

Discrete Choice Models for Static Facial Expression Recognition

Gianluca Antonini¹, Matteo Sorci¹, Michel Bierlaire², and Jean-Philippe Thiran¹

¹ Ecole Polytechnique Federale de Lausanne, Signal Processing Institute
Ecublens, 1015 Lausanne, Switzerland

{Matteo.Sorci, Gianluca.Antonini, JP.Thiran}@epfl.ch

² Ecole Polytechnique Federale de Lausanne, Operation Research Group
Ecublens, 1015 Lausanne, Switzerland
Michel.Bierlaire@epfl.ch

Abstract. In this paper we propose the use of Discrete Choice Analysis (DCA) for static facial expression classification. Facial expressions are described with expression descriptive units (EDU), consisting in a set of high level features derived from an active appearance model (AAM). The discrete choice model (DCM) is built considering the 6 universal facial expressions plus the neutral one as the set of the available alternatives. Each alternative is described by an utility function, defined as the sum of a linear combination of EDUs and a random term capturing the uncertainty. The utilities provide a measure of likelihood for a combinations of EDUs to represent a certain facial expression. They represent a natural way for the modeler to formalize her prior knowledge on the process. The model parameters are learned through maximum likelihood estimation and classification is performed assigning each test sample to the alternative showing the maximum utility. We compare the performance of the DCM classifier against Linear Discriminant Analysis (LDA), Generalized Discriminant Analysis (GDA), Relevant Component Analysis (RCA) and Support Vector Machine (SVM). Quantitative preliminary results are reported, showing good and encouraging performance of the DCM approach both in terms of recognition rate and discriminatory power.

1 Introduction

Facial expressions are probably the most visual method to convey emotions and one of the most powerful means to relate to each other. An automatic system for the recognition of facial expressions is based on a representation of the expression, learned from a training set of pre-selected meaningful features. For unseen expressions, the corresponding representation has to be associated with the correct expression. In this process, two are the key tasks: the choice of the set of features representing the expression and the choice of the classification rule. In [1] the author focuses on optical flow analysis for feature extraction, in order to model muscle activities and estimating the displacements of salient points. This is a dynamic approach, where temporal information is used both in the feature extraction and classification steps, the last performed through an Hidden Markov Models (HMM) scheme. Gabor wavelet based filters have been used in

[2], in order to build templates for facial expressions, over multiple scales and different orientations. Template-based matching is used in order to associate an observed feature vector with the corresponding expression, in a static context. Statistical generative models such as principal and independent component analysis (PCA, ICA) are used in [3] and [4], in order to capture meaningful statistics of face images. Neural Networks (NN) and HMMs are used for the classification step, respectively in static and dynamic frameworks. Recent years have seen the increasing use of feature geometrical analysis ([5, 6]). The Active Appearance Model (AAM, see [7]) is one of these techniques which elegantly combines shape and texture models, in a statistical framework, providing as output a mask of face landmarks.

The contribution of this work is twofold. First, we propose Discrete Choice Models for expression classification. These models have been recently introduced in the computer vision community by [9], in the context of pedestrian modeling and tracking. DCMs are econometric models designed to forecast the behavior of individuals in choice situations, when the set of available alternatives is finite and discrete. In this context, the logic behind the use of DCMs is to model the choice process representing the human observer labelling procedure. The DCM classifier is compared with several other classification methods: LDA, GDA, RCA and SVM. The LDA is a supervised discriminative method to produce the optimal linear classification function. It transforms the data into a lower-dimensional space where it is decided, according to some chosen metric, to which class a given sample x belongs. The GDA is the kernel-based version of the LDA. RCA is a method that seeks to identify and down-scale global unwanted variability within the data. The method performs a projection of the input data into a feature space by means of a linear transformation. In the transformed space, a nearest neighbor classification based on the Euclidean distance is used, in order to assign the new sample to a class (see [10]). Second, we propose a set of *Expression Descriptives Units* (EDU) for static expression representation. They are derived from a set of 55 face landmarks, obtained using an AAM model. The EDUs represent intuitive descriptors of the facial components (eyebrows, eyes, nose and mouth) and the mutual interactions between them. They have been derived taking inspiration from the Facial Action Unit Coding System (FACS) [8] which is a human-observer based system designed to detect subtle changes in facial features. FACS itself is purely descriptive, uses no emotion or other inferential labels and provides the necessary ground-truth with which to describe facial expression. On the other hand, FACS require a huge set of salient facial points, and for most of them a tracking step is required, in order to capture variations over time. EDUs can be considered as a more compact and static counterpart of the FACS.

