Appears in the *15th Int'l Conference on Pattern Recognition*, **ICPR'2000**, Barcelona, Spain, September 3-8, 2000.

Learning the Face Space — Representation and Recognition

Chengjun Liu
Dept. of Math and Computer Science
University of Missouri - St. Louis
St. Louis, MO 63121
cliu@cs.umsl.edu

Harry Wechsler
Dept. of Computer Science
George Mason University
Fairfax, VA 22030
wechsler@cs.gmu.edu

Abstract

This paper advances an integrated learning and evolutionary computation methodology for approaching the task of learning the face space. The methodology is geared to provide a framework whereby enhanced and robust face coding and classification schemes can be derived and evaluated using both machine and human benchmark studies. In particular we take an interdisciplinary approach, drawing from the accumulated and vast knowledge of both the computer vision and psychology communities, and describe how evolutionary computation and statistical learning can engage in mutually beneficial relationships in order to define an exemplar (Absolute)-Based Coding (ABC) multidimensional face space representation for successfully coping with changing population (face) types, and to leverage past experience for incremental face space definition.

1. Introduction

Among the most challenging tasks for visual form ('shape') analysis and object recognition are understanding how people process and recognize each other's face, and the development of corresponding computational models and automated face recognition systems [43]. The enormity of the problem has involved hundreds of scientists in interdisciplinary research but the ultimate solution remains elusive. The face space metaphor [41] provides the starting point for this paper. Our goal here is to advance an integrated learning and evolutionary computation methodology for approaching the task of learning the face space. The methodology is geared to provide a framework whereby enhanced and robust face coding and classification schemes can be derived and evaluated using both machine and human benchmark studies. In particular we take an interdisciplinary approach, drawing from the accumulated and vast knowledge of both the computer vision and psychology communities, and describe how evolutionary computation and statistical learning can engage in mutually beneficial relationships in order to define an exemplar (Absolute)-Based Coding (ABC) multidimensional face space representation for successfully coping with changing population (face) types, and to leverage past experience for incremental face space definition. Evolutionary computation, akin to stochastic search and instantiated through Genetic Algorithms (GAs), allows for the computational efficiency needed to sift through the huge number of possible face spaces.

A "fundamental question about cognition concerns how knowledge about a category is acquired through encounters with examples of the category" [24]. Memory-based reasoning supports the view that categorical concepts are implicitly defined in terms of some of the exemplars encountered and it does not call for stored abstractions and/or prototypes [17]. Knowlton and Squire [24] further view that "category-level knowledge has no special status but emerges naturally from item memory", and that a novel probe "would be endorsed as belonging to a particular category as a function of the similarity between the new item and the exemplars of that category already stored in memory". Categorical effects have also been found to affect perception of faces and Beale and Keil [3] have shown that "subjects discriminate most accurately when face-pairs straddle apparent category boundaries along a continuum of morphed faces". Memory-based reasoning leads one directly to the face space characterization, in terms of face exemplars, labeled as Absolute-Based Coding (ABC). Driven by clustering and density measures, ABC considers typicality, like that for ethnicity, to be dependent on the local density surrounding some face(s). The competing alternative for the face space definition, Norm-Based Coding (NBC), defines typicality in terms of the face distance from the (mean) prototype abstracting known faces. Recent experiments reported by Rhodes et al [37] seem to suggest that the ABC model compares favorably against the NBC model, even that it may not provide a complete account of ethnicity effects in recognition.

The questions now begging an answer are how to define faces as (compact and discriminating) categorical self-contained concepts, and how to measure their similarity. Our answer to this question is that the exemplars referred to earlier have to be represented along some face basis, whose "natural" dimensions evolve over time in response to the spatiotemporal statistics of the face space encountered. The generic solution advanced in this paper unifies the Norm-Based Coding (NBC) and Absolute-Based Coding (ABC), the alternative coding strategies competing to define the face space, as the exemplars are encoded along shared ("prototypical") dimensions. The natural dimensions for the face space bridge front-end image processing and memory-based image retrieval, and combine known perceptual ("representation") and cognitive ("classification") face recognition abilities like those discussed by Burton et al [7] in support of their Interactive Activation and Competition (IAC) model of face recognition. The specific objective criteria addressed by this paper and concerning the face space are then related to (i) the face recognizers used to discriminate among faces or possibly categorize them according to gender and/or ethnic origin, and (ii) the kind of preprocessing face images are subject to prior to encoding and classification. The above two criteria correspond to the cognitive and perceptual processes referred to by the IAC model. Most important, and fundamental to this undertaking is to (iii) find the optimal mix of perceptual and cognitive processes needed to define the face space, and to identify the specific characteristics of robust encoding schemes, i.e., the basis along which the preprocessed face images are projected.

