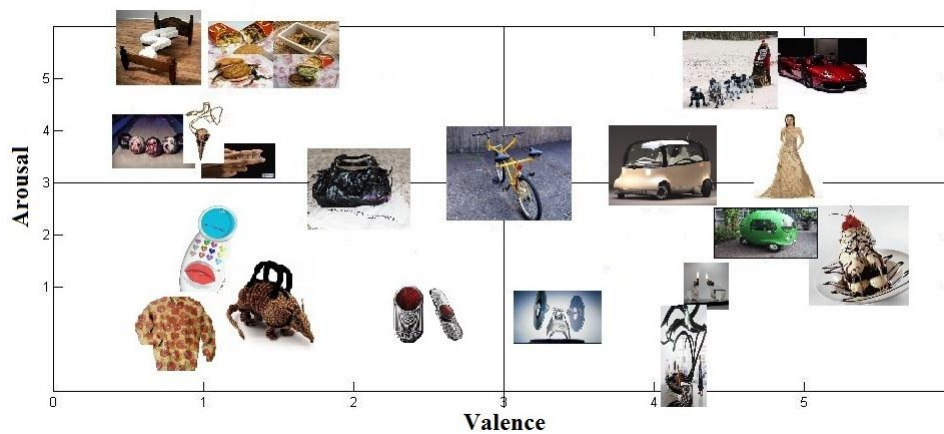


Fig. 4 Representation of product appreciation in the valence and arousal space.

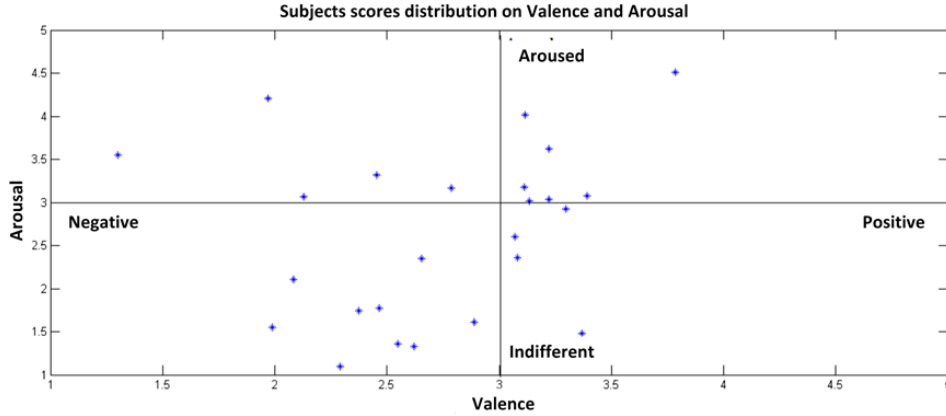


Fig. 5 Subjects distribution in the Russel space based on the products scores.

coordinates for each subject $s_i(V, A)$ are computed using a weighted average (high scores receive a higher weight) of the provided scores for each product:

$$s_i(V, A) = \sum_{j=1}^n w_j * p_j(V, A), \text{ where } \sum_{j=1}^n w_j = 1. \quad (1)$$

The presence of each subject in one of the four quadrants can be regarded as an indication of the general reaction of each of them. Accordingly, the subjects which were positively reacting to most of the products are represented in the (positive-aroused) quadrant, while the ones which were indifferent and did not like the set of products in most of the cases, can be found in the (negative-indifferent) quadrant.

Apart from studying the self-report correlations, we also investigated the types of facial expressions displayed by the participants. In order to select the most meaningful expression for each product and for each participant we used the following procedure. We manually labelled the fiducial points which will be introduced in Section for the first image of the set of facial images corresponding to each product. Next we used the AAM to track the landmark points in the rest of the images. We computed the distance between the first neutral image and the remaining set of images and selected the image with the highest distance.

