

Facial expression recognition in ageing adults: from lab to ambient assisted living

Elena Lozano-Monador¹  · María T. López² · Francisco Vigo-Bustos² · Antonio Fernández-Caballero³

Received: 11 April 2016 / Accepted: 9 February 2017 / Published online: 27 February 2017
© Springer-Verlag Berlin Heidelberg 2017

Abstract This paper introduces an application that uses a webcam and aims to recognise emotions of ageing adults from their facial expression. Six basic emotions (*Happiness, Sadness, Anger, Fear, Disgust* and *Surprise*) as well as a *Neutral* state are distinguished. Active shape models are applied for feature extraction, the Cohn–Kanade, JAFFE and MMI databases are used for training, and support vector machines (*v*-SVM) are employed for facial expression classification. These six basic emotions are classified into three categories (*Happiness, Negative-Emotion* and *Surprise*), as *Sadness, Anger, Fear* and *Disgust* have been grouped into a single expression. The new three categories are found to be relevant for the purpose of detecting abnormalities in the ageing adult’s mood state. In the future, we will transfer the experience in a lab/research setup to a real ambient assisted living environment, which is the elderly person’s home. The application is devised to be the starting point to enhance the mood of the elderly people living alone at their homes by external stimuli.

Keywords Ageing adult · Ambient assisted living · Facial expressions · Emotions · Active shape models · Support vector machines

1 Introduction

Ambient assisted living (AAL) aims at extending the time older people can live in their home environment by increasing their autonomy and assisting them in carrying out activities of daily living. This aim should be achieved by the use of intelligent products and the provision of remote services. The overall objective of the AAL programme is to enhance the quality of life of older people and strengthen the European industrial base through the use of Information and Communication Technologies. In a current context of increasing population, it is a challenge to provide new technologies that can recognise emotions or moods in order to enhance the elderly quality of life (Fernández-Caballero et al. 2016; Castillo et al. 2016; Shan et al. 2009; Castillo et al. 2014; Fernández-Caballero et al. 2014a). Indeed, in recent years, there has been a growing interest in improving all aspects of interaction between humans (including the ageing population) and computers (Gascuña et al. 2014). The emerging field of human-computer interaction has been of interest to researchers from a number of diverse fields, including Computer Science, Psychology, and Neuroscience.

Gaining insight into the state of the user’s mind via facial analysis can provide valuable information for affective sensing systems. Facial expressions reflect not only emotions, but also other mental activities, social interaction and physiological signals. For establishing emotional interactions between humans and computers, a system to recognise human emotions is of a high priority. An automated system that can determine the emotions of a person via his/her expressions provides the system with the opportunity to customise its response (Alugupally et al. 2011; Lozano-Monador et al. 2014).

✉ Antonio Fernández-Caballero
Antonio.Fdez@uclm.es

¹ Instituto de Investigación en Informática de Albacete, Universidad de Castilla-La Mancha, 02071 Albacete, Spain

² Escuela Superior de Ingeniería Informática, Universidad de Castilla-La Mancha, 02071 Albacete, Spain

³ Escuela de Ingenieros Industriales, Universidad de Castilla-La Mancha, 02071 Albacete, Spain

Moreover, emotion recognition using visual cues is receiving a great deal of attention in the past decade. Most of the existing approaches do recognition on six universal basic emotions (*Happiness, Sadness, Anger, Fear, Disgust* and *Surprise*) because of their stability over culture, age and other identity related factors. For instance, an integrated system for emotion detection has been presented, in which only eye and mouth expressions are used for detecting five emotions (all the above minus *Disgust*) (Maglogiannis et al. 2009). Even, an approach to facial expression recognition for estimating patients' emotions is proposed with only two expressions (*Happiness* and *Sadness*) (Wang et al. 2009). Applications that use these techniques are varied, ranging from software able to recognise and act according to the emotions of the user who is using it or systems capable of detecting lies up to applications that allow knowing if a product is liked or not with only analysing the emotional reaction of a user.

A facial expression recognition system is normally composed of four main steps: face detection/tracking, feature extraction, feature selection, and emotion classification. Choosing suitable feature extraction and selection algorithms plays central roles in providing discriminative and robust information (Zhang et al. 2012; Fernández-Caballero et al. 2014b). The selection of features employed for emotion recognition is usually classified into two main categories: geometric features and appearance features. Techniques based on geometrical facial features are the most extended in identifying emotions. In fact, most papers and studies focused on emotion detection use these methods to classify expressions (Tariq et al. 2011; Setyati et al. 2012). There are several reasons for this method to be the most extensively used. First of all, geometrical features of face expression are the best suited to obtain the Facial Action Coding System of the expression (Ekman et al. 2002). This coding system has proven to be greatly useful in the detection of emotions, especially the six universal ones mentioned before. Secondly, alternate methods such as electroencephalography, electrocardiography, and skin impedance, among others, can be certainly useful to get reliable information to predict the mood of a person, but they are very invasive since it is necessary to attach a device to the patient, while analysing the facial expression is a completely user transparent process. Finally, since we want the resulting system to be installed at home, it is desirable that all the components are as cheap as possible. There are alternate non-invasive ways to get information for predicting a person's emotion, such as thermal imaging cameras, but this kind of equipment is more expensive than a standard webcam.

Thus, in this paper, we are interested in geometric features, which are extracted from the shape or salient point locations of important facial components such as mouth

and eyes. Moreover, this paper introduces the extraction of facial features to detect emotions represented by particular facial expressions. This involves a series of steps: (a) the study of techniques for detecting and extracting facial features, as well as the attainment of a model to operate with a response-time that allows the system to analyse as many frames as possible without compromising the accuracy when detecting facial points, and (b) the creation of an emotion detector through implementing the most suited classification techniques.

The rest of the article is organised as follows. Section 2 describes some previous works that use active shape models and support vector machines for facial expression recognition from geometric features. In Sect. 3 our approach to recognition of emotions from face expressions is introduced. Section 4 offers the performance of our approach in form of data used and results obtained. This section inserts the most important challenge of this proposal, namely to assess the possibility of transferring lab/research results to real AAL homes. Therefore, Sect. 4 is divided into two subsections: the first one is dedicated to the results obtained in a lab setup, whilst the second offers preliminary results. Finally, some conclusions are provided in Sect. 5.

2 ASM and SVM for facial expression recognition from geometric features

It has been demonstrated that the active shape model (ASM) is an excellent method for locating facial feature points (Cootes et al. 1996). Generally speaking, ASM fits the shape parameters using optimisation techniques such as gradient descent. On the other hand, support vector machines (SVM) (Cortes and Vapnik 1995) exhibit good classification accuracy even when only a modest amount of training data is available, making them particularly suitable to a dynamic, interactive approach to expression recognition. This is why the tandem ASM-SVM is intensively being used for facial expression recognition.

