14

Subspace Image Representation for Facial Expression Analysis and Face Recognition and its Relation to the Human Visual System

Ioan Buciu[1,2] and Ioannis Pitas[1]

[1] Department of Informatics, Aristotle University of Thessaloniki GR-541 24, Thessaloniki, Box 451, Greece.
pitas@zeus.csd.auth.gr

[2] Electronics Department, Faculty of Electrical Engineering and Information Technology, University of Oradea 410087, Universitatii 1, Romania.
ibuciu@uoradea.ro

Summary. Two main theories exist with respect to face encoding and representation in the human visual system (HVS). The first one refers to the dense (holistic) representation of the face, where faces have "holon"-like appearance. The second one claims that a more appropriate face representation is given by a sparse code, where only a small fraction of the neural cells corresponding to face encoding is activated. Theoretical and experimental evidence suggest that the HVS performs face analysis (encoding, storing, face recognition, facial expression recognition) in a structured and hierarchical way, where both representations have their own contribution and goal. According to neuropsychological experiments, it seems that encoding for face recognition, relies on holistic image representation, while a sparse image representation is used for facial expression analysis and classification. From the computer vision perspective, the techniques developed for automatic face and facial expression recognition fall into the same two representation types. Like in Neuroscience, the techniques which perform better for face recognition yield a holistic image representation, while those techniques suitable for facial expression recognition use a sparse or local image representation. The proposed mathematical models of image formation and encoding try to simulate the *efficient storing, organization* and *coding* of data in the human cortex. This is equivalent with embedding constraints in the model design regarding dimensionality reduction, redundant information minimization, mutual information minimization, non-negativity constraints, class information, etc. The presented techniques are applied as a feature extraction step followed by a classification method, which also heavily influences the recognition results.

Key words: Human Visual System; Dense, Sparse and Local Image Representation and Encoding, Face and Facial Expression Analysis and Recognition.

14.1 Introduction

In the human visual system (HVS), the visual image propagates from retina to the inferotemporal (IT) cortex, where the visual signal is decoded and processed. The question of how the human brain stores the image patterns in its visual cortex and how many pattern-specific neurons are activated and respond to a specific visual stimulus is a fundamental problem of psychology. A huge amount of research has been done in the attempt to understand how information captured by sensory channels is represented in the brain at different levels. Nowadays, this task is not only a concern of psychologists but also of image processing and computer vision experts. The pattern features that must be extracted from the data is a task-dependent matter. Among the huge visual data our eyes are overwhelmed by, facial images receive particular attention, due to their biological and sociological significance. This fact explains why the face analysis enjoys an important status with psychologists, anthropologists, neuroscientists and computer scientists alike. It is well known that in social interaction the human face constitutes the primary source of information for person recognition. As far as the computer scientists are concerned, the development of an automated face recognition system is necessary in order to cope with a large and complex area of applications, such as biometrics for security, surveillance, banking, law enforcement, video indexing, human-computer interaction, etc.

Another aspect closely related to face analysis is provided by facial expressions. Emotions can typically be conveyed by facial expressions. like for face recognition, the recognition of facial expressions is a subject of interdisciplinary research. From the psychological and anthropological perspectives the following questions are addressed: What information does a facial expression typically convey? Can there be emotions without facial expression? Can there be facial expression without emotions? How do individuals differ in their facial expression of emotions? [23]. It is well known among psychologists that the social context is dominated by language. However, the language alone is insufficient when it comes to successful social interaction. Plenty of communication comes through non-verbal communication. As Mehrabian suggested in [40], people express only 7% of the messages through a linguistic language, 38% through voice, and 55% through facial expressions.

A good understanding of the underlying process that governs the appearance of expressions is necessary in order to develop an appropriate facial image representation. In a human-computer interaction task, this constitutes the input to a human facial expression recognition system with satisfactory classification performance and, eventually, to artificial facial expression synthesis on an avatar for friendlier human-computer interface.

This chapter is organized as follows. Face encoding in the HVS from the neuroscience perspective is described in section 14.2. It starts with the analysis of dense, sparse and local face image representation followed by examples of these representations for face and facial expression recognition. A com-

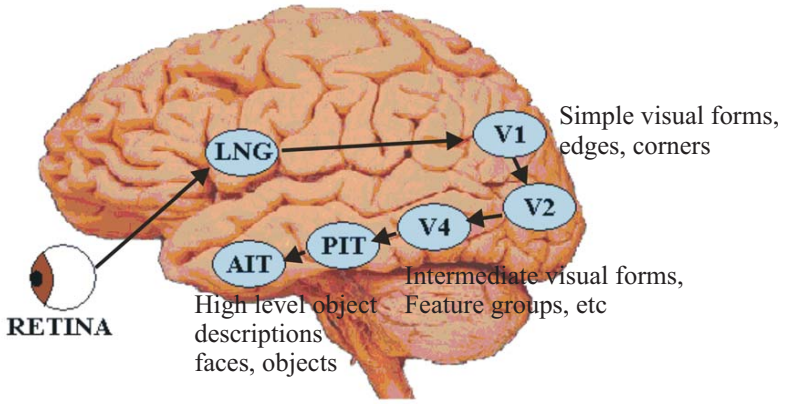

Fig. 14.1. Visual pathway in HVS. Information passes from the retina to the lateral geniculate nucleus (LGN) before arriving in cortical area V1. Further processing occurs in areas V2 and V4 and the posterior and anterior inferotemporal (IT) cortex (PIT and AIT).

puter vision analysis of face and facial expression recognition approaches is undertaken in section 14.3, where both dense and sparse image representation techniques are presented. The chapter ends with a discussion in section 14.4.

