Face analysis : identity *vs.* expressions

Hugo Mercier[1,2] Patrice Dalle[1]

[1]IRIT - Université Paul Sabatier
118 Route de Narbonne, F-31062 Toulouse Cedex 9, France
[2]Websourd – 3, passage André Maurois - BP 41125
31036 Toulouse Cedex 01, France
E-mail: {mercier,dalle}@irit.fr

Abstract

Facial images present important visual variations, due to several parameters. Here, we focus on the identity and expression parts. We study the hypothesis of (image-based) automatic separability of identity from expressions. Indeed, sign language speakers using videos need a tool able to offer anonymity to their sign productions and such a tool has to modify the part of the facial image carrying identity features without degrading expressive part, needed for comprehension. We present here models of the face space and how they can be used to anonymize facial images.

1 Introduction

On Internet, one of the preferred way for sign language speakers to communicate with each other is based on videos. In the current development of a video based web, able to offer sign language contents, a problem is raised concerning signed message anonymity, because a video record of a signer is indistinguishable from the signer's identity.

The most important part of identity is carried by the face, through its unique feature configurations. Facial expressions are an important part of signed languages, carrying a part of the meaning. Offering anonymity to a signer means degrading his identity, to thwart face recognition, without degrading his expression. Here, we propose to study how the face space is organized, with the help of a representative face image set. We use two models of the face space to test the separability hypothesis of the identity from expressions. We compare how these models are able to separate facial identity and expressions and we present an application of automatic facial "anonymization".

In a first section, we introduce the significance of facial expressions in sign language and the need of a face anonymizer able to separate identity and expressions. From a brief recall of past works, we introduce the models used and how they can be used to validate the separability hypothesis on a collection of facial images. We finally show examples of an automatic face anonymizer and conclude, giving directions for future works.

2 Facial expressions

A facial expression is here defined by the activation of a set of facial muscles. The activation of an isolated facial muscle, is called an *action unit*, as stated by Ekman [Ekman and Friesen, 1978]. The meaning associated with the expression is not important in this study. We study a set of facial expressions or isolated action units that seem to occur frequently in sign language, without meaning linked to.

2.1 In Sign Language

In french sign language (FSL), facial expressions are an important subset of the language. They take part in the message formation, at different levels:

- at a lexical level, where some signs are only discriminated by the use of facial expressions;
- at a syntactic level, where some expressions are syntactic markers, like those introducing modal forms;

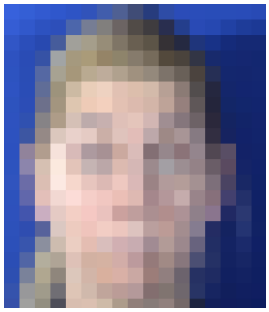


Figure 1. Typical anonymized face image where both identity and expressions are degraded.

- at a semantic level, where facial expressions reflect emotions of the played character (in situations where the signer takes a role, called personal transfer).

2.2 Anonymous videos of sign language

As no written form of sign language is widely used for now, deaf people are used to communicate via videos. A problem raised by the deaf community is how to communicate via video with the warranty of a certain level of anonymity ? Existing video effects treat the face as a whole, by the use of blur filters or mosaicking techniques. However, such methods are unusable with a video containing sign language, because both the identity and expressions of the signer are blurred (see Fig. 1).

The problem is stated as finding an image operator able to degrade only the identity part of a signing face and, at the same time, to leave expressions at a sufficient level of intelligibility.

Here we propose to degrade the identity part by changing it to another one, that may exist or not. From a class of n different identities, an identity id_i is replaced by another identity id_j or by an identity $\bar{id}$ corresponding to the average of all.

3 Past works

We present in this section a recall of what has been done in the literature concerning automatic analysis of facial expressions, separability of facial identity from expressions and how to render an anonymous face.

3.1 Automatic facial expression analysis

Automatic facial expression analysis is an emerging task in the field of computer vision and image processing. Most of the existing work aims at recognizing universal emotions listed by Ekman. It is thus a classification problem: which of the 7 emotions best matches the observed face ? It is an interesting problem, but applications are very limited: a precise context of communication involves expressions specific to this context, and most probably not universally-recognized.

