Gender Classification from Neutral and Expressive Faces

Yasmina Andreu Pedro García-Sevilla Ramón A. Mollineda

Institute of New Imaging Technologies
University Jaume I of Castellón (Spain)

Motivation

Which approach is more suitable for face gender classification when

- ... faces show facial expressions?
- ... acquisition and demographic conditions of the images vary considerably?

Study of the suitability of global and local approaches for addressing automated face gender classification of expressive faces

Main characteristics of this study

- Cross-database experiments involving 3 different databases
- Classifiers: 1-NN, PCA+LDA, SVM
- Features: Grey levels and PCA
- Statistical analysis of the results using several statistical tests

Outline

Methodology I

Global Approach

- Faces are described as a whole
- Predicted gender is provided by the classifier

Local Approach

- Faces are described per patches
- For each test patch its gender is estimated by comparing it with neighbouring patches from the training set
- Predicted gender is obtained by majority voting of local decisions

Methodology II

Face Descriptions: Types of features

- Raw features: Grey Level values of the pixels (GL)
- Transformed features: Principal Components Analysis (PCA)

Classifiers

- 1-Nearest Neighbour (1NN)
- Linear Discriminant Analysis (PCA+LDA)
- Support Vector Machine (SVM) with polynomial kernel

Experimental set-up

Classification Models

Global	(1) 1NN-grey-G	(2) 1NN-pca-G	(3) PCALDA-G	(4) SVM-grey-G	(5) SVM-pca-G
Local	(6) 1NN-grey-L	(7) 1NN-pca-L	(8) PCALDA-L		

Dataset combinations for training (rows) and test (columns)

(a) Set-up 1: Non-expressive faces for training

	FERET	PAL	AR Neutral	AR Happy	AR Angry	AR "Screaming"
FERET		×	×	♦	♦	♦
PAL	×		×	♦	♦	♦
FERET ∪ PAL			×	♦	♦	♦

(b) Set-up 2: Non-expressive vs expressive faces for training

	FERET	PAL	AR Neutral	AR Expressive
AR Neutral	×	×	×	♦
AR Expressive	♦	♦	♦	♦

Experimental Results Set-up 1

Training with non-expressive faces

Training Data Set	Test Data Set	Global					Local		
		NN		PCA+LDA	SVM		NN		PCA+LDA
		Grey Levels	PCA		Grey Levels	PCA	Grey Levels	PCA	
FERET	PAL	66.03	64.98	71.25	66.72	62.55	66.03	62.19	60.80
	AR Neutral	79.17	82.31	77.69	81.54	84.62	86.15	86.92	83.08
	AR Happy	85.50	83.97	81.68	83.97	83.97	88.55	87.79	85.50
	AR Angry	83.97	83.21	79.39	82.44	81.68	86.26	84.73	81.68
	AR “Screaming”	78.63	80.92	79.39	76.34	80.15	87.02	87.02	84.73
PAL	FERET	66.53	65.59	75.22	72.99	70.66	63.16	62.07	77.11
	AR Neutral	81.25	82.31	89.23	92.31	91.54	90.00	90.00	87.69
	AR Happy	81.68	82.44	83.97	84.73	85.50	90.84	87.02	89.31
	AR Angry	82.44	80.92	87.79	89.31	83.21	90.07	88.55	86.26
	AR “Screaming”	76.34	77.10	74.81	76.34	77.10	77.10	77.10	85.50
FERET ∪ PAL	AR Neutral	84.62	84.62	87.79	90.77	91.54	90.00	89.23	86.92
	AR Happy	85.50	83.97	87.79	87.02	88.55	90.84	90.84	85.50
	AR Angry	83.97	83.21	83.97	87.02	87.79	90.08	88.55	87.02
	AR “Screaming”	80.92	80.92	81.68	77.86	83.97	88.55	87.79	86.26

Table 1: Correct classification rates (%) obtained **training with non-expressive faces**

Experimental Results Set-up 1

Statistical Significance Tests

(a) Neutral & Expressive	(b) Neutral	(c) Expressive
$\mathbf{F_F = 5.98}$	$F_F = 1.40$	$\mathbf{F_F = 7.56}$
$F(7, 81)_{0.95} = 2.12$	$F(7, 28)_{0.95} = 2.34$	$F(7, 56)_{0.95} = 2.18$