14.2 Face encoding in human visual system: a neuroscience view point

14.2.1 Dense and sparse image representation

How can we represent facial image information so that it can activate a representation in human memory under various conditions? Is human perception of a facial image based on its parts or it is viewed as a whole? Despite the huge amount of psychological research done in this respect, there is no general consensus in answering these questions. Rather, the answer to the problem of how the visual cortex understands complex objects, and, in particular human faces, is a controversial one. In recent years it has been argued from a visual neuroscience viewpoint that the architecture of the visual cortex suggests a hierarchical organization, in which neurons become selective to progressively more complex aspects of image structure.

Figure 14.1 depicts the visual pathway starting from the retina and ending at the two regions of inferotemporal cortex – IT (PIT and AIT). Multiple representations of the retinal space are mapped onto the cortex in a manner that preserves the visual topology. These representations define the visual modules: V1, V2, V4, IT. Whereas the earliest stages of the human visual system

(e.g. retina and V1 neurons) seem to produce a local distributed image representation, as we step into the higher visual system levels (such as V2, V4 areas or IT), the neurons have increasing receptive field sizes, being able to tackle increasingly complex stimuli [34]. Concerning neuroscience, the type of image encoding is related to the number of neurons that are active (respond) to a certain piece of information represented by a specific sensory stimulus caused by the image. We refer to a *local* image code when only a single individual specific cell is activated. We have a *dense* image code, when a large cell population with overlapping sensory input is activated and contributes to the image representation. The local code is "computed" very fast and occupies little memory. However, it cannot generalize (i.e., when trained with a sufficient number of samples, it achieves satisfactory results when tested on samples from the training set, but performs poorly on new test samples not belonging to the training set) [27]. This is caused by the fact that the input-output unit association (as in single-layer neural networks) is very weak and a new sample cannot be linked with the old association learned during the training process. On the other hand, a system based on a dense code suffers from slow training, requires heavy training and is likely to produce redundant image representations. However, it has a large capacity of making new associations. In between local and dense codes, we have the *sparse* image codes, where only a fraction of a large neuronal population is active. It is a trade-off between dense and local image codes, combining their advantages and trying to eliminate their drawbacks. Dense and local codes are closely related to holistic and local (component, or part-based) image representation and processing. The term *holistic* refers to an image representation which stores a face as a perceptual whole, without explicitly specifying its parts (components). The term *component* describes the separated parts of the face (e.g. eyes, nose, mouth, chin) that are perceived independently as distinct parts of the whole.

Atick and Redlich [1] support the idea of a dense image code within the HVS and argue for compact, densely decorrelated codes for image representation. They have demonstrated that receptive fields of retinal ganglion cells can be viewed as local "whitening" filters that remove second-order correlations between image pixels. Bandpass, multiscale and oriented receptive fields of V1 neurons may also be considered as filters that remove second-order correlation, the way Principal Component Analysis (PCA) does. Regarding human facial images, PCA has a certain appeal as a psychological model of face perception and memory. For example, the application of principal components is consistent with psychological evidence that the PCA of a set of face images accounts for some aspects of human memory performance, as shown by Valentine [56].

Ample evidence for sparse image coding within HVS has been collected by other researchers. They argue for a sparse image representation that leads to "efficient coding" in the visual cortex [26]. Since spatial receptive fields of simple cells (including V1 neurons) have been reasonably well described physiologically as being localized, oriented and bandpass, Olshausen and Field [42]

Fig. 14.2. Thatcher illusion [54]. The eyes and mouth of Margaret Thatcher (former Prime Minister of England whose face is depicted in the left hand image) have been inverted relative to the rest of her face (middle and right hand image). When the picture is viewed upright the face appears (middle image) highly grotesque. This strange distortion is much less evident when the face is turned upside-down (right hand image). Reproduced with permission from [54].

show that efficient image coding can be produced by considering an approach where the image is described by a small number of descriptors. These descriptors can be found by applying principles such as entropy minimization [3], which is equivalent to minimizing the mutual information in a such a way that the higher-order correlation between images is removed. Palmer [45] and Wachsmuth et al. [58] have drawn psychological and physiological evidence for parts-based object representations in the brain. Biederman came up with the theory of recognition-by-components (RBC) [7]. Empirical tests support his idea that complex objects are segmented into components called 'geons', which are further used by humans for image understanding. The "Thatcher illusion" presented in [54] suggests that parts of the face are processed independently. As depicted in figure 14.2 the rotated face seems to be processed by matching parts, which could be the reason why the face looks normal when turned upside-down.