On the other hand, expressions can be viewed as action unit combinations. The goal is thus to measure action units of a face along a video in an automatic way. Existing methods range from those treating the face as a combination of known facial features, each located and characterized by *ad-hoc* image processing operators to those treating the face as any other object, learning what a face is from labelled data.

Here, we use a paradigm called *Active Appearance Models* (AAM - [Cootes et al., 2001, Matthews and Baker, 2003]) where the face and its shape and appearance variations are learned from data. It offers a generic framework to address different applicative contexts : tracking of a previously known individual, including or not head rotations, including or not expressions. An AAM is defined by two sets of deformation vectors : a set of shape deformations and a set of appearance variations. These deformation vectors are learned from a training set. The broader the training set is, the more the AAM is able to model deformations.

3.2 Active Appearance Models

An active appearance model is used to describe an object which may present different visual configurations. It has originally been proposed by Cootes *et al.* [Cootes et al., 2001]. It is particularly used to model face and its deformation, either due to shape or appearance.

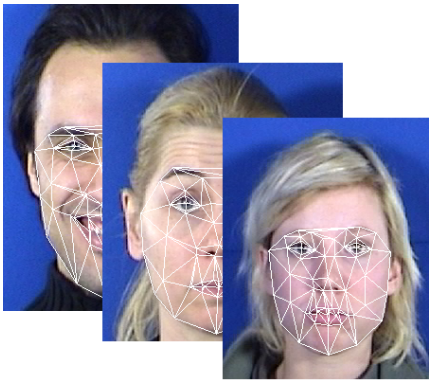


Figure 2. Typical training set used, a collection of facial feature points is manually labelled on each image (a triangulation is displayed here). From each image, a shape and an appearance are extracted.

.

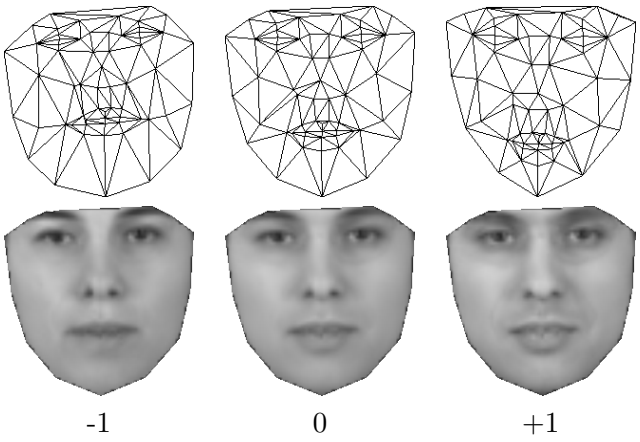


Figure 3. Sample variation modes of an AAM. The first shape deformation vector is on the first row. The second row contains an appearance variation vector. The left column shows generated faces with a negative weight applied to the deformation vector. The weight is positive on the right, and the zero-weighted column shows the mean face.

On a previously hand-labelled training set, a shape and an appearance are extracted from each image (see Fig. 2). Two orthonormal deformation basis (one for the shape and one for the appearance) are then computed by means of a principal component analysis (see Fig. 3).

These two basis form a face space, where all face of the training set may be represented. Such a model is used by fitting and tracking algorithms: it represents the search space. Fitting and tracking algorithms are built upon an optimization method. The goal is, given a first estimate of the observed face, to modify its shape and appearance in such a way that an error measure decreases. A typical error function is the pixel-wise difference between the current face estimate and the image.

The fitting algorithm need some iterations to converge in a configuration visually close to the observed face. The way the current face estimate evolves along time is the most prolific part of the method and is here out of scope.

At converge, it is thus possible to associate the observed face to a point in the face space. In the sequel, we call this point coordinates *face parameters*.

We suppose here that the problem of tracking facial deformations of a signer along a video is resolved. For that purpose, we rely on recent advances in the field of AAM fitting and tracking algorithms. Recent variants [Matthews and Baker, 2003, Baker et al., 2004] are able to track deformations in the presence of occlusions (which occur very often in sign language – the face can be occluded by hands) and some out of plan rotations. Consequently, all the results presented here assume an AAM has been fitted to the face we want to analyze.

