

Real-Time Facial Expression Recognition with Illumination-Corrected Image Sequences

He Li

Department of Computer Science and Engineering, Fudan University, China

demonstrate@163.com

José M. Buenaposada

Dept. Ciencias de la Computación, Universidad Rey Juan Carlos, Spain

josemiguel.buenaposada@urjc.es

Luis Baumela

Dept. Inteligencia Artificial, Universidad Politécnica de Madrid, Spain

lbaumela@fi.upm.es

Abstract

We present a real-time user-independent computer vision system that processes a sequence of images of a front-facing human face and recognises a set of facial expressions at 30 fps. We track the face using an efficient appearance-based face tracker. We model changes in illumination with a user-independent appearance-based model. In our approach to facial expression classification, the image of a face is represented by a low dimensional vector that results from projecting the illumination corrected image onto a low dimensional expression manifold. In the experiments conducted we show that the system is able to recognise facial expressions in image sequences with large facial motion and illumination changes.

1. Introduction

Facial expression classification and recognition is a topic of interest within the Computer Vision community, mainly, because of its applications in the construction of advanced human-computer interaction systems. The problem of facial expression recognition can be divided into three subproblems: face detection, discriminative information extraction and expression classification. Once the position of the face in an image has been estimated, it is analysed to extract discriminative information that will be used to classify the facial expression. Different facial expression recognition algorithms have been introduced in the literature depending on the discriminative information extracted from the image and the classification procedure used [11, 6].

Regarding to facial expressions there is evidence to support the existence of six primary emotions, which are universal across cultures and human ethnicities [5]. Each emotion possesses a distinctive prototypic facial expression. These basic emotions are joy (jo), surprise (su), anger (an), sadness (sa), fear (fe) and disgust (di). Recognising all or a subset of these prototypic facial expressions from images has been a topic of research in computer vision and pattern recognition for the last decade (e.g. [12, 10, 4, 13, 16]).

In this paper we consider all three facial expression recognition subproblems. Our solution represents a good compromise between user and illumination independence, real-time performance and classification error. Our system tracks, in real-time, the rigid motion of a front-facing human face while dealing with illumination changes. We adopt a model-based procedure for tracking. In our approach the appearance of a face is represented by a linear subspace of illumination computed in a user independent way. We fit the user independent appearance model to an incoming image by using the algorithm of Buenaposada et al [2] an efficient and robust fitting algorithm.

We also adopt a model-based approach for facial expression recognition. By tracking a set of 322 image sequences of 92 subjects from the Cohn-Kanade data base [8], we build a user-and-illumination-independent global representation of all facial expressions. In this model, a face image is represented with a point in an n -dimensional space of deformations, where n is the number of face model parameters. The variability of the classes of images associated to the prototypic facial expressions are represented by a set of samples that model a low-dimensional manifold embedded in the n -dimensional space of deformations. Images repre-

sending similar expressions are mapped to nearby points on the manifold. We use a probabilistic procedure [3] to combine the information provided by the image sequence with the information represented in the manifold to estimate a posterior probability for each facial expression.

Our solution differs from previous related approaches [4, 13, 3] in various ways: a) our manifolds are user independent, while those introduced in [4] depend on the identity of the user. Although the manifold used in [3] was also user independent the face fitting procedure was subject dependent; b) we use illumination corrected images (with a user-independent illumination model) as discriminative information for expression classification, whereas Active Wavelet Networks [7], deformation appearance subspace coefficients [2] and Local Binary Pattern features [9] were used respectively in [4], [3] and [13];

2. Face alignment

In this section we describe the face alignment used in our facial expression recognition algorithm. The face is initially located using a face detection algorithm. We use the well known Viola-Jones procedure [15]. Face detection algorithms provide a rough estimate of the location and scale of the face which does not suffice for facial expression recognition. We use then an efficient model-based face alignment procedure to accurately locate the face and compensate illumination effects.

2.1. The face model

We will assume that faces are in frontal view. In this case changes in the appearance of a face are caused by identity and illumination variations. We build a user-independent model of illumination by registering sample faces from different individuals under different illumination conditions. In our experiments we show that this model suffices to track different people and recognise their facial expressions with a high degree of robustness with respect to illumination changes.