2. Background

The idea of learning the face space has been motivated by the natural scene encoding. Encoding natural scenes takes advantage of intrinsic image statistics and seeks also to derive an natural ('universal') basis [21], [32]. The derived basis functions have been found to closely approximate the receptive fields of simple cells in the mammalian primary visual cortex. The receptive fields resemble various derivative-of-Gaussian (DOG) functions, which are spatially localized, oriented and bandpass [32]. Barlow [2] argues that such receptive fields might arise from unsupervised learning, subject to redundancy reduction or minimum entropy coding. Olshausen and Field [32] derive localized oriented receptive fields based on a criterion of sparseness, while Bell and Sejnowski [5] use an independence criterion to derive qualitatively similar results.

The rationale behind a natural basis is that the basis should be complete and low dimensional, and it should thus allow for the efficient derivation of suitable image representations corresponding to the intrinsic structure of sensory signals. The intrinsic structures are essential for processes such as image retrieval and object recognition. Once the natural basis has been derived, no additional training is necessary and both the training and the novel images on future tasks are represented in terms of the already available natural basis. The natural basis, however, also has its drawbacks, i.e., it might be too general to properly encode for a specific task. As for face recognition, the class of objects to be represented is quite specific, human face images, possibly indexed by gender, ethnicity and age, and one thus should seek and learn the face rather than an 'universal' and all encompassing natural basis. This observation also fits with knowledge that the "bias/variance dilemma may be circumvented if we are willing to purposely introduce bias, which then makes it possible to eliminate the variance or reduce it significantly" [22]. Learning low dimensional representations of visual objects with extensive use of prior knowledge has also been recently suggested by Edelman and Intrator [15] who claim that "perceptual tasks such as similarity judgment tend to be performed on a low dimensional representation of the sensory data. Low dimensionality is especially important for learning, as the number of examples required for attaining a given level of performance grows exponentially with the dimensionality of the underlying representation space."

One popular technique capable of deriving low-dimensional representation is Principal Component Analysis (PCA), which has been applied extensively for both face representation and recognition. Kirby and Sirovich [23] showed that any particular face can be economically represented along the eigenpictures coordinate space, and that any face can be approximately reconstructed by using just a small collection of eigenpictures and the corresponding projections ('coefficients'). Applying PCA technique to face recognition, Turk and Pentland [40] developed a well-known eigenfaces method, where the eigenfaces correspond to the eigenvectors associated with the dominant eigenvalues of the face covariance matrix. The eigenfaces thus define a feature space, or "face space", which drastically reduces the dimensionality of the original space, and face detection and identification are then carried out in the reduced space. Under the same PCA scheme, shape and texture ('shape-free' image) coding has recently become prominent for encoding face images in order to encompass more concrete information for face representation [13], [6], [11]. Shape and texture coding, usually used in conjunction with norm based coding, is a two-stage process once the face has been located. Coding starts by annotating the face using important internal and face boundary points. Once these control points are located, they are aligned using translation, scaling and rotation transformations as necessary, and a corresponding mean shape is derived. The next stage then triangulates the annotated faces and warps each face to the mean shape. The first stage yields the shape, while the second stage yields the texture and corresponds to what is known as full anticaricature [13].

3. Face Representation and Classification

Robust indexing and retrieval encoding schemes require both low dimensional feature representations, for data compression and faithful reconstruction purposes, and enhanced discrimination abilities for subsequent image retrieval [31]. Such complementary requirements can be addressed using an approach similar to constrained optimization, where the unifying theme is that of lowering the subspace dimension ('data compression') subject to increased fitness for the discrimination index. The three candidates for possible coding schemes we have recently developed and found most appropriate are the Probabilistic Reasoning Models (PRM) [28], the Enhanced Fisher Models (EFM) [26], and the Enhanced Independent Component Analysis (EICA) [29]. Lowering the subspace dimension for all three coding schemes is done using PCA, possibly whitened. The three coding schemes manipulate the discrimination index using the Bayes classifier, specific trade-offs between PCA and Fisher Linear Discriminant (FLD), and the sparseness or independence of the resulting image representations, respectively. Following PCA data compression, the PRM method employs the Bayes classifier, which yields the minimum error when the underlying probability functions are known, and carry out the recognition task in the compressed subspace using the MAP (maximum a-posteriori) criterion.