3.3 Separability of identity from expressions

Face is known to be a visually varying object, due to many parameters, including illumination, pose, identity and expressions.

Abboud [Abboud, 2004] present methods able to separate identity from expressions. The goal is to generate new expressions from a given face. One of the proposed model, called *scalable linear model*, is used here.

Costen *et al.* [Costen et al., 2002] try to separate the different sources of face image variation : pose, illumination, identity and expressions. The

proposed model is mainly used to remove variation due to pose, illumination and expressions in order to improve face recognition process. Their proposed model, simplified to our two parameter case, is used here.

3.4 Anonymous rendering

To the authors' knowledge, the only work addressing face anonymity problem may be found in [Newton et al., 2005, Gross et al., 2005], where a method is proposed to warranty that a face will not be recognized by a face recognition software. A protocol aiming at checking if an image processing operator is able to thwart recognition software and preserve face details in the same time. The most effective operator used is the average operator. The problem of anonymized face with intelligible expressions is not addressed.

Works partly addressing this problem can be found in the facial animation field. Indeed, if a 3D model is able to mimic expressions of a real face without carrying information related to identity, it can be used as a face anonymizer. However, many of them use intrusive motion capture system or only rely on lip motion.

4 Experimental protocol

To validate the hypothesis of facial identity and expressions separability, we present here two models of what we called the face space and the data used with. From a comparison between these two models, we show examples of a possible face anonymizer.

4.1 Data used

We used a subset of the MMI Face Database [Pantic et al., 2005] as a training set for our experiments. The MMI Face Database contains a set of videos where numerous individuals each performs all of the known facial action units. We choose 15 action units that occur frequently in sign language :

- AU 2 et 4 for the eyebrows;
- AU 5, 6, 7, 43 and 44 for the eyes;
- AU 12, 15, 22 and 27 for the mouth;
- AU 36 and 27 for the tongue;
- AU 33 and 35 for the cheeks.

We choose 6 very different individuals from the database (3 women and 3 men from different ethnicities and one wearing glasses).

Each video contains an individual performing an isolated action unit or set of action units. We extracted from each video the frame corresponding to the peak of the muscle activation and manually labelled the frame with a mesh made of 61 points. The retained mesh is a subset of the one used in the MPEG4 SNHC model [MPEG Working Group on Visual, 2001].

From the labelled images, we computed an Active Appearance Model and obtained a set of shape deformation vectors and a set of appearance variation vectors. We retain enough vectors to explain 99% of the variance, *i.e.* 49 shape vectors and 65 appearance vectors.

4.2 Face space

Each face (an appearance and a shape) can be represented as a point in the face space. We projected each training face onto the face space. The first three dimensions (for illustration purpose) of the face space are plotted on Fig. 4.

We observe that each identity forms a dense cluster, meaning that between two faces, identity variation is more important than expression variation.

4.3 First model

From the observed face space, we propose here to rely on a linear model of the face (originally proposed in [Abboud, 2004]). Any face is modeled as a fixed identity part, independent of expressions and a weighted expression part, independent of the identity.

$$\boldsymbol{p}_{i,e} = \bar{\boldsymbol{p}}_i + r_e \boldsymbol{p}_e$$

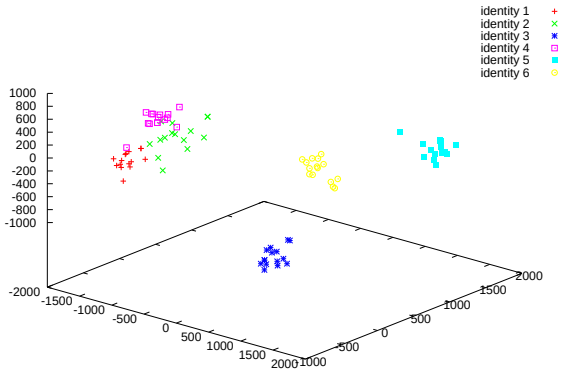


Figure 4. Visualization of the first three dimensions (for illustration purpose) of the face space.