The paper is structured as follows: in Section 2 we review the AAM and introduce the DCM theory. In Section 3 a detailed description of the utility functions is given along with the EDU description and the results of the learning process. We finally report the experiments and a description of the data used to compare the different classifiers with our approach in Section 4. Conclusions and future works are finally reported in Section 5.

2 Background

2.1 Active Facial Appearance Model

The AAM is a statistical method for matching a combined model of shape and texture to unseen faces. The combination of a model of shape variation with a model of texture variation generates a statistical appearance model. The model relies on a set of annotated images. A training set of images is annotated by putting a group of landmark points around the main facial features, marked in each example. The shape is represented by a vector \mathbf{s} brought into a common normalized frame -w.r.t. position, scale and rotation- to which all shapes are aligned. After having computed the mean shape $\bar{\mathbf{s}}$ and aligned all the shapes from the training set by means of a Procrustes transformation, it is possible to warp textures from the training set onto the mean shape $\bar{\mathbf{s}}$, in order to obtain shape-free patches. Similarly to the shape, after computing the mean shape-free texture $\bar{\mathbf{g}}$, all the textures in the training set can be normalized with respect to it by scaling and offset of luminance values. PCA is applied to build the statistical shape and textures models:

$$\mathbf{s}_i = \bar{\mathbf{s}} + \Phi_s \mathbf{b}_{si} \quad \text{and} \quad \mathbf{g}_i = \bar{\mathbf{g}} + \Phi_t \mathbf{b}_{ti} \quad (1)$$

where \mathbf{s}_i and \mathbf{g}_i are, respectively, the synthesized shape and shape-free texture, Φ_s and Φ_t are the matrices describing the modes of variation derived from the training set, b_{si} and b_{ti} the vectors controlling the synthesized shape and shape-free texture. The unification of the presented shape and texture models into one complete appearance model is obtained by concatenating the vectors b_{si} and b_{ti} and learning the correlations between them by means of a further PCA. The statistical model is then given by:

$$\mathbf{s}_i = \bar{\mathbf{s}} + \mathbf{Q}_s \mathbf{c}_i \quad \text{and} \quad \mathbf{g}_i = \bar{\mathbf{g}} + \mathbf{Q}_t \mathbf{c}_i \quad (2)$$

where \mathbf{Q}_s and \mathbf{Q}_t are the matrices describing the principal modes of the combined variations in the training set and \mathbf{c}_i is the appearance parameters vector, allowing to control simultaneously both shape and texture. Fixing the parameters \mathbf{c}_i we derive the shape and the shape-free texture vectors using equations (2). A full reconstruction is given by warping the generated texture into the generated shape. In order to allow pose displacement of the model, other parameters must be added to the appearance parameters \mathbf{c}_i : the pose parameters p_i . The matching of the appearance model to a target face can be treated as an optimization problem, minimizing the difference between the synthesized model image and the target face [7].