Let $I(\mathbf{x}, t)$ be the image acquired at time t , where \mathbf{x} is a vector representing the co-ordinates of a point in the image, and let $\bar{\mathbf{I}}(\mathbf{x}, t)$ be a vector storing the brightness values of $I(\mathbf{x}, t)$. The first order approximation to the grey value of pixel \mathbf{x} can be expressed as $\bar{\mathbf{I}}(\mathbf{x}) + [\mathbf{B}\mathbf{c}_t](\mathbf{x})$, where \mathbf{B} is the basis of the illumination subspace, vector \mathbf{c} is the illumination appearance parameters and $\bar{\mathbf{I}}(\mathbf{x})$ is the average image.

The rigid motion of the face is modelled by $f(\mathbf{x}, \boldsymbol{\mu})$, being $\boldsymbol{\mu}$ the vector of rigid motion parameters. So, the brightness constancy equation is

$$\mathbf{I}(f(\mathbf{x}, \boldsymbol{\mu}_t), t) = \bar{\mathbf{I}}(\mathbf{x}) + [\mathbf{B}\mathbf{c}_t](\mathbf{x}) \quad \forall \mathbf{x} \in \mathcal{F}, \quad (1)$$

where $k = \dim(\mathbf{c}_t)$, and \mathcal{F} represents the set of pixels of the face used for alignment.

We train a generic facial appearance model with the PIE database [14]. Here, each illumination is averaged across all identities. The result is an average image for each illumination direction (see Fig. 1) that will be used for building the illumination subspace. Matrix \mathbf{B} is estimated by selecting the nine¹ directions with highest variance of the eigenspace spanning the set of frontal averaged images of the PIE database (see Fig. 2). We also add a vector of ones to account for global brightness changes.



Figure 1. Aligned images used to build the illumination subspace (61 by 72 pixels each).



Figure 2. Mean of illumination training images (first image on the left). Illumination subspace basis vectors (remaining images).

2.2. Model fitting

We fit the previous model to a target image by estimating the motion, $\boldsymbol{\mu}$, and appearance, \mathbf{c} , parameters which minimise $E(\boldsymbol{\mu}, \mathbf{c}) = \|\mathbf{I}(f(\mathbf{x}, \boldsymbol{\mu}_t), t) - \bar{\mathbf{I}} - [\mathbf{B}\mathbf{c}_t](\mathbf{x})\|^2$. These parameters may be estimated in real-time with Buenaposada et al's fitting algorithm [2]. We make a Taylor series expansion of \mathbf{I} at $(\boldsymbol{\mu}_t, \mathbf{c}_t, t)$, producing a new error function

$$E(\delta\boldsymbol{\mu}, \delta\mathbf{c}) = \|\mathbf{M}\delta\boldsymbol{\mu} + \mathbf{I}(f(\mathbf{x}, \boldsymbol{\mu}_t), t + \delta t) - \bar{\mathbf{I}} - \mathbf{B}(\mathbf{c}_t + \delta\mathbf{c})\|^2, \quad (2)$$

where $\mathbf{M} = \left[\frac{\partial \mathbf{I}(f(\mathbf{x}, \boldsymbol{\mu}_t), t)}{\partial \boldsymbol{\mu}} \right]_{\boldsymbol{\mu}=\boldsymbol{\mu}_t}$ is the $N \times n$ ($n = \dim(\boldsymbol{\mu})$) Jacobian matrix of \mathbf{I} , where N is the number of pixels in appearance model. An efficient solution for estimating the minimum of (2) is given by

$$\delta\boldsymbol{\mu} = -(\boldsymbol{\Sigma}^\top \boldsymbol{\Lambda}_{M1} \boldsymbol{\Sigma})^{-1} \boldsymbol{\Sigma}^\top \boldsymbol{\Lambda}_{M2} \boldsymbol{\mathcal{E}}; \quad \delta\mathbf{c} = \boldsymbol{\Lambda}_B [\mathbf{M}\delta\boldsymbol{\mu} + \boldsymbol{\mathcal{E}}], \quad (3)$$

where $\boldsymbol{\Lambda}_B$ (the \mathbf{B} matrix pseudo-inverse), $\boldsymbol{\Lambda}_{M1}$ and $\boldsymbol{\Lambda}_{M2}$ are constant matrices that can be precomputed off-line and $\boldsymbol{\Sigma}$ is a matrix that depends on $\boldsymbol{\mu}_t$ and \mathbf{c}_t [2].