The EFM models implement the dimensionality reduction with the goal to balance between the need that the selected PCA eigenvalues account for most of the spectral energy of the raw data and the requirement that the (trailing) eigenvalues of the within-class scatter matrix in the reduced PCA subspace are not too small. A successful face recognition methodology also depends heavily on preprocessing and feature extraction. We have recently shown that an integrated shape (vector) and texture ('shape free') preprocessing method [30], [25], [11] compares favorably against existing methods. In particular we have shown that an integrated shape (vector) and texture ('shape-free' image) preprocessing method [30] is most successful when it is coupled to the classifier referred to earlier. Shape encodes the feature geometry of a face while texture provides a normalized shape-free image by warping the original face image to the mean shape, i.e., the average of aligned shapes. The dimensionalities of the shape and the texture spaces are again first reduced using PCA. The corresponding but reduced shape and texture features are then integrated through a normalization procedure to form augmented features. The dimensionality reduction procedure, constrained by EFM (see above) for enhanced generalization, maintains a proper balance between the spectral energy needs of PCA for adequate representation, and the FLD discrimination requirements, i.e., that the eigenvalues of the within-class scatter matrix do not include small trailing values after the dimensionality reduction procedure, as they appear in the denominator.

Independent Component Analysis (ICA), which has emerged recently as a powerful solution to the problem of blind source separation, seeks a linear transformation to express a set of random variables as linear combinations of statistical independent source variables. The search criterion involves the minimization of the mutual information expressed as a function of high order cumulants. PCA considers the 2nd order moments only and it uncorrelates the data, while ICA provides a more powerful data representation, as it accounts for higher order statistics and it also identifies the independent source components from their linear mixtures (the observables).

3.1. From PCA and Eigenfaces to Probabilistic Reasoning Models (PRM)

PCA is an optimal signal representation criterion that offers reduction of a large set of correlated variables to a smaller number of uncorrelated components. Following its application, one derives an orthogonal projection basis that directly leads to dimensionality reduction and possibly to feature selection. Recently, PCA has been commonly applied to representing and recognizing human faces, and one well-known face recognition method is the Eigenfaces method [40].

The Eigenfaces method defines a face space spanned by the eigenvectors (eigenfaces) corresponding to the leading eigenvalues of the face covariance matrix. The features derived by Eigenfaces are thus specified by the following equation:

$$Y = P^t X \tag{1}$$

where X and Y represent the original face image and the derived features, respectively, and P is a projection matrix consisting of the eigenfaces.

As PCA derives eigenfaces based on the observed variations using all the training samples, it enjoys good generalization abilities for image reconstruction when tested with novel images not seen during training [34]. PCA is, however, an optimal signal representation criterion in the sense of mean-square-error (MSE), and such PCA inspired features do not necessarily provide for good discrimination. Our solution integrates PCA, the optimal representation criterion, with the Bayes classifier, the optimal classification criterion. Such an integration leads to our PRM method [28]. The PRM method applies PCA first for dimensionality reduction with the goal of signal approximation, and then utilizes the within-class scatter to estimate the covariance matrix for each class in order to estimate the conditional pdf. Finally, it applies the MAP rule for classification. The

MAP decision rule optimizes the class separability in the sense of Bayes error and improves on PCA and FLD based methods that apply criteria not related to the Bayes error.

3.2. From Linear Discriminant Analysis (LDA) and Fisherfaces to Enhanced Fisher Models (EFM)

LDA, similar to the Fisher Linear Discriminant (FLD), is yet another commonly used criterion in pattern recognition and recently in face recognition [39], [18], [4]. The LDA derives a projection basis that separates the different classes as far as possible and compacts the same classes as close as possible. One representative LDA/FLD - based method is the Fisherfaces method [4].