2.2 Discrete Choice Models

Discrete choice models are known in econometrics since the late 50's. They are defined to describe the behavior of people in choice situations, when the set of available alternatives is finite and discrete (choice set). They are based on the concept of *utility maximization* in economics, where the decision maker is assumed to be *rational*, performing

a choice in order to maximize the utilities she perceives from the alternatives. The alternatives are supposed to be mutually exclusive and collectively exhaustive, while the rationality of the decision maker implies transitive and coherent preferences.³ The utility is a *latent* construct, which is not directly observed by the modeler, and is treated as a random variable. The discrete choice paradigm well matches the labelling assignment process in a classification task. This approach can be interpreted as an attempt to model the decision process performed by an hypothetical human observer during the labelling procedure for the facial expressions.

Given a population of N individuals, the (random) utility function U_{in} perceived by individual n from alternative i , given a choice set C_n , is defined as follows:

$$U_{in} = V_{in} + \varepsilon_{in} \quad (3)$$

with $i = 1, \dots, J$ and $n = 1, \dots, N$. V_{in} represents the deterministic part of the utility, which is a function of alternatives' attributes and socio-economic characteristics of the decision maker. In the context of this paper, we only deal with attributes of the alternatives, represented by combinations of the chosen features. The ε_{in} term is a random variable capturing the uncertainty. Under the utility maximization assumption, the output of the model is represented by the choice probability that individual n will choose alternative i , given the choice set C_n . It is given by:

$$P_n(i|C_n) = P_n(U_{in} \geq U_{jn}, \forall j \in C_n, j \neq i) = \int_{\varepsilon_n} I(\varepsilon_n < V_{in} - V_{jn}, \forall j \in C_n, j \neq i) f(\varepsilon_n) d\varepsilon_n \quad (4)$$

where $\varepsilon_n = \varepsilon_{jn} - \varepsilon_{in}$ and $I(\cdot)$ is an indicator function which is equal to 1 when its argument is satisfied, zero otherwise. Based on Equation 4, in order to define the choice probability, only the difference between the utilities matters. The specification of the utility functions represents the modeler's mean to add her prior knowledge on the choice process (a similar interpretation of the decision theoretic approach can be found in [11]). In this sense, the DCM approach is similar to graphical probabilistic models, such as belief networks and random fields, where the graph topology embeds the prior knowledge, helping designing causal relationships. Different DCMs are obtained making different assumptions on the error terms. A family of models widely used in literature are the GEV (Generalized Extreme Value) models, introduced by [12]. GEV models provide a closed form solution for the choice probability integral, allowing at the same time for a certain flexibility in designing the variance/covariance structure of the problem at hand (i.e., several correlation patterns between the alternatives can be explicitly captured by these models). Assuming the error terms being multivariate type I extreme value distributed⁴, the general expression of the GEV choice probability for

³ Transitive preferences means that if alternative i is preferred to alternative j which is preferred to alternative k , then alternative i is also preferred to k . Coherent preferences means that the decision maker will make the same choice in exactly the same conditions.

⁴ The main reasons for the choice of this kind of distribution derive from its good analytical properties. More details can be found in [13]

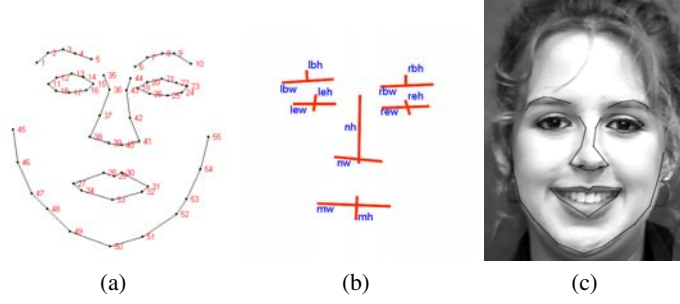


Fig. 1. a)Facial landmarks (55 points);b)Facial components descriptors;c)Expressions Descriptive Units

a given individual to choose alternative i , given a choice set C with J alternatives, is as follows:

$$P(i|C) = \frac{e^{V_i + \log G_i(y_1, \dots, y_J)}}{\sum_{j=1}^J e^{V_j + \log G_j(y_1, \dots, y_J)}} \quad (5)$$

