The final publication is
available at https://link.springer.com/chapter/10.1007/978-3-319-54407-6_17

Emotion understanding using multimodal information based on autobiographical memories for Alzheimer's patients

Juan Manuel Fernandez Montenegro[1], Athanasios Gkelias[2], Vasileios Argyriou[1]

Kingston University[1], Imperial College[2]
J.Fernandezmontenegro@kingston.ac.uk, a.gkelias@imperial.ac.uk,
Vasileios.Argyriou@kingston.ac.uk

Abstract. Alzheimer Disease (AD) early detection is considered of high importance for improving the quality of life of patients and their families. Amongst all the different approaches for AD detection, significant work has been focused on emotion analysis through facial expressions, body language or speech. Many studies also use the electroencephalogram in order to capture emotions that patients cannot physically express. Our work introduces an emotion recognition approach using facial expression and EEG signal analysis. A novel dataset was created specifically to remark the autobiographical memory deficits of AD patients. This work uses novel EEG features based on quaternions, facial landmarks and the combination of them. Their performance was evaluated in a comparative study with a state of the art methods that demonstrates the proposed approach.

1 Introduction

Alzheimer Disease (AD) is a dementia characterized by the decline of various cognitive domains such as memory and learning ability, language expression difficulties or social cognition problems. Based on these symptoms and the importance of AD early detection, many research works are focused on the detection of those cognitive handicaps that characterise Alzheimer disease. This work is focused on the social cognition problems and memory related problems, in particular, these related with emotion expressions.

Many works are focused on studying dementia patients' capability to recognise emotions [10, 11] whereas a minority tries to analyse patients' facial expressions to specific stimulus. In contrast to other dementias such as Lewy Body dementia where there is lack of facial expression, AD patients' facial expression is increased [14]. This work focuses on the automatic detection of emotions to certain stimulus for AD early detection.

Different approaches for automatic emotion recognition are focused on the variety of human interaction capabilities or biological data. For example, the study of speech and other acoustic cues in [18], body movements in [19],Electroencephalogram (EEG) in [25], facial expressions or combinations of previous ones such as speech and facial expressions in [22] or EEG and facial expressions in [20]. Our approach will focus on EEG and facial emotion detection.

The study of facial expression was part of various disciplines since Aristotelian era but it was in 1978 when the first automatic recognition study appeared [8, 16]. Several techniques have been proposed for facial expressions interpretation. The most well known system is the Facial Action Coding System(FACS) [17]. FACS describes facial expressions as action units (AU), where each AU corresponds to a facial configuration. When it comes to the computational side of face analyses the known approaches can be classified as spatial or spatio-temporal and appearance or shape based. The first approach differentiates between methodologies that work with single images or with groups of successive frames. The second approach groups methods that use the appearance features of the face, such as pixel intensity and methods that use a description of the face shape. All of them face the same challenges, such as head-pose and illumination variations, registration errors, occlusions and identity bias. Some of these problems are not included in most of the available databases therefor some of them may not work properly on real conditions.

Several datasets, focusing on different applications, are available for emotion recognition. For example, DEAP dataset provides EEG and face recordings of participants while they watch musical videos just for the analysis of human affective states [9]; SEMAINE database aims to provide voice and facial information to study the behaviour of subjects interacting with virtual avatars [23]; MAHNOB-HCI database was created for the study of emotions while humans are watching multimedia, supplying several data such as audio, an RGB video and five monochrome videos of the face, EEG, ECG, respiration amplitude, skin temperature and eye-gaze data [21]; or CASMEII dataset which studies facial micro-expressions for security and medical applications, requiring cameras of higher frame rate and spatial resolution [24].

It has been proved that for AD patients, semantic, autobiographical and implicit memory are more preserved than recent memory; therefore our work is based on the subjects' autobiographical memory [12, 13, 15]. Thus a novel dataset was created based on these symptoms providing RGB, IR and Depth video data of the participants' faces, EEG and eye-gaze data.

The purpose of this work is to introduce human behaviour and face expression recognition techniques for the detection of early dementia symptoms. Our novel dataset contains recordings of the participants' reactions when specific images, related and unrelated with their personal life stories, are shown. The classification of different reactions related to the images displayed is performed using different data features included in our dataset, such as facial landmarks and EEG signals, as input to supervised learning approaches. Our study analyses expected emotions. Thus our classification is based on the expected emotions according to the images displayed during the test instead of classifying accordingly to the emotions felt (represented on the captured video). This work investigates healthy people to analyse the differences and level of the emotional inputs generated from the available image classes and generate a model that describes the reactions associated to the healthy group of people. Thus, every reaction detected out of this model could be considered as a possible sign of dementia.

The remainder of this paper is organized as follows: section 2 describes previous related work on behaviour and face expression recognition. Section 3 analyses the proposed methodology and in section 4 details on the evaluation process and the obtained results are presented. Section 5 gives some conclusion remarks.

2 Previous Work

In this section current state of the art facial and EEG based emotion recognition approaches are analysed.

2.1 Facial Emotion Recognition Approaches

Images and video sequences of faces are highly utilised as source for emotion recognition. There are several models to represent emotions and they define emotions according to the number of dimensions, such as the three dimension Schlosberg Model: pleasantness-unpleasantness, attention-rejection and sleep-tension [40]. Most of the facial recognition approaches use the Facial Action Coding System (FACS) [17] to describe facial human emotions such as happiness, sadness, surprise, fear, anger or disgust; where each of these emotions is described as a combination of AUs. Other approaches abandon the path of specific emotions recognition and focus on emotions' dimensions, measuring their valence, arousal or intensity [46, 22, 41].

The methods for facial emotion recognition can be classified according to the approaches used during the recognition stages: registration, features selection, dimensionality reduction or classification/recognition [8, 16].