Another sparse model of the neural receptive fields in early visual system was provided by Gabor functions [37]. A Gabor function is a sinusoid windowed with a Gaussian function. Its size, frequency and orientation can be manipulated to produce a wide range of different receptive field models. By convolving the image with the Gabor functions, a new image representation can be achieved with features that are sparse, oriented and localized.

Despite the large number of experiments and investigations, it is still unclear whether holistic/sparse image representations are unique and global or face image processing is a task-dependent [8, 16]. For instance, some evidence has been found that face identification and facial expression recognition are two independent tasks based on different representations and processing mech-

anisms. This hypothesis comes from the dissociation of these two processes found in brain damaged patients. It leads to the hypothesis that multiple representations of faces may reside in the visual cortex. The IT area of the temporal lobe contains neurons whose receptive fields cover the entire visual space. It also contains specialized neurons (face cells) that are selectively tuned to faces. There are dedicated areas in temporal cortical lobe that are responsible for processing information about faces [20], [47], [33]. It was also found that in AIT areas neurons with responses related to facial identity recognition exist, while other neurons (located in the superior temporal sulcus) are specialized to respond only to facial expressions [28].

14.2.2 Face and facial expression recognition

Experimentally, evidence for both sparse and dense face representations in the HVS has been found by neurophysiologists. However, the contribution of each representation depends on the task to be processed. While face recognition seems to favor a dense image representation (hence producing a holistic appearance of the faces), a more sparse (or even local) image representation has been found to account for facial expression analysis. This difference has been noticed in several works. Tanaka and Farah presented evidence in favor of a holistic process involved in face recognition [51]. These findings are stressed by the work of Farah et al. [25], which brings new evidence that part-based shape representation for faces has less impact in recognition than the holistic one. Furthermore, their theory is emphasized by the work of Dailey and Cottrell [19].

Contrary to representation for face identification, the work of Ellison and Massaro [24] has revealed that facial expressions are better represented by facial parts, suggesting non-holistic representation. This is consistent with research results showing that human subjects respond to information around the eyes independently from variation around the mouth and they are able to recognize and distinguish isolated parts of faces. The dissociation between face and facial expression recognition is also noted by Cottrell et al. [16] who found that PCA (which produces eigenfaces) performs well for face recognition but eigeneyes and eigenmouth (nonholistic eigenfeatures) perform better in recognizing expressions than eigenfaces, suggesting that eigenfeatures might transmit facial expression information. One of the techniques successfully applied to classify facial actions related to facial expressions, was Independent Component Analysis (ICA), which looks for components as independent as possible from each other and produces image features that can mimic the output of V1 receptive fields with orientation selectivity, bandpass and scaling properties [6]. In a direct comparison between PCA and ICA, Draper et al [22] found that facial identity recognition performance is better when the features are represented by a holistic approach (PCA) while an approach based on more localized features (ICA) performs better for facial action recognition.

14.3 Face and facial expression analysis: a computer-vision view point

The HVS often serves as an informal standard for evaluating systems. Therefore, not surprisingly, most face analysis approaches rely on biologically inspired models. To be plausible, these computer vision models have to share some characteristics and constraints with their organic models. A common characteristic of the proposed HVS models is the *dimensionality reduction* principle of image space. This physical constraint is easily understood if we consider, for instance, that an image of 64×64 pixels has dimensionality 4096. It is commonly accepted that the intrinsic dimensionality of the space of possible faces is much lower than that of the original image space. Basically, the latent variables incorporated there are discovered by decomposing (projecting) the image onto a linear (nonlinear) low dimensional image subspace. By reference to neuroscience, the receptive fields can be modeled by the basis images of the image subspace and their firing rates can be represented by the decomposition coefficients [42].

In order to generate the subspace image representation produced by the methods presented in this chapter, 164 image samples from the Cohn-Kanade AU-coded facial expression database [32] have been used. The number of basis images (subspace) was chosen to be 49.

14.3.1 Holistic image representations

As already mentioned, one of the most popular techniques for dimensionality reduction is PCA, which represents faces by their projection onto a set of orthogonal axes (also known as principal components, eigenvectors, eigenfaces, or basis images) pointing into the directions of maximal covariance in the facial image data. The basis images corresponding to PCA are ordered according to the decreasing amount of variance they represent, i.e., the respective eigenvalues. PCA-based Representations of human faces give us a dense code and the post-processed images have a holistic ("ghostlike") appearance, as can be seen from the first row of figure 14.3.

The principal components produce an image representation with minimal quadratic error. One of the proposed general organizational principles of the HVS refers to redundancy reduction. In PCA, this is achieved by imposing orthogonality among the basis images, thus redundancy is minimized. The nature of information encoded in the basis images was analyzed by O'Toole et al. [43] and Valentin and Abdi [57]. They found that the first basis images (containing low spatial frequency information) were most discriminative for classifying gender and race, while the basis images with small eigenvalues (corresponding to a middle range of spatial frequencies) contain valuable information for face recognition. This is coherent with findings that the face cells within HVS respond most strongly to face images containing energy within a middle range of spatial frequency between 4 and 32 cycles per image [49].

Fig. 14.3. Holistic subspace image representation. From top to bottom, each row depicts the first 10 basis images (out of 49) corresponding to PCA, FLD, ICA2 and NMF. For PCA the basis images are ordered by decreasing variance, for ICA2 and NMF by decreasing kurtosis.