where $y_i = e^{V_i}$ and $G_i = \frac{\partial G}{\partial y_i}$. The function G is called *generating function* and it captures the correlation patterns between the alternatives. Details about the mathematical properties of G are reported in [12] (differentiable and homogeneous of degree $\mu > 0$, among the others). Several GEV models can be derived from Equation 5, through different specifications of the generating function. In this paper we use a Multinomial Logit Model (MNL), which is largely the simplest and most used discrete choice model in literature. It is obtained assuming the following G function, which implies no correlations between the alternatives:

$$G(y_1, \dots, y_J) = \sum_{j \in C} y_j^\mu \quad (6)$$

where μ is a positive scale parameter. Under these assumptions, the MNL choice probability is given by the following expression

$$P(i|C) = \frac{e^{\mu V_i}}{\sum_{j \in C} e^{\mu V_j}} \quad (7)$$

3 DCM and facial expression classification

3.1 Expressions Descriptive Units

The use of AAM allows to detect facial components in a face (an example is shown in Figure 1(c)). Figure 1(a) shows the 55 landmarks used to build the AAM model, while Figure 1(b) shows the descriptors we use for the facial components: eyes, eyebrows, nose and mouth. These descriptors represent the width and the height of each

EDU1	$\frac{lew+rew}{leh+reh}$	EDU8	$\frac{leh+reh}{lbh+rbh}$
EDU2	$\frac{lbw}{lbh}$	EDU9	$\frac{lew}{nw}$
EDU3	$\frac{rbw}{rbh}$	EDU10	$\frac{nw}{mw}$
EDU4	$\frac{mw}{mh}$	EDU11	EDU2 / EDU4
EDU5	$\frac{nh}{nw}$	EDU12	EDU3 / EDU4
EDU6	$\frac{lew}{mw}$	EDU13	EDU2 / EDU10
EDU7	$\frac{leh}{mh}$	EDU14	EDU3 / EDU10

Table 1. Expressions Descriptive Units

Expressions	edu1	edu2	edu4	edu5	edu6	edu7	edu8	edu10	edu11	edu13	edu14
anger	✓			✓	✓					✓	✓
disgust	✓			✓					✓	✓	✓
fear			✓	✓	✓			✓			✓
happiness	✓		✓	✓			✓	✓			
neutral			✓								
sadness		✓					✓				
surprise			✓			✓					

Table 2. Utility functions: each row corresponds to an expression while the columns are the EDUs included in the utilities.

facial component. In order to give a useful representation of the expression in terms of interactions among those descriptors, we define a set of EDUs, reported in Table 1. The first 5 EDUs represent, respectively, the eccentricity of eyes, left and right eyebrows, mouth and nose. The EDUs from 7 to 9 represent the eyes interactions with mouth and nose, while the 10th EDU is the nose-mouth relational unit. The last 4 EDUs relate the eyebrows to mouth and nose. Differently from other approaches [14, 10] that use the combined AAM vector parameters as facial features, in our framework the 14 EDUs represent the features describing the face. The intuitive interpretation and the reduced number of dimensions make of the EDUs a valid set of descriptors for facial expressions.

3.2 The model

The utility functions are specified using a linear-in-parameters form, combining the expression descriptive units. Each EDU in each utility is weighted by an unknown deterministic coefficient, that has to be estimated. The choice for a linear form is based purely on simplicity considerations, in order to reduce the number of parameters in the estimation process. The general form of the utilities is given by:

$$U_i = \alpha_i + \sum_{k=1}^K I_{ki} \beta_{ki} \text{EDU}_k \quad (8)$$

where $i = 1, \dots, C$ with $C = 7$ is the number of expressions, $K = 14$ is the number of EDUs, I_{ki} is an indicator function equal to 1 if the k -th EDU is included in the utility for