In the experiments conducted we use a RTS (rotation, translation and scale) motion model, so $\boldsymbol{\mu} = (t_u, t_v, \theta, s)$, and $f(\mathbf{x}, \boldsymbol{\mu}) = s\mathbf{R}(\theta)\mathbf{x} + \mathbf{t}$, where $\mathbf{x} = (u, v)^\top$, $\mathbf{t} = (t_u, t_v)^\top$ and $\mathbf{R}(\theta)$ is a 2D rotation. Therefore $\boldsymbol{\Lambda}_{M1}$ is a $4(k+1) \times 4(k+1)$ matrix and $\boldsymbol{\Lambda}_{M2}$ is an $4(k+1) \times N$ ($N = 61 \times 72$ in our experiments).

¹Nine components suffice to represent 97% of the energy in the image[1].

3. Facial Expression Recognition

The classification procedure used for facial expression recognition is based on a user-and-illumination-independent facial expression model. This model is built by tracking a set of sequences from the Cohn-Kanade data base [8]. The last image in each sequence is labelled with the FACS Action Units (AUs) that describe the expression. There is no direct translation from AUs into one of the universal expressions (most of the time the translation is subjective). We have manually translated these AUs into one of the six prototypic expressions. To construct our manifold, we selected only the 322 sequences of 92 subjects for which the prototypic expression could be clearly identified². In the next two subsections we explain the features and the classifier used in our experiments.

3.1. Dimensionality Reduction

We used the tracker introduced in section 2 to process the sequences from the Cohn-Kanade data base. Once motion and illumination parameters have been estimated, we can compute an illumination normalised version of the rectified image, $\mathbf{I}(f(\mathbf{x}, \boldsymbol{\mu}_t), t) - \text{Bc}_t$ (see Fig. 3).



Figure 3. Illumination rectified images. The first row displays the original cropped images. The second row displays the corresponding illumination rectified ones.

The dimension of the illumination normalised images is quite high compared with the amount of data available for training (in the experiments conducted in section 4 this dimension is 61×72). To avoid the curse of dimensionality we use a dimensionality reduction procedure. We further process the set of Cohn-Kanade illumination corrected images by making PCA (we retain 90 eigenvectors) and then LDA. As we have 6 face expressions, we get a 5 dimensional subspace. We are going to call this dimensionality reduction procedure PCA+LDA.

Only six expressions (surprise, fear, joy, sadness, disgust and anger) will be identified. The neutral expressions can be regarded as the case when the system shows no favour for any of the 6 known expressions. Since the information associated to the appearance of the facial expression is represented by the LDA parameters, \mathbf{f} , the expression in the sequence of images $\mathbf{I}_1, \dots, \mathbf{I}_m$ can be identified as a trajectory, $\mathbf{f}_t, t = 1 \dots m$ in the expressions subspace. Trajectories associated to the same prototypic facial expression

represent roughly similar facial deformations and, consequently, will be located in nearby positions in the deformation subspace. Conversely, the trajectories of different expressions will be located in different positions in the subspace (see Fig. 4).

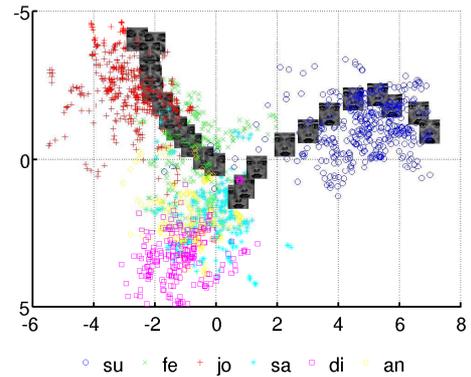


Figure 4. Trajectories of two prototypic facial expressions (joy and surprise) for two different subjects in the subspace spanned by the two directions of PCA+LDA subspace with the largest variance.

Our model of a prototypic facial expression is the manifold that contains the set of trajectories of that expression in the data base. Since all expressions are defined in the common linear space spanned by the PCA+LDA subspace vectors, our facial expression model is the union of the six manifolds associated to each prototypic facial expression (see Fig. 5). In our model we only introduce the last 6 images of each training sequence, since the first images of all sequences contain the neutral expression.