Table 2: **Iman-Davenport's statistic** applied to the results in Table 1 (in bold when statistical differences exist)

(a) Neutral & Expr.	(b) Neutral	(c) Expressive
1NN-grey-L	**SVM-grey-G**	**1NN-grey-L**
1NN-grey-G	1NN-grey-G	PCALDA-G
1NN-pca-G	1NN-pca-G	1NN-pca-G
PCALDA-G	PCALDA-L	1NN-grey-G
SVM-grey-G	1NN-pca-L	SVM-grey-G
PCALDA-L	PCALDA-G	SVM-pca-G
SVM-pca-G	1NN-grey-L	PCALDA-L
1NN-pca-L	SVM-pca-G	1NN-pca-L

Table 3: **Holm's method** applied to the results in Table 1 with a 95% significance level. Models above the double line performed significantly worse than the others

Experimental Results Set-up 1

Statistical Significance Tests

(a) Neutral & Expressive

	1	2	3	4	5	6	7	8
1NN-grey-G (1)	-			○	○	○	○	○
1NN-pca-G (2)		-			○	○	○	○
PCALDA-G (3)			-			○		○
SVM-grey-G (4)				-		○		
SVM-pca-G (5)	●	●			-	○		
1NN-grey-L (6)	●	●	●			-	●	
1NN-pca-L (7)	●	●				○	-	
PCALDA-L (8)	●	●						-

(b) Neutral

	1	2	3	4	5	6	7	8
1	-			○				
2		-						
3			-					
4				-				
5					-			
6						-		
7							-	
8								-

(c) Expressive

	1	2	3	4	5	6	7	8
1	-					○	○	○
2		-				○	○	○
3			-			○	○	○
4				-		○	○	
5					-	○	○	○
6	●	●	●	●	●	-	●	
7	●	●	●	●	●	○	-	
8	●	●	●					-

Table 4: Summary of the Wilcoxon's Signed Rank test applied to the results in Table 1. Above the main diagonal 90% confidence level, and below it 95%. Symbol "●" indicates that the classification model in the row significantly outperforms the model in the column, and "○" indicates that the model in the column outperforms the one in the row

- When training with non-expressive faces → Local approach
- If test only neutral faces → Both approaches perform equally well
- No statistical differences between types of features

Experimental Results Set-up 2

Non-expressive vs expressive faces for training

Training Data Set	Test Data Set	Global					Local		
		NN		PCA+LDA	SVM		NN		PCA+LDA
		Grey Levels	PCA		Grey Levels	PCA	Grey Levels	PCA	
AR Neutral	FERET	76.02	76.86	80.09	**80.83**	77.21	78.90	78.90	78.20
	PAL	73.35	72.30	71.43	**75.09**	70.38	74.39	73.17	65.51
	AR Neutral	83.99	82.46	87.54	**90.42**	98.15	88.92	89.08	86.31
	AR Neutral & Expr.	88.18	87.76	85.66	88.30	**94.65**	89.79	89.45	85.32
AR Neutral & Expressive	FERET	72.59	72.94	76.56	77.66	75.22	80.59	**81.23**	77.41
	PAL	72.47	72.65	72.64	**76.48**	73.52	73.69	73.34	65.85
	AR Neutral	91.23	91.38	91.08	95.93	**96.92**	95.54	94.62	86.15
	AR Neutral & Expr.	91.24	91.24	92.82	95.66	**99.07**	94.22	93.69	86.42

Table 5: Correct classification rates (%) obtained in the additional set of experiments (in bold the highest accuracy of each training-test configuration).

- Wilcoxon's Signed Rank test: training with expressive faces improves performance of classifiers
- When training with expressive faces $\rightarrow$ Global approach

Conclusions

- Training mainly with non-expressive faces, local approaches outperform global ones
- Local approaches can deal better with distorted/unaligned faces
- Test faces only show neutral expressions, both approaches perform equally
- Large number of expressive faces available for training, global approach achieves better classification rates
- No statistical differences were found between raw and transformed features

Thank you!

Any questions?