expression i and 0 otherwise, β_{ki} is the weight for the k -th EDU in alternative i and α_i is an alternative specific constant. The α_i coefficients represent the average value of the unobserved part of the corresponding utility and one of them has to be normalized to 0, in order to be consistent with DCM theory (see [13]). In our case, we normalize with respect to the neutral expression. We summarize in Table 2 what are the EDUs included in the different utilities, i.e. when the $I_{ki} = 1$. Table 2 shows how, during the model specification step, we are free to customize the utilities of the different expressions. This flexibility represents the strength of DCMs; note that the utility expressions reported here are the result of a strong iterative process, where several hypothesis have been tested and validated, starting from a uniform expression for every alternative, including all the EDUs. In the final utility functions, only the EDUs corresponding to statistically significant parameters (t -test statistic against the zero value) are reported, resulting in a final model with 31 unknown parameters (6 α_i and 25 β_{ki}). The parameters have been estimated by maximum likelihood estimation, using the Biogeme package [15]. Biogeme is a freeware, open source package available from roso.epfl.ch/biogeme. It performs maximum likelihood estimation and simulated maximum likelihood estimation of a wide class of random utility models, within the class of mixtures of Generalized Extreme Value models (see [16] for details). The maximization is performed using the CFSQP algorithm (see [17]), using a Sequential Quadratic Programming method. Note that such nonlinear programming algorithms identify local maxima of the likelihood function. We performed various runs, with different starting points (a trivial model with all parameters to zero, and the estimated value of several intermediary models). They all converged to the same solution. Most of the estimated utility parameters are significantly different from zero. Classification is performed running the learned model on the test set, using the BioSim package (available at the same address as Biogeme). BioSim performs a sample enumeration on the test data, providing for each of them the utilities and the choice probabilities for each expression in the choice set. The classification rule consists in associating each sample with the alternative having the maximum probability, whose equation is reported in (7). We can state the classification rule used here as a *soft max* principle based on an *entropy maximization* criterion ([18]). However, such a (only formal) 'equivalence' arises on the base of the specific form of the GEV probability equation. Other discrete choice methods exist (Probit, Logit Kernel, [16, 19]) whose choice probabilities cannot be expressed by an analytical solution, leading to a more general soft max classification scheme, not related with the maximum entropy principle.

Learning results The learned parameters show important consistencies with the common reading of facial expressions in terms of facial component modifications. For space reasons, we report in Table 3 only a subset of β_{ki} estimates.

The parameters β_{5a} represents the coefficient of the 5th EDU (nose eccentricity) for the anger alternative. Its positive value shows a positive impact on the respective utility. It means that increasing nose eccentricity corresponds to higher utilities for the anger alternative. Looking at the definition of this EDU, this is in line with our expectations, showing that for an anger expression the nose width increases while its height decreases, with respect to the neutral expression (the reference one in our model). The parameters β_{1a} represents the eye eccentricity (1st EDU) for the anger expression. A similar inter-

β_{ki}	estimate	t test 0
β_{5a}	+ 1.238	+ 4.298
β_{1a}	+ 2.067	+ 2.018
β_{10f}	- 14.69	- 1.871
β_{10h}	- 42.64	- 3.440
Sample size = 143		
Number of estimated parameters = 30		
Null log-likelihood = - 278.265		
Final log-likelihood = - 88.317		
Likelihood ratio test = 379.896		
$\bar{\rho}^2 = 0.575$		

Table 3. MNL Part of the estimation results

Expressions	Training images	Test images
Neutral	26	15
Happiness	20	18
Surprise	21	20
Fear	18	11
Anger	18	17
Disgust	22	17
Sadness	18	17

Table 4. Number of images in the classification training and test set

Classifiers	Classification Rate(%)
DCM	78.261
SVM	76.522
RCA	70.435
LDA	49.565
GDA	62.609

