

# 2D Principal Component Analysis for Face and Facial-Expression Recognition

*Although it shows enormous potential as a feature extractor, 2D principal component analysis (2DPCA) produces numerous coefficients. Using a feature-selection algorithm based on a multiobjective genetic algorithm to analyze and discard irrelevant coefficients offers a solution that considerably reduces the number of coefficients, while also improving recognition rates.*

Face recognition and facial-expression recognition have been active research fields for several years. Potential application areas include access control, searching mugshots, screening, security monitoring and surveillance systems, human-computer interaction, emotion analysis, and automated tutoring systems. Both face and facial-expression recognition continue to attract researchers from image processing, pattern recognition, machine learning, and computer vision.<sup>1</sup> Several attempts have been made to improve the reliability of these recognition systems. One highly successful approach is eigenfaces, which Matthew Turk and Alex Pentland proposed in 1991<sup>2</sup> based on principal component analysis (PCA). Since then, researchers have been investigating PCA and using it as a basis for developing successful techniques for face and facial-expression recognition.<sup>1</sup>

To solve some of PCA's computational problems, Jian Yang and his colleagues proposed 2D PCA (2DPCA),<sup>3</sup> which also generally surpasses PCA for recognition problems.<sup>4</sup> However, 2DPCA's drawback is that it requires more coefficients for image representation than PCA. Yang and his colleagues argue that this problem can be alleviated by using PCA after 2DPCA for further dimensional reduction. Other researchers<sup>4</sup> proposed (2D)<sup>2</sup>PCA to reduce coefficients by simultaneously considering the original image's row and column directions.<sup>3</sup> Still, although 2DPCA has proven to be a good feature extractor for face recognition, it still generates numerous coefficients.

To address this, we investigated the hypothesis that not all coefficients produced by 2DPCA are relevant to the recognition task. As an alternative to PCA and (2D)<sup>2</sup>PCA, we propose a methodology based on feature selection to find 2DPCA's most discriminant coefficients. Our feature-selection algorithm is based on a multiobjective genetic algorithm that lets us optimize the number of coefficients and the classifier's error rate together. Using a multiobjective algorithm also lets us avoid premature convergence to a single search-space region. As we describe here, our comprehensive experiments show that our approach drastically reduces the number of coefficients while improving the recognition rate.

1521-9615/11/\$26.00 © 2011 IEEE  
COPUBLISHED BY THE IEEE CS AND THE AIP

LUIZ S. OLIVEIRA, ALESSANDRO L. KOERICH,  
AND MARCELO MANSANO

*Federal University of Paraná, Curitiba, Brazil*

ALCEU S. BRITTO, JR.

*Pontifical Catholic University of Paraná, Curitiba, Brazil*

Table 1. Experimental results before and after feature selection.

Database	Images	Resolution	Dim	Baseline		After FS			2D <sup>2</sup> PCA Recognition rate (%)
				Recognition rate (%)		Recognition rate (%)			
				kNN	SVM	Dim	kNN	SVM	Dim
ORL	400	112 × 92	560	91.0	83.1	48	91.5	93.1	130
Yale	165	110 × 100	550	84.0	81.2	55	86.7	90.0	125
AR	1,000	100 × 100	500	88.0	82.0	55	89.2	92.1	116
Feret	600	128 × 128	640	87.0	80.5	73	87.9	89.2	150

### PCA versus 2DPCA

In the PCA-based face-recognition technique, the 2D face-image matrices must be previously transformed into 1D image vectors. The resulting image vectors usually lead to a high-dimensional image vector space in which it's difficult to evaluate the covariance matrix accurately due to its large size and relatively few training samples. Fortunately, we can calculate the eigenvectors (eigenfaces) efficiently using single value decomposition (SVD) techniques, which avoid the process of generating the covariance matrix. Unlike conventional PCA, 2DPCA is based on 2D matrices rather than 1D vectors. That is, the image matrix doesn't need to be previously transformed into a vector. Instead, an image covariance matrix can be constructed directly using the original image matrices. In contrast to PCA's covariance matrix, the image covariance matrix's size using 2DPCA is much smaller. As a result, 2DPCA has two important advantages over PCA. First, it's easier to evaluate the covariance matrix accurately. Second, less time is required to determine the corresponding eigenvectors.<sup>3</sup>