For instance, 58 landmark points are used to construct an ASM for face expressions (Chang et al. 2006). These are then tracked and give facial expressions recognition in a cooperative manner. Introducing a set of more refined features, facial characteristic points around the mouth, eyes, eyebrows, nose, and chin are used as geometric features for emotion recognition (Pantic and Bartlett 2007). A quite recent approach (Hsieh and Jiang 2011) utilises facial components to locate dynamic facial textures such as frown lines, nose wrinkle patterns, and nasolabial folds to classify facial expressions. Adaboost using Haar-like feature and ASM are adopted to accurately detect faces and acquire important facial feature regions. Gabor filters and Laplacian or Gaussian distributions are employed to extract

texture information in the acquired feature regions. These texture feature vectors represent the changes of facial texture from one expression to another expression. Then, SVM is deployed to classify the six facial expression types including *Neutral*, *Happiness*, *Surprise*, *Anger*, *Disgust* and *Fear*. The Cohn–Kanade database is used to test the feasibility of the method.

Recently, an algorithm of face recognition based on ASM and Gabor features of key points has been proposed (Wu and Mei 2013). Firstly, the AdaBoost algorithm detects the face region in an image. Then, the ASM localises the key feature points in the detected facial region. The Gabor features of these points are extracted. Finally, the features are classified using SVM. Preliminary experiments show promising results of the proposed algorithm on The ORL Database of Faces (AT&T Laboratories Cambridge 1994). Another paper describes a method for recognition of continuous facial expression changes in video sequences (Wan et al. 2012). Again, ASM automatically localises the facial feature points in the first frame and then tracks the feature points through the video frames. After that comes the selection of the 20 optimal key facial points, those which change the most with modifications in the facial expression. After building the feature space, SVM is trained for classification and results are tested. Another proposal for geometric feature extraction integrates the distances between face fiducial points and the centre of gravity of the face's ASM shape with the FAU relative facial component deformation distances (Gang et al. 2009). The approach also introduces a multiclass one-against-one ν -SVM for facial expression classification.

Another paper (Zhao et al. 2012) empirically evaluates facial representation based on statistical local features, ASM and local binary patterns (LBP) for person-independent facial expression recognition. AdaBoost-LBP based ASM is used for emotion classification. Lastly, a work system overview is explained next (Cruz and Bhanu 2012). Multi-class AdaBoost with dynamic time warping similarity distance between the feature vector of input facial expression and prototypical facial expression, is used as a weak classifier to select the subset of discriminative feature vectors (Ghimire and Lee 2013). Face region of interest is detected with a boosted cascade of Haar-like features. Dynamic and static information are computed in separate pathways. Dynamic information is quantified with ASM; facial points detected with ASM are used for registration, and appearance features are developed from the registered images. Static information is obtained by estimating a static representation of the face and warping each face to minimise dynamics. Appearance features are generated from this representation. The two approaches are fused at the match-score level and emotion labels are classified with an SVM classifier.

To sum up, the proposals presented in this section can be classified according to the approach taken for facial feature detection and the classifiers applied: (a) proposals that use static approaches for getting feature vectors using geometric features (Chang et al. 2006; Pantic and Bartlett 2007; Hsieh and Jiang 2011; Wu and Mei 2013; Gang et al. 2009; Zhao et al. 2012). (b) Proposals that use dynamic approaches for sequence frames, which translate static feature vectors to temporal feature vectors (Wan et al. 2012; Ghimire and Lee 2013). (c) Classifiers used for emotion detection: SVM (e.g. Hsieh and Jiang 2011; Wu and Mei 2013; Wan et al. 2012; Gang et al. 2009; Ghimire and Lee 2013) and AdaBoost (e.g. Pantic and Bartlett 2007; Zhao et al. 2012; Ghimire and Lee 2013) as weak classifier.

3 Recognition of emotions from face expressions

As explained before, this paper presents a facial expression recognition system based on geometric features (Zhou and Wang 2013). This method first uses ASM to track the fiducial points coarsely and then applies a method based on threshold segmentation and deformable model to correct the mouth fiducial points due to incorrect locations in the presence of non-linear image variations such as those caused by large facial expression changes. The geometric features extracted from the fiducial points are classified in one of the six basic expressions plus *Neutral* by an SVM classifier.

3.1 Facial expression analysis

Today, less intrusive automatic emotion recognition is based on the facial expression of the subject. In recent years, several methods have been developed to extract and analyse facial features. To do this, a complete description of facial expressions is needed. The Facial Action Coding System (FACS) (Ekman et al. 2002) is a system based on human observation to detect changes in facial features. This system encodes all possible facial expressions as action units (AUs) which take place individually or in combination.

Indeed, FACS considers 44 AUs, 30 anatomical which are contractions of certain facial muscles, and 14 miscellaneous ones that involve a change in expression but are not associated with a facial muscle. For each AU there are five levels of intensity, depending on the force the muscle has to exercise. Facial expressions associated with emotions are generally described as a set of AUs. The way to get the AUs of a subject is to locate a series of facial points and compare their distances to know what facial muscles are moving. This approach analyses the changes that occur in the facial expression and relate them to a specific emotion.

Obviously, a reference database is used to associate the facial expressions observed.

Figure 1 and Table 1 show the features that the system is taking into account to predict emotions. As it can be seen, these features are located in the eyebrows, eyes and mouth. The reason to select these specific facial features is that they can be related to one or several AUs, which means that the system is taking advantage of the FACS to train and predict the emotions. For example, feature C5 (distance between the internal ends of the eyebrows is related to AU4

and AU5. If C5 decreases, the person is activating AU4, while when C5 increases it is AU5 that is being activated. Several papers have used these features or similar ones to predict emotions (Hsieh and Jiang 2011; Gang et al. 2009; Ghimire and Lee 2013; Soleymani et al. 2012).

3.2 Facial emotion detection

The approach described in this paper is divided into four steps:

