Facial Expression Recognition from Webcam Based on Active Shape Models and Support Vector Machines

Elena Lozano-Monasor[1], María T. López[1,2], Antonio Fernández-Caballero[1,2], and Francisco Vigo-Bustos[2]

[1] Instituto de Investigación en Informática de Albacete (I3A), 02071-Albacete, Spain
[2] Universidad de Castilla-La Mancha, Departamento de Sistemas Informáticos, 02071-Albacete, Spain
Antonio.Fdez@uclm.es

Abstract. This paper introduces an application that uses a webcam and aims to recognize emotions of an elderly from his/her facial expression in real-time. 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 (ν-SVM) are employed for facial expression classification. In the future, the application is thought to be the starting point to enhance the mood of the elderly by external stimuli.

Keywords: Facial expressions, Emotions, Active shape model, Support vector machines.

1 Introduction

In recent years, there has been a growing interest in improving all aspects of interaction between humans and computers [1]. 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 recognize human emotion 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 customize its response [2].

Now, emotion recognition using visual cues has been 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

L. Pecchia et al. (Eds.): IWAAL 2014, LNCS 8868, pp. 147–154, 2014.

been presented, in which only eye and mouth expressions are used for detecting five emotions (all the above minus `Disgust`) [3]. Even, an approach to facial expression recognition for estimating patients' emotion is proposed with only two expressions (`Happiness` and `Sadness`) [4]. Applications that use these techniques are varied, ranging from software able to recognize and act according to the emotions of the user who is using it, systems capable of detecting lies, up to applications that allow knowing if a product is liked or not with only analyzing 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 the central roles in providing discriminative and robust information [5]. The selection of features employed for emotion recognition are classified into two main categories: geometric features and appearance features. 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 in real-time, and, (b) the creation of an emotion detector through implementing the most suited classification techniques.

2 ASM and SVM for Facial Expression Recognition from Geometric Features

It has been demonstrated that the active shape model (ASM) is a good method for locating facial feature points [6]. Generally speaking, ASM fits the shape parameters using optimization techniques such as gradient descent. On the other hand, support vector machines (SVM) [7] 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 [8]. 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 utilized as geometric features for emotion recognition [9]. A quite recent approach [10] utilizes 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 face and acquire important facial feature regions. Gabor filter and Laplacian of Gaussian 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 [11], an algorithm of face recognition based on ASM and Gabor features of key points has been proposed. Firstly, AdaBoost algorithm detects the face region in an image. Then, the ASM localizes 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" (see `http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html`). Another paper describes a method for recognition of continuous facial expression change in video sequences [12]. Again, ASM automatically localizes 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 changes in 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 center of gravity of the face's ASM shape with the FAU relative facial component deformation distances [13]. The approach also introduces a multiclass one-against-one ν-SVM for facial expression classification.

Another paper [14] 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's system overview is explained next [15]. 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 minimize 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.

3 Real-Time Recognition of Emotions from Face Expressions

This paper presents a real-time facial expression recognition system based on geometric features [16]. 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 the 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 analyze facial features. To do this, a complete description of facial expressions is needed. The Facial Action Coding System (FACS) [17] 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 you have to exert the muscle. 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 analyzes the changes that occur in facial expression and relate them to a specific emotion. Obviously, a reference database is used to associate the facial expressions observed.

3.2 Facial Emotion Detection

The approach described in this paper is divided into 4 steps:

1. **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.
2. **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.
3. **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 in real-time.
4. **Detection of emotions.** At the last step, an emotion detection system is obtained. It is built from the models generated in the previous steps.

Detection of Facial Points and Features. ASMLibrary [18] 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 ASMs 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) [19], which contains the analysis of 37 images of frontal faces. The model is composed of 58 facial points; (b) BioID [20] is a database consisting of 1521 images of frontal faces. Each face is labeled with 20 facial points; (c) Extended Multi Modal Verification for Teleservices and Security (XM2VTS) [21] 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 analyze 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 real-time video input (webcam). 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.

Training, Classification and Detection of Emotions. LibSVM [22] is a library for programming support vector machines (SVMs). The image features belonging to properly labeled 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 [23]) from the two possible alternative approaches. Although this method involves using more classifiers, the employed training time is much lower. It has been decided to use the RBF kernel as it is the one that offers best results in terms of accuracy and training time.

