Machine Vision and Applications
DOI 10.1007/s00138-006-0052-0

ORIGINAL PAPER

Face recognition using localized features based on non-negative sparse coding

Bhavin J. Shastri · Martin D. Levine

Received: 8 March 2006 / Accepted: 10 October 2006

Abstract Neural networks in the visual system may be performing sparse coding of learnt local features that are qualitatively very similar to the receptive fields of simple cells in the primary visual cortex, V1. In conventional sparse coding, the data are described as a combination of elementary features involving both additive and subtractive components. However, the fact that features can 'cancel each other out' using subtraction is contrary to the intuitive notion of combining parts to form a whole. Thus, it has recently been argued forcefully for completely non-negative representations. This paper presents Non-Negative Sparse Coding (NNSC) applied to the learning of facial features for face recognition and a comparison is made with the other part-based techniques, Non-negative Matrix Factorization (NMF) and Local-Non-negative Matrix Factorization (LNMF). The NNSC approach has been tested on the Aleix–Robert (AR), the Face Recognition Technology (FERET), the Yale B, and the Cambridge ORL databases, respectively. In doing so, we have compared and evaluated the proposed NNSC face recognition technique under varying expressions, varying illumination, occlusion with sunglasses, occlusion with scarf, and varying pose. Tests were performed with different distance metrics such as the L_1-metric, L_2-metric, and Normalized Cross-Correlation (NCC). All these experiments involved a large range of basis dimensions. In general, NNSC was found to be the best approach of the three part-based methods, although it must be observed that the best distance measure was not consistent for the different experiments.

Keywords Face recognition · Sparse-coding · Non-negative representations · Distance metrics

B. J. Shastri (✉) · M. D. Levine
Centre for Intelligent Machines and Department of Electrical and Computer Engineering, McGill University,
3480 University Street, Rm. 410, Montreal,
QC, Canada H3A 2A7
e-mail: shastri@cim.mcgill.ca

M. D. Levine
e-mail: levine@cim.mcgill.ca

1 Introduction

1.1 Background and motivation

When humans see, light falls on the retinal sensors at the back of the eye and is converted to electrical pulse trains. After processing by various neural layers in the retina, the signals pass out of the eye via the optic nerve and to the lateral geniculate nucleus (LGN) from where the data enter the visual cortex. At the back of the brain is a structure called the striate cortex, which is the primary visual cortex, V1. The latter encompasses six layers of neurons on the surface of the brain, and the input from the LGN enters at the fourth layer. The neurons in V1 each take input from a number of geniculate neurons, although any individual neuron can only "see" a small portion of the image that the eyes are viewing. This small region is called the receptive field, which can be characterized as being localized, oriented, and bandpass [1]. In general, neurons have both excitatory and inhibitory inputs. Those in V1 directly receiving data from LGN typically have what are called center-surround geometrical configurations. There are two types: ON-center and OFF-center receptive fields. An ON-center receptive field is a neuron that is excited by LGN inputs ("pixels") to V1 near the center of its

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 networks in the visual system could be performing *sparse coding* of learnt features that are qualitatively very similar to the receptive fields of simple cells in V1. In conventional sparse coding [3], the data are described as a combination of elementary features involving *both* additive and subtractive components. However, the fact that features can 'cancel each other out' using subtraction is contrary to the intuitive notion of combining parts to form a whole [4]. Thus, Lee and Seung have recently argued forcefully for *completely* non-negative representations. Other arguments for non-negative representations come from biological modeling, where such constraints are related to the non-negativity of the neural firing rates for both the ON- and OFF-receptive fields [5].

This paper presents Non-Negative Sparse Coding (NNSC) applied to learning facial features for face recognition. NNSC, first proposed by [5], is a data-adaptive representation that is tailored to the statistics of the data and has its roots in neural computation. This approach provides a principled method for using training data to determine the significant parts of an object such as a face. It is well known that the factors affecting the performance of face recognition systems are robustness to scale, head pose, scene illumination, facial expression, and various occlusions. By focusing on parts of a face and not assuming the conventional holistic approach, NNSC offers the possibility of improving the performance of such systems.

Two other similar methods have been proposed recently for spatially localized, part-based feature extraction. Non-negative Matrix Factorization (NMF) [4] imposes a non-negativity constraint on learning basis images as part of a subspace approach. The pixel gray levels of the resulting basis images, as well as coefficients for reconstruction, are all non-negative. Thus NMF has been considered as a procedure for learning part-based representations [4].

Local Non-negative Matrix Factorization (LNMF) [6], inspired by the original NMF, also aims for a part-based representation. In addition to the non-negativity constraint of [4], sparseness is imposed on the coordinates h in the low-dimensional feature space and locality of features is imposed on the basis components.

This paper demonstrates how NNSC is able to learn parts of a face and can be used for face recognition. A direct comparison is made with the earlier part-based techniques, NMF and LNMF. The NNSC approach has been tested on the Aleix–Robert (AR) database [7], the Face Recognition Technology (FERET) database [8], the Yale B database [9], and the Cambridge ORL database [10]. We have compared and evaluated the NNSC face recognition technique under varying expressions, varying illumination, occlusion with sunglasses, occlusion with scarf, and varying pose. In addition, tests were performed with different distance metrics such as the L_1- metric, L_2-metric, and Normalized Cross-Correlation (NCC). All of these experiments involved a large range of basis dimensions.

1.2 Overview

Following this introduction, the idea behind sparse coding is presented in Sect. 2, after which the non-negativity constraints imposed on the sparse model are explained in Sect. 3. The proposed face recognition system is outlined in Sect. 4. In Sect. 5, NNSC is applied to the AR, FERET, Yale B, and Cambridge ORL databases, respectively, to learn appropriate features for face recognition. High recognition rates are obtained using only a small set of basis components. This is in contrast to the other part-based techniques [4,6] that have reported lower rates on the same databases for the same number of basis dimensions. Finally, Sect. 6 concludes the paper and summarizes the results. We note that to date no other literature sources have reported face recognition results using NNSC.

2 The idea behind sparse coding

Olshausen and Field [2] published a paper describing how a simple neural network performs sparse coding of visual features that are qualitatively very similar to the receptive fields of V1 simple cells. They showed how all the basic spatial properties of simple cells could emerge in an unsupervised manner from natural images. This was the first study to investigate this idea. Subsequently, the very closely related model of Independent Component Analysis (ICA) [3] was shown to give similar results. This framework explained the observed receptive fields from an information-theoretic point of view, which had been lacking in previous work and thereby demonstrated how sensory coding was related to the statistics of the environment [11].

The basic idea behind sparse coding is relatively simple. An observed multi-dimensional vector $\mathbf{v}$ is modeled as a linear superposition of features (basis vectors) $\mathbf{W}$.

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 related to the sparse coding basis vectors and the neuron firing rates representing the latent variables $\mathbf{h}_1, \ldots, \mathbf{h}_r$.

3 Non-negativity constraints on sparse coding

One way in which the standard sparse coding image model is unrealistic as a paradigm of V1 simple cell behavior is the fact that each component $\mathbf{h}_1, \ldots, \mathbf{h}_r$ can, in addition to being effectively silent (close to zero), be either positively or negatively active [5,11]. This means that each feature contributes to representing stimuli of opposing polarity. This is in clear contrast to the behavior of simple cells in V1 where the neurons produce pulse-firing rates as outputs. They tend to have quite low background firing rates and, as these cannot go negative, can only represent one half of the output distribution of a signed feature $\mathbf{h}_1, \ldots, \mathbf{h}_r$.