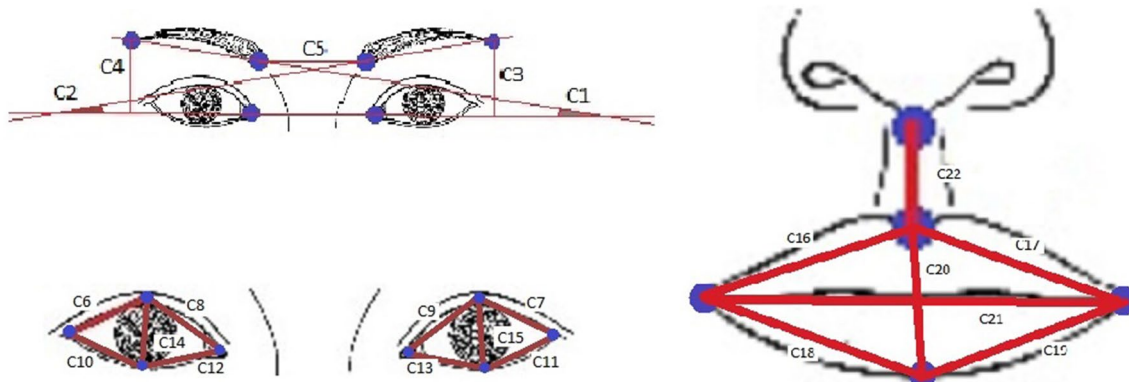


Fig. 1 Geometrical features extracted from the ASM Mask

Table 1 List of geometrical features

Code	Description
C1	Angle created by the line through the outer and inner ends of the right eyebrow and the line connecting the inner ends of the eyes
C2	Angle created by the line through the outer and inner ends of the left eyebrow and the line connecting the inner ends of the eyes
C3	Distance between the outer end of the left eyebrow and the closest point in the line connecting the inner ends of the eyes
C4	Distance between the outer end of the right eyebrow and the closest point in the line connecting the inner ends of the eyes
C5	Distance between the inner ends of the eyebrows
C6	Distance between the upper and outer ends of the right eye
C7	Distance between the upper and outer ends of the left eye
C8	Distance between the upper and inner ends of the right eye
C9	Distance between the upper and inner ends of the left eye
C10	Distance between the lower and outer ends of the right eye
C11	Distance between the lower and outer ends of the left eye
C12	Distance between the lower and inner ends of the right eye
C13	Distance between the lower and inner ends of the left eye
C14	Distance between the upper and lower ends of the right eye
C15	Distance between the upper and lower ends of the left eye
C16	Distance between the upper and right ends of the mouth
C17	Distance between the upper and left ends of the mouth
C18	Distance between the lower and right ends of the mouth
C19	Distance between the lower and left ends of the mouth
C20	Distance between the upper and lower ends of the mouth
C21	Distance between the right and left ends of the mouth
C22	Distance between the upper end of the mouth and the base of the nose

- *Detection of facial points.* Currently, the detection of emotions is based on the analysis of facial expression from different facial points. The first step is to generate points on a facial expression in the simplest possible way. At this early stage it is necessary to perform a series of tests to select the model of facial detection points that best fits the needs and provides better results. Figure 2 shows the input and output for this step, which are the input image (Fig. 2a) and the ASM mask (Fig. 2b), respectively.
- *Feature extraction.* Once the facial points have been obtained, we study what are the most useful features which are obtained from these points for the detection of emotions. It is also detailed how to obtain each of the features. As it can be seen in Fig. 2c, the result of this step is the numeric data of the geometrical features.
- *Training and classification.* The third step consists of the selection of images for training, the choice of the most appropriate SVM kernel function, and the generation of a classification model that operates with the best response-time. In Fig. 3, the input and output for this step are shown. As input the system receives a set of different features that were obtained by applying Step 2 to a given number of facial expressions. A classifying model is achieved as a result of the training.
- *Detection of emotions.* At the last step, an emotion detection system is obtained. It is built from the models generated in the previous steps. The input for this step is

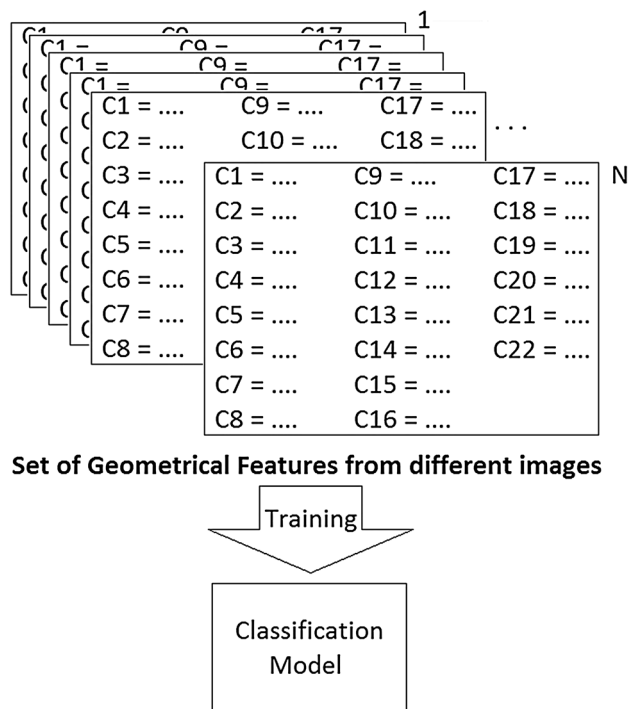
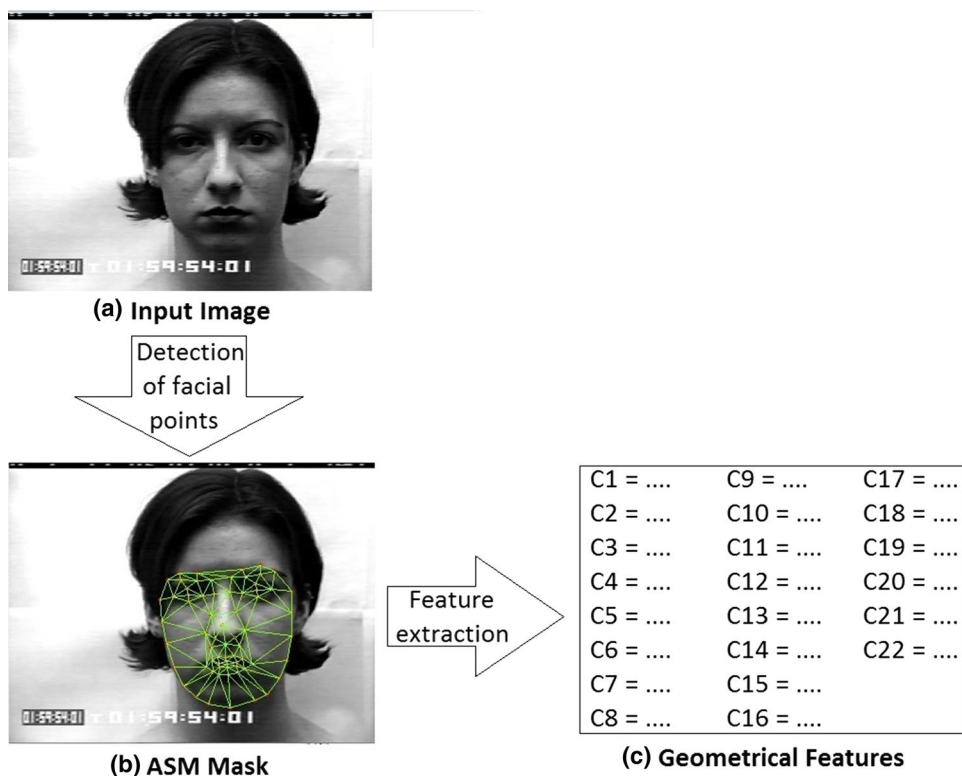