Three different approaches can be used for face registration: whole face, parts or points registration. These registration approaches usually are based on Active Appearance Models (AAM) [27, 28]; a method that matches a statistical model of the face to the images to extract face landmarks and specific face areas. Whole face approaches get the features from the whole face. Littlewort et al [42] get image based features of the whole face, such as Gavor Wavelets, in order to detect AUs for pain recognition. Face parts approaches use face areas that contain the maximum amount of information related to face expressions, such as the eyebrows and the mouth. Nicolle et al [41] propose a method for emotion recognition (valence, arousal, expectancy and power) using a combination of whole face, face parts, points and audio features. This approach gets patches of the face on regions of interest and they use the log-magnitude Fourier spectra and other measures as features. Points based approaches use fiducial points for shape representation. Michel et al [33] use a tracker to get 22 fiducial points and calculate the distance of each point between a neutral and a peak frame. These distances are used as features of an Support Vector Machine (SVM) algorithm in order classify the emotions. Neutral and peak frames are automatically detected when the motion of the points is almost zero. Valstar et al uses Particle Filtering

Likelihoods [39] in order to extract 20 fiducial points, but they still have to select the initial position of these points manually. These points are normalised by respecting a neutral point(tip of the nose) and a scale transformation is also applied. The distances between certain points are used as features to recognise specific AUs using SVM.

When it comes to feature representation, methods can be divided in spatial and spatio-temporal approaches. Spatial approaches include shape representations, low-level histograms or Gabor representations amongst others. For example, Huang et al [43] proposed a spatial shape representation using groups of three fiducial points (triangular features) as input to a neural network classifier; and Sariyanidi et al presented in [44] a low-level histogram representation using local Zernike moments for emotion recognition based on kNN and SVM classifiers. On the other hand, spatio-temporal approaches get the features from a range of frames within a temporal window, detecting more efficiently emotions that cannot be easily differentiated in spatial approaches. Zhao et al [45] proposed a method that uses spatio-temporal local binary patterns as features and SVM for classifying facial expressions.

Once the features are selected, dimensionality reduction techniques such as PCA are used before classification in order to reduce challenges such as illumination variation, registration errors and identity bias.

The results from most of the approaches are not always reliable since many of them are tested on posed datasets such as CK [37] and MMI [38]. Therefore, the results are not reliable on naturalistic conditions regarding illumination, head-pose variations and nature of expressions. Nevertheless, there are non-posed datasets to test naturalistic expressions such as SEMAINE [23], MAHNOB-HCI [21] or DECAF [52]. In these cases the illumination and head-pose variation problems are taken into account depending on the aim of the study.

2.2 EEG Emotion Recognition Approaches

EEG based emotion recognition is a less common approach since the majority uses facial or speech data as source for emotion detection. Considering that these sources are easy to fake [25] amongst other problems, EEG provides an extra source that solves problems such as falseness, illumination or speech impaired subjects. On the other hand, EEG signal deals with other challenges such as noise and biological and non-biological artifacts [20, 30], such as electrooculogram (EOG), electromyogram (EMG) and electrocardiogram (ECG). Nevertheless, these biological artifacts are also affected by emotions and are expected to provide extra information to EEG signal for emotion recognition [20].

Two types of descriptors can be used for EEG signal analysis: simple descriptors such as frequency and amplitude; and more complex ones such as asymmetry metrics, time/frequency analysis, topographic mapping, coherence analysis or covariation measures. These descriptors are used depending on the area of study; for example, asymmetry metrics are usually applied in cognitive neuroscience [30]. In particular, asymmetric hemispheric differences are used for emotion recognition [29, 31]. Furthermore, state of the art methods use techniques such as Independent Component Analysis (ICA) for removing some artifacts, then they extract different features such as amplitude or spectral power and use them in classifiers such as k-Nearest Neighbour (kNN) or Support Vector Machine (SVM). For example, Vijayan et al [26] use DEAP data (data captured using 32 sensors) and filter 50Hz frequency to remove noise. Afterwards, they get the Gamma band from the signal and use auto-regressive modeling to obtain the features that are passed to an SVM classifier.

2.3 Facial and EEG Emotion Recognition Approaches

Few approaches utilise a combination of EEG and facial information to recognise emotions. The work in [20] considers both types of data using the MAHNOB-HCI database [21]. The EEG signal was captured using 32 sensors and the power spectral density was extracted from overlapping one second windows. The facial approach extracts 49 fiducial points and calculates the distance from 38 of these points to a reference point. Finally, they use regression models for emotion detection. As a result, they have obtained better results using the facial data and conclude that the good performance of the EEG results are due to the facial artifacts present in the EEG signal.

In this work a novel multimodal non-posed database is introduced. Due to the nature of our study, the environment where the RGB video is recorded is controlled avoiding illumination variations and occlusions. In addition, the head-pose variations are minimal since the video sequences are recorded while the participants are looking at the screen in front of them. Using this novel multimodal database a method based on expected emotions is presented. These emotions are not defined as specific standard emotions, therefore our approach does not use FACS or any other emotion coding system. The facial modality presented uses geometric based spatio-temporal features. For the EEG data a new feature is introduced based on quaternion principal component analysis using only four channels. Both modalities individually and combined are studied and compared with with state of the art methods.

3 Proposed Dataset And Methodology

This section describes the approach utilised to recognize the spontaneous reactions to specific visual stimulus. Next subsections describe our novel multimodal database and the proposed features used for emotion recognition.

3.1 Spontaneous Emotion Multimodal Database (SEM-db)

SEM database is a multimodal dataset for spontaneous emotional reaction recognition that contains multimodal information of nine participants aged between 30 to 60 years old with different educational background taken while completing cognitive/visual tests. Ten repetitions have been recorded per participant providing a total of 90 instances.

The novelty of SEM dataset is the non-posed reactions to autobiographical and non-autobiographical visual stimulus data. The main contribution of SEM database is the use of personalized images for each participant. These images are photos of themselves or their relatives and friends both from the recent and distant past. Moreover, the participants did not know that those images were used so the reactions were genuine. Additionally, images of famous and unknown to the subjects persons or places were shown. In more details we had the following classes of images with the corresponding expected spontaneous emotions or reactions.

a) 10 images of distant past faces of the subjects and their relatives.
b) 10 images of recent past faces of the subjects and their relatives.
c) 10 images of distant past group of relatives, including themselves.
d) 10 images of recent past group of relatives, including themselves.
e) 10 images of famous people.
f) 10 images of unknown to the subject persons.
g) 10 images of famous places/objects.
h) 10 images of unknown to the subject places/objects.