Also, different components were found to be responsible for encoding identity and facial expression by Cottrell et al. in [15].

PCA has been successfully applied to face recognition [17], [5] and [55], and facial expression recognition, respectively [18], [44] and [14]. One statistical limitation of PCA is that it only decorrelates the input data (second-order statistics) without addressing higher-order statistics between image pixels. It is well known and accepted that, at least for natural stimuli, important information (e.g. lines, edges) is encoded in the higher-order statistics. Another limitation is related to the poor face recognition results for PCA when the faces are recorded under strong illumination variations.

Another holistic subspace image representation is obtained by a class-specific linear projection method based on Fisher's linear discriminant (FLD) [5]. This technique projects the images onto a subspace where the classes are maximally separated by maximizing the between-classes scatter matrix and minimizing the within-class scatter matrix at the same time. The basis images obtained through FLD are depicted in the second row of figure 14.3. This approach has been shown to be efficient in recognizing faces, outperforming PCA. Although this method seems to be more robust than PCA when small variation in illumination conditions appears, it fails in case of strong illumination changes. This is due to the assumption of linear separability of the classes. This assumption is violated, when strong changes in illumination occur. Another drawback of this method is that it needs a large number of training image samples for reasonable performance. Furthermore, the projection onto too few subspace dimensions does not guarantee the linear separability of the classes, hence the method will yield poor performance.

Along with redundancy reduction, another principle of HVS image coding mechanism is given by phase information encoding. It was shown by Field [26]

that methods relying only on second order statistics capture the amplitude spectrum of images but not the phase. The phase spectrum can be captured by employing higher order statistics. This has been proven to be accomplished by extracting independent image components [6]. There are several optimization principles taken into account when extracting independent components. The one described in [6] is based on the maximal information transfer between neurons and, among all the proposed ICA techniques, it seems to be the most plausible approach from the neuroscientific point of view.

Bartlett et al. [4] used two ICA configurations to represent faces for recognition. PCA was carried out prior to ICA for dimensionality reduction. An intermediate step for "whitening" the data has been introduced between PCA and ICA processing. The data were then decomposed into basis images and decomposition coefficients. Their second ICA configuration (ICA2) yields holistic basis images very similar to those produced by PCA. Such basis images are depicted in the third row of figure 14.3. In that case, ICA is applied to the projection matrix containing the principal components. Under this architecture, the linear decomposition coefficients are as independent as possible.

A recently proposed subspace image decomposition technique is Nonnegative Matrix Factorization (NMF) [36], which allows the data to be described as a combination of elementary features that involve only additive parts to form the whole. Both basis images and decomposition coefficients are constrained to be non-negative. Allowing only addition for recombining basis images to produce the original data is justified by the intuitive notion of combining parts to form the whole image. Another argument for imposing non-negativity constraints comes from neuroscience and is related to the non-negative firing rate of neurons. Finally, the positivity constraint arises in many real image processing applications. For example, the pixels in a grayscale image have non-negative intensities. Euclidean distance and Kullback-Leibler (KL) divergence were originally proposed as objective functions for minimizing the difference between the original image data and their decomposition product. Although, theoretically, the decomposition constraints tend to produce sparse image representations of basis images by composing the parts in an additive fashion, this is not always the case. It has been noticed in several works that, for some databases, the NMF decomposition rather produces a holistic image representation [38, 12, 30]. The representation could be affected by the imprecise image alignment procedure performed on the original database prior to NMF. It is known that the subspace techniques are generally sensitive to image alignment (registration). As noted in the last row of figure 14.3, for Cohn-Kanade database images, the basis images retrieved by NMF have a holistic appearance. A measure for quantifying the degree of sparseness in image representations is provided by the normalized kurtosis. If the basis images are stored as columns of a matrix $\mathbf{Z}$ the kurtosis of a base image $\mathbf{z}$ is defined as $k(\mathbf{z}) = \frac{\sum_i (z_i - \overline{z})^4}{(\sum_i (z_i - \overline{z})^2)^2} - 3$, where z_i are the the elements of $\mathbf{z}$ (pixels of base image) and $\overline{z}$ denotes the sample mean of $\mathbf{z}$. The average normalized kur-

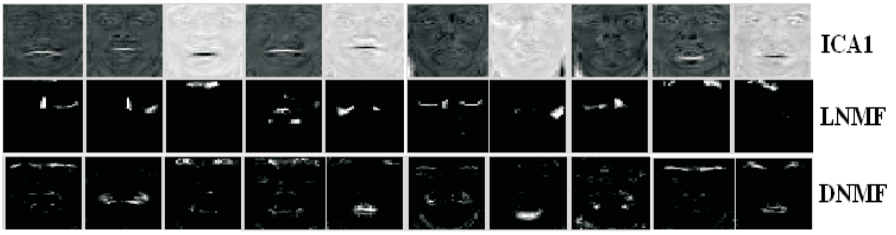

Fig. 14.4. Sparse subspace image representation. From top to bottom, each row depicts the first 10 basis images (out of 49) corresponding to ICA1, LNMF and DNMF, respectively. The basis images are ordered by decreasing kurtosis.

tosis for the 49 basis images are: $\overline{k}_{PCA} = 1.22$, $\overline{k}_{FLD} = 1.23$, $\overline{k}_{ICA2} = 0.93$, $\overline{k}_{NMF} = 5.93$. Thus, by far, NMF is the sparsest representation among ones represented in Figure 14.3.