We extract the features from the 2DPCA matrix using the optimal projection vector axes. The vector's size is given by the image's size and the number of coefficients. If the image size is 112 × 92, for example, then the number of coefficients is 112 ×  $d$ . Researchers have demonstrated experimentally that  $d$  should be set to no less than 5 to satisfy accuracy.<sup>4</sup> This leads us to a numerous coefficients. (More details on this are available elsewhere.<sup>5</sup>)

### Feature Selection

An important issue in constructing classifiers is the selection of the best discriminative features. Many applications represent patterns through hundreds of features. Beyond a certain point, however, including additional features leads to a worse performance in terms of both accuracy and computational complexity.<sup>6</sup>

In practical applications, we can view feature selection as the optimization of a multicriterion function where the criteria are usually the number of features and classification accuracy. Researchers have demonstrated that multiobjective genetic algorithms offer a particularly attractive approach to solve this kind of problem. In addition, Mineichi Kudo and Jack Sklansky have demonstrated that genetic algorithms are suitable for feature selection in high-dimensional space (typically, more than 30 dimensions), surpassing several other feature-selection methods.<sup>6</sup>

We perform feature selection based on a strategy that uses a powerful multiobjective genetic algorithm called the *nondominated sorting genetic algorithm* (NSGA).<sup>7</sup> Unlike a single genetic algorithm, the NSGA produces a set of potential solutions known as Pareto-optimal solution. This lets the user experiment with different trade-offs between the objectives being optimized. The idea behind the NSGA is that we use a ranking-selection method to emphasize good points and use a niche method to maintain stable subpopulations of good points. (More details are available elsewhere.<sup>5,7</sup>)

### Experiments and Analysis

Our goal with these experiments is to show that the 2DPCA provides highly discriminant features, but with an inherent cost of numerous unnecessary coefficients. As a demonstration, we applied our feature-selection method on the feature matrix extracted by the 2DPCA method. We considered both the facial recognition and facial-expression recognition problems.

To show our proposed approach's efficiency, we used two classifiers: k-nearest neighbor (kNN) and a support vector machine (SVM).

### Face Recognition

For our face-recognition experiments, we used four databases: Olivetti Research Laboratory (ORL), Yale, Facial Recognition Technology (Feret), and AR. Table 1 shows the number of images used