Furthermore, four different well-known image databases are selected to carry out the training of the SVM: (1) JAFFE (Japanese Female Facial Expression) database [24], (2) IMM facial expressions database [25], (3) Cohn-Kanade (CK) database [26], and, (4) Cohn-Kanade extended (CK+) database [27]. Finally, the ν-SVM algorithm is used due to the ease of adjustment of the ν parameter. The classification features and values used are shown in Table 1.

Table 1. Features of the ν-SVM model

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

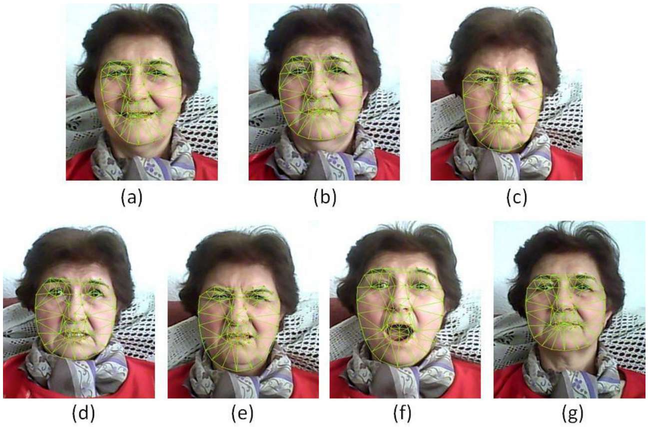

Fig. 1. Example of webcam capture where the detected emotion is (a) `Joy`. (b) `Sadness`. (c) `Anger`. (d) `Fear`. (e) `Disgust`. (f) `Surprise`. (g) `Neutral`.

4 Data and Results

In order to validate our proposal, we capture video in real-time 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. So, the tests are performed for each of the seven classes that the system has to distinguish, that is, `Joy`, `Sadness`, `Anger`, `Fear`, `Disgust`, `Surprise` and `Neutral`.

The emotions that offer better results are `Surprise`, `Joy`, `Sadness` and `Anger` (all of them with a hit rate over 0.95). The results are acceptable when the facial expression reflects `Surprise` and `Fear` (hit rate over 0.8). Emotion `Disgust` provides the most false positive results (around a hit rate of 0.5), probably because it contains features that can lead to confusion with other emotions, such as frowning (characteristic of emotion `Anger`) or lips down (characteristic of `Sadness` and `Fear`). The main reasons that probably explain the prediction errors are: (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.

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 in real-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 in real-time. Tests have been performed with several models, and the best results were obtained with the 68 facial points model. Also, we have studied support vector machines for classification and 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.

This has led to the construction of an application capable of distinguishing emotions of older people in real-time 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 are `Surprise`, `Joy`, `Sadness` and `Anger`. The results are acceptable, especially when the facial expression reflects `Surprise`. For other emotions, such as `Disgust` and `Fear`, the system tends to get confused because emotions have very similar facial features. The work described in this paper is the first step in developing a system to improve mood in the elderly by external non-intrusive stimuli.

Acknowledgements. This work was partially supported by Spanish Ministerio de Economía y Competitividad / FEDER under TIN2013-47074-C2-1-R and TIN2010-20845-C03-01 grants.

References

1. Gascueña, J.M., Castillo, J.C., Navarro, E., Fernández-Caballero, A.: Engineering the development of systems for multisensory monitoring and activity interpretation. International Journal of Systems Science 45(4), 728–740 (2014)
2. Alugupally, N., Samal, A., Marx, D., Bhatia, S.: Analysis of landmarks in recognition of face expressions. Pattern Recognition and Image Analysis 21(4), 681–693 (2011)
3. Maglogiannis, I., Vouyioukas, D., Aggelopoulos, C.: Face detection and recognition of natural human emotion using Markov random fields. Personal and Ubiquitous Computing 13(1), 95–101 (2009)
4. Wang, L., Gu, X., Wang, Y., Zhang, L.: Happy-sad expression recognition using emotion geometry feature and support vector machine. In: Köppen, M., Kasabov, N., Coghill, G. (eds.) ICONIP 2008, Part II. LNCS, vol. 5507, pp. 535–542. Springer, Heidelberg (2009)
5. Zhang, L., Tjondronegoro, D.W., Chandran, V.: Discovering the best feature extraction and selection algorithms for spontaneous facial expression recognition. In: Proc. 2012 IEEE Conference on Multimedia and Expo, pp. 1027–1032 (2012)
6. Cootes, T.F., Taylor, C.J., Coper, D.H., Graham, J.: Active shape models - their training and application. Computer Vision and Image Understanding 61(1), 38–59 (1996)
7. Cortes, C., Vapnik, V.: Support-vector networks. Machine Learning 20(3), 273–297 (1995)
8. Chang, Y., Hu, C., Feris, R., Turk, M.: Manifold based analysis of facial expression. Image and Vision Computing 24(6), 605–614 (2006)