Let us examine Eq. (1) in the context of NNSC and extend it from representing a single image $\mathbf{v}$ to a collection of images $\mathbf{V}$. Let a dataset of m training images be given as an $n \times m$ matrix $\mathbf{V}$ with each column consisting of the n non-negative pixel values of an image. This matrix is approximately factorized into the basis matrix $\mathbf{W}$ and an $r \times m$ matrix $\mathbf{H}$. Each facial image can then be represented as a linear combination of the basis images using the approximate factorization:

$$\mathbf{V} \approx \tilde{\mathbf{V}} = \mathbf{W}\mathbf{H} \tag{2}$$

As stated earlier, a column of matrix $\mathbf{W}$ contains a basis vector, while each column of $\mathbf{H}$ contains the weights needed to approximate the corresponding columns in $\mathbf{V}$ using the bases from $\mathbf{W}$. The rank r of factorization is generally chosen so that [4]:

$$r < \frac{nm}{n+m} \tag{3}$$

Combining the goal of small reconstruction error with that of sparseness, the following objective function to be minimized can be arrived [5]

$$\Theta_{\mathrm{NNSC}}(\mathbf{W},\mathbf{H}) = \sum_{i=1}^{n}\sum_{j=1}^{m}\left(V_{ij} - \sum_{l=1}^{r} W_{il}H_{lj}\right)^2 + \lambda\sum_{k=1}^{r}\sum_{j=1}^{m} f(H_{kj}) \tag{4}$$

subject to the constraints $\forall i,j,k,l : W_{il} \geq 0, H_{lj} \geq 0$, and $|w_j| = 1$, where w_j denotes the jth column of $\mathbf{W}$. The sparseness parameter λ is a constant that controls the tradeoff between accurate reconstruction and sparseness.[3] With λ equal to zero this objective function reduces to the squared error version of NMF [4]. The form of f defines how sparseness is measured and Hoyer [5] suggests that a typical choice for f is $f(H_{kj}) = |H_{kj}|$. We observe that the NNSC of a non-negative data matrix $\mathbf{V}$ is given by the minimization of the objective function $\Theta_{\mathrm{NNSC}}(\mathbf{WH})$ subject to the same constraints as that of NMF [4].

The NNSC *learning* algorithm includes two parameters, the sparseness factor $\lambda \geq 0$ and the iterative step size μ for the projected gradient descent. The objective function is non-increasing under the following update rules [5]:

$$H_{kj} \leftarrow H_{kj}\sum_{i=1}^{n} \frac{[W^T V]_{kj}}{[W^T W H]_{kj} + \lambda} \tag{5}$$

$$W \leftarrow W^T - \mu\Big(W^T H - V\Big)H^T \tag{6}$$

$$W_{kl} \leftarrow \frac{W_{kl}}{\sum_{k=1}^{n} W_{kl}} \tag{7}$$

[1] The right-hand side of (1) closely equals the left-hand side.

[2] That is, it favors representations where only a few coefficients $\mathbf{h}_1, \ldots, \mathbf{h}_r$ are significantly active (non-zero) for any given input $\mathbf{v}$.

[3] Hoyer [11] discusses how to select the sparseness measure. In our experiments, the best set of basis functions that provides a truly part-based representation was for $\lambda = 0.75$ In this case, the term $\mathbf{V}-\mathbf{WH}$ converges the fastest and also to the lowest value.

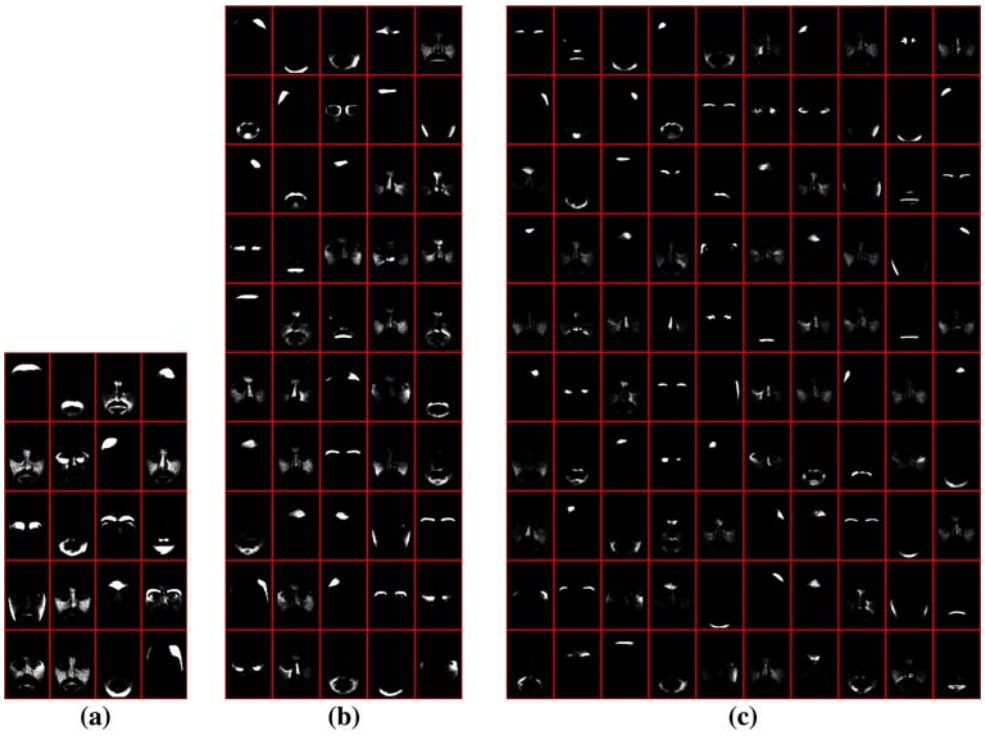

Fig. 1 Sparse and part-based basis components of dimensionality 20, 50, and 100 obtained by the NNSC technique. It can be observed that the basis components tend to become more localized as the dimensionality is increased from 20 to 100. Also, histogram plots of the basis images reveal that as their 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 $\mathbf{h}_{q1}, \ldots, \mathbf{h}_{q1}$ are extracted. The closest match is found by performing a search that computes distance measures and uses nearest neighbor classification. These steps are explained in more detail in Sects. 4.2 and 4.3.

The decision, f made by a machine can then be given by

$$f = \begin{cases} 1, & \text{if } v_q \text{ is an authorized face} \\ 0, & \text{if } v_q \text{ is an unauthorized face} \end{cases} \tag{9}$$

where $\mathbf{v}_q$ is a query face.

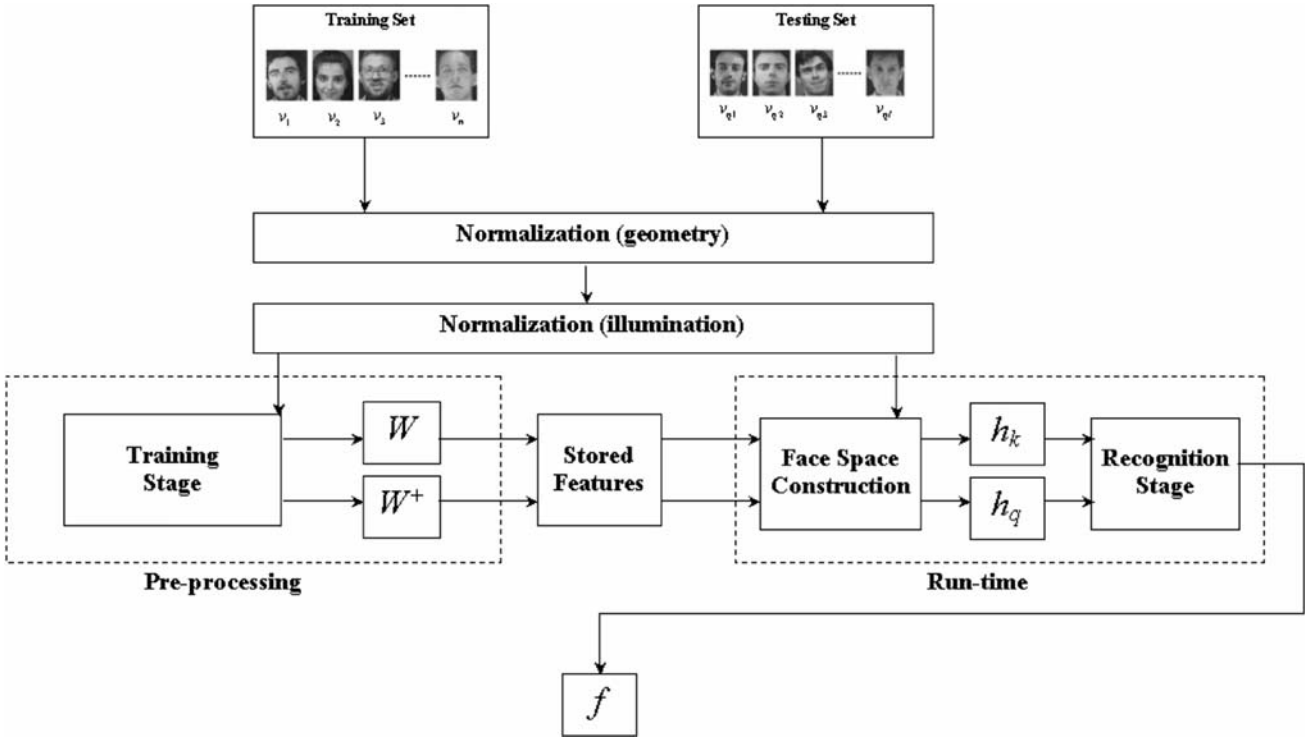

Fig. 2 Steps involved in the proposed face recognition approach. The *symbols* used are explained in Sects. 4.2 and 4.3