Fig. 3 Training and classification input (set of features obtained from different images and labeled according to their representative emotion) and output (SVM model to classify the emotions)

presented in Fig. 4. It is composed of the features from one image that were obtained using Step 2, and the clas-

Fig. 2 Facial points detection and resulting geometrical data



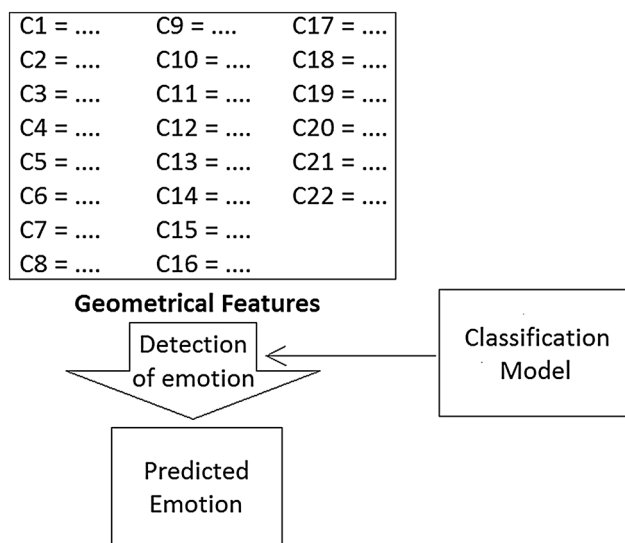


Fig. 4 Detection of emotions input (facial features and classification model) and output (emotion prediction)

sification model resulting from Step 3. The predicted emotion is obtained by means of these two parameters.

3.2.1 Detection of facial points and features

ASMLibrary (Wei 2009) is a library that easily creates an ASM from an image database and the images' corresponding log files. ASMLibrary is used in our case for generating ASM masks that will later detect facial points. In order to construct a valid model, a series of face image files are needed along with an annotation file attached to each image. The coordinates of each of the image points of interest are annotated in the log file. On the other hand, several models are generated from databases prepared for this purpose. This way, the advantages and disadvantages of each of them are studied before choosing the best model. The major database repositories are: (a) Informatics and Mathematical Modeling (IMM) (Stegmann 2002), which contains the analysis of 37 images of frontal faces. The model is composed of 58 facial points; (b) BioID (Cootes et al. 2005) is a database consisting of 1521 images of frontal faces. Each face is labelled with 20 facial points; (c) Extended Multi Modal Verification for Teleservices and Security (XM2VTS) (Chan 2000) consists of 2360 images which have been marked-up 68 facial features.

Three ASMs are created with the images and log files that make up the above mentioned databases. The objective is to analyse new images and verify that the facial point detection is performed correctly. Each model is checked in terms of its performance for still images, recorded videos

and webcam video input. The model of the XM2VTS database, with 68 facial points, is the most complete with respect to the other two models in terms of reliability of point detection. Furthermore, it allows a more accurate alignment of the face.

So, geometric features are extracted based on the positions of the facial points. The extracted features are: (a) eyebrows: angles between the horizontal line connecting the inner corners of the eyes and the line that connects inner and outer eyebrow, the vertical distances from the outer eyebrows to the line that connects the inner corners of the eyes, distances between the inner eyebrows' corner, (b) eyes: distances between the outer eyes' corner and their upper eyelids, distances between the inner eyes' corner and their upper eyelid, distances between the outer eyes' corner and their lower eyelids, distances between the inner eyes' corner and their lower eyelids and vertical distances between the upper eyelids and the lower eyelids, and (c) mouth and nose: distances between the upper lip and mouth corners, distances between the lower lip and mouth corners, distances between the mouth corners, vertical distance between the upper and the lower lip, distance between the upper lip and the lower nose.

3.2.2 Training, classification and detection of emotions

LibSVM (Chang and Lin 2011) is a library for programming support vector machines (SVMs). The image features belonging to properly labelled emotion databases are extracted in order to generate the file that is used in training the SVM. The method used for classification is a multiclass SVM, because we aim at distinguishing among seven classes. We have chosen the one-vs-one method for multiclass SVM (see Wu et al. 2004) from the two possible alternative approaches. Although this method involves using more classifiers, the employed training time is much lower. Given a set of instance-label training pairs, (x_i, y_i) , SVM kernels $K(x_i, y_i)$ are used, which can be either $K(x_i, y_i) = x_i^T y_i$ (linear SVM), $K(x_i, y_i) = (x_i^T y_i + \theta)^d$ (polynomial SVM of degree d), $K(x_i, y_i) = \tanh(\gamma x_i^T y_i + \theta)$ (sigmoid SVM), or $K(x_i, y_i) = \exp(-\gamma \|x_i - y_i\|^2)$ (RBF SVM), where γ and θ are constants. It has been decided to use the Radial Basis Function (RBF) kernel as it is the one that offers best results in terms of accuracy and training time for the task of face and/or facial expression recognition (e.g. Shan et al. 2009; Chittora and Mishra 2012; Suk and Prabhakaran 2014). LibSVM implements two types of classifiers (C -SVM and ν -SVM). Here, the ν -SVM algorithm is used due to the ease of adjustment of the ν parameter in the range $(0,1]$. Furthermore, four different well-known image databases are selected to carry out the training of the SVM: (1) JAFFE (Japanese Female Facial Expression) database (Lyons et al. 1997), (2) IMM facial expressions database

(Valstar and Pantic 2010), (3) Cohn–Kanade (CK) database (Kanade et al. 2000), and (4) Cohn–Kanade extended (CK+) database (Lucey et al. 2000). The classification features and values used are shown in Table 2.

4 Data and results

4.1 Response-time

As stated at the beginning of this paper, the aim is to create a system capable of analysing a continuous stream of frames that are captured from a camera and to obtain predictions about the emotion shown by the monitored person at a speed that allows the system to scan as many frames as possible without compromising the accuracy of the results. The development of the emotion detection system described in this article has been performed on an HP Pavilion g6 Notebook PC with the following features: (a) Intel Core TM i3-2350M 2.3 GHz processor; (b) 6GB RAM; (c) Intel HD Graphics Family 1024MB card; (d) Microsoft Windows 7 64-bit operating system.

Figure 5 shows the time spent by the system to predict the emotion of a facial expression from the moment it

receives the frame input to be analysed. The tests were performed with one image that was reduced and expanded to cover a range of sizes from 100 to 300 pixels (px).