The recorded data is provided in different data modalities: HD RGB, depth and IR frames of the face, EEG signal and eye gaze data; which were recorded using 4 different devices: a 30fps HD RGB camera, IR/Depth sensors (Kinect), an eye tracker (Tobii eye tracker) and EEG sensors (Emotiv headset)(see Fig. 1). The recording of the data has been done in done in a controlled environment e.g. an office. The participants were asked to put on the EEG headset and they were seated in a comfortable chair in front of the test screen, the RGB camera, the Kinect sensor and the eye tracker. The height of their chair was adjusted in order the eye tracker to detect their eye movements (see Fig. 2). Once the eye tracker is detecting the participants' eyes and all the EEG sensors are receiving good quality signal the test begins. The instructions of the test are provided before they start and at the beginning of each test, a red image is displayed for synchronization.

3.2 Emotion Recognition using fiducial points and EEG Quaternion based supervised learning.

Our approach intends to classify the reaction of the participants using two data modalities: the EEG data and the fiducial points obtained from the RGB face images. Using each modality and combining them (see Fig. 3), two binary classifications have been performed trying to recognise spontaneous reactions from distant and recent memories that were triggered during our experiments (see Table 1). The main reaction to be detected is the 'positive recognition' reaction versus the 'indifference' reaction. Additionally, it is expected a stronger recognition reaction when the participant watches images from the distant past.

Our approach extracts features from both data modalities: EEG and Facial points. The facial fiducial points were obtained using Baltru et al approach [28]

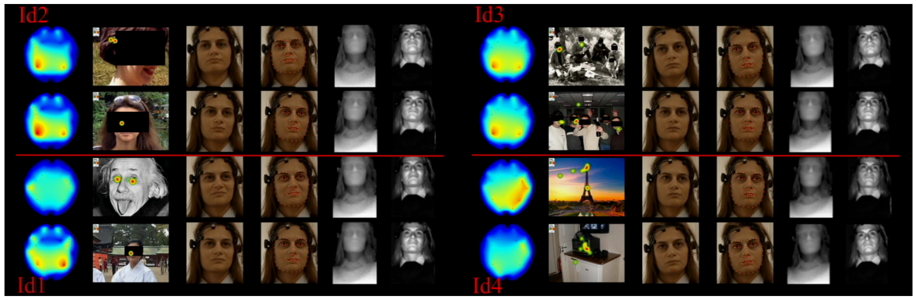

Fig. 1. Data modalities contained in the database and the related classes analysed in our approach (see Table 1 for the emotion definitions). The left figure shows from top to bottom, images of people from distant vs recent past; and famous vs unknown people. The right figure shows from top to bottom images of group of people from distant vs recent past; and famous vs unknown places. The different modalities from left to right in each case are EEG, gaze tracked heat map, RGB, facial landmarks, depth and IR.

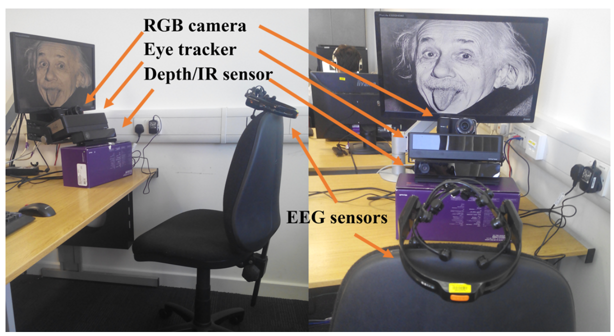

Fig. 2. Location of the devices during the recording of the database.

Table 1. Classes chosen for recognition and the expected reaction.

Id	Class 1	Class 2	Expected emotion
1	Famous faces	Unknown faces	Recognition vs Neutral reaction
2	Distant past images of the participant family and friends faces	Recent past images of the participant family and friends faces	Long term memory recognition vs short term memory recognition
3	Distant past images of group of people including the participant, family and friends faces	Recent past images of group of people including the participant, family and friends faces	Long term memory recognition vs short term memory recognition
4	Famous places, objects/brands	Unknown places and objects	Recognition vs Neutral reaction

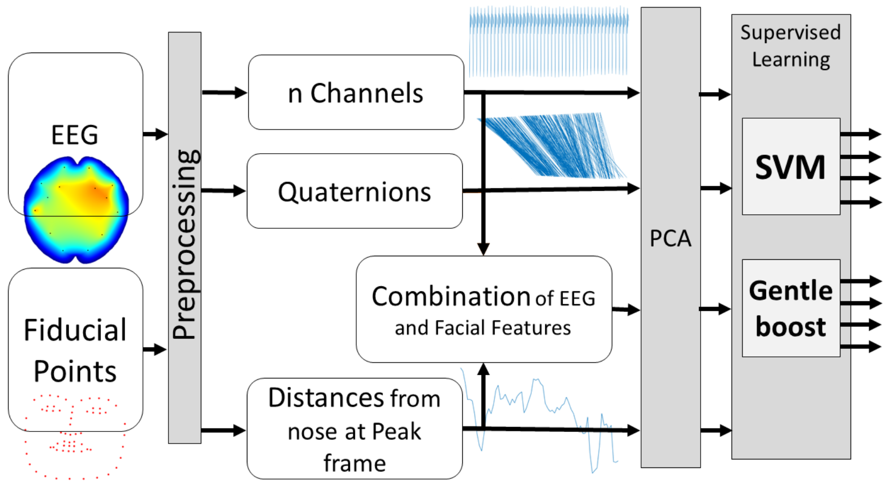

Fig. 3. Diagram demonstrating our approach. The two modalities of data used (Fiducial Points, EEG) go through the process independently. The combination of the features concatenates the features extracted from the fiducial points with the EEG features, including all combinations.

from a 30 frame rate video. This approach obtains 68 fiducial points per frame. The coordinates of the fiducial points have been preprocessed normalizing them according to a neutral face point (i.e. nose) [32] to obtain rigid head motions invariant features. The EEG data was recorded using a EEG headset (Emotiv Epoc) which collects EEG from 14 sensors at 128Hz.