4.2 Feature extraction

For a set of m facial training images, the average face $\bar{\mathbf{v}}$ is given by

$$\bar{\mathbf{v}} = \frac{1}{m}\sum_{i=1}^{m} \mathbf{v_i} \tag{10}$$

Then each face $\mathbf{v}_k$ differs from the average face by

$$\xi_k = \mathbf{v}_k - \bar{\mathbf{v}}; \quad k = 1, \ldots, m \tag{11}$$

A face space is constructed by projecting each individual face $\mathbf{v}_k$ of the training set to the low dimensional feature space to obtain an r-dimensional vector

$$\mathbf{h}_k = \mathbf{W}^{+}\xi_k \tag{12}$$

where $\mathbf{W}^{+} = (\mathbf{W}^T\mathbf{W})^{-1}\mathbf{W}^T$. Each of these feature vectors is then used as a prototype feature point. A query face to be classified is represented by its projection into the space as

$$\mathbf{h}_q = \mathbf{W}^{+}\mathbf{v}_q \tag{13}$$

4.3 Nearest neighbor classification

The next step is to compute the distance $d(\mathbf{h}_q, \mathbf{h}_k)$ between the query and each prototype. The query is classified as the class to which the closest prototype belongs. For the experiments carried out in this paper, recognition rates were obtained using three different distance metrics: L_1-metric, L_2-metric, and NCC. These are calculated as follows:

1. *L_1-metric:* Also known as the sum norm. It sums the absolute differences between the two vectors and is given by

$$d(\mathbf{h}_q, \mathbf{h}_k) = L_1(\mathbf{h}_q, \mathbf{h}_k) = \sum_{j=1}^{r} |\mathbf{h}_q - \mathbf{h}_k| \tag{14}$$

2. *L_2-metric:* This is also known as the Euclidean norm or the Euclidean distance when its square root is calculated. It sums up the squared differences between two vectors and is given by

$$d(\mathbf{h}_q, \mathbf{h}_k) = L_2(\mathbf{h}_q, \mathbf{h}_k) = \|\mathbf{h}_q - \mathbf{h}_k\| = \sqrt{\sum_{j=1}^{r} (\mathbf{h}_q - \mathbf{h}_k)^2} \tag{15}$$

3. **NCC:** The basic principle behind this distance metric is to compare two vectors of the same size by measuring their correlation. Correlation of any two vectors can lie in the range $[-1, +1]$, where '+1' indicates maximum correlation and '−1' indicates that

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

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

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] and for illumination, as in [13].

5.1 Face databases

The four databases used are the AR database, the FERET database, the Cambridge ORL database, and the Yale B database. One of the main motivations for the choice of the Cambridge ORL and AR databases is that Li et al. [6] and Guillamet et al. [14] have used them for their part-based experiments with NMF and LNMF, respectively. We note that no results have yet been reported in the literature using the NNSC technique.

5.1.1 Aleix–Robert (AR) database

The AR database was collected at the Computer Vision Center in Barcelona and contains images of 116 individuals (63 males and 53 females). The original images are 768 × 576 pixels in size with a 24-bit color resolution. This database is interesting because subjects were recorded twice at a 2-week interval in which each session captured 13 conditions with varying facial expressions, illumination, and occlusion. For example, some individuals are captured wearing glasses during the first session but not in the next. Due to the high dimensionality of the original images, each image was reduced to 181 × 121. The images were also cropped within an elliptical region to remove any influence of the background, hair, and ears. Furthermore, all the images were aligned with the eye positions to remove any effect of the (slight) in-plane rotations.

5.1.2 Face Recognition Technology (FERET) database

The well-known FERET facial database has become the de facto standard for evaluating face recognition technologies. It consists of a total of 14,126 images of 1,199 individuals and 365 duplicate[6] sets of images. The size of each image is 512 × 768 pixels with some images having either a 24-bit or an 8-bit color resolution. Further, for some subjects, more than 2 years have elapsed between the first and most recent sittings, with some subjects being photographed multiple times [8]. The database displays diversity across gender, ethnicity, and age. Since images are acquired during different photo sessions, the illumination conditions, facial expression, and size of the face may vary. This implies a relatively high level of variation in appearance. For example, some individuals are captured wearing glasses during the first session but not in the next. Each image was reduced to 181 × 121 due to the high dimensionality of the original images. As stated earlier, the images were also cropped to remove any influence of the background, hair, and ears. Also, all the images were aligned with the eye positions to remove any effect of the (slight) in-plane rotations.

5.1.3 Yale B database

The Yale B database contains 5,760 single light source images of ten subjects each seen under 576 viewing conditions (9 poses × 64 illumination conditions). For every subject in a particular pose, an image with ambient (background) illumination was also captured. Hence, the total number of images is 5,850 $(5,760 + 90)$. This database has subjects of different ages and races, with different hairstyles and facial hair. The illumination direction contains the azimuth and elevation of the single light source direction. The azimuth angle varies from +130° to −130° while the elevation angle varies from +19° to +40°.

5.1.4 Cambridge ORL database

The Cambridge ORL face database contains ten different images of 40 distinct subjects. The size of each image is 112 × 92. The images are taken at different times, with only slightly varying illumination, different facial expressions (open/closed eyes, smiling/non-smiling), and facial details (glasses/no-glasses). All the images are

[5] In-plane rotation normalization is achieved by first aligning the left and right eyes so that they lie on a line parallel to the horizontal axis. Next, each image is resized so that the distance between the eyes is always 50 pixels. Finally, each image is cropped with an elliptical mask so that only the face is visible and the hair, ears, and background are all removed.

[6] A duplicate set is a second set of images of a person already in the database and which were usually taken on a different day [8].

taken against a dark homogeneous background. The faces are in an upright position in frontal view, with slight left–right out-of-plane rotation. This database includes both males and females. Each image was linearly stretched to the full range of pixel values [0, 255].

5.2 Experiments using the AR database

In this section, we apply sub-space learning algorithms that are part-based in nature (NMF, LNMF, and NNSC) to the problem of face recognition by first learning the specific facial features of each individual. The methods are then compared for their ability to perform face recognition. Also, the best distance metric (L_1-metric, L_2-metric, and NCC) for nearest neighbor classification for each of these methods is proposed. A total of four experiments were performed: recognition across expression, illumination, occlusion with scarf, and occlusion with sunglasses.

The results obtained are also compared with two leading techniques, FaceIt and Bayesian, used in the computer vision community. FaceIt is a commercial face recognition system and is primarily based on the Local Feature Analysis (LFA) [15]. The Bayesian technique was developed by Moghaddam and Pentland [16] in order to model large non-linear variations in facial appearance due to self-occlusion and self-shading, and uses a PCA approach [17] as a probability density estimation tool. Gross et al. [18] present a detailed comparison of these two techniques on the AR database. The methods were compared by considering a wide set of situations: natural occlusions, changes in expressions, changes in lighting conditions.

5.2.1 Learning basis components

Part-based (NMF, LNMF, NNSC) representations were obtained for 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 150, and 200 basis components.[7] This was done to determine how the results are affected by the number of dimensions of the feature space. Guillamet et al. [14] limit their dimensionality experiments to only 50, 100, and 200.

Each approach (NMF, LNMF, NNSC) was analyzed as a face classifier for all dimension sizes. The training images consisted of two neutral poses of each individual that were captured on two separate days (AR 01). A set of NNSC basis images of dimensionality 20, 50, and 100, is shown in Fig. 1. It can be seen that the basis components are both sparse and part-based.

5.2.2 Results

In order to determine how each technique deals with expressions, illumination, occlusion with sunglasses, and occlusion with scarf, four major experiments were performed. For the expression experiment, images labeled as AR 02, AR 03, and AR 04 were used for testing because they contain smile, anger, and scream expressions, respectively. Images labeled AR 05, AR 06, and AR 07 were used as test images for the illumination experiment as they have neutral expressions with varying illumination. For the occlusion experiments with sunglasses, images labeled AR 08, AR 09, and AR10 were used, whereas the occlusion experiments with scarf used AR 11, AR 12, and AR 13. The recognition accuracy, defined as the percentage of correctly recognized faces, was taken as the performance measure.