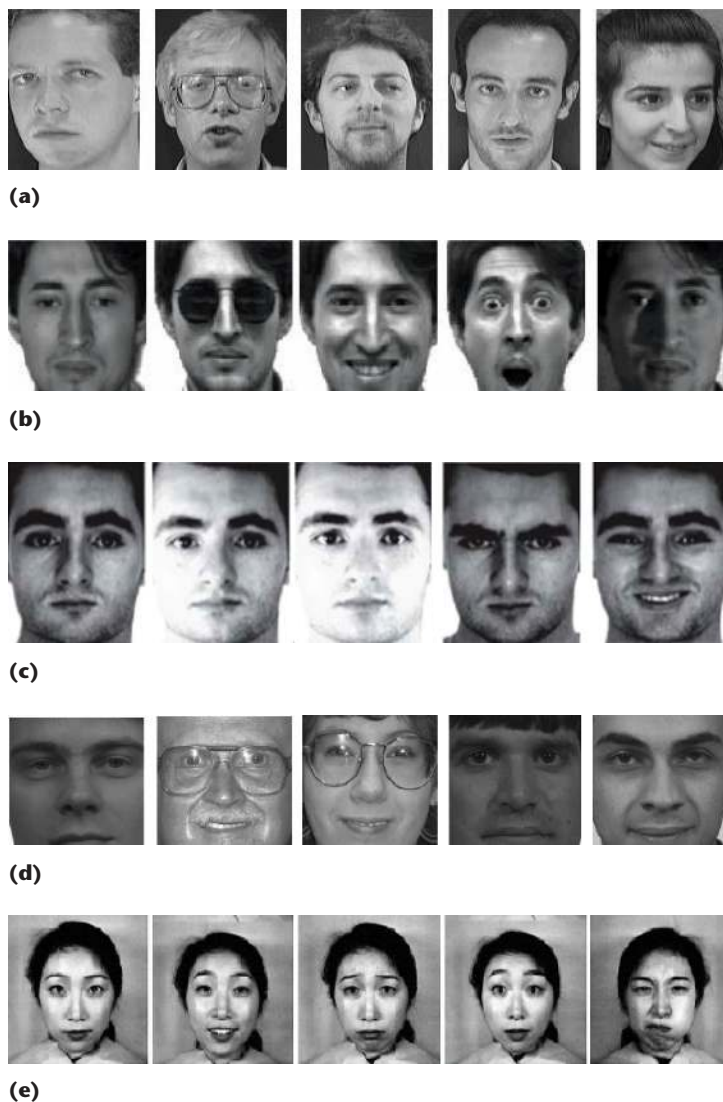


Figure 1. Samples extracted from the databases used in the experiments. Rows 1 through 4 show examples from four face-recognition databases: Olivetti Research Laboratory (ORL), Yale, Facial Recognition Technology (Ferret), and AR. Row 5 shows examples from the Japanese Female Facial Expression (Jaffe) database.

in the experiments and their resolutions. The first four rows in Figure 1 show examples from these databases, highlighting different facial expressions and lighting conditions.

In all facial-recognition experiments, we divided the databases into four subsets: training (40 percent), validation (10 percent), searching (10 percent), and testing (40 percent). As the name suggests, we used the training set to train the models. We used the searching set to compute the fitness during the search and the validation set to find the best solution in the Pareto front while avoiding the overfitted ones. Finally, we used the testing set to evaluate the final classification accuracy.

The first strand of experiments was carried out to provide some baseline to further evaluate the feature selection's impact on the classification accuracy. Thus, we reproduced the results reported elsewhere,<sup>3</sup> but using both kNN and SVM. We tried different values for  $d$ ; the minimum value we found without losing performance was  $d = 5$ . The baseline in Figure 1 shows the recognition rates we have achieved using  $d = 5$ . We tried different kernels for the SVM, but the linear kernel produced better results in our experiments.

The NSGA used for feature selection is based on bit representation, one-point crossover, bit-flip mutation, and roulette wheel selection (with elitism). We used the following parameter settings: population size was 100, there were 200 generations, the crossover probability was 0.8, the mutation probability was 0.002, and the niche distance was 0.5.

In these experiments, our first objective was to minimize the number of coefficients and the second was to minimize the error rate on classification. The latter was computed on the searching set using the kNN algorithm. We replicated the experiments 10 times to better compare the results.

In our previous work,<sup>5</sup> we demonstrated that the population converges to a specific search-space region with few coefficients, offering good performance. The NSGA also finds some solutions with few coefficients but poor performance. This is due to the concept of nondominated sorting used during the search. However, those solutions are easily discarded by using the validation set.

The best solutions we obtained by applying the validation set on all solutions in the Pareto-front contain 48, 55, 55, and 73 coefficients for ORL, Yale, AR, and Ferret, respectively. Table 1 also shows the recognition rates on the testing set and the feature vectors' dimensionality for all databases (see the "After FS" columns).

The feature selection was performed using only the kNN classifier. To show the validity of the solution found during the optimization process—and show that it could generalize well with other classifiers—we used it to train the SVM classifier.

Table 1 shows two interesting aspects. First, the feature selection process can drastically reduce the feature vector's size (by 10 times, on average). Second, the SVM classifiers' performance, as expected, was slightly better than the kNN classifiers. However, when trained with the original feature vectors (before feature selection), SVM performance was worse than kNN. This can be explained by the feature vectors'

high dimensionality as well as the training data set's small size. So far, these results corroborate our claim that the 2DPCA works fine as a feature extractor but that it generates numerous unnecessary or even correlated coefficients.

Figure 2 compares the number of coefficients and the performance of the 2DPCA and the 2DPCA with feature selection on the ORL database. As we noted earlier, the best setup we found was for  $d = 5$ . Beyond such a value, the performance gets worse. The same behavior is observed for the performance of the 2DPCA with feature selection. Figure 2 shows the performance for the feature selection (dashed line) just for  $d < 5$ . This is because the initial feature set we used for feature selection consisted of 560 coefficients ( $d = 5$ ).

