FACIAL EXPRESSION RECOGNITION WITH RELEVANCE VECTOR MACHINES

D.Datcu　　*L.J.M.Rothkrantz*

Delft University of Technology
Man-Machine Interaction Group
Mekelweg 4, 2628 CD Delft, The Netherlands

ABSTRACT

For many decades automatic facial expression recognition has scientifically been considered a real challenging problem in the fields of pattern recognition or robotic vision. The current research aims at proposing Relevance Vector Machines (RVM) as a novel classification technique for the recognition of facial expressions in static images. The aspects related to the use of Support Vector Machines are also presented. The data for testing were selected from the Cohn-Kanade Facial Expression Database. We report 90.84% recognition rates for RVM for six universal expressions based on a range of experiments. Some discussions on the comparison of different classification methods are included.

1. INTRODUCTION

It is the human nature that we can estimate a person's psychological state following the observation on his face. Nonverbal communication channels are typically set during common interpersonal relations and visual messages are processed in a transparent manner. The general tendency is to construct robotic systems that are able to understand the environmental world and to interact with the existent actors. Human-computer interfaces play an essential role in the perception and feedback the system
is capable of. In this context, the advantage of making machines to read human facial expressions is tremendous. Facial expressions reveal internal characteristics of the expresser. To address the problem of facial expression recognition, in our approach we extract parametric information with high discrimination power from facial feature space and use it in a data-driven classification environment. The current paper primarily focuses on the aspects related to the classification methods for facial expression recognition. The classifiers are aimed at solving the universal problem of classification. We begin with Support Vector Machines SVM [1] that is based on a solid mathematical foundation. Then the novel Relevance Vector Machines RVM [2] is introduced as an enhanced SVM. The difficulty of the automatic analysis of facial expressions [3] resides in the variety of characteristic appearance with respect to both individuality and face anatomic dynamics. The inner complexity makes from the processes of feature oriented expression recognition and feature detection difficult tasks. To our knowledge, this is the first research that involves Relevance Vector Machines for facial expression recognition.

2. RELATED WORK

The recognition of facial expressions implies finding solutions to three distinct types of problems. The first one relates to detection of faces in the image. Once the face location is known, the second problem is the detection of the salient features within the facial areas. The final analysis consists in using any classification model and the extracted facial features for identifying the correct facial expression. For each of the processing steps described, there have been developed lots of methods to tackle the issues and specific requirements. Depending on the method used, the facial feature detection stage involves global or local analysis. The internal representation of the human face can be either 2D or 3D. In the case of global analysis, the connection with certain facial expressions is made through features determined by processing the entire face. The efficiency of methods as Artificial Neural Networks or Principal Component Analysis is greatly affected by head rotation and special procedures are needed to compensate the effects of that. On the other hand, local analysis performs encoding of some specific feature points and uses them for recognition. The method is actually used in the current paper. However, other approaches have been also performed at this layer. One method for the analysis is the internal representation of facial expressions based on collections of Action Units (AU) as defined in Facial Action Coding System (FACS) [4] [5]. It is one of the most efficient and commonly used methodologies to handle facial expressions. Some

The research reported here is part of the Interactive Collaborative Information Systems (ICIS) project, supported by the Dutch Ministry of Economic Affairs, grant nr: BSIK03024.

0-7803-9332-5/05/$20.00 ©2005 IEEE

attempts to automatically detect the salient facial features implied computing descriptors such as scale-normalized Gaussian derivatives at each pixel of the facial image and performing some linear-combinations on their values. It was found that a single cluster of Gaussian derivative responses leads to a high robustness of detection given the pose, illumination and identity [6]. A representation based on topological labels is proposed [7]. It assumes that the facial expression is dependent on the change of facial texture and that its variation is reflected by the modification of the facial topographical deformation. The classification is done by comparing facial features with those of the neutral face in terms of the topographic facial surface and the expressive regions. Some approaches firstly model the facial features and then use the parameters as data for further analysis such as expression recognition. The system proposed by [8] is based on a 2D generative eye model that implements encoding of the motion and fine structures of the eye and is used for tracking the eye motion in a sequence. As concerning the classification methods, various algorithms have been developed, adapted and used during time [3]. Neural networks have been used for face detection and facial expression recognition [9] [10]. The second reference directs to a system called Facial Expression Dictionary (FED) [10] that was a first attempt to create an online nonverbal dictionary. Other classifiers included Bayesian Belief Networks (BBN) [11], Expert Systems [12] or Support Vector Machines (SVM) [4]. Other approaches have been oriented on the analysis of data gathered from distinct multi-modal channels. They combined multiple methods for processing and applied fusion techniques to get to the recognition stage [13].