This method was employed for automatically selecting the apex facial expression for each product. The cases when this approach failed were due to errors introduced in the tracking process, but also due to different head poses or occlusions by the subject's hand. After correcting all the mentioned errors, we obtained an image set of 1000 pictures, which was labelled in several categories. We need to mention that only expressive frontal facial images were selected for a further analysis. The agreement between two annotators on positive vs. negative facial expressions was very high, reaching a value of 0.89. A more refined sub-division into seven emotional classes (*admiration, like, dislike, disagreement, happiness, disgust, and neutral*) obtained a lower correlation of 0.78, mainly due to the different labels assigned to

the *dislike* and *disagreement* classes. An example of the displayed facial expression by the participants is depicted in Fig. 6. Next, we present the methodology used to analyse product related facial expressions.

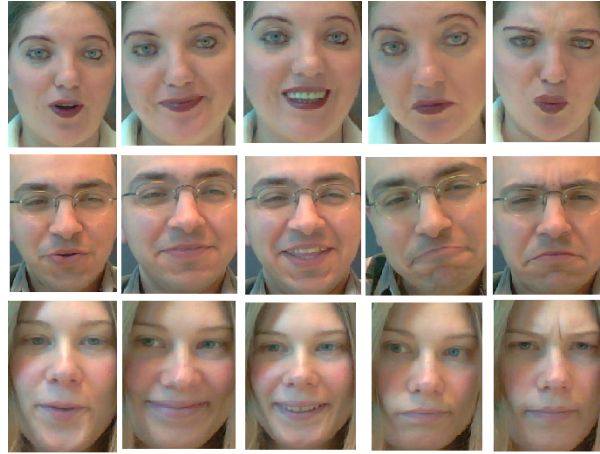


Fig. 6 Facial expressions elicited by products, (from left to right: admiration, like, happiness, dislike, and disagreement).

4. Recognition algorithms and methods

This section describes the algorithms and methods used for the recognition of the different product related emotions. The system flow of activities is presented in Fig. 7. The system receives as input video sequences containing individuals’ reaction to different product images. The pre-processing step is represented by the detection of the facial region. We adopted a commonly used face detection method, introduced in [22], based on Haar-wavelet features and a cascade of weak classifiers trained by Adaboost. Next, for each video frame in which a facial region was detected, we apply the next step consisting of feature extraction. Finally, the different feature sets are fed to either an unsupervised or a supervised learning module.

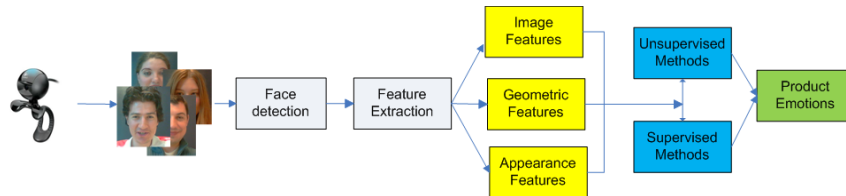


Fig. 7 Flow of activities of a facial expression recognition system.

4.1 Feature extraction

In the facial feature extraction step, a set of characteristics are derived from the original face images which are useful at representing the salient facial information. In this study we used three representations which are presented next.

4.1.1 Image features

As a first representation we used the raw gray-scale images of the facial region. Initially, all images have to be rescaled to the same dimensions to compensate for possible differences in scale changes. Next we constructed a feature vector by concatenating the image pixels. To reduce the data dimensionality we used Principal Component Analysis (PCA), by projecting the initial feature points to different principal components. This representation was initially chosen due its simplicity, as no further processing is required. In case a more refined analysis around the important facial landmarks is needed, new types of feature sets have to be computed.

4.1.2 Geometric features

The second chosen representation deals with the shape and location of the facial landmarks (e.g. mouth, eyes, eyebrows, nose, and chin). Different facial expressions correspond to different shape deformations of the facial landmarks. The location of the fiducial points was chosen according to the facial model proposed by [10] and is shown in Fig. 8a.