Also, the selected coefficients are uniformly distributed among the family of projected feature vectors. This shows that complementary information can be found in the entire feature space.

After analyzing the misclassified images, we can observe that the features extracted by 2DPCA aren't powerful enough to absorb distinct lighting conditions, even after feature selection. An example of this weakness is shown in the last face of the Yale database in Figure 1: half the face is covered by shadow due to variation in illumination. All the faces in the database that had this kind of shadow were misclassified, but such images weren't presented in the training set.

The last row in Table 1 compares our results with the ones produced by  $(2D)^2PCA$ , showing that our methodology achieves better results with feature vectors considerably smaller than  $(2D)^2PCA$ .

### Facial-Expression Recognition

We applied our coefficient selection scheme on facial-expression recognition to further demonstrate its feasibility on a different but related problem. In this case, we selected the widely used Japanese Female Facial Expression (Jaffe) database for the experiments.

The Jaffe database contains images of 10 women with 213 different facial expressions. Each image has a  $256 \times 256$  pixel resolution. The number of images corresponding to each of the seven categories of expression (neutral, happiness, sadness, surprise, anger, disgust, and fear) is almost the same. The last row in Figure 1 shows samples from the Jaffe database.

We applied the same protocol here as we used in the face-recognition experiments. In this case, the face images were cropped to  $150 \times 150$  using the face-detector algorithm proposed by Paul Viola and Michael Jones.<sup>8</sup> In addition to a drastic

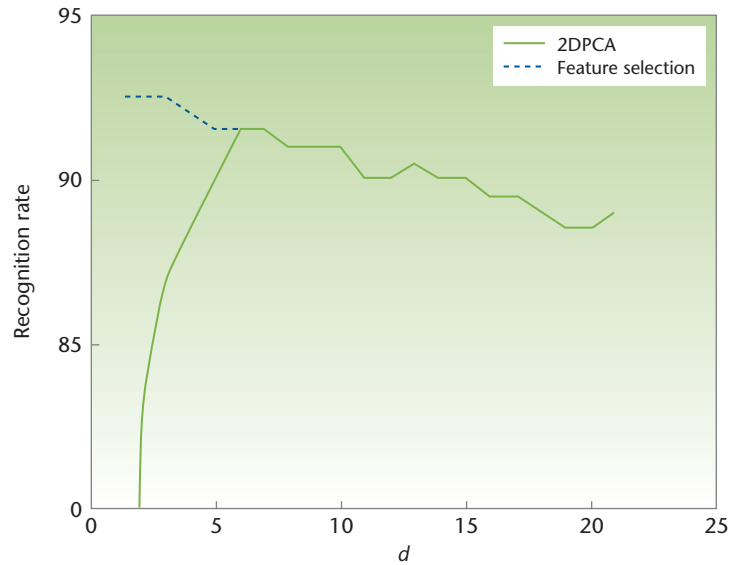


Figure 2. Recognition rate versus  $d$  values for the conventional 2DPCA and the 2DPCA with feature selection (ORL database).

Table 2. 2DPCA results before and after feature selection for facial-expression recognition.

Strategy	Coefficients	Recognition rate (%)	
		kNN	SVM
2DPCA	900	82.0	78.8
2DPCA + feature selection	58	91.0	94.0

reduction in the number of coefficients, we observed a considerable improvement in the recognition rate relative to the face-recognition experiment. This could be explained by the fact that facial-expression recognition is less complex, at least when considering the number of classes involved and because we have a few more samples per class. Table 2 summarizes the experiments on the Jaffe database testing set.

Our results compare favorably to previous literature,<sup>9–11</sup> which reports recognition rates around 90 percent and much higher dimensional feature vectors, which are based on Gabor filters and local binary patterns.

**O**ur future work in this area will involve evaluating the selected feature subset's performance on other standard facial-expression databases and developing a new feature set to detect mixed emotions—such as happiness and surprise. Also, implementing a reliable and fast FPGA-based embedded system is a challenge we plan to face very soon.