9. Pantic, M., Bartlett, M.: Machine analysis of facial expressions. In: Face Recognition, ch. 20, pp. 978–973 (2007) ISBN 978-3-902613-03-5
10. Hsieh, C.-C., Jiang, M.-K.: A facial expression classification system based on active shape model and support vector machine. In: Proc. 2011 International Symposium on Computer Science and Society, pp. 311–314 (2011)
11. Wu, J., Mei, L.: A face recognition algorithm based on ASM and Gabor features of key points. In: Proceedings of SPIE 8768, Article number 87686L (2013)
12. Wan, C., Tian, Y., Liu, S.: Facial expression recognition in video sequences. In: Proc. 10th World Congress on Intelligent Control and Automation, pp. 4766–4770 (2012)
13. Gang, L., Xiao-hua, L., Ji-liu, Z., Xiao-gang, G.: Geometric feature based facial expression recognition using multiclass support vector machines. In: Proc. 2009 IEEE International Conference on Granular Computing, pp. 318–321 (2009)
14. Zhao, X., Zhang, H., Xu, Z.: Expression recognition by extracting facial features of shapes and textures. Journal of Computational Information Systems 8(8), 3377–3384 (2012)
15. Cruz, A., Bhanu, B.: A biologically inspired approach for fusing facial expression and appearance for emotion recognition. In: Proc. 19th IEEE International Conference on Image Processing, pp. 2625–2628 (2012)
16. Zhou, Q., Wang, X.: Real-time facial expression recognition system based-on geometric features. Lecture Notes in Electrical Engineering, vol. 212, pp. 449–456 (2013)
17. Ekman, P., Friesen, W.V., Hager, J.C.: Facial Action Coding System (FACS) (2002), `http://face-and-emotion.com/dataface/facs/new_version.jsp`
18. Wei, Y.: Research on facial expression recognition and synthesis. Master Thesis. Department of Computer Science and Technology, Nanjing University (2009), `http://code.google.com/p/asmlibrary`
19. Stegmann, M.B.: Analysis and segmentation of face images using point annotations and linear subspace techniques. Technical Report IMM-REP-2002-22 (2002), `http://www2.imm.dtu.dk/pubdb/views/publication_details.php?id=922`
20. Cootes, T.F., Cristinacce, D., Babalola, K.: BioID face database (2005), `http://www.bioid.com/index.php?q=downloads/software/bioid-face-database.html`
21. Chan, C.H.: The XM2VTS Database (2000), `http://www.ee.surrey.ac.uk/CVSSP/xm2vtsdb/`
22. Chang, C.C., Lin, C.J.: LIBSVM: a library for support vector machines. ACM Transactions on Intelligent Systems and Technology 2(27), 1–27 (2011)
23. Wu, T.F., Lin, C.J., Wang, R.C.: Probability estimates for multi-class classification by pairwise coupling. Journal of Machine Learning Research 5, 975–1005 (2004)
24. Lyons, M.J., Kamachi, M., Gyoba, J.: Japanese Female Facial Expressions (JAFFE). Database of Digital Images (1997), `http://www.kasrl.org/jaffe_info.html`
25. Valstar, M.F., Pantic, M.: Induced disgust, happiness and surprise: an addition to the IMM facial expression database. In: Proc. International Conference on Language Resources and Evaluation, Workshop on Emotion, pp. 65–70 (2010)
26. Kanade, T., Cohn, J., Tian, Y.L.: Comprehensive database for facial expression analysis. In: Proc. 4th IEEE International Conference on Automatic Face and Gesture Recognition, pp. 46–53 (2000)
27. Lucey, P., Cohn, J.F., Kanade, T., Saragih, J., Ambadar, Z., Matthews, I.: The Extended Cohn-Kanade Dataset (CK+): A complete facial expression dataset for action unit and emotion-specified expression. In: Proc. 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, pp. 94–101 (2010)