receptive field and inhibited by LGN inputs near the periphery of its receptive field. An OFF-center receptive field is the reverse. Thus, these neurons are sensitive to either light or dark spots at the center of their receptive fields, but in both cases their outputs are characterized by the firing rate of pulses of the *same* type. The neurons in subsequent layers in V1 are more specific in their behavior to the local visual input patterns.

Recent research [2] has indicated that neural net-

That is, let an n-dimensional input image $\mathbf{v}$ be represented by

$$\mathbf{v} = \mathbf{W}\mathbf{h} = \sum_{a=1}^{r} \mathbf{W}_{ia}\mathbf{h}_a \tag{1}$$

The r columns of $\mathbf{W}$ are the basis images and each of these is an n-dimensional vector. Thus, $\mathbf{W}$ is an $n \times r$ matrix. The hidden components that provide the specific contribution of each basis vector in the input vectors are $\mathbf{h}_1, \ldots, \mathbf{h}_r$. These latent variables are stochastic and differ for each observed input $\mathbf{v}$. They are represented as an r-dimensional vector $\mathbf{h}$. The crucial assumption in the sparse coding framework is that the hidden variables exhibit *sparseness* [11]. The goal is to select a set of basis components so that $\mathbf{v}$ can (by proper choice of $\mathbf{h}_1, \ldots, \mathbf{h}_r$) be represented *accurately*[1] and *sparsely*.[2] However, note that even though some of the elements in a basis component are zero, they are actually all needed because the particular set of *active* coefficients changes from input to input.

Hoyer suggested that, when this sparse model is learnt from image data, the learnt basis components have the properties of the spatial receptive fields of simple cells in V1. Thus the neural interpretation of the model is that simple cells in V1 perform sparse coding on the visual input they receive, with the receptive fields being closely

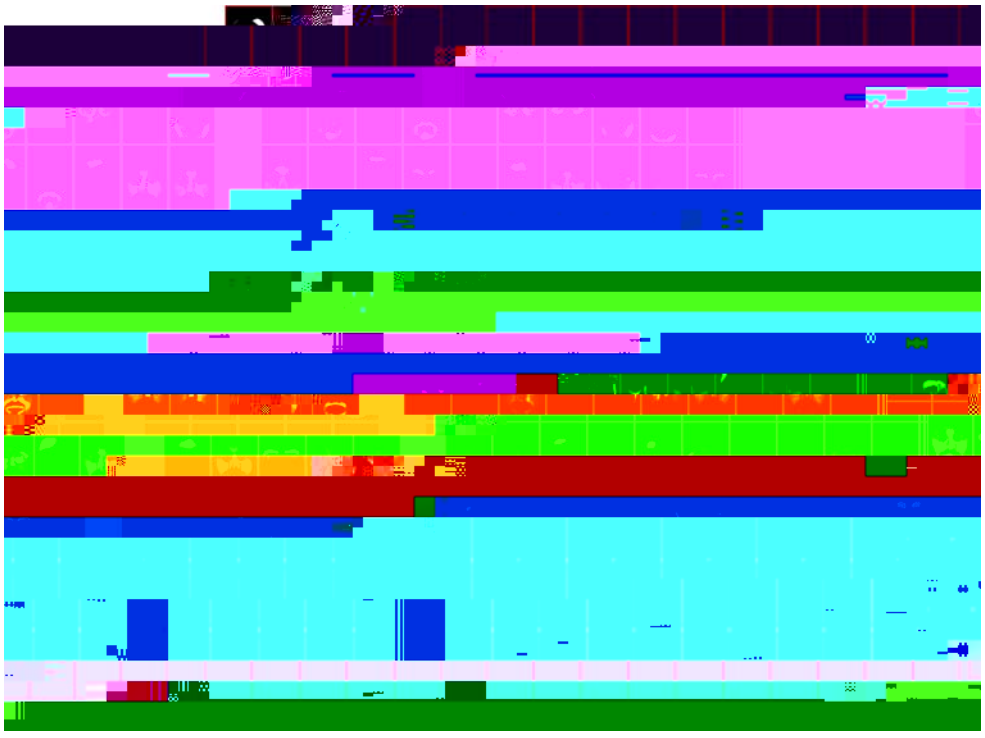


Fig. 1 Sparse and part-based basis components of dimensionality 20, 50, and 100 obtained by the NNSC technique. It can be observed that the basis components tend to become more localized as the dimensionality is increased from 20 to 100. Also, histogram plots of the basis images reveal that as their dimensionality is increased, the basis components become more binary in nature (either 0 (*black*) or 255 (*white*))

where $[]_{kj}$ indicates that the noted divisions and multiplications are computed element by element. This projected gradient decent step is guaranteed to decrease the objective function if the step size $\mu \geq 0$ is small enough [5]. However, there is no guarantee of reaching the global minimum, due to the non-convex constraints:

$$\Theta_{\text{NNSC}}\left(\mathbf{W}^{(t+1)}, \mathbf{H}^{(t+1)}\right) \leqslant \Theta_{\text{NNSC}}\left(\mathbf{W}^{(t)}, \mathbf{H}^{(t)}\right); \quad t \geqslant 0 \tag{8}$$