5.2.3 Expression

The recognition accuracy obtained by the part-based techniques as a function of basis dimension for different distance metrics for smile, anger, and scream expression are shown in Figs. 3, 4, and 5, respectively. For all these experiments, the recognition rates for NNSC, NMF, and LNMF, improve as the basis dimension increases, eventually saturating. However, for NNSC, the rates tend to improve at a much faster rate giving much higher recognition accuracies for relatively fewer basis dimensions. This explanation for this phenomenon is detailed in Sect. 5.2.7.

NCC seems to be the best distance measure for NMF. This contradicts the results in Guillamet et al. [14] who suggested that the L_1-metric was the best choice. For LNMF, the L_1-metric is the best choice. NNSC gives the best recognition rates (in some cases reaching 100%) for both expressions. Accuracies of up to 98 and 87% are obtained by NMF and LNMF, respectively, for the smile and anger expressions. With FaceIt, 96 and 93%, respectively, were obtained for these expressions while the Bayesian approach performed poorly with rates of 72 and 67%. The PCA technique performed below average relative to the part-based methods with rates of 83 and 90%.

[7] As indicated in Sect. 3, the rank r of factorization is generally chosen lower than the upper bound $(n+m)r < nm$ where n is the dimension of each image (number of pixels) and m is the number of training images in a given dataset. Thus, for the AR datasets: $n = 121 \times 181$, $m = 116$, giving $r < 115$. Similarly, for the Yale B dataset: $n = 181 \times 121$, $m = 10$ (only the neutral pose is used for training) giving $r < 9$. Hence, for experiments with the AR dataset, a basis dimension up to 115 would have sufficed and similarly for the Yale B dataset, a basis dimension up to 9 would have been appropriate. However, to observe the effect on the recognition rates for $r > r_{\text{optimum}}$, we have extended the experiments to 200 for the AR database and 30 for the Yale B database.

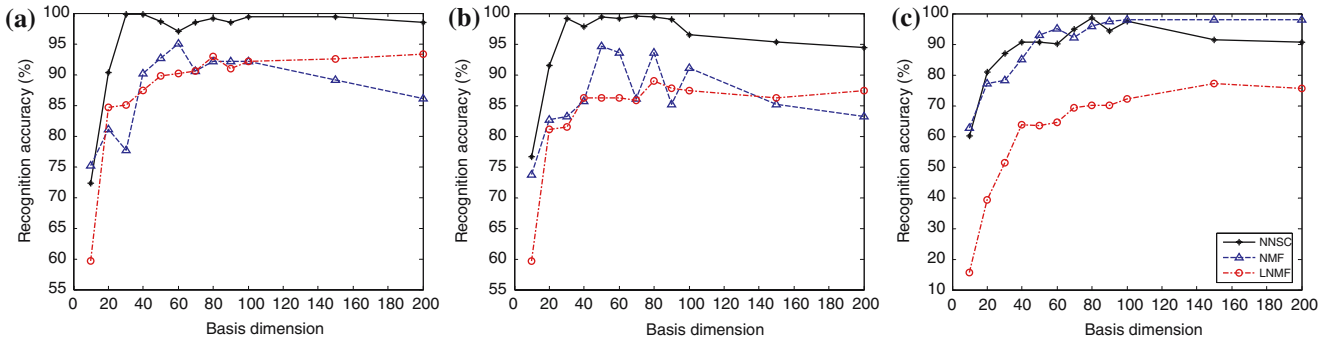

Fig. 3 Smile expression results (AR 02). Comparison of the part-based representation techniques (NMF, LNMF, and NNSC) with different distance metrics (L_1-metric, L_2-metric, and NCC) over a varying number of basis dimensions

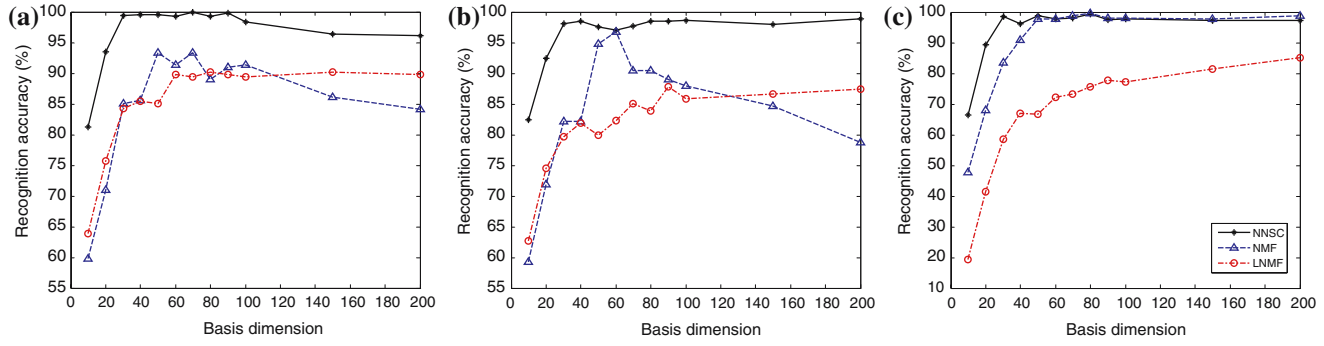

Fig. 4 Anger expression results (AR 03). Comparison of the part-based representation techniques (NMF, LNMF, and NNSC) with different distance metrics (L_1-metric, L_2-metric, and NCC) over a varying number of basis dimensions

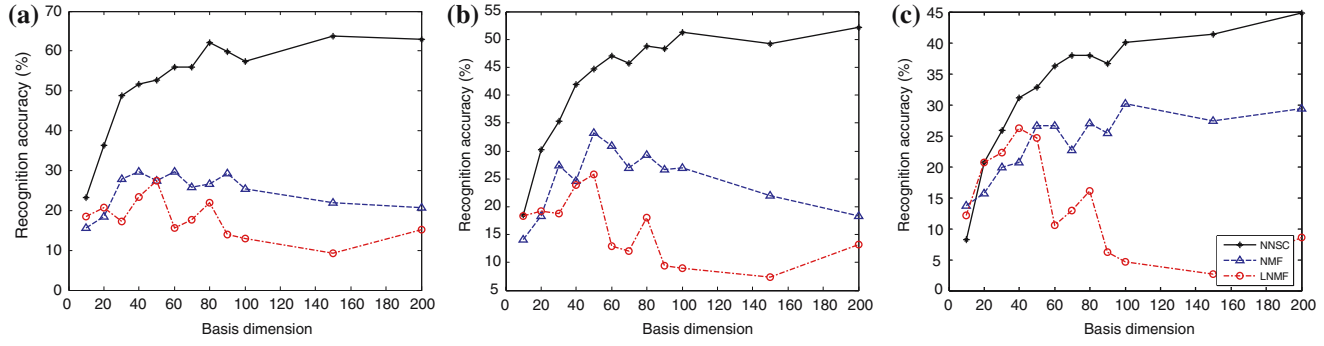

Fig. 5 Scream expression results (AR 04). Comparison of the part-based representation techniques (NMF, LNMF, and NNSC) with different distance metrics (L_1-metric, L_2-metric, and NCC) over a varying number of basis dimensions

The results for the scream expression demonstrate that it is a very difficult image to recognize. FaceIt gave the best results in this case. NNSC, although not as high, yielded 63% recognition accuracy, which is much higher than Bayesian, PCA, and the other (NMF and LNMF) part-based techniques. These results are summarized in Table 1. Note that the rates given in Table 1 reflect the "best" recognition rates obtained for the respective technique. They do not indicate which combination of distance measure and number of basis dimensions produced this rate. For example, with NNSC, the highest accuracy (100%) is obtained for a basis dimension of 80 with the L_1-metric. However for NMF, the highest accuracy (99%) is obtained with NCC as a distance measure and a basis dimension of 200.

Table 2 provides a summary of which distance metric is best suited for which part-based techniques for the expression considered. A measure of "1", "2", or "3" is assigned to a technique for a given distance metric, with "1" being the best and "3" the worst.

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), %
FaceIt	96	93	78
Bayesian	72	67	41
PCA	83	90	29
NMF	98	99	30
LNMF	93	90	27
NNSC	100	100	63

Table 2 A summary of the 'best' distance measures (see text for explanation) for the part-based techniques with 1 being the best choice and 3 the worst

Technique	Best distance metric (in the order 1, 2, 3) for expressions								
	AR 02			AR 03			AR 04		
	L_1	L_2	NCC	L_1	L_2	NCC	L_1	L_2	NCC
NMF	2	3	1	2	3	1	3	2	1
LNMF	1	2	3	1	2	3	1	2	3
NNSC	1	2	3	1	2	3	1	2	3

5.2.4 Illumination

Illumination conditions are an important factor to take into account in a face recognition framework. These conditions include: illumination from an azimuth of $+90°$ (right side), $-90°$ (left side), and $0°$ (frontal lighting) and are labeled as AR 05, AR 06, and AR 07.