Once the data was collected and preprocessed, the features were extracted. The spatio-temporal facial features studied were based on Michel et al work [33]. The distance of each coordinate to the nose point was measured. The first frame of each subject data was considered as neutral face since at the beginning of the test a neutral pose is expected. Each frame was compared to the neutral face, calculating the frame that varies most from the neutral face. That frame was selected as the peak frame and used as a p points long feature vector. For EEG, three variations of features were analysed: (i) a combination of the 14 channels, (ii) a combination of the four frontal channels and (iii) novel features combining the 4 frontal channels into a quaternion representation based on quaternion principal component analysis.

Quaternion principal component analysis (Quaternion PCA) is based on the fact that a vector can be decomposed in linearly independent components, such that they can be combined linearly to reconstruct the original vector. However, depending on the event that changes the vector, correlation between the components may exist from the statistical point of view (i.e. two uncorrelated variables are linearly independent but two linearly independent variables are not uncorrelated). In most of the cases during the feature extraction process complex or hyper-complex features are generated but decomposed to be computed by a classifier. For example, normals and gradients in 2D/3D are features that are consisted by more than one element and this decomposition can imply a loss of information.

To do so, vectorial features can be represented more precisely using a complex or hyper-complex representation [47, 48]. Since, in our case and many similar scenarios, vectorial features such as a location, speed, gradients or angles, are the primary source of information, a hyper-complex representation of these features is more efficient allowing better correlation between these channels [47–49]. The proposed method exploits the hyper-complex (quaternion) representation capturing the dependencies within the EEG sensors located on the sides of the head and the ones over the eyes, [51, 50].

Quaternion PCA is applied in order to reduce the number of the selected hyper-complex features without increasing the complexity. In more details, the quaternion representation was introduced in [36, 35] as a generalization of the complex numbers. A quaternion $q \in \mathcal{H}$ has four components:

$$q = q_r + q_i i + q_j j + q_k k \tag{1}$$

where $q_r, q_i, q_j, q_k \in \Re$ and i, j, and k satisfy

$$\begin{aligned} i^2 = j^2 = k^2 = -1,\ ij = -ji = k \\ jk = -kj = i, ki = -ik = j \end{aligned} \tag{2}$$

Conjugation of quaternions denoted by H is analogous to conjugation of complex numbers elements and is defined as:

$$q^H = q_r - q_i i - q_j j - q_k k. \tag{3}$$

The square of the norm of a quartenion is defined as

$$||q||^2 = q_r^2 + q_i^2 + q_j^2 + q_k^2 = q^H q. \tag{4}$$

with $(q_1 q_2)^H = q_2^H q_1^H$ and the four components (q_r, q_i, q_j, q_k) to correspond to the available four frontal EEG channels (AF3, AF4, F7 and F8).

Let quaternion column vector $\mathbf{q} = [q_1, \ldots, q_F]^T \in \mathcal{H}^F$ where T denotes simple transposition be the EEG values over time. The conjugate transpose of vector $\mathbf{q}$ is denoted by $\mathbf{q}^H$. There is an isomorphy between a quaternion and a complex 2×2 matrix defined as

$$\mathbf{Q} = \begin{bmatrix} q_r + q_i i & q_j + q_k i \\ -q_j + q_k i & q_r - q_i i \end{bmatrix} \tag{5}$$

Let $\mathbf{x}_l$ be the F-dimensional vector obtained by writing in lexico- graphic ordering and form $\mathbf{X} = [\mathbf{x}_1 | \cdots | \mathbf{x}_N] \in \mathcal{H}^{F \times N}$. Also we denote by $\bar{\mathbf{x}} = \frac{1}{N}\sum_{i=1}^{N} \mathbf{x}_i$ and $\overline{\mathbf{X}}$ the sample mean and the centralized sample matrix $\mathbf{X}$, respectively. A projection vector is denoted by $\mathbf{u} \in \mathcal{H}^F$ and by $y_i = \mathbf{u}^H \mathbf{x}_i$ the projection of $\mathbf{x}_i$

onto $\mathbf{u}$. We want to maximize the (sum of the) variances of the data assigned to a particular class

$$\begin{aligned} E(\mathbf{u}) &= \sum_{l=1}^{N} ||y_l - \tilde{m}||^2 = \sum_{l=1}^{N} ||\mathbf{u}^H(\mathbf{x}_i - \mathbf{m})||^2 \\ &= \mathbf{u}^H \sum_{l=1}^{N} (\mathbf{x}_l - \mathbf{m})(\mathbf{x}_l - \mathbf{m})^H \mathbf{u} \\ &= \mathbf{u}^H \mathbf{S} \mathbf{u} \end{aligned} \tag{6}$$

where $\mathbf{S} = \bar{\mathbf{X}}\bar{\mathbf{X}}^H$. It can be easily proven that matrix $\mathbf{S}$ is a quaternion Hermitian matrix i.e., $S_{ij} = S_{ji}^H$.

In order to find K projections $\mathbf{U} = [\mathbf{u}_1| \dots |\mathbf{u}_k] \in \mathcal{H}^{F\times K}$ we may generalize $E(\mathbf{U})$:

$$\begin{aligned} \mathbf{U}_o &= \arg\max_{\mathbf{U}\in\mathbb{H}^{F\times p}} E(\mathbf{U}) \\ &= \arg\max_{\mathbf{U}\in\mathbb{H}^{F\times p}} \mathrm{tr}[\mathbf{U}^H\mathbf{S}\mathbf{U}] \\ \text{s.t.} \quad & \mathbf{U}^H\mathbf{U} = \mathbf{I}. \end{aligned} \tag{7}$$

We aim at solving the above noted problem by using the isomorphic complex form that can be reformulated as

$$\begin{aligned} \tilde{\mathbf{U}}_o &= \arg\max_{\tilde{\mathbf{U}}} \mathrm{tr}[\tilde{\mathbf{U}}^H\tilde{\mathbf{S}}\tilde{\mathbf{U}}] \\ \text{s.t.} \quad & \tilde{\mathbf{U}}^H\tilde{\mathbf{U}} = \mathbf{I}. \end{aligned} \tag{8}$$