The Fisherfaces method specifies a face space by combining the PCA and the FLD projections:

$$Q = RS \tag{2}$$

where R is the PCA projection matrix similar to P in Eq. 1, and S is the FLD projection matrix, which is derived by maximizing the ratio of the between- and within-class scatters in the transformed space [4]. The features defined are as follows:

$$Z = Q^t X \tag{3}$$

Unlike PCA, FLD differentiates between the within- and between-class scatters and derives a class specific feature space. The Fisherfaces, known also as the Most Discriminating Features (MDF) space, is superior to the Eigenfaces encoding scheme, known as the Most Expressive Features (MEF) space for face recognition, only when the training images are representative of the range of face (class) image variations; otherwise, the performance difference between the MEF and MDF is not significant [39]. The FLD procedure, when implemented in a high dimensional PCA space, leads often to overfitting [31]. Overfitting is more likely to occur for the small training sample size scenario, which is the typical one for face recognition [35]. One possible remedy for this drawback is to artificially generate additional data and thus increase the sample size [18]. Another solution, to analyze the reasons for overfitting and propose new models with improved generalization abilities, led to our EFM method [31]. The EFM method first addresses (concerning PCA) the range of principal components used and how it affects performance, and (regarding FLD) the reasons for overfitting and how to avoid it. It presents, then, specific models to improve generalization by balancing the spectral energy criterion for sufficient representation (perception) and the eigenvalue spectral requirement for good generalization. This requires that the selected PCA eigenvalues account for most of the spectral energy of the raw data, while the (trailing) eigenvalues of the within-class scatter matrix in the reduced PCA subspace are not too small. The range of principal components can be specified according to the eigenvalues spectrum. As an example, for one PRM model, we utilized less than 7% of the total principal components.

3.3. Shape and Texture Representations

The feature geometry of a face, shape (vector), comprises a set of control points that underscore the important face features such as eyebrows, eyes, bridge of the nose, nose, mouth, and the contour of the face. The shapes of all the training face images are aligned with respect to one another under the transformations of translation, rotation and scaling, and the mean shape is defined as the average of these aligned shapes. Texture is derived by warping the original face image to the mean shape. Under the constraints of representation and generalization, the shape and texture features are integrated through a normalization procedure to form augmented features [30].

The augmented features encoding both shape and texture information are defined as follows:

$$Y = \left(\frac{Y_1^t}{\| Y_1 \|} \quad \frac{Y_2^t}{\| Y_2 \|} \right)^t \tag{4}$$

where Y_1 and Y_2 are the lower dimensional shape and texture features derived using PCA under the constraints of the EFM for enhanced generalization performance. For comparison purposes we also compute the shape images and masked images. Shape images undergo the same alignment procedure as the shapes do, but preserve the intensity information within the outlines of the faces only. Masked images are derived by first using the centers of two eyes as control points for alignment, and then putting a mask on them.

Our comparative study shows that the augmented shape and texture features carry the most discriminant information followed in order by textures, shape vectors, masked images, and shape images. It also shows that our shape and texture based Enhanced Fisher classifier Model (EFM) performs the best among the Eigenfaces method using the L_1 or L_2 distance measure, and the Mahalanobis distance classifiers using a common covariance matrix for all the classes [25], [12] or a pooled within-class covariance matrix [16].

3.4. Enhanced Independent Component Analysis (EICA)

ICA (Independent Component Analysis) is not restricted to 2nd order statistics and as a result it reduces statistical dependencies and produces a sparse code useful for subsequent pattern discrimination and associative recall [32]. The role ICA plays is to seek non-accidental and sparse feature codes analogue to the goal of sensory systems "to detect redundant features and to form a representation in which these redundancies are reduced and the independent features and objects are represented explicitly" [19].

The ICA of a random vector X factorizes the covariance matrix, $Cov(X)$, into the following form:

$$Cov(X) = F \Delta F^t \tag{5}$$

where Δ is diagonal real positive and F transforms the original random vector X into a new one Z, where $X = FZ$, such that the components of the new random vector Z are independent or "the most independent possible" [10]. To derive the ICA transformation F, Comon [10] developed an algorithm that consists of three operations: whitening, rotation, and normalization.