Examination of the recognition rates obtained by the part-based techniques as a function of basis dimension (for different distance metrics) leads to the conclusion that these rates are almost identical with the rates obtained for the smile expression (AR 02) images discussed in Sect. 5.2.3. Table 3 provides a summary of the recognition accuracies achieved by the part-based techniques as well as by PCA, Bayesian, and FaceIt.[8] Bayesian and PCA performed the worst. This implies that they (Bayesian and PCA) cannot deal with illumination changes as well as the part-based techniques. The rates obtained by FaceIt are good, but when compared with the part-based techniques they are lower by about 2–10%. Rates obtained by NNSC and NMF are the best in this category, with LNMF lower by 2% in the case of AR05 and 10% for AR 07, but still comparable with and better than FaceIt.

Table 3 A summary of recognition rates with varying illumination

Technique	Recognition accuracy		
	Left light (AR 05), %	Right light (AR 06), %	Both lights (AR 07), %
FaceIt	95	93	86
Bayesian	77	74	72
PCA	86	87	71
NMF	99	99	98
LNMF	97	98	86
NNSC	99	99	96

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

Again, we observe that NNSC performs better and obtains very high recognition rates with a low number of basis dimensions. This is not the case with NMF and LNMF (or even PCA) where the recognition results improve with an increase in basis dimension (eventually saturating) but at a much lower rate. These observed trends and the obtained recognition rates are consistent with the experiments in Sect. 5.2.3. This issue will be discussed in Sect. 5.2.7.

In addition, we note that the best distance measure for NMF in the case of varying illumination is the L_1-metric and for LNMF the L_2-metric, whereas for NNSC it is NCC. These results are summarized in Table 4.

Table 4 A summary of the 'best' distance measures for the part-based techniques with 1 being the best choice and 3 the worst

Technique	Best distance metric (in the order 1, 2, 3) for illumination								
	AR 05			AR 06			AR 07		
	L_1	L_2	NCC	L_1	L_2	NCC	L_1	L_2	NCC
NMF	2	1	3	1	2	3	2	1	3
LNMF	2	3	1	1	2	3	2	1	3
NNSC	2	3	1	3	2	1	2	3	1

5.2.5 Occlusion with sunglasses

In this section, we experiment with a set of faces subject to occlusion by sunglasses (AR 08). In this case, the subjects are wearing sunglasses that occlude both eyes, including a significant region around the eyes as well. AR 09 and AR 10 are similar except with left and right illumination, respectively.

The recognition rates obtained by the part-based techniques as a function of basis dimension (for different distance metrics) are shown in Figs. 6, 7, and 8, respectively.

[8] As the recognition rates for the Bayesian, FaceIT, and PCA methods are reported in [14], but the method used for illumination compensation is not indicated.

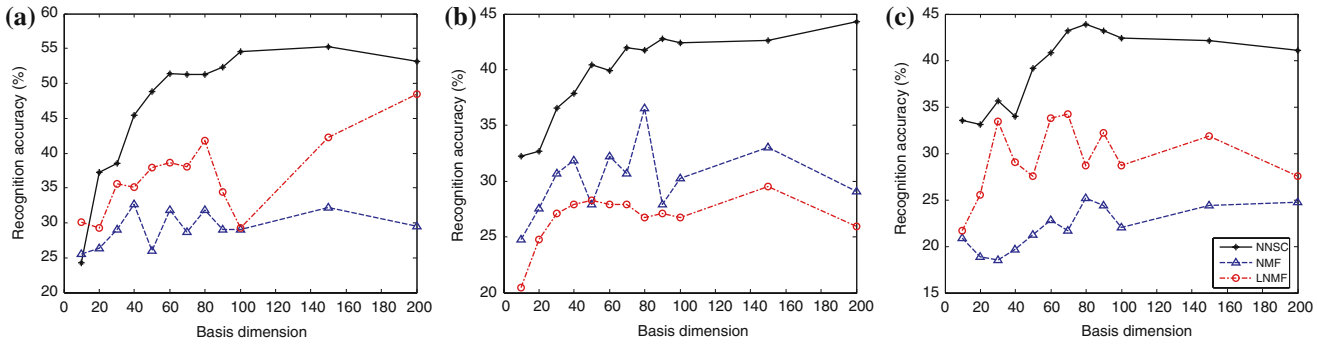

Fig. 6 Occlusion (sunglasses) results (AR 08). Comparison of the part-based representation techniques (NMF, LNMF, and NNSC) with different distance metrics (L_1-metric, L_2-metric, and NCC) over a varying number of basis dimensions

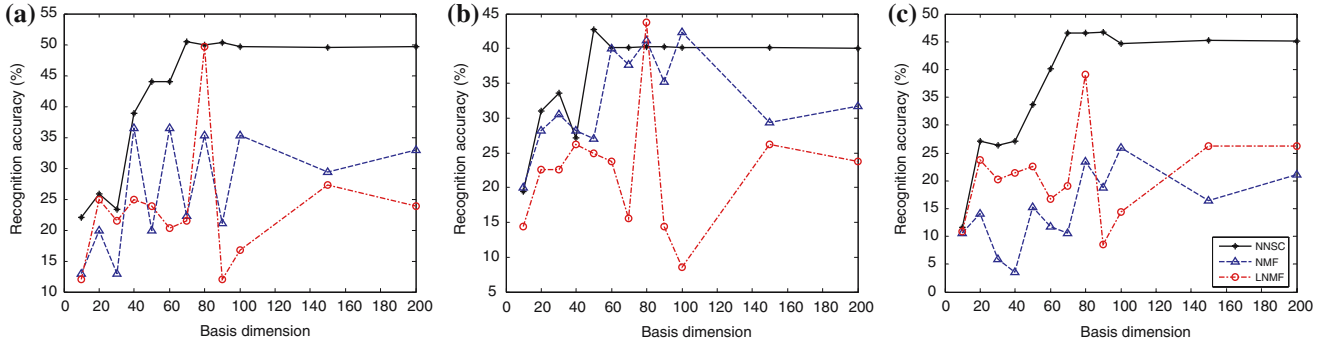

Fig. 7 Occlusion (sunglasses) results (AR 09). Comparison of the part-based representation techniques (NMF, LNMF, and NNSC) with different distance metrics (L_1-metric, L_2-metric, and NCC) over a varying number of basis dimensions

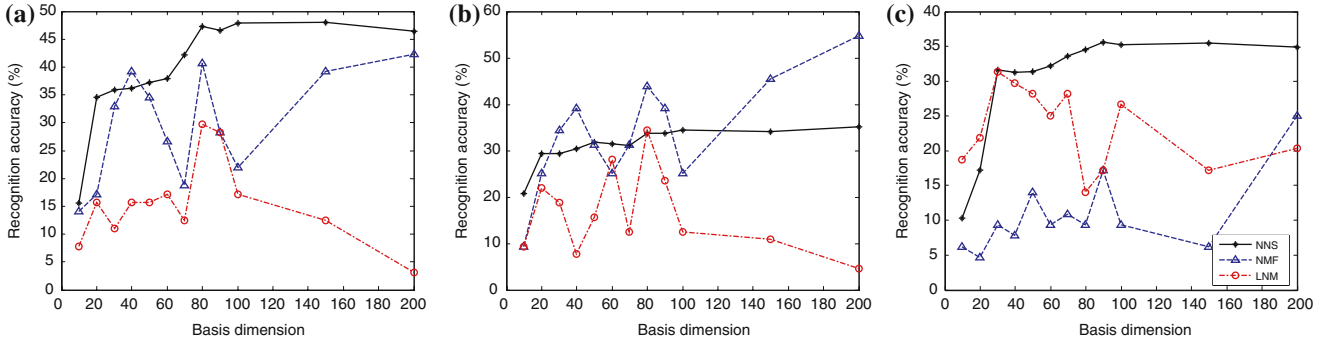

Fig. 8 Occlusion (sunglasses) results (AR 10). Comparison of the part-based representation techniques (NMF, LNMF, and NNSC) with different distance metrics (L_1-metric, L_2-metric, and NCC) over a varying number of basis dimensions

As observed in the graphs, the presence of sunglasses significantly reduces the recognition rates, corroborating the importance of eyes for classifying faces.