All the $\boldsymbol{p}$ vectors are face parameters (either shape or appearance). The vector $\bar{\boldsymbol{p}}_i$ correspond to the face having identity id_i and a neutral expression. It is obtained by averaging all the faces of identity id_i over expressions. $\boldsymbol{p}_e$ is learned from the training set. The system to be resolved is :

$$\boldsymbol{P} = \boldsymbol{E}\boldsymbol{p}_e$$

The matrix $\boldsymbol{P}$ is formed by the stacking of the face parameters of all the n expressive faces and by the subtraction of the neutral faces. The corresponding row of $\boldsymbol{E}$ codes the amount of expression: 0 for none and 1 for an intensive expression. Here, we form the system by the face parameters of all expressive faces, took at their maximum activation, corresponding to an amount of 1.

The general solution is obtained with:

$$\boldsymbol{p}_e = (\boldsymbol{E}^{\boldsymbol{T}}\boldsymbol{E})^{-1}\boldsymbol{E}^{\boldsymbol{T}}\boldsymbol{P}$$

On a new face $\boldsymbol{q}$, assuming we know its corresponding neutral face $\bar{\boldsymbol{q}}$ and the expression displayed e, we can extract the amount of expression r_e and thus change the identity part to the identity id_j.

1. $r_e = (\boldsymbol{q} - \bar{\boldsymbol{q}})\boldsymbol{p}_e^+$
2. $\hat{\boldsymbol{q}} = \bar{\boldsymbol{p}}_j + r_e\boldsymbol{p}_e$

The major drawback of this model is that we need to know what is the expression displayed. It is only convenient for applications where we need to analyze separate expressions, like *emotional* expressions.

4.4 Second model

The previous model does not take mixed expressions into account. To address this problem, we propose to consider a face as being an identity part and a weighted sum of expressions:

$$\boldsymbol{p} = \bar{\boldsymbol{p}}_i + \sum_{i=1}^{m} e_i\boldsymbol{p}_{e_i}$$

Or, in matrix notation:

$$\boldsymbol{p} = \bar{\boldsymbol{p}}_i + \boldsymbol{P}\boldsymbol{e}$$

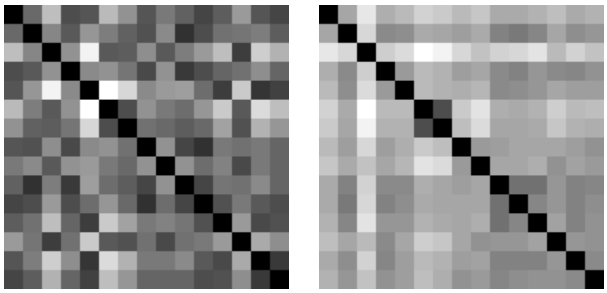
where $\boldsymbol{P}$ is a matrix containing all the $\boldsymbol{p}_{e_i}$ and $\boldsymbol{e}$ is a vector containing all the e_i.

Uniquely resolving this system need the addition of some constraints. We thus constrain the system in such a way that all the p_{e_i} are orthonormal. Consequently, it can be solved using principal component analysis, where we retain all the eigenvectors (see [Costen et al., 2002] for details).

On a new face $\boldsymbol{q}$, assuming we know its corresponding neutral face $\bar{\boldsymbol{q}}$, we can extract the amount of expression e and thus change the identity part to the identity id_j.

1. $e = \boldsymbol{P}^T(\boldsymbol{q} - \bar{\boldsymbol{q}})$
2. $\hat{\boldsymbol{q}} = \bar{\boldsymbol{p}}_j + \boldsymbol{P}\boldsymbol{e}$

The main drawback of this approach is that the extracted $\boldsymbol{p}_{e_i}$ may not represent a physical deformation. On a training set containing m different action units, we will get m vectors. Some of these vectors may represent a combination of action units. It is thus difficult to interpret what we observe. However, the interpretation is not needed when modifying a face for an anonymous rendering.



(a) (b)

Figure 5. Visual representation of the confusion matrix for (a) the first model and (b) the second model. The grey level on column i and row j represents the similarity between the expression ex_i and ex_j. The closer are two expressions, the darker is the cell.