14.3.2 Sparse image representations

The first ICA (ICA1) configuration produces independent basis images [4]. In this case, ICA is applied to the projection coefficients of PCA. Entropy minimization leads to a highly kurtotic distribution of basis image pixels, most of them having zero value, thus producing a sparse representation, as can be seen in the first row of Figure 14.4. Another representation, Local Non negative Matrix Factorization (LNMF) [38] enhances the sparseness of basis images by generating much more sparse, even localized and oriented image features. The extremely sparse basis images resulting from the LNMF approach are depicted in the second row of figure 14.4. The proposed approach uses the KL divergence as objective function to be minimized [38]. In addition to the non negativity constraints imposed for both decomposition factors, the redundant information is minimized by adding orthogonality constraints in the basis images formation. Furthermore, two more terms are added to the objective function for maximizing both sparseness and total activity and retaining only the most "expressive" image components [38]. In direct comparison for face recognition LNMF outperformed NMF [38, 12]. Buciu and Pitas further modified the LNMF algorithm in [11] and proposed a supervised NMF approach called Discriminant Nonnegative Matrix Factorization (DNMF) for facial expression classification. Besides the common constraints borrowed from NMF and LNMF, its underlying objective function also contains terms referring to discriminant class information. The basis images found by running the algorithm on image samples are sparse, oriented and localized, as can be seen in the last row of figure 14.4. DNMF is differentiated from the other NMF algorithms in that its facial basis images emphasize the salient facial features (eyes, eyebrows, mouth), when the images are labeled according to facial ex-

pression. These features convey the most discriminative information and are of great relevance for facial expression recognition. As can be seen by visual comparison of the basis images in figure 14.4, DNMF preserves local spatial information of salient facial features (that are almost absent in the case of LNMF) while it discards information less important for expression analysis (e.g., nose and chin, which is not the case for NMF) by incorporating class information. The average normalized kurtosis for the 49 basis images are: $\overline{k}_{ICA1} = 17.26$, $\overline{k}_{LNMF} = 49.16$, $\overline{k}_{DNMF} = 31.69$. The preservation of the spatial facial topology correlates well with the findings of Tanaka et al. [52], who argued that some face cells require the correct spatial feature configuration in order to be activated for facial expression recognition. Interestingly, DNMF seems to resemble many characteristics of the neural receptive fields [13]. The three NMF approaches have been applied to classify facial expressions [11] and to face recognition [10]. The DNMF approach was found to perform best for facial expression recognition, a fact that is indicating the role of sparse image representations. However, for face recognition, DNMF did not achieve the best performance compared to the other two approaches.

Local Feature Analysis (LFA) is another biologically inspired method that retrieves local image features [46]. Its biological motivation comes from the same redundancy minimization principle stating that, among the tremendous amount of neural receptors in the human retina, only a small fraction are active — corresponding to natural stimuli that are statistically redundant. To exploit this redundancy, LFA is used to extract a set of topographic local features defined by kernel filters that are optimally matched to the second-order statistics from the global PCA modes. They are found by minimizing the image reconstruction error and by using a process called sparsification [46]. To achieve this, a LFA neural network is employed, where the active units found by LFA are sparsely distributed. The selection of the spatial support of 100 filters found by LFA is shown in Figure 14.5 over a mean face from the experiment database.

One of the two most popular techniques for face recognition are known as "elastic graph matching" [35], and its relative, named "elastic bunch graph matching" [59]. "Elastic bunch graph matching" is based on applying a set of Gabor filters to special representative landmarks on the face (corners of the eyes and mouth, the contour of the face). Gabor filters represent the multi-scale nature of receptive fields, as each component has a unique combination of orientation, frequency tuning and scale. The face is represented by a list of values that comprise the amount of contrast energy that is present at spatial frequencies, orientations and scales included in the jet. For recognition, each face is compared with any other one with a similarity metric that takes into account the spatial configuration of the landmarks. It has been noted that the similarity metric involved in this approach is in line with the one used by the HVS [8]. Similar Gabor filters have been used by Würtz in [60] to extract local features that are robust to translations, deformations, and background changes. Each image is convolved with a set of different Gabor kernels, fol-

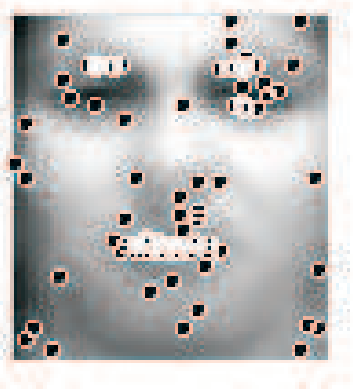

Fig. 14.5. A set of 100 optimally localized topographic kernel filters found by Local Feature Analysis (LFA) . The clusters of these filters are located around the fiducial points represented by eyes, eyebrows and mouth of the mean face. The algorithm from [46] has been applied to samples from the Cohn-Kanade database.