Table 5. Classification rates

pretation holds for this coefficient, in line with observations: the eye movement leads to a lower eye’s height and a higher eye’s width, with respect to the reference alternative. The other two parameters relate the nostrils width with the mouth width. Their negative sign induces a negative impact on the utilities of fear and happiness. This is coherent with the data, where for these two expressions we note a characterizing increase in the mouth width, leading to a decreasing nostril-mouth interaction parameter. The coefficient estimates are significantly different from zero at 95%, with the exception of the β_{10f} (significant at 90%). We finally report some interesting statistics. The log-likelihood corresponding to a trivial model (all the coefficients equal to zero) is consistently increased after the estimation process, rising its value from -278.265 to -88.317. The likelihood ratio test and the $\bar{\rho}^2$ coefficient are also reported, showing the good fitting of the estimated model.

4 Experiments

In order to test the proposed approach we use the Cohn-Kanade Database [20]. The database consists of expression sequences of subjects, starting from a neutral expression

DCM/SVM	happiness	surprise	fear	anger	disgust	sadness	neutral	Overall(%)
happiness	16/16	0/0	2/1	0/0	0/1	0/0	0/0	88.88/88.88
surprise	0/0	19/19	0/0	1/0	0/0	0/0	0/1	95.00/95.00
fear	1/4	0/0	7/4	2/2	0/0	1/0	0/1	63.64/36.36
anger	0/0	0/0	1/2	10/9	3/1	2/3	1/2	58.82/52.94
disgust	0/1	0/0	0/0	2/2	12/12	0/1	3/1	70.58/70.58
sadness	0/0	0/0	0/0	2/0	0/1	15/14	0/2	88.24/82.35
neutral	0/1	0/0	2/0	1/0	0/0	1/0	11/14	73.34/93.33

Table 6. DCM and SVM confusion matrices

and ending most of the time in the peak of the facial expression. There are 104 subjects in the database, but only for few of them the six expressions are available. From the database we extrapolate 3 data sets:

- *AAM training set*: it consists of 300 images from 11 different subjects; it is composed by 48 neutral images and 42 images for each of the 6 primary emotions.
- *Classifiers training and test set*: they consist respectively of 143 and 115 appearance masks, as reported in Table 4.

The appearance model is built using 49 shapes modes and 140 texture modes leading to 84 appearance modes, capturing the 98% of the combined shape and texture variation. The shape-free texture vector \mathbf{g} is composed of 38310 pixels and the shape vector dimension is 55. Concerning the implementation, we use the AAM C++ code available at <http://www2.imm.dtu.dk/~aam/>. The classifiers used in the comparison procedure all share the same training and test sets. Their input consists in the EDUs built on the matched appearance masks. For the other classifiers we implemented the related state of the art. The SVM (using **libsvm** with radial basis functions) and DCM classifiers have been tuned on the common training set. The test experiments reported in Table 5, although preliminaries, could be interpreted as a better transferability of the learned DCM model over unseen samples. The intuition explaining this behaviour could lie on the more flexible hypotheses at the base of the DCM approach. The verification of this intuition will be part of our further investigations.

With the exception of RCA, the performance of the other nearest neighbor classifiers are significantly lower than SVM and DCM. For this reason we report in Table 6 only the confusion matrices for the two best performing methods.

A second empiric measure for classifiers performance comparison, related only to their discriminatory power and not to the recognition rates, has been computed. It is described in [21] and we report here a short explanation. For the various methods to be compared, let $m(wc)$ the mean probability assigned to well classified samples and $std(wc)$ the relative standard deviation. Similarly, let $m(bc)$ and $std(bc)$ the same values for the bad classified samples.⁵ Good classification and bad classification thresholds are defined as:

⁵ For the nearest neighbor classifiers the sample distances from the classes have been normalized, in order to sum up to one.