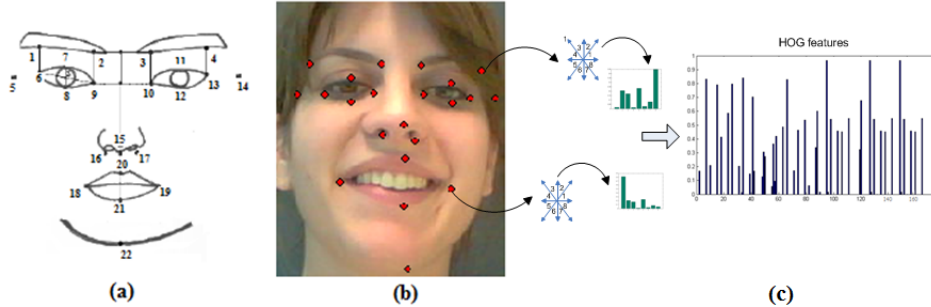


Fig. 8 (a) Facial model depicting the position of the chosen fiducial points. (b) Example of points position on a test subject face. (c) HOG feature extraction.

An alternative is to use the fiducial point coordinates as features in the classification process, but this representation achieves poor performance, as it is not able to capture the variations between different individuals. For increasing the discriminative ability of the feature set, we compute geometric features, which may be represented by segments, perimeters, or areas of the figures formed by the fiducial points. In the current approach, we computed relative Euclidean distances between fiducial points, after applying a normalization step. The considered distance set is

illustrated in Table I. Finally all distances are concatenated and form the geometric feature set: $f_G = (D_1, D_2, \dots, D_{16})$, formed of 16 values.

Nr.	Meaning	Definition	Nr.	Meaning	Definition
D_1	Mouth width	$\ x_{18} - x_{19}\ $	D_9	Left eye height	$\ x_7 - x_8\ $
D_2	Mouth height	$\ x_{20} - x_{21}\ $	D_{10}	Right eye height	$\ x_{11} - x_{12}\ $
D_3	Chin to mouth distance	$\ x_{22} - x_{21}\ $	D_{11}	Outer left eyebrow distance to eye	$\ x_1 - x_6\ $
D_4	Chin to nose distance	$\ x_{22} - x_{15}\ $	D_{12}	Inner left eyebrow distance to eye	$\ x_2 - x_9\ $
D_5	Left eyebrow width	$\ x_1 - x_2\ $	D_{13}	Outer right eyebrow distance to eye	$\ x_4 - x_{13}\ $
D_6	Right eyebrow width	$\ x_3 - x_4\ $	D_{14}	Inner right eyebrow distance to eye	$\ x_3 - x_{10}\ $
D_7	Left eye width	$\ x_6 - x_9\ $	D_{15}	Nose distance to left eye	$\ x_9 - x_{15}\ $
D_8	Right eye width	$\ x_{10} - x_{13}\ $	D_{16}	Nose distance to right eye	$\ x_{10} - x_{15}\ $

Tab. I *Definitions of Geometric features.*

4.1.3 Appearance features

A complementary set of features usually employed for facial expression recognition is represented by appearance features. Between their characteristics, we mention robustness to head pose, illumination changes, and also to inter-subject variability such as facial structure or skin tone. From the existing alternatives (e.g. LBP, SIFT, SURF) we chose Histograms of Oriented Gradients (HOG) [5], as they proved to be successful in the object recognition field. The difference from the usually employed method in object recognition, is that instead of detecting salient points using the SIFT descriptor, in our case the points of interest were introduced in Section 4.1.2 as the mouth corners, eyes corners, eyebrows corners and nose corners. Next, in each region, an 8-bin normalized histogram of HOG is constructed. Finally, the values corresponding to each region are merged into the appearance feature vector as shown in Fig. 8c, resulting into a 152 dimensional feature vector (8×19 points of interest).

4.1.4 Fused feature vector

For exploiting the complementarity of the proposed feature types, we also employed a fused feature vector, composed of geometric and appearance features. First the two types of features were normalized and then the fused feature vector was obtained by concatenating them, resulting into a 168 dimensional vector. As the proportion of the appearance features (90%) was much higher than the one of geometric features (10%), we projected the appearance features into a lower dimensional space using PCA. The number of components which accounted for 95% of the total variation was 20, resulting into a new appearance feature vector composed of 20 features and a new fused feature vector composed of 36 features. In this way the contribution of both types of features to the fused feature vector was similar.

4.2 Learning methods

Given the wide range of available classification approaches, we aim at finding the best model in terms of learning from instances (a set of examples in the training set) and also generalization to new instances. The choice of which specific algorithm to use is highly dependent on the dataset at hand, being a critical step. Classification techniques can be divided into two groups, namely supervised and unsupervised.