## References

1. W. Zhao et al., "Face Recognition: A Literature Survey," *ACM Computing Surveys*, vol. 35, no. 4, 2003, pp. 399–458.
2. M. Turk and A. Pentland, "Eigenfaces for Recognition," *J. Cognitive Neuroscience*, vol. 3, no. 1, 1991, pp. 71–86.
3. J. Yang et al., "Two-Dimensional PCA: A New Approach to Appearance-Based Face Representation and Recognition," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 26, no. 1, 2004, pp. 131–137.
4. D. Zhang and Z-H. Zhou, "2D<sup>2</sup>PCA: 2-Directional 2-Dimensional PCA for Efficient Face Representation and Recognition," *NeuroComputing*, vol. 69, nos. 1–2, 2005, pp. 224–231.
5. A.S. Britto, Jr., L.S. Oliveira, and A.L. Koerich, "Face Recognition Using Selected 2DPCA Coefficients," *Proc. 17th Int'l Conf. Systems, Signals and Image Processing*, Editora da Universidade Federal Fluminense, 2010, pp. 490–493; [www.ic.uff.br/iwssip2010/Proceedings/nav/papers/paper\\_162.pdf](http://www.ic.uff.br/iwssip2010/Proceedings/nav/papers/paper_162.pdf).
6. M. Kudo and J. Sklansky, "Comparison of Algorithms that Select Features for Pattern Classifiers," *Pattern Recognition*, vol. 33, no. 1, 2000, pp. 25–41.
7. L.S. Oliveira et al., "A Methodology for Feature Selection Using Multi-Objective Genetic Algorithms for Handwritten Digit String Recognition," *Int'l J. Pattern Recognition and Artificial Intelligence*, vol. 17, no. 6, 2003, pp. 903–930.
8. P. Viola and M. Jones, "Robust Real-Time Face Detection," *Int'l J. Computer Vision*, vol. 57, no. 2, 2004, pp. 137–154.
9. S. Bashyal and G.K. Venayagamoorthy, "Recognition of Facial Expressions Using Gabor Wavelets and Learning Vector Quantization," *Eng. Applications of Artificial Intelligence*, vol. 28, no. 7, 2008, pp. 1056–1064.
10. A. Koutlas and D.I. Fotiadis, "An Automatic Region-Based Methodology for Facial Expression Recognition," *Proc. IEEE Int'l Conf. Systems, Man and Cybernetics*, IEEE Press, 2008, pp. 662–666.
11. C. Shan, S. Gong, and P.W. McOwan, "Facial Expression Recognition Based on Local Binary Patterns: A Comprehensive Study," *Image and Vision Computing*, vol. 27, no. 6, 2009, pp. 803–816.

**Luiz S. Oliveira** is head of the Department of Informatics' graduate program in computer science at the Federal University of Paraná in Curitiba, Brazil. His research interests include computer vision, machine learning, pattern recognition, and evolutionary computation. Oliveira has a PhD in computer science from the University of Quebec. Contact him at [lesoliveira@inf.ufpr.br](mailto:lesoliveira@inf.ufpr.br).

**Marcelo Mansano** is a graduate student in the Department of Informatics at the Federal University of Paraná in Curitiba, where he's working on gender classification. His research interests include computer vision and pattern recognition. Mansano has a BS in computer engineering from the State University of Ponta Grossa in Paraná, Brazil. Contact him at [mrmansano@gmail.com](mailto:mrmansano@gmail.com).

**Alceu de Souza Britto, Jr.**, is a professor in the Post-graduate Program in Applied Informatics at the Pontifical Catholic University of Paraná, where he previously directed the undergraduate computer science program. His research interests include pattern recognition, computer vision, document analysis, and handwriting recognition. Souza Britto has a PhD in computer science from the Pontifical Catholic University of Paraná. Contact him at [alceu@ppgia.pucpr.br](mailto:alceu@ppgia.pucpr.br).

**Alessandro L. Koerich** is a professor in the Department of Computer Science at the Pontifical Catholic University of Paraná, Brazil, and an associate professor at the Department of Electrical Engineering at the Federal University of Paraná. His research interests include machine learning and computer vision. Koerich has a PhD in automated manufacturing engineering from the University of Quebec. He is a member of IEEE and the ACM. Contact him at [alekoe@ppgia.pucpr.br](mailto:alekoe@ppgia.pucpr.br).



Selected articles and columns from IEEE Computer Society publications are also available for free at <http://ComputingNow.computer.org>.