Classifiers	opp
DCM	- 6.92
SVM	- 3.60
RCA	- 3.64
LDA	- 4.51
GDA	- 3.27

Table 7. opp values for the compared classifiers

$$gctr = m(wc) + \text{std}(wc) \quad bctr = m(bc) - \text{std}(bc)$$

Based on these values, an overall performance parameter is defined as:

$$opp = \frac{gctr - bctr}{m(bc) - m(gc)} \quad (9)$$

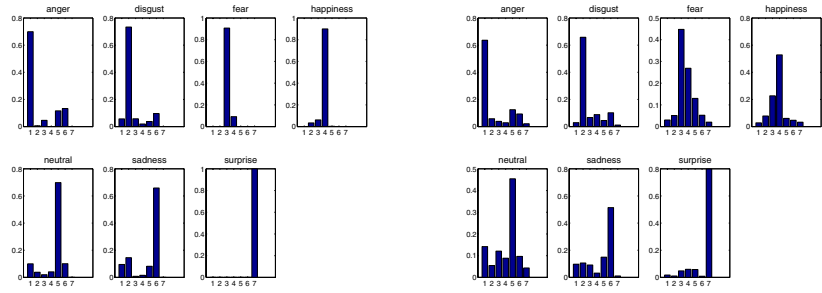
measuring how well a classifier discriminates. For a robust method we expect an opp value as low as possible. If the value of opp is negative, the *gctr* and *bctr* thresholds are well separated. In Table 7 we report the opp values for the tested classifiers. In order to visualize the discriminatory power for DCM and SVM, it is worth to show, for each set of correctly classified samples, the mean values of the probabilities assigned to each of the other classes, as shown in Figure 2. The plots confirm the better class discrimination performed by the DCM classifier, resulting in a sharper shape of the output probabilities.

5 Conclusions and future works

In this paper we propose a new classifier based on discrete choice analysis and a set of expression descriptive units for facial expression representation. Both the feature set and the modeling approach are motivated by the research of methods able to bring into the process the modeler prior knowledge. The set of the proposed EDUs is suitable because intuitively related to static expressions, describing the salient facial components and their mutual interactions. The DCM modeling approach redefines facial expression classification as a discrete choice process, which well matches the human observer labelling procedure. Prior knowledge can be included in the process customizing the utilities. The result of the DCM is a set of probabilities assigned to the alternatives, represented by the possible expressions. The one with the maximum probability is chosen for classification. We compared the DCM classifier with several other methods, finding that only the SVM has comparable performance. However, the more flexible properties of DCM lead to better results of this approach both in terms of classification rate on new data and discriminatory power. We are currently working to include in the model both the expression dynamics and the variation in a population of individuals performing the labelling task. Based on our experience, we think that a subjective component biases the labelling process, requiring a detailed statistical analysis on collected data from an heterogeneous population of human observers. DCMs, coming from econometric, provide a strong statistical framework to include such a heterogeneity.