3. VISUAL FEATURE MODEL

Although local analysis is sensitive to identity and partial occlusions, we overcome that by increasing the variability of data for training the model and by increasing the parameter redundancy. The variability is handled by the face characteristics in the Cohn-Kanade database [14]. The redundancy assumes the use of feature parameters on an asymmetric model, i.e. in case of occlusion or low visibility for one eye, the recognition is done by taking into account the data from the other eye. The task prior to classification stage is aimed at preparing the feature data associated with the input face. Depending on the type of the classifier involved in the next step, the data have to be transformed to a certain format. In the current approach a transform $\Gamma : \theta \to \vartheta$ converts the facial feature images θ to some parameters $p_i \subseteq \vartheta, i = 1,...,L$ of an intermediate model. The parameterization of the facial features has the advantage of providing the classifier with data that encode the most important aspects of the facial expressions. Furthermore, it acts as a dimensionality reduction procedure since the dimension of the feature space is lower than the dimension of the image space. An advantage of the model is that it also can handle a certain degree of asymmetry by using some parameters for both left and right sides of the face. Each facial feature $\kappa \in \theta$ θ = {Left/Right eye, Left/Right Eyebrow, Mouth, Chin} is extracted at the previous processing stage from distinct processing channels. The transform Γ firstly extracts the location of each FCP from the input facial feature κ.

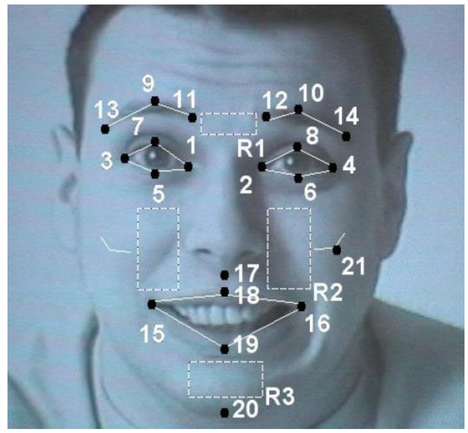

Figure 1: FCP set

Eventually the feature parameters p_i are computed as values of some angles and/or Euclidean distances between key points assumed to reflect the location of the facial features. The key points are defined as Facial Characteristic Points (FCPs) and the FCP-set (Figure 1) is based on an extension of Kobayashi & Hara model [15]. The final step of preprocessing was related to scale all the distances so as to be invariant to the size of the image.

4. CLASSIFIERS

If X denotes the space of input variables representing the face images and T the space of output variables i.e. the facial expression label, then f is the associated deterministic function and $t_n = f(x_n) + \varepsilon_n$ represent the possible overlapping target values. The training database is $Z = \{(x_i, t_i) \in X \times T \mid i = 1,..., M\}$. The learning step implies the use of the training database Z together with any other prior knowledge for finding a function $\hat{f}$ out of a class F of functions that encodes the estimated dependency. The process can be seen as a transform Ψ of the initial data v into some internal knowledge ν of the interest phenomena and so $\Psi : v \rightarrow v$. From the notations used above, $Z \subseteq v$ and $\hat{f} \subseteq v$. The result function $\hat{f}$ is assumed to produce an efficient prediction of the facial expression label t given a new feature vector x.