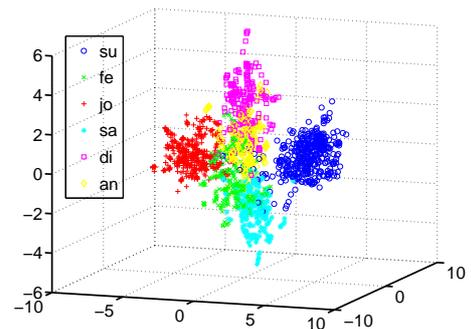


Figure 5. Manifold of facial expressions (we show only the three directions of the subspace with the largest variance).

3.2. Sequential Classification

We use the the probabilistic facial expression recognition procedure introduced in [3] but changing the features used by the ones described on section 5. It combines the prior information stored in the expression manifold with the incoming data obtained from a temporally ordered sequence of images of a face.

²<http://www.dia.fi.upm.es/~pcr/downloads.html>

Let $\mathbf{I}_1, \dots, \mathbf{I}_t$ be a temporally ordered image sequence of a face wearing one or more facial expressions and $\mathbf{f}_1, \dots, \mathbf{f}_t$ be the temporally ordered set of co-ordinates of the image sequence in the facial expression subspace, which we will denote $\mathcal{F}_{1:t}$. Let $G_t = \{g_1, g_2, \dots, g_c\}$ be a discrete random variable representing the facial expression at time t and F_t be a continuous random variable associated to the co-ordinates in the facial expression subspace (PCA+LDA subspace) of the image acquired at time t . We will denote by $P(g_i) \equiv P(G_t = g_i)$ the probability that the discrete random variable G_t takes value g_i and by $p(\mathbf{f}) \equiv p(F_t = \mathbf{f})$ the probability density function (p.d.f.) of the continuous variable \mathbf{f} at time t .

The facial expression $g(t)$ at time instant t is obtained as the maximum of the posterior distribution of G_t given the sequence of images up to time t

$$g(t) = \arg \max_i \{P(G_t = g_i | \mathcal{F}_{1:t})\}.$$

Alternatively, the facial expression may also be described as a probabilistic blending of the c primary facial expressions.

We will estimate the posterior distribution using a recursive Bayesian filter. For the first image in the sequence the problem can be immediately solved by $P(G_1 | \mathbf{f}_1) \propto p(\mathbf{f}_1 | G_1)P(G_1)$. Now, if we have a temporal sequence of images $\mathcal{F}_{1:t}$, and assume that measurements depend only on the current state and that our system is Markovian, then $P(G_t | \mathcal{F}_{1:t}) \propto p(F_t | G_t)P(G_t | \mathcal{F}_{1:t-1})$, where

$$P(G_t | \mathcal{F}_{1:t-1}) = \sum_{i=1}^c P(G_t | G_{t-1} = g_i)P(G_{t-1} = g_i | \mathcal{F}_{1:t-1}), \quad (4)$$

and $P(G_t | G_{t-1})$ is the expression transition probability.

In contrast to previous approaches (e.g. [4, 13]), which try to estimate the probability of transition between two facial expressions, we believe that all expression transitions are equally probable and use the following definition

$$P(G_t = g_j | G_{t-1} = g_i) = \begin{cases} h & \text{if } j = i \\ \frac{1-h}{c-1} & \text{if } j \neq i, \end{cases} \quad (5)$$

where $0 \leq h \leq 1$ is a smoothing parameter that controls how G_{t-1} influences the predictions about G_t . In our recognition system the parameter h acts as a forgetting factor. The closer h is to 1, the less we forget about the information provided by all previous images in the sequence. In extreme cases, when $h = 1$, all images in the sequence are taken into account, and when $h = \frac{1}{c}$, the recognition is performed exclusively on the basis of the last image acquired.