Since $\mathbf{S}$ is a quaternion Hermitian matrix, $\tilde{\mathbf{S}}$ is a complex Hermitian. Also, given that $\tilde{\mathbf{S}}$ is a positive semidefinite Hermitian matrix (i.e., it has only non-negative eigenvalues) the solution $\tilde{\mathbf{U}}_0$ is given by the p eigenvectors of $\tilde{\mathbf{S}}$ that correspond to p largest eigenvalues. We want an efficient algorithm for performing eigen-analysis to $\tilde{\mathbf{S}}$, which is a complex $2F \times 2F$ matrix and can be written as $\tilde{\mathbf{S}} = \tilde{\mathbf{X}}\tilde{\mathbf{X}}^H$ where $\tilde{\mathbf{X}} \in \mathbb{C}^{2n\times F}$ and needs $O((2F)^3)$ time.

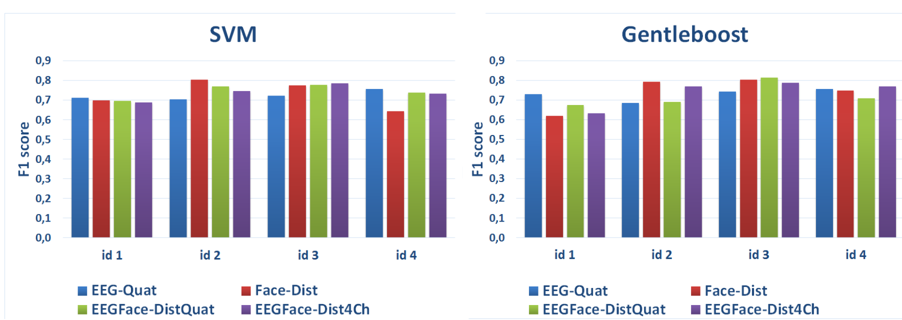

Fig. 4. Results of the features that provide the best results for each classification using SVM and Gentleboost classifier: EEG quaternion, Face distance and EEGFace distances plus quaternion.

In general, given a quaternion Hermitian matrix $\mathbf{A}$ then it has n nonnegative real eigenvalues (due to the non-commutative multiplication property of quaternions, there exist two kinds of eigenvalue; in this paper we are interested only

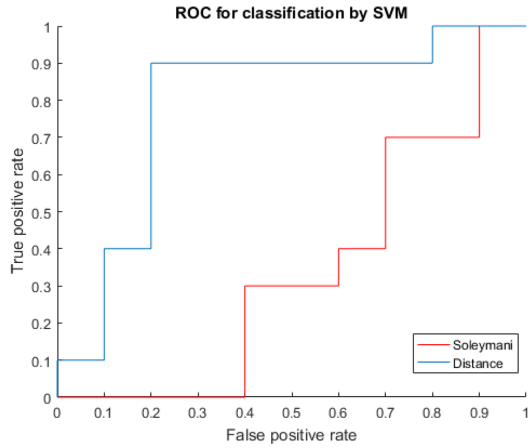

Fig. 5. ROC curve of the proposed method based on facial features in comparison to the ones proposed by Soleymani.

on the left eigenvalues) $\mathbf{l} = [\sigma_1, \ldots, \sigma_n]$. Let $\tilde{\mathbf{A}}$ be its complex form

$$\tilde{\mathbf{A}} = \begin{bmatrix} \mathbf{A}_r + i\mathbf{A}_i & \mathbf{A}_j + i\mathbf{A}_k \\ -\mathbf{A}_j + i\mathbf{A}_k & \mathbf{A}_r - i\mathbf{A}_i. \end{bmatrix}$$

then the eigenvalues of $\mathbf{l}_{2n} = [\sigma_1, \sigma_1, \ldots, \sigma_n, \sigma_n]$. Representing $\mathbf{A} = \mathbf{B}\mathbf{B}^H$, where $\mathbf{B}$ is a quaternion matrix, and considering $\tilde{\mathbf{A}}$ and $\tilde{\mathbf{B}}$ to be the complex forms of matrices $\mathbf{A}$ and $\mathbf{B}$, respectively, then, $\tilde{\mathbf{A}}$ will be given by $\tilde{\mathbf{A}} = \tilde{\mathbf{B}}\tilde{\mathbf{B}}^H$. So, based on this analysis, we can write $\tilde{\mathbf{S}} = \tilde{\mathbf{X}}\tilde{\mathbf{X}}^H$. Also by defining matrices $\mathbf{A}$ and $\mathbf{B}$ such that $\mathbf{A} = \mathbf{\Gamma}\mathbf{\Gamma}^H$ and $\mathbf{B} = \mathbf{\Gamma}^H\mathbf{\Gamma}$ with $\mathbf{\Gamma} \in \mathcal{C}^{m \times r}$, and considering $\mathbf{U}_A$ and $\mathbf{U}_B$ to be the eigenvectors corresponding to the non-zero eigenvalues $\mathbf{\Lambda}_A$ and $\mathbf{\Lambda}_B$ of $\mathbf{A}$ and $\mathbf{B}$, respectively, we finally obtain $\mathbf{\Lambda}_A = \mathbf{\Lambda}_B$ and $\mathbf{U}_A = \mathbf{\Gamma}\mathbf{U}_B\mathbf{\Lambda}_A^{-\frac{1}{2}}$.

Thus, according to the above, in a classification problem, we may represent the quaternion Hermitian matrix (descriptor) providing a subspace analysis method in the quaternion domain. Assuming that we have a quaternion matrix P with dimension $m \times n$, we consider n to be the total number of the captured data and m the number of the actual hyper-complex features. A quaternion PCA of P, as it was analysed above, seeks a solution that contains r $(r < m, n)$ linearly independent quaternion eigenvectors in the columns of Q $(m \times r)$ such that $P = QA$; where the rows of A $(r \times n)$ contain the r quaternion principal component (QPC) series. As a result, a solid representation of the selected quaternion features is obtained, while the computational complexity is low.