As was the case with non-occluded experiments (Sects. 5.2.3 and 5.2.4), the recognition accuracy for NNSC, NMF, and LNMF, increases as the number of basis dimensions increase, and eventually saturate (at a relatively low accuracy when compared to the non-occluded cases) However, NNSC achieves relatively higher recognition rates for a smaller number of basis dimensions . This issue will be discussed in Sect. 5.2.7.

Table 5 presents a summary of the recognition rates obtained by each technique. An immediate conclusion can be drawn about PCA and the part-based techniques by considering the rates. The part-based techniques have a much higher recognition rate than PCA. This behavior is expected and is consistent with the theory that PCA is based on a global transformation of the original space and the part-based techniques are local in nature. Thus, it turns out that when considering local effects, such as occlusion, changes in expression or even changes in illumination, PCA is not able to represent them as well

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

In this set of experiments, faces are occluded by a scarf (AR 11) that covers the mouth region. AR 12 and AR 13 consider the same situation but including left and right illumination.

The recognition rates by the part-based techniques for this case as a function of basis dimension (for different distance metrics) indicate that the recognition rates are much lower than when no occlusion exists. Furthermore, the behavior of the rates obtained by NNSC as a function of basis dimension is consistent with partial occlusion due to sunglasses (Sect. 5.2.5).

The rates obtained by FaceIt are the best in this set of experiments, whereas Bayesian gives the lowest rates. It is interesting to note here that as the number of basis dimensions increases, the rates obtained by LNMF and NNSC stay more or less constant over the whole range. However, the recognition accuracy for NMF increases with basis dimension. When the lighting in the scene is not extreme (AR 11), NMF can have a high recognition rate even compared to FaceIt, but when it is severe (AR 12 and AR 13), these recognition rates decrease. This implies that part-based methods cannot deal with more global changes in scene illumination as well as FaceIt. Table 7 shows a summary of the recognition rates obtained by each technique. These results are summarized in Table 8.

5.2.7 Summary of results using the AR database

In this section, we have applied sub-space learning algorithms that are part-based in nature (NMF, LNMF, and NNSC) to the problem of face recognition using the AR facial database. Experiments were also performed with respect to the distance metric [L_1-metric, L_2-metric, and Normalized Cross-Correlation (NCC)] for nearest neighbor classification and the best method was found in each case. The results obtained were compared with PCA as well as to two leading techniques used by the computer vision community: FaceIt and Bayesian. A total of four experiments were performed: recognition

Table 7 A summary of recognition rates with occlusion (with scarf) and varying light direction

Technique	Recognition accuracy		
	Scarf (AR11), %	Scarf/left light (AR 12), %	Scarf/right light (AR13), %
FaceIt	81	73	71
Bayesian	46	43	40
PCA	62	47	48
NMF	78	76	59
LNMF	69	51	51
NNSC	75	52	55

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

Table 6 A summary of the 'best' distance measures for the part-based techniques with 1 being the best choice and 3 the worst

Technique	Best distance metric (in the order 1, 2, 3) for sunglass occlusion								
	AR 08			AR 09			AR 10		
	L_1	L_2	NCC	L_1	L_2	NCC	L_1	L_2	NCC
NMF	2	1	3	2	1	3	2	1	3
LNMF	1	2	3	1	2	3	1	2	3
NNSC	1	3	2	1	3	2	1	3	2

Table 8 A summary of the 'best' distance measures for the part-based techniques with 1 being the best choice and 3 the worst

Technique	Best distance metric (in the order 1, 2, 3) for scarf occlusion								
	AR 11			AR 12			AR 13		
	L_1	L_2	NCC	L_1	L_2	NCC	L_1	L_2	NCC
NMF	1	2	3	1	2	3	1	2	3
LNMF	1	2	3	1	3	2	1	2	3
NNSC	2	1	3	1	2	3	1	2	3

across expression, illumination, occlusion with sunglasses, and occlusion with scarf.

It was found that part-based techniques perform much better for smiling and angry faces (with NNSC being the best) than FaceIt, Bayesian, and PCA. However, for images that portray a scream, FaceIt obtained the best results, with NNSC being the second best. The best distance measure for NMF is NCC, whereas for LNMF and NNSC it is the L_1-metric. Also, we note here that the high recognition accuracies obtained by NNSC are with a relatively *small* number of basis dimensions (less than 80). This is not the case with NMF and LNMF (or even PCA) where the recognition results improve with an increase in the number of basis dimensions[9] but at a much slower rate.

To explain why a relatively small number of basis components yield high recognition rates for NNSC, that is, why $r_{\text{optimum}}^{\text{NNSC}} << \left\{r_{\text{optimum}}^{\text{NMF}}, r_{\text{optimum}}^{\text{LNMF}}\right\}$, where r_{optimum} is given by Eq. 3, consider the objective function, $\Theta_{\text{NNSC}}(\mathbf{W},\mathbf{H})$, in Eq. 4. In this equation, the parameter λ controls the tradeoff between accurate reconstruction and sparseness. With $\lambda \to 0$, Eq. 4 reduces to the same objective function used for NMF. Hence, it should be noted that for NNSC, after the optimization, $\Theta_{\text{NNSC}}(\mathbf{W},\mathbf{H})$ is not close to zero, even if $\mathbf{V}$ may be equal to the product of $\mathbf{WH}$. That is, $\Theta_{\text{NNSC}}(\mathbf{W},\mathbf{H}) > \{\Theta_{\text{NMF}}(\mathbf{W},\mathbf{H})\Theta_{\text{LNMF}}(\mathbf{W},\mathbf{H})\}$ which is not the case for NMF and LNMF, where $\mathbf{V} \approx \mathbf{WH}$ after the optimization process. Also, it is to be noted that this "extra" additive term in Eq. 4 has an outer summation "controlled" by the value of r. It follows that the upper bound of r for the case of NNSC can then be given by making a small modification to Eq. 3 such that

$$r < \varphi\frac{nm}{n+m}; \quad 0 < \varphi \leqslant 1 \tag{17}$$

where $\phi 1$ is an unknown scalar. The upper bound obtained by Lee and Seung (Eq. 3) occurs when $\mathbf{V} - \mathbf{WH} \approx 0$, but because for NNSC, $\mathbf{V} - \mathbf{WH} > 0$, the scalar ϕ comes into play. This is also consistent with the definition of sparseness in footnote 3 which states that "... it favors representations where only a *few* coefficients $\mathbf{h}_1, \ldots, \mathbf{h}_r$ are significantly active (non-zero) for any given input $\mathbf{v}$". The interested reader is referred to Fig. 7.2 of [11] for more details on sparseness.

Having explained why only a small number of basis components is required to achieve high recognition rates for NNSC, we now consider why the recognition accuracies saturate for a larger number of basis components ($r > r_{\text{optimum}}$).

Rewriting Eq. 4 as $\Theta_{\text{NNSC}}(\mathbf{W},\mathbf{H}) = A + \lambda \cdot B$, we can plot the terms A (reconstruction component) and $\lambda \cdot B$ (sparseness component) as a function of the number of basis dimensions obtained after optimization (learning the basis components) using the update rules given in Eqs. 5, 6, and 7. These plots are depicted in Fig. 9. It can be observed that as the number of basis dimensions increases, the terms A and $\lambda \cdot B$ attain their steady state values for just a *small* number of basis dimensions. Moreover, there is no further improvement with an increasing number of basis dimensions. The consequence of this is shown is Fig. 9c, where the objective function has negligible improvement for $r > r_{\text{optimum}}$, that is $\Theta_{\text{NNSC}}(\mathbf{W},\mathbf{H})|_r \approx \Theta_{\text{NNSC}}(\mathbf{W},\mathbf{H})|_{r_{\text{optimum}}}$ Therefore, the resulting basis components learnt by optimization for $r > r_{\text{optimum}}$, will show no more improvement, giving $\mathbf{W}|_r \approx \mathbf{W}|_{r_{\text{optimum}}}$. Hence, we can conclude that the recognition accuracies tend to saturate for $r > r_{\text{optimum}}$.

The recognition rates obtained by varying illumination with FaceIt, Bayesian, and PCA, indicate more or less the same trends as those obtained for the expression experiments. Bayesian and PCA cannot deal with illumination variations as well as the part-based techniques or FaceIt, which may explain why they (Bayesian and PCA) provide the poorest rates. The rates obtained by FaceIt are good, but when compared with the part-based techniques, they are slightly lower by about 2–10%. Rates obtained by NNSC and NMF are the best.