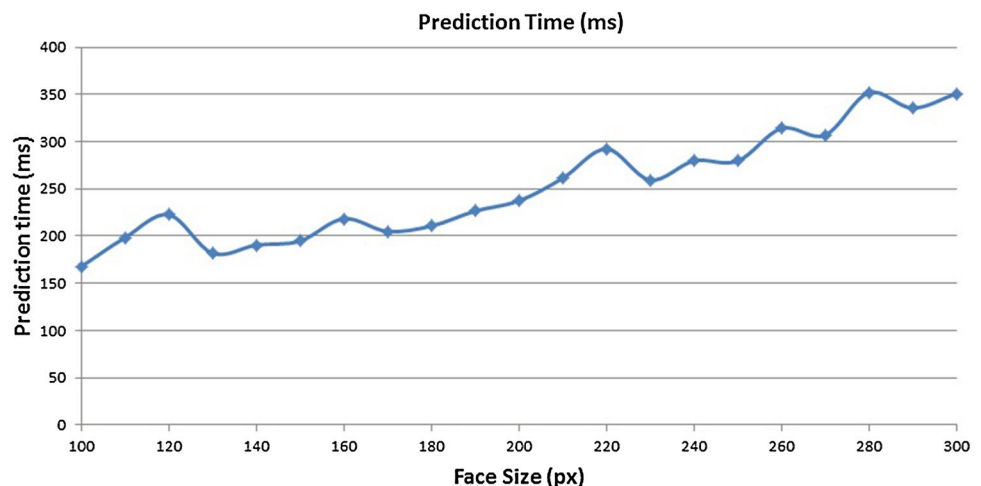
As it can be seen, the response-time is 351.05 ms in the worst case, which means that the speed of this system allows to analyse around three frames per second in that case. The system could analyse more images if they were smaller, since the ASM model needs less time to adapt to a face when it is smaller. However, the adjustment in faces smaller than 150px is not recommended because some facial points that are very close, such as the orbital points, would not be properly located due to the low resolution of the image. In order to get a proper ASM adjustment, the face should cover more than 170px. An example of this adjustment requirement is shown in Fig. 6. In column (a) there are the original images, in column (b) the adjustment done by the system when faces are rescaled to 140px, and in column (c) when they are resized to 190px. In the first set of images, it can be seen that the eyes and eyebrows are not correctly detected in the 140px image. In the 190px image these are correctly found, although the points in the internal side of the eyebrows are slightly below their real landmarks. In the second set of images the detection of the landmarks of the chin are wrong for the 140px image. However, the 190px image is correctly adjusted.

Also, during emotion prediction, some issues have been found regarding inaccuracies due to by different reasons. This is the case for a sudden movement of the person’s head, which makes the ASM mask to be wrongly fit for a couple of frames. Another problem occurs when a face expression is not representative of any of the seven emotions considered; for example, a transition face between one emotion and another. In order to mitigate the possible contingencies that lead to incorrect prediction, the system takes into account not only the frame it is analysing, but the last ten frames of the video input, to make the emotion prediction. This way, since the inaccuracies derived from a

Table 2 Features of the ν -SVM model

Feature	Value
Type of SVM	ν -SVM
Type of Kernel	Radial Basis Function (RBF)
Parameter ν	0.52
Parameter γ	0.12
Number of classes	7
Number of support vectors	237

Fig. 5 Prediction time depending on face size



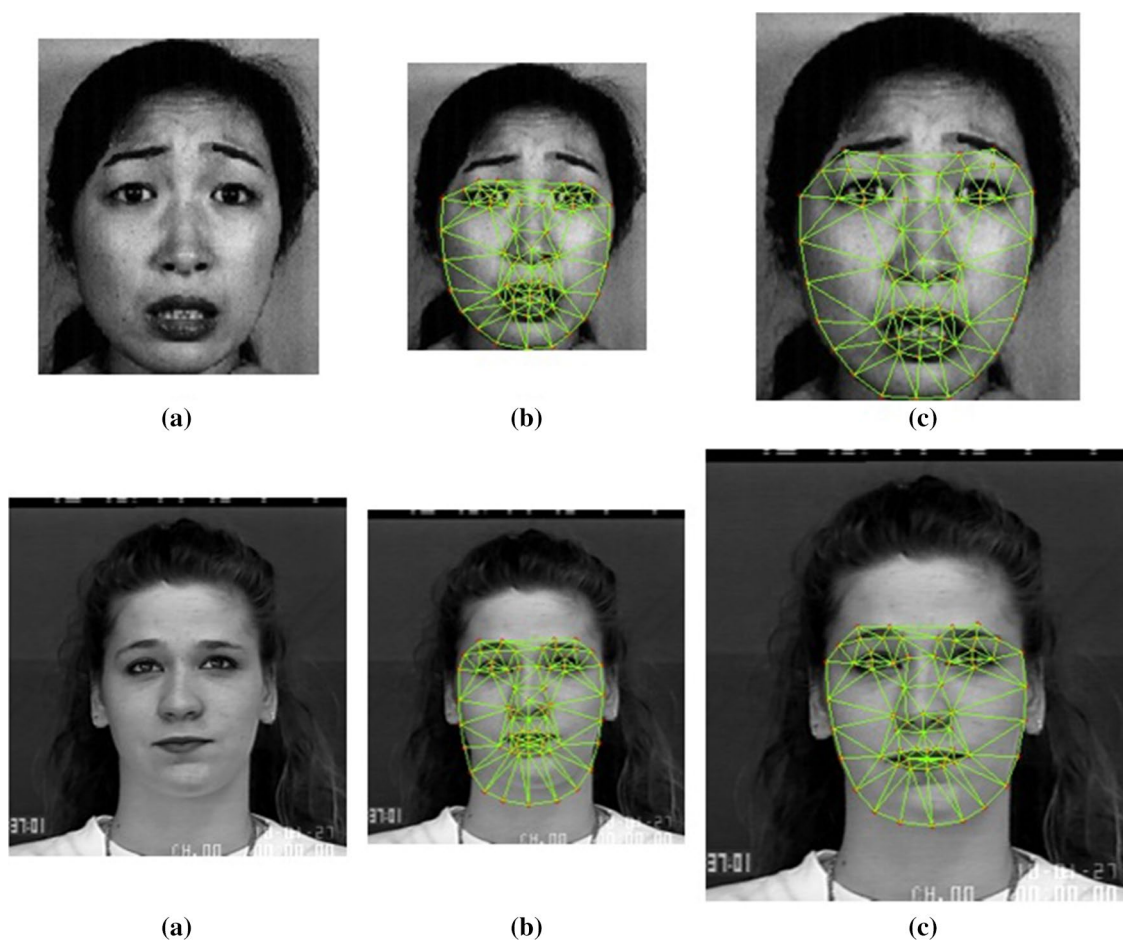


Fig. 6 ASM fitting examples. **a** Original images. **b** Adjustment on images scaled to 140px. **c** Adjustment on images scaled to 190px

sudden movement or from an emotion transition only takes place during two or three frames, the system can “discard” those frames that suppose a change of the emotion for an insignificant amount of time.