Besides the individual features modalities, a combination of the aforementioned EEG and facial features has been also analysed. This combination comprises the attachment of the EEG features vector to the facial one. Once the features are structured properly, dimensionality reduction is applied using PCA and the reduced features are used as input to two supervised learning algorithms:

Table 2. F1 scores obtained using SVM. See Table 1 for id definitions.

SVM		id 1	id 2	id 3	id 4	Overall
			EEG			
Soleymani [20]		0.6002	0.5677	0.6194	0.7122	0.6249
Proposed	14 Ch	0.5972	0.6882	0.6507	0.6704	0.6516
	4 Ch	0.6637	0.6965	0.7177	0.6725	0.6876
	Quaternion	**0.7105**	**0.7043**	**0.7225**	**0.7553**	**0.7232**
			Face			
Soleymani [20]		0.6235	0.6699	0.6722	**0.6942**	0.6650
Proposed	Dist	**0.6987**	**0.8028**	**0.7750**	0.6438	**0.7301**
			EEGFace			
Soleymani [20]		0.6429	0.7090	0.6502	0.6461	0.6620
Proposed	Dist +14 Ch	**0.6950**	0.6825	0.7452	0.7313	0.7135
	Dist +4 Ch	0.6887	0.7470	**0.7843**	0.7319	0.7380
	Dist +Quaternion	**0.6945**	**0.7699**	0.7774	**0.7372**	**0.7448**

SVM and GentleBoost. A leave one out approach is used so the features obtained from N-1 of the participants, being N the number of participants, are used for training and the remaining participant data are used for testing. Moreover, k-fold cross-validation has been applied so the final results are the average of all the folds.

4 Results

This section shows and analyses the classification results obtained using the EEG and Facial approaches presented in the previous section using SVM and gentleboost classifiers. The results are represented by the F1 score which is a measure of accuracy that takes into account the precision and recall. A leave one out approach and a k-fold cross validation is applied for all the participants in our database. These results are compared with the ones obtained using as features the suggested in [20].

Tables 2 and 3 show the F1 scores for all the modalities and both classifiers, SVM and gentleboost, respectively. Also, the precision and recall values are shown in table 4, while an overview of the best outcomes is presented in figure 4. Furthermore, the ROC curves of the proposed method based on facial features in comparison to the ones proposed by Soleymani is shown in figure 5. The results of both individual modalities (EEG and facial) are coherent and adequate for the detection of emotions with overall F1 values around 70%. Comparing both data modalities, facial fiducial landmarks provide slightly better results than EEG signal for both classifiers; and the combination of both modalities only improves slightly the results using the gentleboost classifier. These results are in alignment

Table 3. F1 scores obtained using Gentleboost. See Table 1 for id definitions.

	Boost	id 1	id 2	id 3	id 4	Overall
			EEG			
Soleymani	[20]	0.6891	0.6843	0.6515	0.7540	0.6947
Proposed	14 Ch	0.7030	0.6508	0.6901	0.6670	0.6777
	4 Ch	0.7061	0.6622	0.7792	0.7035	0.7128
	Quaternion	**0.7297**	**0.6861**	**0.7439**	**0.7565**	**0.7291**
			Face			
Soleymani	[20]	**0.7068**	0.7362	0.7295	0.6579	0.7076
Proposed	Dist	0.6200	**0.7934**	**0.8024**	**0.7481**	**0.7410**
			EEGFace			
Soleymani	[20]	**0.7146**	0.7296	0.6711	0.7412	0.7141
Proposed	Dist +14 Ch	0.6444	0.7084	0.7841	0.7148	0.7129
	Dist +4 Ch	0.6327	**0.7694**	0.7871	**0.7680**	**0.7393**
	Dist +Quaternion	0.6753	0.6911	**0.8145**	0.7077	0.7222

Table 4. Best precision and recall values of our approach. corresponding to EEG Quaternion. Facial distance and Distance plus Quaternion features. in comparison with the ones obtained by [20] features.

		SVM					**Boost**				
		id 1	id 2	id 3	id 4	OA	id 1	id 2	id 3	id 4	OA
				EEG					**EEG**		
[20]	Prec	0.664	0.687	0.733	0.735	0.705	0.749	0.721	0.714	0.787	0.743
	Rec	0.644	0.622	0.655	0.716	0.659	0.705	0.694	0.672	0.761	0.708
Quat	Prec	0.758	0.737	0.746	0.784	**0.757**	0.751	0.723	0.783	0.780	**0.759**
	Rec	0.722	0.711	0.727	0.761	**0.730**	0.733	0.694	0.750	0.761	**0.734**
				Face					**Face**		
[20]	Prec	0.716	0.713	0.727	0.807	0.741	0.820	0.749	0.806	0.762	0.784
	Rec	0.661	0.722	0.705	0.722	0.702	0.738	0.777	0.750	0.705	0.743
Dist	Prec	0.749	0.849	0.838	0.710	**0.787**	0.684	0.823	0.860	0.786	**0.788**
	Rec	0.711	0.816	0.794	0.683	**0.751**	0.644	0.800	0.816	0.755	**0.754**
				EEGFace					**EEGFace**		
[20]	Prec	0.797	0.730	0.708	0.711	0.737	0.751	0.742	0.790	0.808	**0.773**
	Rec	0.688	0.733	0.688	0.683	0.698	0.744	0.772	0.705	0.755	**0.744**
Dist +	Prec	0.764	0.806	0.806	0.767	**0.786**	0.722	0.736	0.847	0.735	0.760
Quat	Rec	0.711	0.777	0.783	0.744	**0.754**	0.688	0.705	0.822	0.716	0.733

with the results obtained by [8]. On the other hand, the emotions related with unknown and known people or places have been recognised with higher accuracy using EEG features. We assume that this is due to a minimal difference on facial expressions during the recognition of famous, but not personally related, versus the unknown people or places.