In the case of partial occlusion, the part-based techniques have much higher recognition rates than PCA. This behavior is expected and is consistent with the theory that the PCA technique is based on a global transformation of the original space and the NMF, LNMF, and NNSC are part-based and local in nature. As expected, using non-uniform illumination degrades the results for the latter.

We note that, as expected, the recognition rates for occluded faces are lower than those obtained for non-occluded ones. An important point in this regard is that *training* for all experiments discussed in this paper was done under strict conditions; that is, only one image per person in neutral pose and expression, and under ambient lighting. Therefore, when considering the occluded case, the faces used for *testing* the recognition performance lacked certain facial features.

When considering occlusion with sunglasses, the recognition rates obtained are low but again, those achieved by NNSC are relatively better. In this experiment, the L_2-metric was the most suited for NMF whereas for LNMF and NNSC, the L_1-metric was best. For occlusions as a result of a scarf, FaceIt obtained the best

[9] This behavior of the recognition rates versus the number of basis dimensions in the case of NNSC is also consistent with the experiments performed on the FERET, Cambridge ORL, and YaleB face databases.

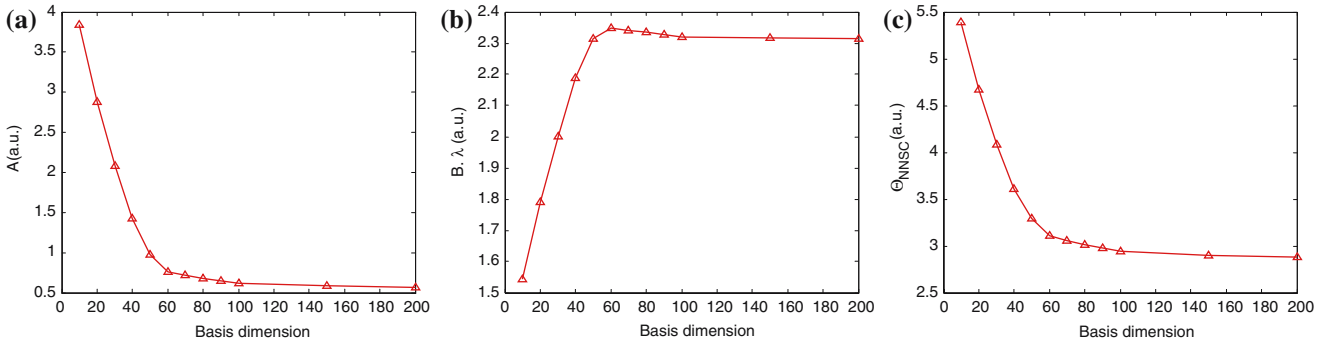

Fig. 9 9 Plots of the **a** A (reconstruction component) and **b** λ. B (sparseness component) of Eq. 4 as a function of the number of basis dimensions; **c** is the optimized objective function of Eq. 4 (that is, the sum of **a** and **b**)

rates and Bayesian the lowest. Comparable results were obtained by the part-based techniques with NNSC being the closest. Here, the L_1-metric was the best-suited distance measure for the part-based methods.

5.3 Experiments on the FERET database

As earlier mentioned, the FERET database is used as a de facto standard for evaluating face recognition technologies. For the purposes of experimentation, the data used for the experiments described here consists of 1,010 subjects in the frontal pose. Each subject appears in the database as a pair ("fa" and "fb"), and can appear more than once ("duplicate"). There is a total of 1,762 images in the "fa" pose and 1,518 images in the "fb" pose which were partitioned into training and testing sets. The training set contained only one image per subject in the "fa" (neutral) pose while the rest of the images corresponding to the subject were used in the testing phase.

Part-based representations with 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 150, and 200 basis components were computed using the training set. Figure 10 shows the recognition rates obtained by each of the NMF, LNMF, and NNSC part-based learning techniques with different metrics as a similarity measure for a varying number of basis dimensions.

From the plots it can be observed that the recognition rate obtained by the NNSC technique is as high as 97%, whereas NMF and LNMF yield maximum recognition rates of 94.48 and 96.32%, respectively. As was observed with the experiments on the AR database, the NNSC technique gives a relatively high recognition rate for a fewer number of basis components. The L_1-metric was found to be the most suitable for part-based techniques.

5.4 Experiments on the Yale B database

The Yale B database contains images of 10 individuals in nine poses. The experiment was performed under constant ambient illumination conditions, neutral expression, and no occlusion. Poses 1, 2, 3, 4, and 5 are about 12° off the optical axis (i.e., from Pose 0), while poses 6, 7, and 8 are about 24°. The images were preprocessed to remove the background, hair, and ears. Other poses were also generated by taking mirror images of poses 2, 3, 4, 6, 7, and 8 to obtain poses 2', 3', 4', 6', 7', and 8'. The frontal view (pose 0) was selected for training only, while the rest of the poses were used for testing.

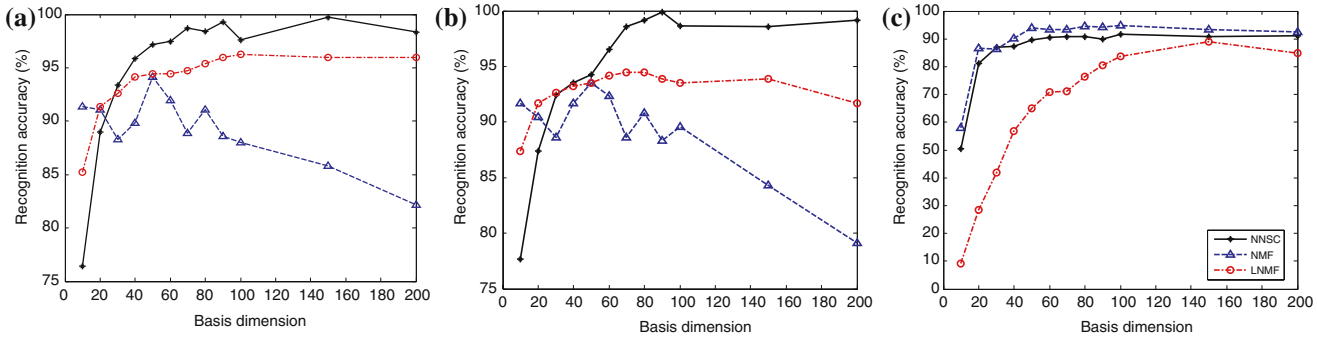

Fig. 10 Face recognition results on the FERET database. Comparison of the part-based representation techniques (NMF, LNMF, and NNSC) with different distance metrics (L_1-metric, L_2-metric, and NCC) over a varying number of basis dimensions

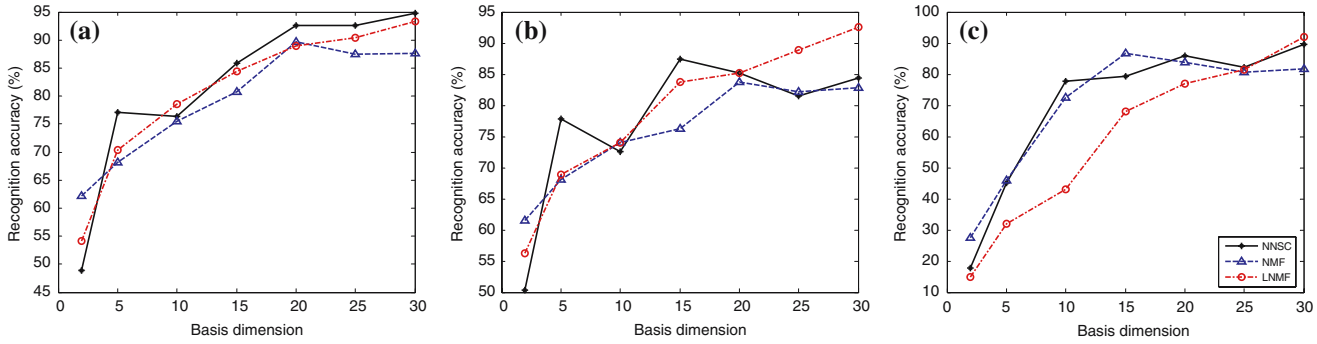

Fig. 11 Face recognition results on the Yale B database. Comparison of the part-based representation techniques (NMF, LNMF, and NNSC) with different distance metrics (L_1-metric, L_2-metric, and NCC) over a varying number of basis dimensions

Figure 11 shows the recognition rates obtained by each of the three part-based learning techniques using different similarity metrics and over varying number of basis dimensions. We observe that NNSC outperformed NMF and LNMF and obtained recognition rates as high as 94.81%, whereas NMF and LNMF yielded 89.63 and 93.33%, respectively.