4.1. SVM

The Support Vector Machine (SVM) algorithm has been successfully used in classification related problems since it was introduced by Vapnik [1] in the late 1970's. The idea was that given the collection of input-target pairs Z with $X \subseteq R^n$ and $T = \{-1,+1\}$, a hyperplane $f \in F_h$, $F_h = \{f : X \rightarrow R \mid (w^T x) + b\}$ with the maximal margin has to be found as a solution of an optimization problem. The distribution of the two classes is such that they are linearly separable. The constraints aim at determining the model parameters $\{w,b\}$ that fit the training data and minimize the complexity of the decision function in the same time. The result is a classifier with a certain level of robustness to over-fitting. The margin represents a measure of class separation efficiency and is defined as the Euclidean distance between the data and the separating hyperplane. Non-linearity that is specific to facial expression representation is handled through kernel methods (non-linear SVM) that first preprocess the data by non-linear mapping $\phi : R^n \rightarrow \varepsilon$ and then apply the linear algorithm in the image space ε. The image space is a vector space of functions $\varepsilon = \{f \mid f : X \rightarrow R\}$. The positive definite kernel function $k : X \times X \rightarrow \phi$ acts as a dot-product over ϕ and the mapping is expressed as $\phi(x) = k(\cdot, x)$.

4.1. RVM

Tipping [2] introduced the Relevance Vector Machine (RVM) as a probabilistic sparse kernel model based on the support vector machine theory. Each of the model's weights has associated a prior that is characterized by a set of hyperparameters whose values are determined during the learning process. Following the above notations $p(y \mid x)$ is assumed to be Gaussian $N(y \mid f(x), \sigma^2)$ and the mean of the distribution is computed as specified for the SVM. The overfitting effect that occurs while determining the parameter values of the likelihood $p(y \mid w, \sigma^2)$ is confined by including an ARD Gaussian prior over the weights w_i as $p(w \mid \alpha) = \prod_{i=0}^{N} N(w_i \mid 0, \alpha_i^{-1})$.

The dataset likelihood is expressed by using logistic functions, in the form:

$$P(y \mid w) = \prod_{n=1}^{N} \sigma\{f(x_n)\}^{y_n} [1 - \sigma\{f(x_n)\}]^{1-y_n} \quad (1)$$

The parameters α_i and w_i of the model are computed through an iteration procedure until convergence is achieved. One advantage over SVM is that for comparable generalization performance, it uses fewer kernel functions. This requires less memory and time for processing and so makes possible for the usage of RVM as a real-time classifier. By using RVM, the relevance vectors stand for representative training samples of the emotional classes rather than data points closer to the separation hyperplane in SVM model.

5. EXPERIMENTS

The data used for training the models for the experiments have been selected from the Cohn-Kanade database. For the experiments only 485 images have been used, each image representing a sample data for the model. Some specific steps were passed to extract and prepare the data to comply with the requirements of the classification process. The algorithms are set to perform classification on the 6 universal expressions (Happy, Anger, Sad, Surprise, Disgust, Fear). The classification results of the facial expression recognition can be analyzed by looking at the mismatch rate. The method used for computing the error was leave-5-out cross validation. As it can be seen (Table 1 and Table 2), the error rate in the case of RVM (9.16%) is comparable to that of SVM (10.15%) classifier. One important aspect is that in case of RVM classifier the number of relevance vectors (156) is smaller than the number of support vectors (276) of SVM. The effect is a decrease of the number of kernel functions and so of the complexity of the model. In terms of practical characteristics that means less processing time and also less memory for using this type of classifier. Nevertheless, the analysis of facial expressions in static images has its own limitations. That can be mainly explained by the dynamics characteristics of the salient features involved in facial expressions' structure. An important improvement for the recognition system may include also the encoding and usage of the knowledge over these elements [11].

6. CONCLUSION

The goal of the current paper is to highlight the potential of the Relevance Vector Machines as a facial expression classifier. The exemplifications start from the idea of the Support Vector Machines and address the issues concerning the use of two types of classifiers in the context of facial expression recognition problem. The RVM is a relatively new classification method and this work is the first one that uses it as a recognition engine for facial expressions. The fundamental aspects are described on both theoretical and practical sides. Each classifier model presents certain advantages and limitations and has been designed so as to perform prediction on the static images. We focused on the recognition of only universal expressions triggered by six basic emotions. Further research aims at including more emotional classes for analysis. The results for RVM show that it is suitable for facial expression classification in static images and it leads to a decrease of complexity comparing to SVM.