In EEG, the use of 4 channels provides similar results with the 14 ones. The proposed quaternion based features improves the overall results by more than 1%. The proposed facial features also provide better F1 scores than the ones used in [20] in most of the classification scenarios. On the other hand, the results of the combined features are not always consistent in terms of which combination is the best one.

5 Conclusion

A novel database (SEM-db) has been created focusing on natural reactions to specific autobiographical and non-autobiographical stimulus that intend to elicit different emotions. This database provides facial videos and EEG signals, amongst other information, that can be used for emotion recognition. Using this database this work presents an approach for expected emotion recognition based on novel feature descriptors. The novel quaternion EEG and facial features result accurate classification rates. The overall results demonstrate that facial features outperform the EEG ones for emotion recognition.

References

References

1. Alpher, A.: Advances in Frobnication. J. of Foo **12** (2002) 234–778
2. Alpher, A., Fotheringham-Smythe, J.P.N.: Frobnication revisited. J. of Foo **13** (2003) 234–778
3. Herman, S., Fotheringham-Smythe, J.P.N., Gamow, G.: Can a machine frobnicate? J. of Foo **14** (2004) 234–778
4. Smith, F.: *The Frobnicatable Foo Filter*. GreatBooks, Atown (2009)
5. Wills, H.: Frobnication tutorial. Technical report CS-1204, XYZ University, Btown (1999)
6. Rosler, A., Mapstone, M. E., Hays A. K., Mesulam, M., Rademaker, A., Gitelman D. R., Weintraub S.: Alterations of visual search strategy in Alzheimer's disease and aging. Neuropsychology **12** (2000) 398–408
7. Pereira, M. L., Camargo, M. V. Z. A., Aprahamian, I., Forlenza, O. V.: Eye movement analysis and cognitive processing: detecting indicators of conversion to Alzheimer's disease. Neuropsychiatric disease and treatment **12** (2014) 1273–1285
8. Alpher, A.: Automatic analysis of facial affect: A survey of registration, representation, and recognition. IEEE transactions on pattern analysis and machine intelligence 37(6) **12** (2015) 1113–1133
9. Koelstra, S., Muehl C., Soleymani, M., Lee, J.S., Yazdani, A., Ebrahimi, T., Pun, T., Nijholt, A., Patras, I.: Deap: A database for emotion analysis; using physiological signals. IEEE Transactions on Affective Computing 3(1) **12** (2012) 18–31

10. Sapey-Triomphe, L. A., Heckemann, R. A., Boublay, N., Dorey, J. M., Hnaff, M. A., Rouch, I., Padovan, C.: Neuroanatomical Correlates of Recognizing Face Expressions in Mild Stages of Alzheimers Disease. PLOS ONE 10(12) **12** (2015)
11. Van den Stock, J., De Winter, F. L., de Gelder, B., Rangarajan, J. R., Cypers, G., Maes, F., Sunaert, S., Goffin, K., Vandenberghe, R., Vandenbulcke, M.: Impaired recognition of body expressions in the behavioral variant of frontotemporal dementia. Neuropsychologia 75 **12** (2015) 496–504
12. Han, K. H., Zaytseva, Y., Bao, Y., Pppel, E., Chung, S. Y., Kim, J. W., Kim, H. T.: Impairment of vocal expression of negative emotions in patients with Alzheimers disease. Frontiers in aging neuroscience 6(101) **12** (2014) 1–6
13. Irish, M., Hornberger, M., Lah, S., Miller, L., Pengas, G., Nestor, P. J., Hodges, J.R., Piguet, O.: Profiles of recent autobiographical memory retrieval in semantic dementia, behavioural-variant frontotemporal dementia, and Alzheimers disease. Neuropsychologia 49(9) **12** (2011) 2694–2702
14. Seidl, U., Lueken, U., Thomann, P. A., Kruse, A., Schrder, J.: Facial Expression in Alzheimers Disease Impact of Cognitive Deficits and Neuropsychiatric Symptoms. American journal of Alzheimer's disease and other dementias 27(2) **12** (2012) 100–106
15. American Psychiatric Association: Diagnostic and statistical manual of mental disorders (DSM-5). American Psychiatric Association Publishing **12** (2013)
16. Bettadapura, V.: Face expression recognition and analysis: the state of the art. Tech Report arXiv:1203.6722 **12** (2012) 1–27
17. Ekman, P., Friesen, W.V.: The Facial Action Coding System: A Technique for The Measurement of Facial Movement. Consulting Psychologists Press, San Francisco **12** (1978)
18. Weninger, F., Wllmer, M., Schuller, B.: Emotion Recognition in Naturalistic Speech and LanguageA Survey. Emotion Recognition: A Pattern Analysis Approach **12** (2015) 237–267
19. Chowdhuri, M. A. D., Bojewar, S.: Emotion Detection Analysis through Tone of User: A Survey. Emotion 5(5) **12** (2016) 859–861
20. Soleymani, M.,Asghari-Esfeden, S., Fu, Y., Pantic, M.: Analysis of EEG signals and facial expressions for continuous emotion detection. IEEE Transactions on Affective Computing 7(1) **12** (2016) 17–28
21. Soleymani, M., Lichtenauer, J., Pun, T., Pantic, M.: A multimodal database for affect recognition and implicit tagging. IEEE Transactions on Affective Computing 3(1) **12** (2012) 42–55
22. Nicolaou, M. A., Gunes, H., Pantic, M.: Continuous prediction of spontaneous affect from multiple cues and modalities in valence-arousal space. IEEE Transactions on Affective Computing 2(2) **12** (2011) 92–105
23. McKeown, G., Valstar, M., Cowie, R., Pantic, M., Schroder, M.: The semaine database: Annotated multimodal records of emotionally colored conversations between a person and a limited agent. IEEE Transactions on Affective Computing 3(1) **12** (2012) 5–17
24. Yan, W. J., Li, X., Wang, S. J., Zhao, G., Liu, Y. J., Chen, Y. H., Fu, X.: CASME II: An improved spontaneous micro-expression database and the baseline evaluation. PloS one 9(1) **12** (2014) e86041
25. Lokannavar, S., Lahane, P., Gangurde, A., Chidre, P.: Emotion recognition using EEG signals. Emotion 4(5) **12** (2015) 54–56
26. Vijayan, A. E., Sen, D., Sudheer, A. P.: EEG-based emotion recognition using statistical measures and auto-regressive modeling. IEEE International Conference on