The system takes into account the last ten frames. The emotion that was predicted more times during these last ten frames is the resulting prediction. In case of a tie, the system chooses, from the tied predictions, the emotion that was more recently predicted. The system works with the same basic emotions that were shown in the training data (*Happiness*, *Sadness*, *Anger*, *Fear*, *Disgust* and *Surprise*) and the *Neutral* state.

4.2 Facial expression recognition at lab setup

In order to validate our proposal with the described databases, tests are performed for each of the seven classes that the system has to distinguish, that is, *Happiness*, *Sadness*, *Anger*, *Fear*, *Disgust*, *Surprise* and *Neutral*. Table 3 shows the hit rate for each facial expression:

The first column of Table 3 shows the labelled emotion for the frame, which means that the prediction should be of that same class. In the first row of the table we have all the possible predictions. The cells that are marked in bold show the hits for each emotion. The rest of the cells are wrong predictions. For example, the cell located in column *Sadness* and row *Happiness* shows the percentage of images in which the prediction should have been *Happiness* but the system predicted *Sadness* instead. As it can be observed, the results are really outstanding.

4.3 Facial expression recognition at home setup

In order to validate our proposal at a simple home setup, we capture video from a webcam situated in front of older persons. The tests are performed by requesting each elderly to pretend the facial expression associated with a particular emotion. The user receives no other external stimulus. The tests were carried out in this way to exempt the ageing adults from any emotional stress that could harm them. So, the tests are performed for each of the seven classes that

Table 3 Results when using benchmarks datasets

Prediction	Happiness (%)	Sadness (%)	Anger (%)	Fear (%)	Disgust (%)	Surprise (%)	Neutral (%)
Happiness	87.5	2.5	0	5	0	0	5
Sadness	7.89	78.42	2.63	0	7.89	0	3.16
Anger	2.5	0	72.5	10	12.5	0	2.5
Fear	2.78	2.78	2.78	83.33	5.56	2.78	0
Disgust	2.5	2.5	2.5	7.5	82.5	0	2.5
Surprise	0	2.5	0	0	0	97.5	0
Neutral	5	5	7.5	0	5	0	77.5

Bold values indicate correct prediction

the system has to distinguish, that is, *Happiness*, *Sadness*, *Anger*, *Fear*, *Disgust*, *Surprise* and *Neutral*. Altogether, ten older people have performed the tests by using this method.

Emotions that offer better results are *Surprise* and *Happiness* (all of them with a hit rate over 0.85). Results are acceptable when the facial expression reflects *Fear*, *Disgust* and *Neutral* (hit rate over 0.75). The worst results appear because some emotions contain features that can lead to confusion with other emotions. For example, *Sadness* shares characteristics with other emotions like *Anger* and *Fear*. Clearly, the *Sadness* emotion includes frowning (feature also related to *Anger*) or lips down (related both to *Sadness* and *Fear*). The main reasons why there are inaccuracies in the system’s predictions are (Fig. 7): (a) the ASM adjustment is incorrect, (b) the pretended emotion is clearly not representative of the expected emotion, (c) the features

between two emotions are very similar, and (d) the transition from one emotion to another causes troubles during a short interval of time.

Obviously the results (see Table 4) are worse than those obtained in Table 3. There are various reasons for the lower hit rates. During training static images from classic databases were used, which have good quality and representative displays of the emotions. In this test we are using a webcam video as an input, which offers worse resolution than the database images. So, the frame quality is affecting the ASM adjustment. Moreover, most subjects in the classic databases are people between 20 and 40 years old. The representation of elderly people is reduced to four or five subjects altogether. It is possible that the model does not work so well on the elderly due to the scarcity of samples. Besides, the person who is being recorded live usually

Fig. 7 Some examples of frames incorrectly predicted. **a** ASM mask incorrectly adjusted. The emotion predicted should be *Surprise*, but the result is *Disgust*. **b** Frame which is supposed to be *Sadness*. The prediction is *Disgust*. **c** Picture representing *Fear*, but the system’s prediction is *Surprise*, since both *Fear* and *Surprise* include widely open eyes. **d** Transition frame between *Disgust* and *Neutral*. The emotion predicted is *Sadness*

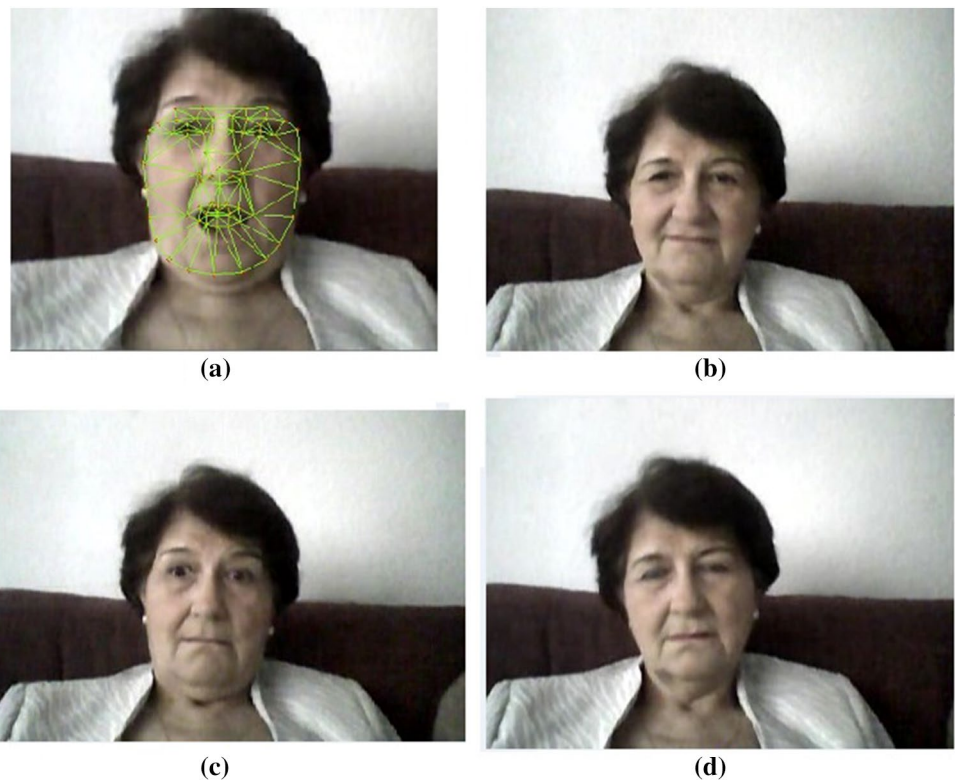


Table 4 Results when using a webcam

Facial expression	Hit rate
<i>Happiness</i>	85.6
<i>Sadness</i>	73.30
<i>Anger</i>	70.27
<i>Fear</i>	78.35
<i>Disgust</i>	75.17
<i>Surprise</i>	96.51
<i>Neutral</i>	75.2

makes sudden movements and covers part of his/her face, thus making it impossible for the system to obtain the mask. In addition, since we are continuously recording and predicting emotions, there are frames where the subject is changing from one emotion to another. These frames are not representative of any emotion, and the prediction turn out to be usually wrong in this case. We have also observed that ageing adults are not so good when pretending some emotions.