3.3. Estimating $p(F|G)$

The p.d.f. of an image \mathbf{I} with face expression co-ordinates \mathbf{f} when subject is wearing facial expression g_i is denoted by $p(\mathbf{f} | g_i)$. Our goal here is to estimate this p.d.f. from the data

in the facial expression manifold computed in section 3.1. We will use a k-nearest neighbour approach. Let k be the number of elements in the nearest neighbour set of \mathbf{f} , $k_i(\mathbf{f})$ the number of elements in the nearest neighbour set that belong to facial expression g_i ($k = \sum_{i=1}^c k_i(\mathbf{f})$) and n_i the number of samples in the manifold of facial expression g_i . Then

$$p(\mathbf{f} | g_i) = \frac{k_i^r(\mathbf{f})}{n_i \mathcal{V}(k)} \propto \frac{k_i^r(\mathbf{f})}{n_i},$$

where $\mathcal{V}(k)$ is the volume of the neighbourhood enclosing the k nearest neighbours and

$$k_i^r(\mathbf{f}) = \begin{cases} \eta & \text{if } k_i(\mathbf{f}) = 0, \\ k_i(\mathbf{f}) & \text{otherwise,} \end{cases}$$

is the regularised the number of elements in the nearest neighbour set that belong to facial expression g_i ($k = \sum_{i=1}^c k_i^r(\mathbf{f})$). Parameter $0 \leq \eta \leq 1$ models the amount of regularisation introduced for a facial expression with no neighbour. We use this regularised nearest-neighbour definition to avoid the so-called *veto effect*³.

4. Experiments

In this section we are going to explain the experiments carried out using the 322 selected sequences from the Cohn-Kanade database.

4.1. Quantitative evaluation

In each cross validation fold the sequences from one subject are tested against the model trained with all sequences from other subjects. We believe that this strategy, leave-one-subject-out, is better, to test for classifier generalisation ability, than the conventional k -fold cross validation, leave-one-sequence-out. Because in the leave-one-sequence-out scheme we still train the classifier using the remaining sequences for the same subject the information of the testing sequences is partially contained in the model.

The classifier described in section 3 has three parameters: smoothing parameter h , neighbourhood size k and η to avoid the veto effect. In the experiments, η is manually set to 0.3. We used an exhaustive a search procedure to select the best h in $\{1/6, 0.2, 0.4, 0.6, 0.8, 1\}$ and k in $\{1, 3, 5, \dots, 59\}$. Figure 6 shows the impact on the recognition of different classifier parameters. The upper-left panel shows the 3D view of errors w.r.t. k and h . The upper-right and lower-left panels display plots of errors v.s. h and k respectively. The gray level of points in the plot represents the impact on the errors from variation of the other parameter. The best parameter configuration, $h = 0.2$, $k = 31$, yields a recognition rate of 86.025% (see table 1).

³If there is a single image in the sequence, \mathbf{I}_r , such that $k_i(\mathbf{f}_r) = 0$, then $P(g_i | \mathcal{F}_{1:t}) = 0$, no matter what the values of this probability for all preceding time instants were.

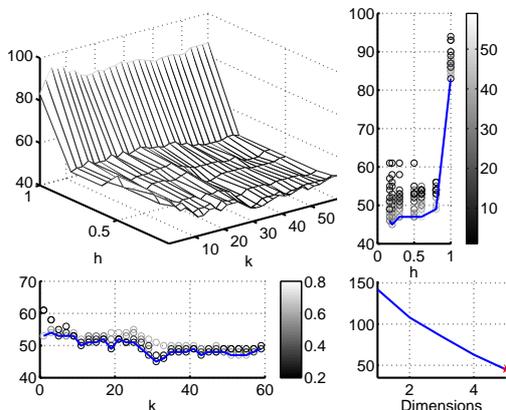


Figure 6. Recognition error curves for different classifier parameters.

	su	fe	jo	sa	di	an	total
su	100.00	0.00	0.00	0.00	0.00	0.00	
fe	0.00	80.00	7.87	6.12	6.52	3.03	
jo	0.00	5.71	87.64	2.04	0.00	3.03	
sa	0.00	8.57	1.12	79.59	8.70	9.09	
di	0.00	0.00	0.00	6.12	78.26	6.06	
an	0.00	5.71	3.37	6.12	6.52	78.79	
total							86.025

Table 1. Confusion matrix (expressed in percentage) for $h = 0.2$ and $k = 31$.

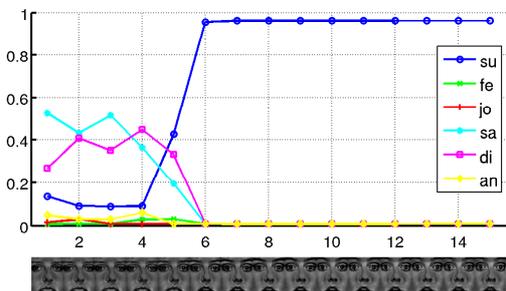


Figure 7. Example of expressions classification.