4.2.1 Unsupervised methods

In the case of unsupervised methods, the data analysis has an exploratory purpose and aims at finding the general characteristics or structure of the data, in absence of the data labels. Cluster analysis is applied in many disciplines, such as image segmentation, bioinformatics, information retrieval, or market segmentation. From the clustering algorithms used in pattern recognition, we mention just a few: mean-shift algorithm, pairwise clustering, fuzzy-clustering, path-based clustering, and spectral clustering. The k -means and EM-clustering methods are probably the best known algorithms in this field.

In the case, when data samples can be represented using a mixture of Gaussian distributions, the Expectation Maximization (EM) clustering algorithm introduced in [1] represents a valid alternative. The prior probability for each Gaussian is the fraction of points in the cluster defined by that Gaussian. These parameters can be initialized by randomly selecting means of the Gaussians, or by using the output of the k -means algorithm for initial centres. All clustering algorithms try to satisfy two constraints: the distances between all data objects belonging to the same cluster are as small as possible, while the distances between different clusters are as large as possible. Next, we introduce the supervised classification techniques considered in our study.

4.2.2 Supervised methods

A wide range of supervised learning algorithms are available, each with its strengths and weaknesses. There is no single learning algorithm that works best on all supervised learning problems. Therefore, we consider several methods which proved their efficiency at facial expression recognition. A possible separation of the supervised methods can be in spatial methods (e.g. Support Vector Machines (SVM), Fisher, or k -Nearest Neighbour (k -NN)) and spatio-temporal ones (e.g. Hidden Markov Models (HMMs) or Conditional Random Fields (CRF)) as presented in [7]. The decision of which specific learning algorithm we should use is a critical step and is mainly driven by the properties of the input data. As we aim at recognizing the facial expression from a single image, the methods from the spatial group will be considered for a detailed investigation. In the following section we present the experimental results obtained on the recorded database.

5. Experimental results

For the analysis of the facial expressions elicited by products, we investigated two different approaches. First, we present our findings using clustering techniques, followed by the description of the classification results.

5.1 Clustering approach

The purpose of applying clustering techniques was to investigate if any structure can be found in the input data. We analysed first the feature sets, proposed in Section 4.1 and the best result was obtained for the fused feature vector of geometric and appearance features, as they contain complementary information. As performance measure, we compared the obtained clusters with the annotated ground truth data. Our main research question was if we can distinguish between positive and negative facial expressions. Therefore, the chosen number of clusters k , was initially 2. The results obtained for the different clustering methods are presented in Table II. For separating the data into training and testing sets we used 10-fold cross validation and the average results over all folds are reported in the table below together with their standard deviation.

Clustering algorithm	Accuracy Exp. 1 [%]	Accuracy Exp. 2 [%]	Student's t -test
k -means	78 ± 2.7	84 ± 2.2	$p = 0.0^* < 0.05$
EM-clustering	78 ± 2.5	85 ± 2.1	$p = 0.0^* < 0.05$
hierarchical clustering	47 ± 0.6	48 ± 0.5	$p = 0.65 > 0.05$

Tab. II Clustering results and the associated Student's t -test.

Inspecting the results of the different clustering algorithms (k -means and EM-clustering) revealed an interesting finding. The unsupervised method was able to group together facial expressions of like and happiness in one cluster (93%) and of dislike and disagreement in the other one (91%). Still, samples corresponding to admiration (which is considered a positive emotion) were assigned to the negative cluster in 36% of the cases and some samples from the disgust emotional class were assigned to the positive cluster (30%).

Even though initially this result seems incorrect, the confusion done by the clustering algorithm can be explained by the similarity between different emotional classes around certain fiducial points. For the samples existing in our database, on one hand, both happiness and disgust are characterized by an open mouth, while the most important difference can be noticed around the eyes and eyebrows region; on the other hand, both examples from the dislike and admiration classes are characterized by a similar mouth shape.