lowed by amplitude thresholding and discarding all units influenced by the background. Once the local features are extracted, four matching approaches (namely, multidimensional template matching, global matching, mapping refinement and phase alignment) are employed and combined to form correspondence maps used further for the face recognition task. To remove the weak correspondence points a relative similarity threshold is introduced, thus a final correspondence map is obtained. The combination of these matching approaches leads to a hierarchical structure of the algorithm with several decision levels, where the correspondence maps obtained by this method was proved to be very reliable. Furthermore, the Gabor functions have been successfully used for facial expression synthesis or recognition. The convolution of images with the set of Gabor filters can be performed either at the location of fiducial points (landmarks) [64] or, alternatively, the Gabor filters can be applied to the entire face image instead to specific face regions [9]. Figure 14.6 presents the result of the convolution of a set of 40 Gabor filters (5 frequencies and 8 orientations) with a sample image from the Cohn-Kanade database. The features extracted by the Gabor filters are localized and oriented [9].

One drawback of this feature extraction technique is the manual annotation of landmarks when the Gabor filters are applied to specific fiducial points. To overcome this issue, Heinrichs et al. [29] reduce the manual annotation to only one single image from which a self-organizing selection strategy builds up the bunches by adding the most similar face to the bunch graph and then the matching is recomputed. Another improvement is the replacement of the resulting Gabor wavelet bunches by principal components of the nodes of all training images. An enhancement in both the precision of landmark local-

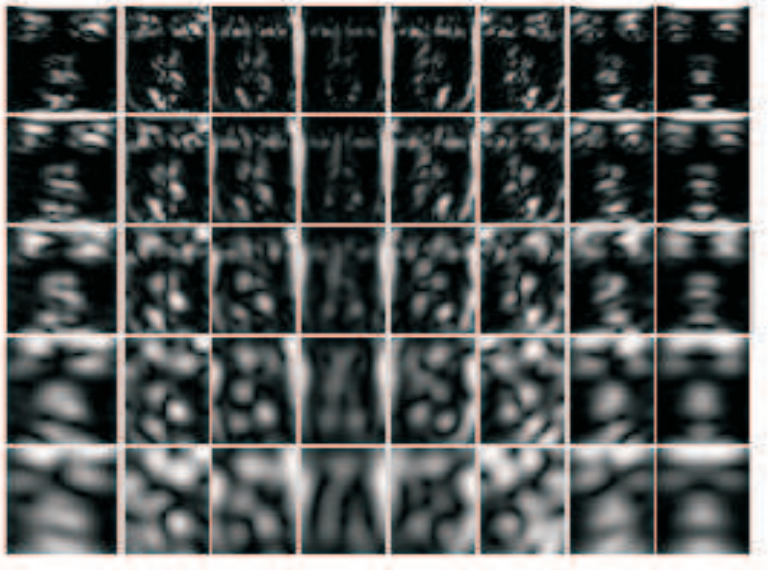

Fig. 14.6. The output of an image sample from the database convolved (filtered) with 40 Gabor filters (5 frequencies and 8 orientations).

ization and face recognition accuracy was obtained with this new approach. An extension of "elastic bunch graph matching" with a new application was recently proposed by Tewes et al. in [53]. They have developed a flexible object model using Gabor-wavelet-labeled graphs to synthesize facial expression. The graphs are then parameterized to allow flexible facial expression generation, where the expression parameters are viewed as a graph function. An overview of "elastic bunch graph matching" approach and its relationship to the Organic Computing paradigms is presented in [61].

A representative research work for facial expression recognition was conducted by Donato et al. [21], who investigated several holistic and sparse image representation techniques and measured their performance. Their work shows that the extraction of sparse features from the entire face space by convolving each image with a set of Gabor filters having different frequencies and orientations can outperform other methods that invoke the holistic representation of the face, when it comes to classify facial actions, closely related to facial expressions. They achieved the best recognition results by using ICA and Gabor filters. However, they also found that other local spatial approaches, like local PCA and PCA jets provide worse recognition accuracy than, for example, Fisher Linear Discriminant (FLD), which is a holistic approach.