Nonetheless, as our aim is to detect negative emotions in elderly people, we have considered to group emotions *Sadness*, *Anger*, *Fear* and *Disgust* into a single emotion called *Negative-Emotion*. This decision coincides with the studied problem, which is the detection of an emotionally abnormal status in the elderly people. The final objective of this system is to be able to predict and enhance the emotion of and ageing adult. This is why it is not really important to know if the subject is feeling *Sadness* or *Anger*. Rather, the system has to detect if he/she is experiencing an emotion that is leading him/her to a bad mood, and to react accordingly to this prediction. By grouping all the negative emotions in one single class, the percentage of hits is increased without losing the pursued functionality of the system. Indeed, a failed prediction that stated a specific frame to be a negative emotion, will now be a correct prediction. It is evident that *Sadness*, *Anger*, *Fear* and *Disgust* compose the negative and undesired emotions.

After grouping the negative emotions, the tests are performed for each of the new four classes that the system has to distinguish, that is, *Happiness*, *Negative-Emotion*, *Surprise* and *Neutral*. The new results are shown in Table 5. As you can observe, now the results are very relevant and competitive except for *Neutral* class which success rate

Table 5 Results when using a webcam after grouping negative emotions

Facial expression	Hit rate
<i>Happiness</i>	85.6
<i>Negative-Emotion</i>	90.86
<i>Surprise</i>	96.51
<i>Neutral</i>	75.2

remains 0.75. Figure 8 shows some examples of the emotions detected.

Since the system is now working with four classes of emotions instead of seven, Table 6 shows the confusion matrix after using benchmark datasets. As it can be seen, the results are also certainly improved in the testing.

Also, a comparison between the hit rates obtained using non-elderly (Table 6) and elderly people (Table 5). As it can be seen, the results obtained for elderly people are slightly worse than for non-elderly people. Since the training was performed by using classical databases, which contain only non-elderly people, it is logical to get better results when the tests are performed with a non-elderly population. Besides, as it has been previously stated in this paper, ageing adults are not as good as younger people at pretending emotions. They feel considerably more uncomfortable when they are aware of being recorded.

5 Conclusions

This article has described the steps followed to study some facial feature extraction and detection techniques, as well as methods that allow the recognition of emotions within an appropriate response-time. This has allowed choosing an appropriate face recognition system and establishing the most suitable features to discriminate emotions. We have implemented an emotion detector that uses the techniques studied. In this sense, we have studied models for automatic acquisition of facial features. We have decided to use an active shape model, due to its good performance and response-time. Tests have been performed with several models, and the best results were obtained with the 68 facial point model. Also, we studied support vector machines for classification and we used a ν -SVM with RBF kernel. Thus, we have obtained a suitable classification system to work with the six basic emotions, namely *Happiness*, *Sadness*, *Anger*, *Fear*, *Disgust* and *Surprise*, plus the *Neutral* emotion.

Table 6 Results when using benchmarks datasets after grouping all negative emotions in one

Prediction	<i>Happiness</i> (%)	<i>Negative-Emotion</i> (%)	<i>Surprise</i> (%)	<i>Neutral</i> (%)
<i>Happiness</i>	87.5	7.5	0	5
<i>Negative-Emotion</i>	3.92	93.34	0.7	2.04
<i>Surprise</i>	0	2.5	97.5	0
<i>Neutral</i>	5	17.5	0	77.5

Bold values indicate correct prediction



Fig. 8 Example of facial emotion classification, where the detected emotions are **a** *Happiness*, **b–e** *Negative-Emotion*, **f** *Surprise*, and **g** *Neutral*

This has led to the construction of an application which objective is to distinguish emotions of older people from their facial expressions captured by a webcam. It has been found that LibSVM and ASMLibrary libraries are adequate tools for programming such a system. The emotions that offer better results in terms of detection in a real AAL environment are *Surprise*, *Happiness*, *Sadness* and *Anger*. For other emotions, such as *Disgust* and *Fear*, the system tends to get confused because emotions have very similar facial features.

As the results are not convincing for the initial purpose of detecting abnormal emotional states in the elderly, we have taken a new direction. Indeed, as the focus of interest is to centre on the detection of negative emotions in the elderly, the six basic emotions have been reduced to four: *Happiness*, *Negative-Emotion*, *Surprise*, plus the *Neutral* emotion. This way the efficiency of the system has augmented in terms of hits in emotion detection.

Lastly, let us highlight that the emotions explored in this paper are pretended. Thus, the work described in this paper is just another step in developing a system to improve mood in the ageing adult by external non-intrusive stimuli.

Acknowledgements This work was partially supported by Spanish Ministerio de Economía y Competitividad/European Regional Development Fund under TIN2015-72931-EXP and DPI2016-80894-R grants.

Author contributions All authors have made substantial contributions in the definition of the research line, as well as in experimentation, data analysis, and manuscript preparation.

Compliance with ethical standards

Conflict of interest The authors declare no conflict of interest.

References

- AT&T Laboratories Cambridge (1994). <http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html>. Accessed 21 Sept 2016
- Alugupally N, Samal A, Marx D, Bhatia S (2011) Analysis of landmarks in recognition of face expressions. *Pattern Recognit Image Anal* 21:681–693
- Castillo JC, Castro-González Á, Fernández-Caballero A, Latorre JM, Pastor JM, Fernández-Sotos A, Salichs MA (2016) Software architecture for smart emotion recognition and regulation of the ageing adult. *Cognit Comput* 8:357–367
- Castillo JC, Fernández-Caballero A, Castro-González Á, Salichs MA, López MT (2014) A framework for recognizing and regulating emotions in the elderly. *Lect Notes Comput Sci* 8868:320–327
- Chan CH (2000) The XM2VTS Database. <http://www.ee.surrey.ac.uk/CVSSP/xm2vtsdb/>. Accessed 21 Sept 2016
- Chang CC, Lin CJ (2011) LIBSVM: a library for support vector machines. *ACM Trans Intell Syst Technol* 2:1–27
- Chang Y, Hu C, Feris R, Turk M (2006) Manifold based analysis of facial expression. *Image Vis Comput* 24:605–614