Based on the findings of the first experiment, we applied the proposed clustering algorithms to a reduced set of data samples. In the second experiment we used only the samples corresponding to the (like, happiness, dislike, and disagreement) emotional classes resulting in a reduced dataset of 716 images. The obtained results (see Table II, Exp. 2) improved by 7%, showing that our assumption regarding the confusing emotional classes was correct.

In order to investigate whether the two sets of data were statistically different, we computed the Student's t -test between *accuracy Exp. 1* and *accuracy Exp. 2* for all the results in Tab. II. We obtained the test decision which rejects the null hypothesis at the 5% significance level for the first two clustering methods (k -means and EM-clustering), meaning that the sets of data samples corresponding to the two experiments were statistically significant different to means, while for

the hierarchical clustering method the null hypothesis was accepted with a p-value of 0.65, result explained as this method fails to separate the data into two clusters and mainly assigns all the samples to only one cluster.

Next, we also investigated different number of clusters, but the results were not as meaningful as in the first two experiments, leading to overlapping between different emotional classes. We also performed the clustering analysis, using a person-dependant approach. In this case, the results were better, leading to a clear separation between different emotional classes. The applicability of the unsupervised learning methods proved to be successful for the discrimination between positive and negative facial expressions, representing a good starting point. For the cases when additional information is needed regarding particular emotional classes, supervised methods are more appropriate to be used as they can exploit the labelled information.

5.2 Classification approach

The selection of the feature set used to describe the facial information and the chosen classification method represent two important decisions, which influence directly the success rate of the facial expression recognition system.

Regarding the validation of the results, we adopted the same testing approach (10-fold cross validation) for each experiment. Firstly, we investigated the discriminative power of the feature sets introduced in Section 4.1. The comparative accuracy results with their standard deviation are presented in Tab. III for the set of chosen classification methods applied to the recognition of positive and negative facial expressions.

Classifier	Image features [%]	Geometric features [%]	Appearance features [%]	Geometric Appearance features [%]
Parzen	71 ± 1.6	91 ± 2.1	92 ± 1.4	94 ± 1.8
Fisher	91 ± 1.8	87 ± 2.2	93.5 ± 2.3	95 ± 2.1
<i>k</i> -NN (<i>k</i> =2)	88 ± 0.9	91 ± 1.8	93 ± 1.7	95 ± 1.5
LDC	57 ± 1.5	87 ± 1.7	93 ± 1.6	95 ± 2.2
SVM (rbf kernel)	91 ± 1.8	85 ± 2.1	94 ± 0.9	95 ± 1.6
Adaboost	92 ± 1.9	90 ± 1.8	92 ± 2.4	91 ± 2.3

Tab. III Comparative accuracy results for positive vs. negative facial expression recognition.

The fusion of geometric and appearance features leads to the best recognition rate of 95%, as their properties are complementary and encapsulate maximum of facial information. Regarding the best performing classifiers, Fisher, KNN, LDC and SVM achieve the best accuracy, while the error analysis highlighted a similar conclusion as the one reached in the previous section, namely that a fraction of the admiration and disgust samples are assigned to the wrong emotional class.

A possible solution to overcome this limitation would be to increase the number of samples corresponding to the two classes, given that they are not so well

represented in the training set. The number of samples corresponding to each emotional class is given in Fig. 9a. The receiver operating characteristic (ROC) curves associated with each classifier are depicted in Fig. 10a, showing the true positive rate against the false positive rate at different thresholds and highlighting that all selected classification methods have a similar performance, except Adaboost, which is hindered by the limited size of the dataset.

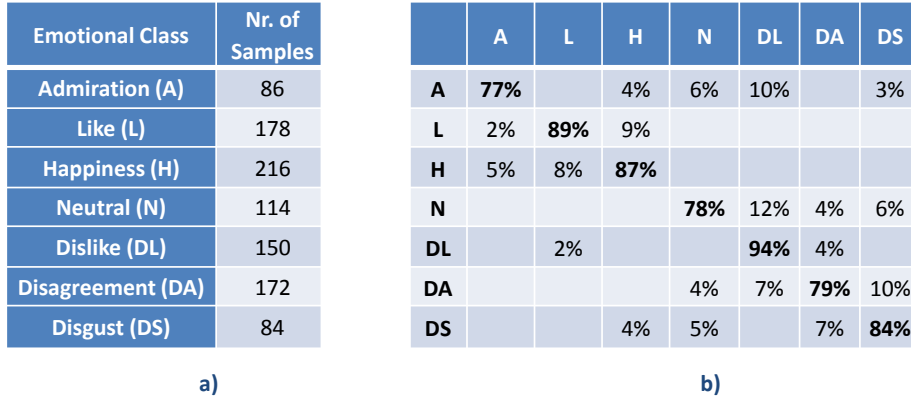


Fig. 9 (a) Number of data samples corresponding to each emotional class. (b) Confusion matrix for the seven emotional classes.