A set of NNSC basis components (computed for the neutral pose in the AR database—see Sect. 5.2) of dimensionality 20, 50, and 100 is shown in Fig. 1.[4] It can be seen that the basis components are both sparse and part-based. In the next section we discuss NNSC in the context of face recognition.

[4] The original facial images were 768×576 pixels. However, after background removal and geometrical normalization (including scaling down the image size by 4:1 while maintaining the original aspect ratio), as explained in detail in Sect. 5, the basis components are 181×121 pixels in size.

4 Face recognition in the part-based subspace

4.1 Basic algorithm

The task of automatically recognizing human faces using NNSC is depicted in Fig. 2. There is a training stage in which the facial codes, $\mathbf{h}_{k1}, \ldots, \mathbf{h}_{km}$ for the individual database images (a total of m) are extracted and stored for later use. This can be thought of as creating a model, which is obviously necessary even for human beings; we perform correlation between what is seen and what is already known in order to actually achieve recognition [12]. At run-time, a test input (a total of l images) is presented to the system, and its facial codes $\mathbf{h}$

the vectors are totally uncorrelated (opposite). The correlation is given by

$$d(\mathbf{h}_q, \mathbf{h}_k) = \mathrm{corr}(\mathbf{h}_q, \mathbf{h}_k) = \frac{r\sum_{j=1}^{r} \mathbf{h}_{qi}\mathbf{h}_{ki} - \sum_{i=1}^{r} \mathbf{h}_{qi}\sum_{i=1}^{r} \mathbf{h}_{ki}}{\sqrt{\left[r\sum_{i=1}^{r} \mathbf{h}_{qi}^2 - \left(\sum_{i=1}^{r} \mathbf{h}_{qi}\right)^2\right]\left[r\sum_{i=1}^{r} \mathbf{h}_{ki}^2 - \left(\sum_{i=1}^{r} \mathbf{h}_{ki}\right)^2\right]}} \tag{16}$$

5 Experiments and results

This section begins with a brief overview of the facial databases used for testing and highlights the differences among them. Then we present and discuss the experiments and results. In all cases, images were first normalized geometrically[5] as in [12

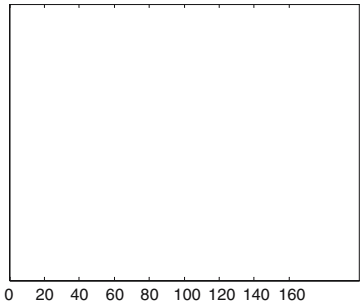

0 20 40 60 80 100 120 140 160

Table 1 A summary of recognition rates with varying expressions (smile, anger, scream). FaceIt, Bayesian, PCA, and the part-based techniques are compared

Technique	Recognition accuracy		
	Smile (AR 02), %	Anger (AR 03), %	Scream (AR 04), %

Table 5 A summary of recognition rates with occlusion (sunglasses) and varying light direction

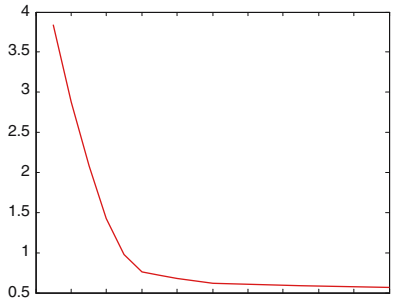
Technique	Recognition accuracy		
	Sunglasses (AR 08), %	Sunglasses/ left light (AR 09), %	Sunglasses/ right light (AR10), %
FaceIt	8	10	6
Bayesian	35	34	28
PCA	16	26	24
NMF	42	36	54
LNMF	50	48	30
NNSC	55	51	50

FaceIt, Bayesian, PCA, and the part-based techniques are compared

as part-based methods. Part-based methods also yield higher recognition rates than FaceIt and Bayesian techniques, while NNSC outperforms both NMF and LNMF. Clearly, as one would expect, part-based features are superior to holistic ones.

Finally, it is noted that for occlusions of this type, the L_1-metric is best suited for NNSC and LNMF, whereas for NMF the L_2-metric is the most suitable. These findings are summarized in Table 6.