A comparative assessment regarding ICA sensitivity to the dimension of the space where it is carried out led us to develop the EICA method [29]. Specifically, as a result of the sensitivity analysis regarding enhancing the ICA performance, EICA is carried out in a compressed and whitened PCA space where the small trailing eigenvalues are discarded. As an example, in our experiments we kept 5% - 7% of the total principal components. The reason for this aspect of EICA is that during whitening the eigenvalues of the covariance matrix appear in the denominator and that the small trailing eigenvalues mostly encode noise. As a consequence the whitening component, if used in an uncompressed image space, would fit for misleading variations and thus generalize poorly to new data. We have also assessed the performance of the EICA alone or when combined with other discriminant criteria such as the Bayesian framework or FLD. Discriminant analysis has shown that the ICA criterion, when carried out in the properly compressed and whitened space, performs better than the Eigenfaces and Fisherfaces methods for face recognition, but its performance deteriorates when augmented by additional criteria such as the MAP rule of the Bayes classifier or the FLD. The reason for the last finding is that the Mahalanobis distance embedded in the MAP classifier duplicates to some extent the whitening component, while using FLD is counter to the independence criterion intrinsic to EICA.

4. Exploratory Pursuit

The problem we address now is that of learning the face space(s) from large and diverse populations using evolution as the driving force. The dimensions ('BASIS') for the (ABC) face space, to be evolved using GAs, are such that their "fitness" is enhanced and driven by the classification/discrimination ('cognitive') and representational ('perceptual') factors referred to earlier, and the interplay between the complexity, the cost ("dimension") and the (categorical) density of the face space on one hand, and the trade-offs between faithful face reconstruction ("representation") and the expected classification accuracy ("guaranteed risk") for the face classifier, on the other hand. The quality of the face space can be also driven by the diversity encountered while learning the face space. Characteristic of both co-evolution and active learning methods, challenging training samples could be boosted and thus given extra weight while assessing the fitness of the face space.

The fundamental problem, that of finding the proper mix of cognitive ("classification") and perceptual ("preprocessing") processes, and in the process deriving the optimal projection basis for face encoding, can be addressed using Evolutionary Pursuit (EP) [27]. In analogy to (exploratory) pursuit methods from statistics, EP seeks to learn an optimal face space for the dual purpose of data compression and pattern classification. The challenges that EP has successfully met on limited population types, are characteristic of sparse functional approximation and statistical learning theory. Specifically, EP increases the generalization ability of the face recognizer as a result of handling the trade-off between minimizing the empirical risk encountered during training ("performance accuracy"), and narrowing the predicted risk ("confidence interval") for reducing the guaranteed risk during future testing on unseen probe images. The prediction risk, corresponding to the penalty factor from regularization methods, measures the generalization ability of the object classifier, and it is driven by the regularization index corresponding to class separation. EP implements strategies characteristic of genetic algorithms (GAs) for searching the space of possible solutions in order to determine the optimal projection basis. EP starts by projecting the original images into a lower dimensional and whitened PCA space. Directed but random rotations of the basis vectors in this space are then searched by GAs where evolution is driven by a fitness function defined in terms of performance accuracy ('empirical risk') and class separation ('confidence interval').

Evolutionary computation represents an emerging methodology motivated by the natural selection. Evolution takes place by maintaining one or more populations of individuals, each of them a candidate solution, and competing for limited resources in terms of placing offsprings in future generations. The competition is implemented via selection mechanisms that choose from the dynamically changing populations due to the birth and death of individuals. The selection mechanisms evaluate the fitness value of individuals based on some fitness criteria (fitness functions), while the population evolves via genetic operators that reflect the concept of inheritance (offsprings resemble their parents). When the fitness functions lack an analytical form suitable for gradient descent or the computation involved is prohibitively expensive, as it is the case when the solution space is too large to search it exhaustively, one alternative is to use (directed) stochastic search methods for nonlinear optimization and variable selection. The unique exploration (variations farther away from an existing population) and exploitation (minor variations of the more fit patents) ability of evolutionary computation guided by fitness values has made possible to analyze very complex search spaces.

Learning the face space requires EP to search through a large number of possible subsets of rotated axes in a properly whitened PCA space. The rotation angles (represented by strings of bits) and the axis indicators (indicating whether the axes are chosen or not) constitute the form of the search space whose size (2 to the power of the length of the whole string) is too large to search exhaustively. The number and choice of (non-orthogonal) axes in the subsets and the angles of rotations are evolved using genetic algorithms (GAs). GAs work by maintaining a constant-sized population of candidate solutions known as individuals ('chromosomes'). The power of genetic algorithms lies in their ability to exploit, in a highly efficient manner, information about a large number of individuals. The search underlying GAs is such that breadth and depth — exploration and exploitation — are balanced according to the observed performance of the individuals evolved so far. By allocating more reproductive occurrences to above average individual solutions, the overall effect is to increase the population's average fitness.