Table 1: Mismatch rate by SVM classifier

Expression	Mismatch rate	Number of support vectors
Surprise	14.32 ± 1.80%	63
Sadness	3.06 ± 2.78%	23
Anger	5.91 ± 1.55%	46
Happy	3.16 ± 2.47%	38
Disgust	9.54 ± 1.97%	34
Fear	24.97 ± 2.07%	72

Table 2: Mismatch rate by RVM classifier

Expression	Mismatch rate	Number of relevance vectors
Surprise	6.25 ± 1.51%	21
Sadness	12.29 ± 2.57%	34
Anger	5.00 ± 1.87%	15
Happy	7.92 ± 1.71%	23
Disgust	8.54 ± 1.91%	25
Fear	15.00 ± 2.38%	38

The still image analysis is very restrictive with respect to the subtle dynamics of the facial features. Additional research has been initiated to encode temporal behavior in the classification models so as to make possible the use of the recognition systems to run on image sequences. Another idea for increasing the capabilities and efficiency is to make use of fusion techniques to handle multiple audio/video modalities.

11. REFERENCES

[1] V.Vapnik, "The nature of statistical learning" *Springer, New York*, 1995.

[2] M. E. Tipping, "The relevance vector machine," *2000*, In Sara A Solla, Todd K Leen, and Klaus-Robert Muller, editors, Advances in Neural Information Processing Systems 12. Cambridge, Mass: MIT Press.

[3] M.Pantic L.J.M.Rothkrantz, "Automatic analysis of facial expressions: the state of the art," *IEEE Trans. PAMI*, 2000, 22(12).

[4] M.S.Bartlett G.Littlewort C.Lainscsek I.Fasel J.Movellan, "Machine learning methods for fully automatic recognition of facial expressions and facial actions," *Proceedings of IEEE SMC*, pp. 592–597, October 2004.

[5] P.Ekman W.V.Friesen, "Facial action coding system: investigator's guide," *Consulting Psychologists Press*, 1978.

[6] N.Gourier D.Hall J.L.Crowley, "Facial feature detection robust to pose, illumination and identity," *Proceedings of IEEE SMC 2004*, pp. 617–622, October 2004.

[7] L.Yin J.Loi W.Xiong, "Facial expression analysis based on enhanced texture and topographical structure," *Proceedings of IEEE SMC*, pp. 586–591, October 2004.

[8] T.Moriyama J.Xiao J.F.Cohn T.Kanade, "Meticulously detailed eye model and its application to analysis of facial image," *Proceedings of IEEE SMC*, pp.580–585, October 2004.

[9] I.O.Stathopoulou G.A.Tsihrintzis, "An improved neural-network-based face detection and facial expression classification system," *Proceedings of IEEE SMC*, pp. 666–671, October 2004.

[10] E.J.deJong L.J.M.Rothkrantz, "Fed - an online facial expression dictionary," *Proceedings of Euromedia 2004*, pp. 115–119, April 2004.

[11] D.Datcu L.J.M.Rothkrantz, "Automatic recognition of facial expressions using bayesian belief networks," *Proceedings of IEEE SMC*, pp. 2209–2214, October 2004.

[12] M.Pantic L.J.M.Rothkrantz, "Self-adaptive expert system for facial expression analysis," *IEEE International Conference on Systems, Man and Cybernetics (SMC '00)*, pp. 73–79, October 2000.

[13] N.A.Fox R.B.Reilly, "Robust multi-modal person indentification with tolerance of facial expression," *Proceedings of IEEE SMC*, pp. 580–585, October 2004.

[14] T.Kanade J.Cohn Y.Tian, "Comprehensive database for facial expression analysis," *Proc. IEEE Int'l Conf. Face and Gesture Recognition*, pp. 46–53, 2000.

[15] H.Kobayashi F.Hara, "Recognition of mixed facial expressions by neural network," *IEEE International workshop on Robot and Human Communication*, pp. 381–386, 1972.