Computational Intelligence & Communication Technology (CICT) 14(1) **12** (2015) 587–591
27. Cootes, T. F., Edwards, G. J., Taylor, C. J.: Active appearance models. IEEE Transactions on pattern analysis and machine intelligence 23(6) **12** (2001) 681–685
28. Baltru, T., Robinson, P., Morency L. P.: OpenFace: an open source facial behavior analysis toolkit. IEEE Winter Conference on Applications of Computer Vision (WACV) **12** (2016) 1–10
29. Sohaib, A. T., Qureshi, S., Hagelbck, J., Hilborn, O., Jeri, P.: Evaluating classifiers for emotion recognition using EEG. International Conference on Augmented Cognition. Springer Berlin Heidelberg **12** (2013) 492-501
30. Mller-Putz, G. R., Riedl, R., Wriessnegger, S.C.: Electroencephalography (EEG) as a research tool in the information systems discipline: Foundations, measurement, and applications. Communications of the Association for Information Systems 37(46) **12** (2015) 911-948
31. Petrantonakis, P. C., Hadjileontiadis, L. J.: Emotion Recognition from Brain Signals Using Hybrid Adaptive Filtering and Higher Order Crossings Analysis. IEEE Transactions on Affective Computing 1 **12** (2010) 81-97
32. Valstar, M. F., Patras, I., Pantic, M.: Facial action unit detection using probabilistic actively learned support vector machines on tracked facial point data. IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) 3 **12** (2005) 76–84
33. Michel, P., El Kaliouby, R.: Real time facial expression recognition in video using support vector machines. In Proceedings of the 5th international conference on Multimodal interfaces **12** (2003) 258–264
34. Hamilton, W.R.: On quaternions, or on a new system of imaginaries in algebra. Philosophical Magazine 25(3) **12** (1844) 489-495
35. Chen, M., Meng, X., Wang, Z.: Quaternion Fisher Discriminant Analysis for Bimodal Multi-feature Fusion. Chap Adv Intel Syst & Comp 370 **12** (2015) 479–487
36. Le Bihan, N., Sangwine, S.J.: Quaternion principal component analysis of color images. International Conference on Image Processing (ICIP) 1 **12** (2003) I-809-12
37. Kanade, T., Cohn, J. F., Tian, Y.: Comprehensive database for facial expression analysis. Fourth IEEE International Conference on Automatic Face and Gesture Recognition **12** (2000) 46–53
38. Pantic, M., Valstar, M., Rademaker, R., Maat, L.: Web-based database for facial expression analysis. IEEE international conference on multimedia and Expo **12** (2005) 317-321
39. Patras, I., Pantic, M.: Particle filtering with factorized likelihoods for tracking facial features. Sixth IEEE International Conference on Automatic Face and Gesture Recognition **12** (2004) 97–102
40. Izard, C. E.: Human emotions. Springer Science & Business Media **12** (2013)
41. Nicolle, J., Rapp, V., Bailly, K., Prevost, L., Chetouani, M.: Robust continuous prediction of human emotions using multiscale dynamic cues. 14th ACM international conference on Multimodal interaction **12** (2012) 501–508
42. Littlewort, G. C., Bartlett, M. S., Lee, K.: Automatic coding of facial expressions displayed during posed and genuine pain. Image and Vision Computing 27(12) **12** (2009) 1797–1803
43. Huang, K. C., Huang, S. Y., Kuo, Y. H.: Emotion recognition based on a novel triangular facial feature extraction method. 2010 International Joint Conference on Neural Networks (IJCNN) **12** (2010) 1–6
44. Sariyanidi, E., Gunes, H., Gkmen, M., Cavallaro, A.: Local Zernike Moment Representation for Facial Affect Recognition. British Machine Vision Conf **12** (2013)

45. Zhao, G., Pietikinen, M.: Boosted multi-resolution spatiotemporal descriptors for facial expression recognition. Pattern recognition letters 30(12) **12** (2009) 1117–1127
46. Wllmer, M., Eyben, F., Reiter, S., Schuller, B., Cox, C., Douglas-Cowie, E., Cowie R.: Abandoning emotion classes-towards continuous emotion recognition with modelling of long-range dependencies. Interspeech **12** (2008) 597–600
47. Adali, T., Schreier, P.J., Scharf, L.L.: Complex-valued signal processing: The proper way to deal with impropriety. IEEE Trans. Signal Processing (overview paper) 59(11) **12** (2011) 5101-5123
48. Li, X.-L., Adali, T., Anderson, M.: Noncircular principal component analysis and its application to model selection. IEEE Sig. Proc. 59(10) **12** (2011) 4516-4528i
49. Chai, Z., Ma, K. K., Liu, Z.: Complex Wavelet-Based Face Recognition Using Independent Component Analysis. Fifth Intern. Conf. on Intelligent Information Hiding and Multimedia Signal Proc. **12** (2009) 832-835
50. Bonita, J. D., Ambolode II, L. C. C., Rosenberg, B. M., Cellucci, C. J., Watanabe, T. A. A., Rapp, P. E., Albano, A. M.: Time domain measures of inter-channel EEG correlations: a comparison of linear, nonparametric and nonlinear measures. Cognitive Neurodynamics 8(1) **12** (2014) 1–15
51. Li, K., Sun, G., Zhang, B., Wu, S., Wu, G.: Correlation Between Forehead EEG and Sensorimotor Area EEG in Motor Imagery Task. Dependable, Autonomic and Secure Computing, 2009. DASC '09. Eighth IEEE International Conference on **12** (2009) 430-435
52. Abadi, M. K., Subramanian, R., Kia, S. M., Avesani, P., Patras, I., Sebe, N.: DECAF: MEG-based multimodal database for decoding affective physiological responses. IEEE Transactions on Affective Computing 6(3) **12** (2015) 209–222