Evolution is driven by a fitness function formulated as follows:

$$\varsigma(F) = \varsigma_a(F) + \lambda\varsigma_g(F) \tag{6}$$

Where F encompasses the parameters (such as the number of axes and the angels of rotations defining each chromosome solution) subject to learning, the first term $\varsigma_a(F)$ records performance accuracy, i.e., the empirical risk, the second term $\varsigma_g(F)$ is the generalization index, i.e., the predicted risk, and λ is a positive constant that indicates the importance of the second term relative to the first one. Accuracy indicates the extent to which learning has been successful so far, while the generalization index gives an indication of the expected fitness on future trials. By combining those two terms together with a proper weight factor λ, GA can evolve balanced results with good recognition performance and generalization abilities. The fitness function has a similar form to the cost functional used by the regularization theory [36] and to the cost function used by the sparse coding [32]. The cost functional of the former method exploits a regularization parameter to control the compromise between a term of the solution's closeness to the data and a term indicating the degree of regularization ('quality') of the solution, while the cost function of the latter method uses a positive constant to achieve the balance between a term of information preserving and a term assessing the sparseness of the derived code.

We consider now the application of the EP method to learning the face space for face recognition [27]. The experimental data consists of a FERET subset of 1,107 images corresponding to 369 subjects such that there are three frontal images for each subject. The variety of the subset is such that for the first 200 subjects the third image is acquired at low illumination, while for the remaining 169 subjects the face images are acquired during different photo sessions and the later acquired images are referred to as duplicates. Two images of each subject are used for training with the remaining image for testing. In other words, the training set includes 738 images while the test set 369 images. The images are cropped to the size of 64x96 and the eye coordinates are manually detected. The background is uniform and the face images are not masked. The reasoning behind not masking the face images is that "at least on some occasions, the processing performed by the visual system to judge identity is better characterized as 'head recognition' rather than 'face recognition"' [38]. Masking as it has been usually implemented deletes the face outline, and the effect of such deletions on recognition performance is discussed in our recent paper [30]. Shape-free face recognition methods avoid this problem by using the shape of the outline encoded by a number of control points for subsequent alignment and normalization [11].

Starting from the 30 dimensional PCA space, the EP method derives 26 vectors as the optimal basis for the learned face space. Note that while for PCA the basis vectors have a natural order, this is not the case with the projection basis derived by EP due to the rotations involved during the evolutionary process. The natural order characteristic of the principal components reflects the representational aspect of PCA and its relationship to spectral decomposition. The very first principal components encode global image characteristics, in analogy to low-frequency components. EP, on the other hand, is a procedure geared primarily towards recognition and generalization. It also worth to point out that while PCA derives orthogonal basis vectors, EP's basis vectors are usually not orthogonal. Orthogonality is a constraint for optimal signal representation, but not a requirement for pattern recognition. Actually, non-orthogonality has been known to have great functional significance in biological sensory systems [14].

The recognition performance, for the EP face space learned using 26 basis vectors, yields 92% while testing on unseen face images, unavailable during training, and it compares favorably against two popular face recognition methods, the Eigenfaces and Fisherfaces methods [27]. To assess the statistical significance of the experimental results, we implemented the McNemar's test [1] to determine whether or not there is strong statistical evidence to indicate that the EP method improves the face recognition performance over the Eigenfaces and the Fisherfaces methods. Based on the statistical testing results, we found that the EP method significantly improves the face recognition performance over Eigenfaces and Fisherfaces.

5. Conclusions

This paper has reviewed recent developments regarding face representation and classification. In analogy to exploratory pursuit methods, which seek interesting projections (defined according to some projection indices) of high dimensional data, we have also advocated Evolutionary Pursuit (EP) as a novel evolutionary framework for object retrieval and recognition, and have shown its application to learning the face space, i.e., deriving its system of coordinates. EP allows for different types of bases, as some statistical methods do, but it updates the dictionary of choices ('kernels') simultaneously as neural networks do. The criterion EP exploits is similar in form to the cost functional used by the regularization theory and to the cost function used by the sparse coding. As this criterion underscores the recognition performance and the generalization abilities of the derived basis, EP is different from the basis pursuit method developed by Chen and Donoho [9], whose criterion is based on signal approximation and minimizing the L1 norm of the corresponding coefficients. It is also interesting to note that, in analogy to statistical learning theory [42], the generalization index encompassed in the criterion is conceptually similar to the capacity of the classifier and its use is to prevent overfitting. As a result, EP provides a new methodology for both functional approximation and pattern classification problems.