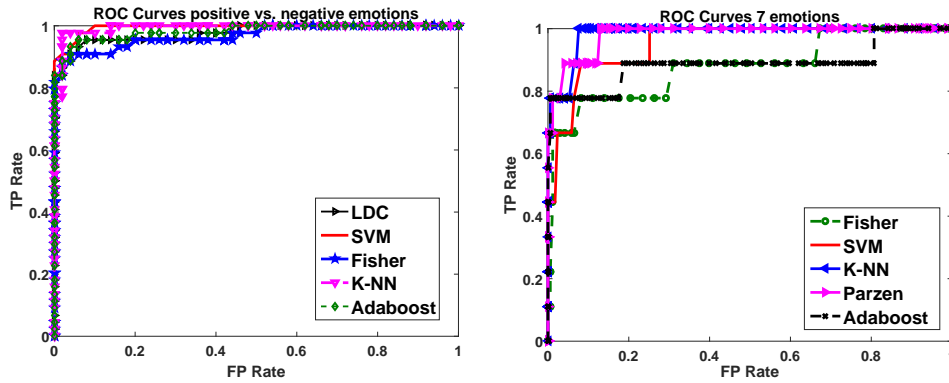


Fig. 10 (a) ROC curves for classifying positive vs. negative facial expressions. (b) ROC curves for classifying the seven emotional classes.

Another experiment aimed at refining the analysis of the facial expressions recognition problem to a discrete number of emotional classes. We employ the same feature sets and classification methods, while the achieved results are presented in Table IV.

Classifier	Image features [%]	Geometric features [%]	Appearance features [%]	Geometric Appearance features [%]
Parzen	67 ± 2.6	80 ± 2.1	82 ± 1.8	84 ± 1.6
Fisher	74 ± 1.8	62 ± 2.2	77 ± 2.1	81 ± 3.2
<i>k</i> -NN (<i>k</i> =2)	69 ± 1.8	80 ± 1.6	82.5 ± 1.4	84 ± 1.8
LDC	78 ± 3.1	67 ± 3.4	80 ± 2.8	82 ± 2.2
SVM (rbf kernel)	74 ± 1.9	68 ± 2.6	79 ± 2.2	82 ± 1.8
Adaboost	70 ± 2.4	76 ± 2.2	78 ± 2.1	81 ± 1.9

Tab. IV Comparative results for seven emotional classes.

The best results were obtained for the fused feature vector and the *k*-NN and Parzen classifiers. The ROC curves associated with the set of employed classifiers are depicted in Fig. 10b. For a better understanding of the overall results, for each particular emotional class, we provide the associated confusion matrix for all folds in Fig. 9b. It is interesting to notice that the confused emotional classes are usually inside the same group (positive or negative) and less between groups, which is a helpful information, if we are only interested in the positive vs. negative facial expression classification problem. Another observation obtained from the confusion matrix shows that the same emotional classes which had a lower agreement between the annotators are also difficult to be discriminated by the classifier (e.g. dislike and disagreement). A possible solutions to this problem was already mentioned, such as enlarging the database size. Furthermore, in both experiments, the *k*-NN classifier achieved the best performance, in spite of its simplicity, observation which can be explained as it operates in a low dimensional space, where the distance metric principle was useful at discriminating between different emotional labels. On the other hand, the properties of more complex classifiers, such as SVM will be better exploited in case of a higher dimensional space and a much larger training set. However, we are satisfied with the result achieved so far, which could enable the development of better solutions for on-line shopping, as the customers could benefit by means of the proposed system of better tailored applications around their needs and preferences.