(a) (b) (c)

Figure 6. Face anonymization illustration with the first model when it fails (a), when it succeeds (b) and with the second model (c). The first row contains the faces to anonymize, the results with the modified identity part are on the second row.

4.5 Comparison

To test the ability of each model to separate identity and expressions, we compare how they can uniquely extract expression information of a face: if, for a given expressive face, the expression extracted can be confounded with the one extracted from another expressive face, the model is unable to distinguish between identity and expression.

For a given face of identity id_i and expression ex_j, the expression parameters are extracted using each model. It is then compared (with a distance measure) to the faces of all others expressions ex_k. The distances are averaged across all the identities id. The result is a confusion matrix containing the similarity between all the expression pairs.

The confusion matrix of the two models are showed on Fig. 5. As it can be observed, the confusion matrix of the second model is more homogeneous than the first and close to the ideal confusion matrix, entirely filled with white except on its black diagonal. Except expressions 6 and 7 which are hard to distinguish, because they are visually hard to distinguish, the second model is much more able to discriminate expressions.

For a qualitative comparison, examples of anonymized faces are shown on Fig. 6. The first model tends to reconstruct "averaged" expressions, while the second tends to reconstruct an expression more specific to an identity.

5 Conclusion and future works

We have presented an experimental protocol aiming at verifying the hypothesis of facial identity and expressions separability. From a training set of previously hand-labelled facial images containing some identities and a set of facial expressions that often occur in sign language, we applied a separation of identity and expressions, based on two models of the face space. We compared how each model is able to distinguish expressions. The second model proposed here is able to separate identity from expressions and can be used as an image processing based face anonymizer, assuming the neutral face is known.

We used the uniqueness of extracted expression as a comparison criterion. An efficient face anonymizer has also to warranty the inability of identity recognition. In particular, the second model used here tends to generate expressions specific to an identity. Such visual features could be used to recognize the original identity. We plan to include face recognition methods in the comparison criterion. Moreover, because it is known

that motion helps the recognition task, results of anonymization have to be tested on an image sequence.

6 Acknowledgement

Portions of the research in this paper use the MMI Facial Expression Database collected by M. Pantic & M.F. Valstar.

References

[Abboud, 2004] Abboud, B. N. (2004). *Analyse d'expressions faciales par modèles d'apparence*. PhD thesis, Université Technologique de Compiègnes.

[Baker et al., 2004] Baker, S., Gross, R., and Matthews, I. (2004). Lucas-kanade 20 years on: A unifying framework: Part 4. Technical Report CMU-RI-TR-04-14, Robotics Institute, Carnegie Mellon University, Pittsburgh, PA.

[Cootes et al., 2001] Cootes, T. F., Edwards, G. J., and Taylor, C. J. (2001). Active appearance models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23:681–685.

[Costen et al., 2002] Costen, N., Cootes, T. F., Edwards, G. J., and Taylor, C. J. (2002). Automatic extraction of the face identity-subspace. *Image Vision Comput.*, 20:319–329.

[Ekman and Friesen, 1978] Ekman, P. and Friesen, W. V. (1978). *Facial Action Coding System (FACS) : Manual*. Palo Alto: Consulting Psychologists Press.

[Gross et al., 2005] Gross, R., Airoldi, E., Malin, B., and Sweeney, L. (2005). Integrating utility into face de-identification. In *Workshop on Privacy-Enhanced Technologies*.

[Matthews and Baker, 2003] Matthews, I. and Baker, S. (2003). Active appearance models revisited. Technical Report CMU-RI-TR-03-02, Robotics Institute, Carnegie Mellon University, Pittsburgh, PA.

[MPEG Working Group on Visual, 2001] MPEG Working Group on Visual (2001). International standard on coding of audio-visual objects, part 2 (visual), ISO/IEC 14496-2:2001.

[Newton et al., 2005] Newton, E., Sweeney, L., and Malin, B. (2005). Preserving privacy by de-identifying facial images. *IEEE Transactions on Knowledge and Data Engineering*, 17:232–243.

[Pantic et al., 2005] Pantic, M., Valstar, M. F., Rademaker, R., and Maat, L. (2005). Web-based database for facial expression analysis. In *Proceddings of the IEEE International Conference on Multimedia and Expo (ICME'05)*.