One important dimension to be further explored is the ability of the face space to accommodate invariance to different view-points and non-rigid facial expressions. The invariance aspect could be addressed by developing corresponding face space "manifolds" suitable for both face indexing and retrieval. One could also consider the real possibility that more than one face space is needed for face classification and recognition, and that semantically loaded attributes related to gender, ethnicity, and age, could determine the characteristics of each such face space. Another relevant question to be addressed in the future is to what extent and in what form motion information can augment in a beneficial way the face space for classification purposes. The derived face space(s) could also provide for a better understanding of the face space metaphor when one needs to define typicality and its density correlate in a high-dimensional space [8].

The interplay between complexity and training performance is of paramount importance for pattern recognition and it has been addressed here using the EP (Evolutionary Pursuit) approach. As an example, when concept formation and categorization is of interest for face recognition, using EP corresponds to the model selection problem in terms of the expected number of projection axes and their specific characteristics. Model selection thus considers robust mixture (of faces) decomposition and their sparse functional approximation. For large search spaces, EP could become the candidate of choice for sparse functional approximation and the derivation of local kernels. Sparse approximation using kernel methods has been recently shown recently to be equivalent to support vector machines (SVM) [20]. Finding equivalences between EP, kernel methods and SVM, and basis pursuit [9] and their relative merits are subject for future research.

References

[1] A. Agresti, editor. *Categorical Data Analysis*. John Wiley & Sons, 1990.

[2] H. Barlow. Unsupervised learning. *Neural Computation*, 1:295–311, 1989.

[3] J. M. Beale and F. C. Kiel. Categorical effects in the perception of faces. *Cognition*, 57:217–239, 1995.

[4] P. Belhumeur, J. Hespanha, and D. Kriegman. Eigenfaces vs. Fisherfaces: Recognition using class specifi c linear projection. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 19(7):711–720, 1997.

[5] A. Bell and T. Sejnowski. An information-maximization approach to blind separation and blind deconvolution. *Neural Computation*, 7:1129–1159, 1995.

[6] D. Beymer. Vectorizing face images by interleaving shape and texture computations. A.I. memo No. 1537, Artifi cial Intelligence Laboratory, MIT, 1995.

[7] M. A. Burton, V. Bruce, and P. Hancock. From pixels to people: a model of familiar face recognition. *Cognitive Science*, 23(1):1–31, 1999.

[8] M. A. Burton and J. R. Vokey. The face-space typicality paradox: Understanding the face-space metaphor. *The Quarterly Journal of Experimental Psychology*, 51A(3):475–483, 1998.

[9] S. Chen and D. Donoho. Basis pursuit. Technical report, Department of Statistics, Stanford University, 1996.

[10] P. Comon. Independent component analysis, a new concept? *Signal Processing*, 36:287–314, 1994.

[11] T. Cootes, G. Edwards, and C. Taylor. Active appearance models. In *Proc. the Fifth European Conference on Computer Vision*, University of Freiburg, Germany, June 2-6, 1998.

[12] I. Craw, N. Costen, and T. Kato. How should we represent faces for automatic recognition? *IEEE Trans. Pattern Analysis and Machine Intelligence*, 21(8):725–736, 1999.

[13] I. Craw and D. Tock. The computer understanding of faces. In V. Bruce and M. Burton, editors, *Processing Images of Faces*. Ablex Publishing Corporation, 1992.

[14] J. Daugman. An information-theoretic view of analog representation in striate cortex. In E. Schwartz, editor, *Computational Neuroscience*, pages 403–424. MIT Press, 1990.

[15] S. Edelman and N. Intrator. Learning as extraction of low-dimensional representations. In D. Medin, R. Goldstone, and P. Schyns, editors, *Mechanisms of Perceptual Learning*. Academic Press, 1990.