5.2.6 Occlusion with scarf


4
3.5
3
2.5
2
1.5
1
0.5

Non-negative Sparse Coding (NNSC) on various databases such as the AR, the FERET, the Yale B, and the Cambridge ORL database. We have compared and evaluated each of the part-based techniques under varying illumination, expression, occlusion, and pose factors. In addition, the part-based representation techniques were tested with different distance metrics such as the L_1-metric, L_2-metric, and Normalized Cross-Correlation (NCC). All the experiments were performed over a large range of basis dimensions. The experiments can be grouped into five main categories: recognition across varying expression, varying illumination, occlusion with sunglasses, occlusion with scarf, and varying pose.

The AR database was used for the first four experiments (varying expression, illumination, and occlusions), and the results obtained were compared to the well-known principal component analysis (PCA) holistic approach. The results were also compared with two leading techniques used by the computer vision community: FaceIt and Bayesian. The Yale B database was utilized for the experiments with varying pose. The Cambridge ORL database was used to validate and compare our results obtained with NNSC with those in the literature addressing the NMF and LNMF techniques. The same sets of experiments were performed but the objective here is to make a direct comparison.

For the experiments with varying expression (anger, smile, scream), it was found that part-based techniques performed much better (with NNSC being the best), giving rates of up to 100%. The FaceIt, Bayesian, and the PCA approaches did not perform as well, with rates in the 80% range. We note that the best distance measure for NMF is NCC, whereas for LNMF and NNSC, it is the L_1-metric.

The recognition accuracies obtained by varying illumination show that Bayesian and PCA cannot deal with illumination changes as well as the part-based techniques or FaceIt. Again, the part-based techniques produced rates of up to 100%, with NNSC being the best. FaceIt yielded rates in the 90% range, while Bayesian and PCA were in the 70% range. The best distance measure for NMF was the L_2-metric, whereas for LNMF and NNSC it was NCC.

Part-based techniques gave much higher recognition rates than PCA when considering partial occlusion and are consistent with the theory of part-based and holistic methods. The L_2-metric was best suited for NMF whereas for LNMF and NNSC, it was the L_1-metric.

The FERET database was also used to test the part-based techniques. The experiments were performed on frontal poses and the recognition rates were obtained for a varying number of basis dimensions, using different distance metrics as the similarity measure. The L_1-metric was observed to be the best distance metric for the part-based techniques. The recognition rates were 97% for the NNSC technique, whereas the NMF and the LNMF techniques gave accuracies of 95 and 96%, respectively. This relatively high recognition rate was obtained by NNSC using only a sm4(n)-7380.4(e)]Tr r95

6. Li, S.Z., Hou, X.W., Zhang, H.J., Cheng, Q.S.: Learning spatially localized, parts-based representation. IEEE Comput. Vis. Pattern Recognit. **1**, 207–212 (2001)
7. Martinez, A., Benavente, R.: The AR face database. Technical report 24, Computer Vision Center (CVC), Barcelona, Spain (1998)
8. Phillips, P.J., Wechsler, H., Huang, J., Rauss, P.: The FERET database and evaluation procedure for face-recognition algorithms. Image Vis. Comput. **16**, 295–306 (1998)
9. Georghiades, A.S., Belhumeur, P.N., Kriegman, D.J.: From few to many: illumination cones models for face recognition under variable lighting and pose. IEEE Trans. Pattern Anal. Mach. Intell. **23**, 643–660 (2001)
10. ORL Database of Faces. AT&T Laboratories, Cambridge, UK. Web address: http://www.orl.co.uk/facedatabase.html
11. Hoyer, P.O.: Probabilistic Models of Early Vision. Ph.D. Dissertation, Department of Computer Science, Helsinki University of Technology, Finland (2002)
12. Hafed, Z.M., Levine, M.D.: Face recognition using the discrete cosine transform. Int. J. Comput. Vis. **43**, 167–188 (2001)
13. Gandhi, M.: Aging adult human faces. M. Eng. Thesis, Dept. Elect. Eng., McGill University, Montreal, Canada (2004)
14. Guillamet, D., Vitria, J.: Determining a Suitable Metric when using Non-Negative Matrix Factorization. In: Proceedings of the 16th International Conference Pattern Recognition, vol. 2, pp. 128–131 (2002)
15. Penev, P.S., Atick, J.J.: Local feature analysis: a general statistical theory for object representation. Neural Syst. **7**, 477–500 (1996)
16. Moghaddam, B., Pentland, A.P.: Probabilistic visual learning of object representation. IEEE Trans. Pattern Anal. Mach. Intell. **19**, 696–710 (1997)
17. Turk, M.A., Pentland, A.P.: Face Recognition using Eigenfaces. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp 586–591 (1991)
18. Gross, R., Shi, J., Cohn, J.F.: Quo vadis Face Recognition?,