5.5 Experiments on Cambridge ORL database

The Cambridge ORL database is relatively simple in the sense that it contains images of individuals with not much variation in head pose, expression, and the images are taken under the same (ambient) lighting conditions. However, these experiments were performed to validate our results and for comparative analysis. The only form of preprocessing done on this database was geometric normalization and illumination compensation. Li et al. [6] used the Cambridge ORL database for their experiments with the LNMF technique and compared it to NMF. We have performed the same experiments and a comparison was also made with the NNSC technique.

As mentioned earlier, the database contains ten images for each of the 40 subjects. The set of ten images for each person was randomly partitioned into a training subset of five images and a test set with the other five. The training set was then used to learn the basis components, whilst the test set was used for evaluation.

Part-based representations with 25, 36, 49, 64, 81, 100, 121 basis components were computed using the training sets. Li et al. chose this same set of basis dimensions for their face recognition experiments using LNMF. However, they do not explain why these specifically were selected. Nevertheless, for comparison purposes, we have used the same set for our experiments. The recognition rate, defined as a percentage of correctly recognized faces, was used as the performance measure.

Li et al. obtained recognition rates of around 95 and 91% using the LNMF and NMF technique, respectively, on this same database. These rates were more or less constant over all basis dimensions. From the experiments we have conducted, NNSC produced recognition rates of 96.7% for fewer basis components. However, as the number of basis components was increased, the rates tended to saturate for a higher number of basis components. This trend is also consistent with the experiments performed on the AR, FERET, and Yale B databases. We also found that the L_1-metric is the most suitable for part-based techniques when tested on the Cambridge ORL database.

6 Conclusions

The problem for automatic machine recognition of human faces has been addressed in this paper. We have designed and implemented a system that is based on learning facial features that provide a part-based description. A subspace approach was employed since it helps reveal the low dimensional structures of the patterns observed in a high dimensional space. We have shown that this dimensionality reduction is achieved by projecting a face that is essentially in a high dimensional space to a low dimensional feature space by using a linear type of mapping. Furthermore, reconstruction from the lower dimensional space can still be achieved by using a set of basis components. We note that no other literature source has so far reported face recognition results using the NNSC technique.

The research objective reported in this paper was to demonstrate how the NNSC technique is able to learn parts of a face and how it can be used for face recognition. We have tested the part-based techniques: Non-negative Matrix Factorization (NMF), Local Non-negative Matrix Factorization (LNMF), and

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 small number of basis components.

For the experiments with varying pose using the Yale B database, it was found that the L_2-metric provided the best distance measure for all part-based approaches. These yielded high recognition rates, with NNSC being the best with 95%, whereas NMF and LNMF gave accuracies of 90 and 93%, respectively.

Similar results were obtained for the experiment on the Cambridge ORL database. NNSC obtained recognition rates of 97%, whereas NMF and LNMF gave 95 and 91%, respectively. The best distance measure in this case was the L_1-metric for all the part-based techniques.

With regard to recognition rates as a function of basis dimension, two comments may be made. Firstly, as the number of basis dimensions was increased, the recognition accuracy improved and asymptotically converged to a fixed value for NNSC, NMF, and LNMF, for $r > r_{optimum}$. Secondly, NNSC obtained higher recognition accuracies with a relatively smaller number of basis dimensions, that is $r^{\mathrm{NNSC}}_{\mathrm{optimum}} << \left\{ r^{\mathrm{NMF}}_{\mathrm{optimum}} r^{\mathrm{LNMF}}_{\mathrm{optimum}} \right\}$.

The final issue to be addressed is, which method produced the best performance? NNSC was in general the superior approach, although it must be observed that the best distance measure was not consistent with respect to the different experiments. Further experimentation will be required to settle this issue.

Acknowledgments The authors would like to thank the National Sciences and Engineering Research Council of Canada for its financial assistance. They would also like to thank the reviewers for their comments and questions, which ultimately forced us to reconsider our results. The recomputed data led to more meaningful graphs in some cases and are presented in the final version of this paper.

References

1. Hubel, D.H., Wiesel, T.N.: Receptive fields and functional architecture of monkey striate cortex. J. Physiol. **195**, 215–243 (1968)
2. Olshausen, A., Field. D.J.: Emergence of simple-cell receptive field properties by learning a sparse code for natural images. Nature **381**, 607–609 (1996)
3. Bell. A.J., Sejnowski. T.J.: The "Independent Components" of natural scenes are edge filters. Vis. Res. **37**, 3327–3338 (1997)
4. Lee, D.D., Seung, H.S.: Learning the parts of objects by non-negative matrix factorization. Nature **401**,788–791 (1999)
5. Hoyer, P.O.: Non-Negative Sparse Coding. In: Proceedings of IEEE Workshop on Neural Networks for Signal Processing, pp. 557–565 (2002)

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?, Technical Report, Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, USA (2001)

Author Biographies

Bhavin J. Shastri (S'03) was born in Nairobi, Kenya in 1981. He received his Honors B. Eng. degree in Electrical and Software Engineering with distinction from McGill University, Montreal, Canada, in 2004. He is currently a graduate student at the Photonic Systems Group of McGill University and is working towards his M. Eng. Degree in Electrical Engineering. His research interests include computer vision, face recognition, optoelectronics, and VLSI systems.

His undergraduate thesis entitled Face Recognition Using Localized Features was done under the supervision of Dr. Levine. In the summer of 2002, he worked as a Research Assistant at the Center for Intelligent Machines (CIM), McGill University. He worked as a Systems Engineer at InterDigital Communications Corporation, Canada, during his internship year in 2003. In the summer of 2004, he worked as a Research Assistant under the supervision of Dr. David V. Plant at the Photonic Systems Group of McGill University.

Mr. Shastri is the recipient of the IEEE Computer Society Lance Stafford Larson Outstanding Student Paper Award and the IEEE Canada Life Member Award for the Best Student Paper in 2004.

Martin D. Levine (S'59-M'66-SM'74-F'88) received the B. Eng. and M. Eng. degrees in Electrical and Computer Engineering from McGill University, Montreal, in 1960 and 1963, respectively, and the Ph.D. degree in Electrical Engineering from the Imperial College of Science and Technology, University of London, London, England, in 1965. He is currently a Professor in the Department of Electrical and Computer Engineering, McGill University and served as the founding Director of the McGill Center for Intelligent Machines (CIM) from 1986 to 1998. During 1972–1973 he was a member of the Technical Staff at the Image Processing Laboratory of the Jet Propulsion Laboratory, Pasadena, CA. During the 1979–1980 academic year, he was a Visiting Professor in the Department of Computer Science, Hebrew University, Jerusalem, Israel.

His research interests include computer vision, image processing, and artificial intelligence, and he has numerous publications to his credit on these topics. As well, he has consulted for various government agencies and industrial organizations in these areas. Dr. Levine is a founding partner of AutoVu Technologies Inc. and VisionSphere Technologies Inc. for which he was the Chief Scientific Officer. He was a member of the Scientific Board of ART Advanced Research Technologies Inc. and is presently on the Technical and Advisory Committees of AutoVu Technologies Inc.

Dr. Levine has authored the book entitled Vision in Man and Machine and has co-authored Computer Assisted Analyses of Cell Locomotion and Chemotaxis. Dr. Levine is an Area Editor and is on the Editorial Board of the journal Computer Vision and Understanding, having also served on the Editorial Boards of the IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE and Pattern Recognition. He was the Editor of the Plenum Book Series on Advances in Computer Vision and Machine Intelligence. He was the General Chairman of the Seventh International Conference on Pattern Recognition held in Montreal during the summer of 1984 and served as President of the International Association of Pattern Recognition during 1988–1990. He was also the founding President of the Canadian Image Processing and Pattern Recognition Society.

Dr. Levine was elected as a Fellow of the Canadian Institute for Advanced Research in 1984. During the period 1990–96 he served as a CIAR/PRECARN Associate. He is a Fellow of the IEEE and the International Association for Pattern Recognition. Dr. Levine was presented with the 1997 Canadian Image Processing and Pattern Recognition Society Service Award for his outstanding contributions to research and education in Computer Vision.