- Chittora A, Mishra O (2012) Face recognition using RBF kernel based support vector machine. *Int J Future Comput Commun* 1:280–283
- Cootes TF, Cristinacce D, Babalola K (2005) BioID face database. <http://www.bioid.com/index.php?q=downloads/software/bioid-face-database.html>. Accessed 21 Sept 2016
- Cootes TF, Taylor CJ, Coper DH, Graham J (1996) Active shape models their training and application. *Comput Vis Image Underst* 61:38–59
- Cortes C, Vapnik V (1995) Support-vector networks. *Mach Learn* 20:273–297
- Cruz A, Bhanu B (2012) A biologically inspired approach for fusing facial expression and appearance for emotion recognition. In: *Proceedings of the 19th IEEE international conference on image processing*, Lake Buena Vista, FL, USA, October 2012, pp 2625–2628
- Ekman P, Friesen WV, Hager JC (2002) Facial Action Coding System (FACS). http://face-and-emotion.com/dataface/facs/new_version.jsp. Accessed 21 Sept 2016
- Fernández-Caballero A, Martínez-Rodrigo A, Pastor JM, Castillo JC, Lozano-Monazor E, López MT, Zangróniz R, Latorre JM, Fernández-Sotos A (2016) Smart environment architecture for emotion recognition and regulation. *J Biomed Inform* 64:55–73
- Fernández-Caballero A, Latorre JM, Pastor JM, Fernández-Sotos A (2014) Improvement of the elderly quality of life and care through smart emotion regulation. *Lect Notes Comput Sci* 8868:348–355
- Fernández-Caballero A, López MT, Serrano-Cuerda J (2014) Thermal-infrared pedestrian ROI extraction through thermal and motion information fusion. *Sensors* 14:6666–6676
- Gang L, Xiao-hua L, Ji-liu Z, Xiao-gang G (2009) Geometric feature based facial expression recognition using multiclass support vector machines. In: *Proceedings of the 2009 IEEE international conference on granular computing*, Nanchang, China, August 2009, pp 318–321
- Gascueña JM, Castillo JC, Navarro E, Fernández-Caballero A (2014) Engineering the development of systems for multisensory monitoring and activity interpretation. *Int J Syst Sci* 45:728–740
- Ghimire D, Lee J (2013) Geometric feature-based facial expression recognition in image sequences using multi-class AdaBoost and support vector machines. *Sensors* 13:7714–7734
- Hsieh C-C, Jiang M-K (2011) A facial expression classification system based on active shape model and support vector machine. In: *Proceedings of the 2011 international symposium on computer science and society*, Kota Kinabalu, Malaysia, July 2011, pp 311–314
- Kanade T, Cohn J, Tian YL (2000) Comprehensive database for facial expression analysis. In: *Proceedings of the fourth IEEE international conference on automatic face and gesture recognition*, Grenoble, France, March 2000, pp 46–53
- Lozano-Monazor E, López MT, Fernández-Caballero A, Vigo-Bustos F (2014) Facial expression recognition from webcam based on active shape models and support vector machines. *Lect Notes Comput Sci* 8868:147–154
- Lucey P, Cohn JF, de Kana T, Saragih J, Ambadar Z, Matthews I (2010) The extended Cohn–Kanade Dataset (CK+): a complete facial expression dataset for action unit and emotion-specified expression. In: *Proceedings of the 2010 IEEE Computer Society conference on computer vision and pattern recognition workshops*, San Francisco, CA, USA, June 2010, pp 94–101
- Lyons MJ, Kamachi M, Gyoba J (1997) Japanese Female Facial Expressions (JAFFE), Database of digital images. http://www.kasrl.org/jaffe_info.html. Accessed 21 Sept 2016
- Maglogiannis I, Vouyioukas D, Aggelopoulos C (2009) Face detection and recognition of natural human emotion using Markov random fields. *Pers Ubiquitous Comput* 13:95–101
- Pantic M, Bartlett M (2007) Machine analysis of facial expressions. In: *Face recognition*, chapter 20. I-Tech Education and Publishing, Vienna, Austria, pp 377–416
- Setyati E, Suprpto YK, Purnomo MH (2012) Facial emotional expressions recognition based on active shape model and radial basis function network. In: *Proceedings of the 2012 IEEE international conference on computational intelligence for measurement systems and applications (CIMSA)*, Tianjin, China, July 2012, pp 41–46
- Shan C, Gong S, McOwan PW (2009) Facial expression recognition based on local binary patterns: a comprehensive study. *Image Vis Comput* 27:803–816
- Soleymani M, Lichtenauer J, Pun T, Pantic M (2012) A multi-modal affective database for affect recognition and implicit tagging. *IEEE Trans Affect Comput* 3:42–55
- Stegmann MB (2002) Analysis and segmentation of face images using point annotations and linear subspace techniques. Technical report IMM-REP-2002-22. Informatics and Mathematical Modelling, Technical University of Denmark
- Suk M, Prabhakaran B (2014) Real-time mobile facial expression recognition system—a case study. In: *Proceedings of the 2014 IEEE conference on computer vision and pattern recognition workshops*, Columbus, Ohio, USA, June 2014, pp 132–137
- Tariq U, Lin K, Li Z, Zhou X, Wang Z, Le V, Huang TS, Lv X, Han TX (2011) Emotion Recognition from an ensemble of Features. In: *Proceedings of the 2011 IEEE international conference on automatic face and gesture recognition and workshops*, Santa Barbara, CA, USA, March 2011, pp 872–877
- Valstar MF, Pantic M (2010) Induced disgust, happiness and surprise: an addition to the IMM facial expression database. In: *Proceedings of the international conference on language resources and evaluation, workshop on emotion*, Valletta, Malta, May 2010, pp 65–70
- Wan C, Tian Y, Liu S (2012) Facial expression recognition in video sequences. In: *Proceedings of the 10th World congress on intelligent control and automation*, Beijing, China, July 2012, pp 4766–4770
- Wang L, Gu X, Wang Y, Zhang L (2009) Happy-sad expression recognition using emotion geometry feature and support vector machine. *Lect Notes Comput Sci* 5507:535–542
- Wei Y (2009) Research on facial expression recognition and synthesis. Master Thesis, Department of Computer Science and Technology, Nanjing University
- Wu J, Mei L (2013) A face recognition algorithm based on ASM and Gabor features of key points. In: *Proceedings of SPIE* 2013, vol 8768, article number 87686L
- Wu TF, Lin CJ, Wang RC (2004) Probability estimates for multi-class classification by pairwise coupling. *J Mach Learn Res* 5:975–1005
- Zhang L, Tjondronegoro DW, Chandran V (2012) Discovering the best feature extraction and selection algorithms for spontaneous facial expression recognition. In: *Proceedings of the 2012 IEEE conference on multimedia and expo*, Melbourne, Australia, July 2012, pp 1027–1032
- Zhao X, Zhang H, Xu Z (2012) Expression recognition by extracting facial features of shapes and textures. *J Comput Inf Syst* 8:3377–3384
- Zhou Q, Wang X (2013) Real-time facial expression recognition system based on geometric features. *Lect Notes Electr Eng* 212:449–456