Figure 7 shows an example of the evolution of expression probabilities in an image sequence. In the first five frames, since the face is neutral, all facial expressions have a similar probability. As the expression evolves towards the apex, the probability of surprise grows closer to one.

4.2. Qualitative evaluation

In this section we analyse the performance of our system in a realistic image sequence (see Fig. 8) of a face moving and performing several facial expressions. In this case we have augmented the number of expressions with the neutral one. We have selected the classifier parameters $h = 0.6$ and $k = 39$, which produce the best classification results on the Cohn-Kanade database for seven expressions (including the neutral one).

The sequence starts with a neutral face until frame 57 in which we have a correctly detected surprise expression (see

Fig. 9) that finishes at frame 73. The tracker is re-started by the face detector (the blue rectangles in Fig. 8) whenever it detects a tracking failure (corresponding to a too high or too low value for equation (2)). This normally occurs during a very fast motion or an out of camera plane head rotation. There is another surprise expression from frame 80 to 87 that is also correctly detected. In frames 95 to 117 the user performs a fear expression that is also properly recognised. In frames 123 to 138 a fear-like expression is performed, but it is estimated as a surprise. This is due to the eyebrows not being frown and the mouth not displaying the exact shape of the fear expressions in the Cohn-Kanade database. A correctly detected smile is performed between frames 168 and 227. Between frames 272 and 277 the face is looking up, the lips are contracted, and illumination is uneven. This is the reason of getting a wrong anger expression estimation. In the interval of frames from 307 to 345 a disgust expression is performed. Only from frame 326 the disgust probability is greater than the neutral probability because of the mouth taking a shape closer to the training disgust expressions. One of the difficult expressions to recognise is sadness, because in our PCA+LDA space it is very close to anger. From frame 449 to 557 the user performs a sad expression that is either confused with the anger expression or not estimated at all. From frame 565 to 630 there are three surprise expressions all of them correctly identified. Finally, between frames 643 and 672, there is a fear-like expression which was not performed exactly as the training ones. It was consequently estimated as a mixture of surprise and fear.

5. Conclusions

In this paper we have presented a system that is able to track a face and classify the facial expression in real-time. The models used both in tracking and in classification are user independent. Our recogniser can be reconfigured, in the sense that it can be used to recognise any other set of expressions just by changing the training data.

It achieves an 86% recognition success on a set of 322 sequences from the Cohn-Kanade data base. Although our results are not as good as the best reported so far for this data base [16] (96%), we have a complete system which detects, tracks and recognises a facial expression in real-time.

Our system has several shortcomings that we are currently trying to solve. First, user and illumination independence is achieved by using as model an average face. The performance of our system degrades as the appearance of the user departs from the average face. We are currently considering the possibility of simultaneously reconstructing the appearance of the user and recognising the facial expression.

Also, our solution, like most previously introduced [4, 13, 16], follows a holistic approach. It uses information

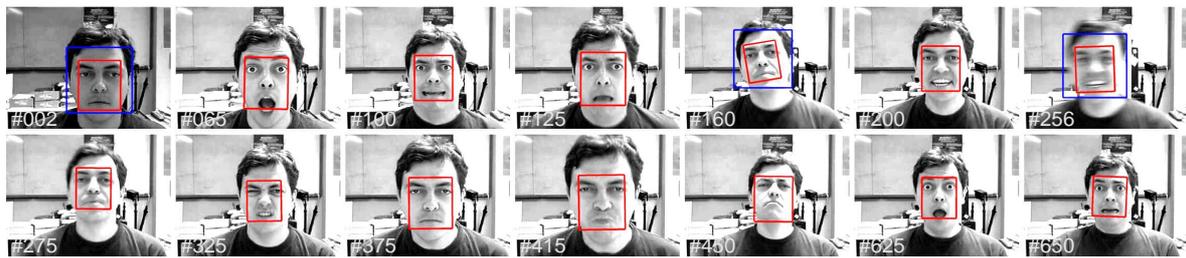


Figure 8. Tracking results for a realistic video.