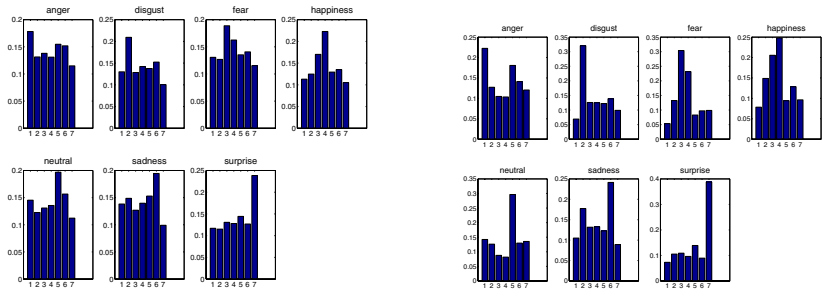
References

1. Lien, J.: Automatic recognition of facial expressions using hidden markov models and estimation of expression intensity (1998)
2. Ye, J., Zhan, Y., Song, S.: Facial expression features extraction based on gabor wavelet transformation. In: IEEE International Conference on Systems, Man and Cybernetics. (2004) 10–13
3. Padgett, C., Cottrell, G.: Representing face images for emotion classification. MIT Press, Cambridge, MA (1997)
4. Bartlett, M.: Face image analysis by unsupervised learning and redundancy reduction (1998)
5. Cootes, T.F., Edwards, G.J., Taylor, C.J.: Active appearance models. Pattern Analysis and Machine Intelligence, IEEE Transactions on **23** (2001) 681–685
6. Lanitis, A., Taylor, C.J., Cootes, T.F.: Automatic interpretation and coding of face images using flexible models. IEEE Trans. Pattern Anal. Mach. Intell. **19** (1997) 743–756
7. Stegmann, M.B.: Active appearance models: Theory, extensions and cases. Master's thesis, Informatics and Mathematical Modelling, Technical University of Denmark, DTU, Richard Petersens Plads, Building 321, DK-2800 Kgs. Lyngby (2000)
8. Ekman, P., Friesen, W.V.: Facial Action Coding System Investigator's Guide. Consulting Psychologist Press, Palo Alto, CA (1978)
9. Antonini, G., Venegas, S., Bierlaire, M., Thiran, J.P.: Behavioral priors for detection and tracking of pedestrians in video sequences. To appear in International Journal of Computer Vision (2005)
10. Sorci, M., Antonini, G., Thiran, J.P.: Relevant component analysis for static facial expression recognition. Technical Report TR_ITS_2005.33, Signal Processing Institute, Ecole Polytechnique Federale de Lausanne (2005)
11. Horvitz, E.J., Breese, J.S., Henrion, M.: Decision theory in expert systems and artificial intelligence. International Journal of Approximate Reasoning **2** (1988) 247–302
12. McFadden, D.: Modelling the choice of residential location. In A. Karlquist *et al.*, ed.: Spatial interaction theory and residential location, Amsterdam, North-Holland (1978) 75–96
13. Ben-Akiva, M.E., Lerman, S.R.: Discrete Choice Analysis: Theory and Application to Travel Demand. MIT Press, Cambridge, Ma. (1985)
14. Abboud, D., Davoine, F.: Appearance factorization based facial expression recognition and synthesis. In: ICPR (4). (2004) 163–166
15. Bierlaire, M.: BIOGEME: a free package for the estimation of discrete choice models. In: Proceedings of the 3rd Swiss Transportation Research Conference, Ascona, Switzerland (2003) www.strc.ch.
16. Train, K.: Discrete Choice Methods with Simulation. Cambridge University Press, University of California, Berkeley (2003)
17. Lawrence, C.T., Zhou, J.L., Tits, A.: A c code for solving (large scale) constrained nonlinear (minimax) optimization problems, generating iterates satisfying all inequality constraints. Technical Report TR-94-16rl, Institute for Systems Research, University of Maryland, College Park, MD 20742 (1997)
18. Cover, T.M., Thomas, J.A.: Elements of information theory. Wiley (1991)
19. Ben-Akiva, M., Bolduc, D.: Multinomial probit with a logit kernel and a general parametric specification of the covariance structure (1996) Working Paper, Department of Civil Engineering, MIT.
20. Kanade, T., Cohn, J., Tian, Y.L.: Comprehensive database for facial expression analysis. In: Proceedings of the 4th IEEE International Conference on Automatic Face and Gesture Recognition (FG'00). (2000) 46 – 53
21. Sujith, K.R., Ramanan, G.V.: Procrustes analysis and moore-penrose inverse based classifiers for face recognition. In: IWBRIS. (2005) 59–66



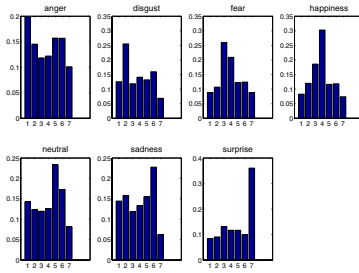
(a) DCM

(b) SVM



(c) RCA

(d) LDA



(e) GDA

Fig. 2. Mean class probabilities for correct classification