6. Conclusions and future work

We presented in this paper an approach towards understanding customers' appreciation of products using facial expression analysis. As no database suitable for our purpose was available, we recorded a database of 25 test subjects' reaction to forty product images. We defined by experiments a set of product images which elicited a broad range of emotions in the valence and arousal space. We analysed the obtained facial expressions using different feature sets (image features, geometric features, and appearance features). Furthermore, we investigated the use of unsupervised and supervised learning methods. By applying clustering techniques we were able to find a separation between positive and negative facial expressions. A more detailed analysis of the displayed facial expressions was obtained by em-

ploying different classification techniques. The best performance of 84% for seven emotional classes (e.g. admiration, like, happiness, dislike, disagreement, disgust, and neutral) was obtained using a fused feature vector of geometric and appearance features and a k -NN or a Parzen classifier. The discrimination between positive and negative facial expressions was successful (95%), proving that it is possible to assess whether customers like or dislike the product. As future work, we plan to test our system on real life data of shopping people interacting with products and to increase the system accuracy by fusing multiple modalities and by taking into account context information.

Acknowledgement

This work was supported by the Netherlands Organization for Scientific Research (NWO) under Grant No. 018.003.017.

References

- [1] BILENKO M., BASU S., MOONEY R.J. Integrating constraints and metric learning in semi-supervised clustering. In: *Proc. 21st Int. Conf. on Machine Learning*, 2004, doi: [10.1145/1015330.1015360](https://doi.org/10.1145/1015330.1015360).
- [2] BRADLEY M.M., LANG P.J. Measuring emotion: the Self-Assessment Manikin and the Semantic Differential. *J. of Behavior Therapy and Experimental Psychiatry*. 1994, 25(1), pp. 49–59, doi: [10.1016/0005-7916\(94\)90063-9](https://doi.org/10.1016/0005-7916(94)90063-9).
- [3] CARENINI G., POOLE D. Constructed Preferences and Value-focused Thinking: Implications for AI research on Preference Elicitation, In: *AAAI-02 Workshop on Preferences in AI and CP: symbolic approaches*, Edmonton, Canada, 2002.
- [4] COOTES T.F., EDWARDS G.J., TAYLOR C.J. Active appearance models. In: *IEEE Tran. on Pattern Analysis and Machine Intelligence*, 2001, 23(6), pp. 681–685, doi: [10.1109/34.927467](https://doi.org/10.1109/34.927467).
- [5] DALAL N., TRIGGS B. Histograms of oriented gradients for human detection. *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR'05)*, 2005, 1, pp. 886–893, San Diego, California, doi: [10.1109/CVPR.2005.177](https://doi.org/10.1109/CVPR.2005.177).
- [6] DESMET P.M.A., HEKKERT P. The basis of product emotions. In: *W. Green and P. Jordan (Eds.), Pleasure with Products, Beyond usability*, London: Taylor and Francis, 2002, pp. 60–68, doi: [10.1201/9780203302279.ch4](https://doi.org/10.1201/9780203302279.ch4).
- [7] DUDA R.O., HART P.E., STORK D.G. *Pattern Classification, Second Edition*, John Wiley & Sons Inc. 2001, doi: [10.1007/s00357-007-0015-9](https://doi.org/10.1007/s00357-007-0015-9).
- [8] EKMAN P., FRIESEN W. *Facial Action Coding System*, Consulting Psychologists Press, Inc., Palo Alto California, USA, 1978.
- [9] HABIBU R., MASHOHOR S., HAMIRUCE M.M, SARIPAN M.I. UPM-3D Facial Expression Recognition Database (UPM-3DFE). In: *PRICAI 2012*. 2012, pp. 470–479, doi: [10.1007/978-3-642-32695-0_42](https://doi.org/10.1007/978-3-642-32695-0_42).
- [10] KOBAYASHI H., HARA F. Facial interaction between animated 3D face robot and human beings. In: *IEEE Int. Conf. on System, Man and Cybernetics*, 1997, pp. 3732–3737, doi: [10.1109/ICSMC.1997.633250](https://doi.org/10.1109/ICSMC.1997.633250).
- [11] LANG P.J., BRADLEY M.M., CUTHBERT B.N. International Affective Picture System (IAPS): Affective ratings of pictures and instruction manual. *Technical Report A-8*, University of Florida, Gainesville, FL, 2007.
- [12] LI X., RUAN Q., MING Y. 3D Facial expression recognition based on basic geometric features. In: *IEEE Int. Conf. on Signal Processing (ICSP)*, 2010, pp. 1366–1369, doi: [10.1109/ICOSP.2010.5656891](https://doi.org/10.1109/ICOSP.2010.5656891).