[16] G. Edwards, T. Cootes, and C. Taylor. Face recognition using active appearance models. In *Proc. the Fifth European Conference on Computer Vision*, University of Freiburg, Germany, June 2-6, 1998.

[17] W. K. Estes. Concepts, categories, and psychological science. *Psychological Science*, 4(3):143–153, 1993.

[18] K. Etemad and R. Chellappa. Discriminant analysis for recognition of human face images. *J. Opt. Soc. Am. A*, 14:1724–1733, 1997.

[19] P. Foldiak. Forming sparse representations by local anti-Hebbian learning. *Biol. Cybernetics*, 4:165–170, 1990.

[20] F. Girosi. An equivalence between sparse approximation and support vector machines. *Neural Computation*, 10:1455–1480, 1998.

[21] P. Hancock, R. Baddeley, and L. Smith. The principal components of natural images. *Network*, 3:61–70, 1992.

[22] S. Haykin. *Neural Networks — A Comprehensive Foundation*. Macmillan College Publishing Company, Inc., 1994.

[23] M. Kirby and L. Sirovich. Application of the Karhunen-Loeve procedure for the characterization of human faces. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 12(1):103–108, 1990.

[24] B. Knowlton and L. S. Squire. The learning of categories: Parallel brain systems for item memory and category knowledge. *Science*, 262:1747–1749, 1993.

[25] A. Lanitis, C. Taylor, and T. Cootes. Automatic interpretation and coding of face images using flexible models. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 19(7):743–756, 1997.

[26] C. Liu and H. Wechsler. Enhanced Fisher linear discriminant models for face recognition. In *Proc. the 14th International Conference on Pattern Recognition*, Queensland, Australia, August 17-20, 1998.

[27] C. Liu and H. Wechsler. Face recognition using evolutionary pursuit. In *Proc. the Fifth European Conference on Computer Vision*, University of Freiburg, Germany, June 2-6, 1998.

[28] C. Liu and H. Wechsler. Probabilistic reasoning models for face recognition. In *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Santa Barbara, California, USA, June 23-25, 1998.

[29] C. Liu and H. Wechsler. Comparative assessment of Independent Component Analysis (ICA) for face recognition. In *Proc. the 2nd International Conference on Audio- and Video-based Biometric Person Authentication*, Washington D. C., March 22-24, 1999.

[30] C. Liu and H. Wechsler. Face recognition using shape and texture. In *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Fort Collins, Colorado, USA, June 23-25, 1999.

[31] C. Liu and H. Wechsler. Robust coding schemes for indexing and retrieval from large face database. *IEEE Trans. Image Processing*, 9(1):132–137, 2000.

[32] B. Olshausen and D. Field. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, 381(13):607–609, 1996.

[33] A. O'Toole, H. Abdi, K. A. Deffenbacher, and D. Valentin. Low-dimensional representation of faces in higher dimensions of the face space. *Journal of The Optical society of America A*, 10:405–411, 1993.

[34] P. Penev and J. Atick. Local feature analysis: A general statistical theory for object representation. *Network: Computation in Neural Systems*, 7:477–500, 1996.

[35] P. Phillips. Matching pursuit filters applied to face identification. *IEEE Trans. Image Processing*, 7(8):1150–1164, 1998.

[36] T. Poggio, V. Torre, and C. Koch. Computational vision and regularization theory. *Nature*, 317:314–319, 1985.

[37] G. Rhodes, S. Carey, G. Byatt, and F. Proffitt. Coding spatial variations in faces and simple shapes: a test of two models. *Vision Research*, 38:2307–2321, 1998.

[38] P. Sinha and T. Poggio. I think I know that face... *Nature*, 384:404, 1996.

[39] D. Swets and J. Weng. Using discriminant eigenfeatures for image retrieval. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 18(8):831–836, 1996.

[40] M. Turk and A. Pentland. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 13(1):71–86, 1991.

[41] T. Valentin. A unified account of the effects of distinctiveness, inversion, and race in face recognition. *The Quarterly Journal of Experimental Psychology*, 43A(3):161–204, 1991.

[42] Y. Vapnik. *The Nature of Statistical Learning Theory*. Springer-Verlag, 1995.

[43] H. Wechsler, P. J. Phillips, V. Bruce, F. F. Soulie, and T. S. Huang, editors. *Face Recognition: From Theory to Applications*. Springer-Verlag, 1998.