14.4 Discussion

It has been argued that the tuning of the temporal cortex neurons that respond preferentially to faces represents a trade-off between fully distributed encoding (holistic or global representation, as PCA, FLD, ICA2, NMF result) and a grandmother cell type of encoding (local representation, achieved by LNMF) [50]. Psychophysiologically, one single pixel representation is similar to having a grandmother cell where a specific image is represented by one neuron. Among the approaches presented in this chapter, Gabor, ICA and DNMF turns out to be the most suitable biologically plausible models used in computer vision. However, it has to be noted that the DNMF approach is a relatively a new method and yet insufficiently investigated. In recent studies [11], it showed superior facial expression classification performance, when compared to Gabor, NMF, LNMF, and ICA approaches. However, its performance was not the one expected when applied to face recognition. More research has to be done in this regard. As we show in this chapter, one of the most popular techniques used for face and facial recognition tasks is PCA. When higher-order statistics are to be extracted and processed, ICA is chosen over PCA, which also seems to resemble the neuroscientific paradigms. However, a question related to ICA and PCA for face and facial expression recognition arises. Is ICA really better for these tasks than PCA? First of all, regardless of the feature extraction technique, the recognition results are not solely dependent on subspace image representation. It is also up to the classifier involved in the final step of the recognition task. A nearest neighbor classifier is usually chosen employing various similarity measures (distance metrics), such as L_1 (city block), L_2 (Euclidean), Mahalanobis or cosine distance. Several metrics favor the holistic representations, while others favor the sparse ones. This is mainly the reason, why, in the PCA-ICA debate, several works reported ICA outperforming PCA [21, 4, 22, 62, 39], while in other works ICA was found inferior to PCA [2], or, finally, no difference was found between them [41, 31]. Recently, new results on the ICA - PCA debate for face recognition have been revealed through the work conducted by Yang et al. [63]. They have repeated the experiments from [4] and have found that it is the "whitening" process (the intermediate step between PCA and ICA) that is responsible for the difference in the classification performance. Thus, as conclusion, ICA has an insignificant effect on the performance of face recognition . ICA was applied to cope with face recognition assuming that important information to discriminate between identities is contained in high-order image statistics, statistics that the PCA cannot retrieve. Interesting evidence that supports the observation that the elimination of high-order correlations between image pixels could not be so important for the neural receptive fields was brought by Petrov and Li [48]. They investigated local correlation and information redundancy in natural images and found that the removal of higher-order correlations between the image pixels increased the efficiency of image representation insignificantly. Accordingly, their results

suggest that the reduction of higher-order redundancies than the second-order ones is not the main cause of receptive field properties of neurons in V1.

Still two main questions remain: Are the holistic subspace image representations more appropriate for face recognition and the sparse subspace image representation more suitable for facial expression recognition? And, if it is so, which similarity measures should these representations be combined with in order to achieve the best recognition performance? Unfortunately, one of the shortcomings in neuroscience literature on face analysis is that no psychological measures for similarity of face image features (neither holistic nor sparse) exist. A large number of psychological studies are required in order to validate an existing subspace image representation model in combination with the optimal choice of the similarity metric.

Acknowledgments

This work has been conducted in conjunction with the "SIMILAR" European Network of Excellence on Multimodal Interfaces of the IST Program of the European Union (www.similar.cc).

References

1. J. J. Atick and A. N. Redlich. What does the retina know about the natural scene? *Neural Computation*, 4:196–210, 1992.
2. K. Baek, B. A. Draper, J. R. Beveridge, and K. She. PCA vs. ICA: A comparison on the FERET data set. *Joint Conference on Information Sciences, Durham, N.C.*, 2002.
3. H. Barlow. Unsupervised learning. *Neural Computation*, 1(3):295–311, 1989.
4. M. S. Bartlett, J. R. Movellan, and T. J. Sejnowski. Face recognition by independent component analysis. *IEEE Trans. Neural Networks*, 13(6):1450-1464, 2002.
5. P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman. Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 19(7):711-720, 1997.
6. A. J. Bell and T. J. Sejnowski. The "independent components" of natural scenes are edge filters. *Vision Research*, 37:3327–3338, 1997.
7. I. Biederman. Recognition-by-components: a theory of human image understanding. *Psychological Review*, 94(2):115–147, 1987.
8. I. Biederman and P. Kalocsai. Neurocomputational bases of object and face recognition. *Philosophical Transactions of the Royal Society of London*, 352(10):1203–1219, 1997.
9. I. Buciu, C. Kotropoulos, and I. Pitas. ICA and Gabor representation for facial expression recognition. in *Proc. 2003 IEEE Int. Conf. Image Processing*, pages 855–858, 2003.