- [13] LIU X., KRAHNSTOEVEER N., YU T., TU P. What are customers looking at? *IEEE Conf. on Advanced Video and Signal Based Surveillance*. 2007, pp. 405–410, doi: [10.1109/AVSS.2007.4425345](https://doi.org/10.1109/AVSS.2007.4425345).
- [14] MATSUMOTO D., HWANG H.S., SKINNER L., FRANK M.G. Evaluating truthfulness and detecting deception. In: *FBI Law Enforcement Bulletin*, 2011, pp. 1–11.
- [15] PANTIC M., VALSTAR M.F., RADEMAKER R., MAAT L. Web-based Database for Facial Expression Analysis, *Proc. IEEE Int'l Conf. Multimedia and Expo (ICME'05)*, Amsterdam, The Netherlands, 2005, doi: [10.1109/ICME.2005.1521424](https://doi.org/10.1109/ICME.2005.1521424).
- [16] PARDAS M., BONAFONTE A. Facial Animation Parameters Extraction and Expression Detection Using HMM. In: *Signal Processing: Image Communication*, 2002, 17, pp. 675–688, doi: [10.1016/S0923-5965\(02\)00078-4](https://doi.org/10.1016/S0923-5965(02)00078-4).
- [17] PETERMANS A., VAN CLEEMPOEL K. Retail Design and the Experience Economy: Where are we going? In: *Design Principles and Practices*, 2009, 3(1), pp. 171–182.
- [18] POPA M.C., ROTHKRANTZ L.J.M, SHAN C., GRITTI T., WIGGERS P. Semantic Assessment of Shopping Behavior Using Trajectories, Shopping Related Actions, and Context Information, *Pattern Recognition Letters*. 2012, doi: [10.1016/j.patrec.2012.04.015](https://doi.org/10.1016/j.patrec.2012.04.015).
- [19] RUSSEL J.A. A circumplex model of affect. *Journal of Personality and Social Psychology*, 1980, 39, pp. 1167–1178, doi: [10.1037/h0077714](https://doi.org/10.1037/h0077714).
- [20] SHAN C., GRITTI T. Learning Discriminative LBP-Histogram Bins for Facial Expression Recognition. *Proc. British Machine Vision Conference (BMVC'08)*, Leeds, UK, 2008, doi: [10.5244/C.22.27](https://doi.org/10.5244/C.22.27).
- [21] TAN C.T., ROSSER D., BAKKES S., PISAN Y. A feasibility study in using facial expressions analysis to evaluate player experiences. In: *Proc. of the 8th Australasian Conf. on Interactive Entertainment*, 2012, pp. 1–10, doi: [10.1145/2336727.2336732](https://doi.org/10.1145/2336727.2336732).
- [22] VIOLA P., JONES M. Rapid object detection using a boosted cascade of simple features. In: *IEEE Conf. on Computer Vision and Pattern Recognition(CVPR)*, 2001, pp. 511–518, doi: [10.1109/CVPR.2001.990517](https://doi.org/10.1109/CVPR.2001.990517).
- [23] WELLS W.D., LO SCIUTO L.A. Direct Observation of Purchasing Behavior, *Journal of Marketing Research*. 1966, 3(3), pp. 227–233, doi: [10.2307/3149897](https://doi.org/10.2307/3149897).
- [24] YEASIN M., BULLOT B., SHARMA R. Recognition of Facial Expressions and Measurement of Levels of Interest from Video. *IEEE Trans. on Multimedia*. 2006, 8(3), pp. 500–508, doi: [10.1109/TMM.2006.870737](https://doi.org/10.1109/TMM.2006.870737).