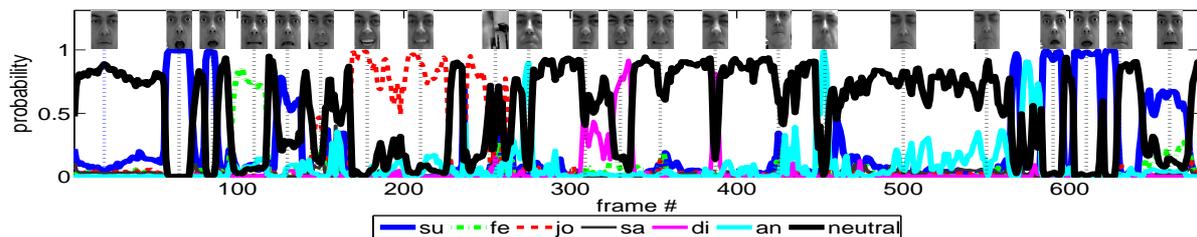


Figure 9. Classification results for a realistic video.

extracted from the whole face to classify the expression. This means that, to correctly recognise an expression, the combination of facial movements in all parts of the user's face must match those in the data base. Since all samples in the Cohn-Kanade data base follow quite a strict expression layout, the performance of the system apparently degrades when it is used by untrained users. So, future research venues must consider tracking and recognising facial expressions using motion data extracted from local face deformations.

References

- [1] R. Basri and D. W. Jacobs. Lambertian reflectance and linear subspaces. *Trans. on PAMI*, 25(2):218–233, February 2003. 2
- [2] J. Buenaposada, E. Muñoz, and L. Baumela. Efficient illumination independent appearance-based face tracking. *Image and Vision Computing*, (doi:10.1016/j.imavis.2008.04.015), 2008. 1, 2
- [3] J. Buenaposada, E. Muñoz, and L. Baumela. Recognising facial expressions in video sequences. *Pattern Analysis and Applications*, 11(1):101–116, 2008. 2, 3
- [4] Y. Chang, C. Hu, and M. Turk. Probabilistic expression analysis on manifolds. In *Proc. of CVPR*, volume 2, pages 520–527, 2004. 1, 2, 4, 5
- [5] P. Ekman. Strong evidence for universals in facial expressions: a reply to russell's mistaken critique. *Psychological Bulletin*, 115(2):268–287, 1994. 1
- [6] B. Fasel and J. Luetttin. Automatic facial expression analysis: a survey. *Pattern Recognition*, 36:259–275, 2003. 1
- [7] C. Hu, R. Ferris, and M. Turk. Active wavelet networks for face alignment. In *Proc. BMVC*, 2003. 2
- [8] T. Kanade, J. Cohn, and Y.-I. Tian. Comprehensive database for facial expression analysis. In *Proc. of International Conference on Automatic Face and Gesture Recognition*, pages 46–53, 2000. 1, 3
- [9] T. Ojala, M. Pietikainen, and T. Menpp. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *Trans. on PAMI*, 24(7):971–987, 2002. 2
- [10] N. Oliver, A. Pentland, and F. Bérard. Lafter: a real-time face and lips tracker with facial expression recognition. *Pattern Recognition*, 33:1369–1382, 2000. 1
- [11] M. Pantic and L. J. Rothkrantz. Automatic analysis of facial expressions: The state of the art. *Trans. on PAMI*, 22(12):1424–1445, December 2000. 1
- [12] M. Rosenblum, Y. Yacoob, and L. S. Davis. Human expression recognition from motion using radial basis function network architecture. *IEEE Trans. on Neural Networks*, 7(5):1121–1138, September 1996. 1
- [13] C. Shan, S. Gong, and P. W. McOwan. Dynamic facial expression recognition using a bayesian temporal manifold model. In *Proc. BMVC*, volume 1, pages 297–306, 2006. 1, 2, 4, 5
- [14] T. Sim, S. Baker, and M. Bsat. The cmu pose, illumination, and expression database. *Trans. on PAMI*, 25(12):1615–1618, 2003. 2
- [15] P. Viola and M. J. Jones. Robust real-time face detection. *International Journal of Computer Vision*, 57(2):137–154, May 2004. 2
- [16] G. Zhao and M. Piettikäinen. Dynamic texture recognition using local binary patterns with an application to facial expressions. *Trans. on PAMI*, 29(6):915–928, June 2007. 1, 5