10. I. Buciu, N. Nikolaidis, and I. Pitas. A comparative study of NMF, DNMF, and LNMF algorithms applied for face recognition. in Proc. *Second IEEE-EURASIP International Symposium on Control, Communications, and Signal Processing (ISCCSP)*, 2006.
11. I. Buciu and I. Pitas. A new sparse image representation algorithm applied to facial expression recognition. In *Proc. IEEE Workshop on Machine Learning for Signal Processing*, pages 539–548, 2004.
12. I. Buciu and I. Pitas. Application of non-negative and local non-negative matrix factorization to facial expression recognition. *Int. Conf. on Pattern Recognition*, pages 228–291, 2004.
13. I. Buciu and I. Pitas. DNMF modeling of neural receptive fields involved in human facial expression perception. *Journal of Visual Communication and Image Representation*, in press, 2006.
14. A. J. Calder, A. M. Burton, P. Miller, A. W. Young, and S. Akamatsu. A principal component analysis of facial expressions. *Vision Research* , 41:1179–1208, 2001.
15. G. Cottrell, K. Branson, and A. J. Calder. Do Expression and Identity Need Separate Representations ? In *Proc. of the 24th Annual Cognitive Science Conference*), pages 238–243, Dordrecht, Kluwer, 1990.
16. G. W. Cottrell, M. N. Dailey, C. Padgett, and R. Adolphs. Is all face processing holistic? *Philosophical Transactions of the Royal Society of London*, 352(10):1203–1219, 1997.
17. G. Cottrell and M. Fleming. Face recognition using unsupervised feature extraction. *Proc. of Int. Neural Network Conference*), pages 322–325, Dordrecht, Kluwer, 1990.
18. G. Cottrell and J. Metcalfe. Face, gender and emotion recognition using holons. *Advances in Neural Information Processing Systems*, 3:564–571, 1991.
19. M. N. Dailey and G. W. Cottrell. Organization of face and object recognition in modular neural network models. *Neural Networks* 12:1053–1073, 1999.
20. R. Desimone. Face selective cells in the temporal cortex of monkey. *Journal of Cognitive Neuroscience*, 3:1–8, 1991.
21. G. Donato, M. S. Bartlett, J. C. Hager, P. Ekman, and T. J. Sejnowski. Classifying facial actions. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 21(10): 974–989, 1999.
22. B. A. Draper, K. Baek, M. S. Bartlett, and J. R. Beveridge, Recognizing faces with PCA and ICA . *Computer vision and image understanding*, 91:115–137, Special issue on Face Recognition, 2003.
23. P. Ekman. Facial expression and emotion. *American Psychologist*, 48(4):384–392, 1993.
24. J. W. Ellison and D. W. Massaro. Featural evaluation, integration, and judgment of facial affect. *Journal of Experimental Psychology: Human Experimental Psychology: Human Perception and Performance*, 23(1):213–226, 1997.
25. M. J. Farah, K. D. Wilson, M. Drain, and J. N. Tanaka. What is special about face perception. *Psychological Review*, 103(3):482–498, 1998.
26. D. Field. What is the goal of sensory coding ? *Neural Computation*, 6(4):559–601, 1994.
27. P. Foldiak. Sparse coding in the primate cortex. *The Handbook of Brain Theory and Neural Networks*, Second Edition, pages 1064–1068, MIT Press, 2002.

28. M. E. Hasselmo, E. T. Rolls, G. C. Baylis, and V. Nalwa. The role of expression and identity in the face-selective responses of neurons in the temporal visual cortex of the monkey. *Behavioral Brain Research*, (32):203–218, 1989.
29. A. Heinrichs, M. K. Müller, A. H. J. Tewes, and R. P. Würtz. Graphs with Principal Components of Gabor Wavelet Features for Improved Face Recognition. in *Information Optics: 5th International Workshop on Information Optics; WIO'06*, pages 243–252, 2006.
30. P. O. Hoyer. Non-negative Matrix Factorization with sparseness constraints. *Journal of Machine Learning Research*, (5):1457–1469, 2002.
31. Z. Jin and F. Davoine. Orthogonal ICA representation of images, *Proceedings of 8th International Conference on Control, Automation, Robotics and Vision*, pages 369–374, 2004.
32. T. Kanade, J. Cohn, and Y. Tian. Comprehensive database for facial expression analysis. In *Proc. IEEE Inter. Conf. on Face and Gesture Recognition*, pages 46–53, 2000.
33. N. Kanwisher, J. McDermott, and M. M. Chun. The fusiform face area: A module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience*, 17:4302–4311, 1997.
34. E. Kobatake and K. Tanaka. Neuronal selectivities to complex object features in the ventral visual pathway of the macaque cerebral cortex. *Journal of Neurophysiology*, 71(3):856–867, 1994.
35. M. Lades, J. C. Vorbrüggen, J. Buhmann, Jörg Lange, C. von der Malsburg, R. P. Würtz, and W. Konen. Distortion invariant object recognition in the dynamic link architecture. *IEEE Trans. on Computers*, 42(3):300–311, 1993.
36. D. D. Lee and H. S. Seung. Learning the parts of the objects by non-negative matrix factorization. *Nature*, 401:788–791, 1999.
37. T. Lee. Image representation using 2D Gabor wavelets. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 18(10):959–971, 1996.
38. S. Z. Li, X. W. Hou, and H. J. Zhang. Learning spatially localized, parts-based representation. *Int. Conf. Computer Vision and Pattern Recognition*, pages 207–212, 2001.
39. C. Liu and H. Wechsler. Independent component analysis of Gabor features for face recognition. *IEEE Trans. Neural Networks*, 14(4):919–928, 2003.
40. A. Mehrabian. Communication without words. *Psychology Today*, 2(4):53–56, 1968.
41. B. Moghaddam. Principal manifolds and probabilistic subspaces for visual recognition. *IEEE Trans. Pattern Anal. Machine Intell.*, 24(6):780–788, 2002.
42. B. A. Olshausen and D. J. Field. Natural image statistics and efficient coding. *Network Computation in Neural Systems*, 7(2):333–339, 1996.
43. A. O'Toole, H. Abdi, K. 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(3 pages 405–411, 1993.
44. C. Padgett and G. Cottrell. Representing face images for emotion classification. *Advances in Neural Information Processing Systems*, 9):894–900, 1997.
45. S. E. Palmer. Hierarchical structure in perceptual representation. *Cognitive Psychology*, (9):441–474, 1977.
46. P. S. Penev and J. J. Atick. Local feature analysis: A general statistical theory for object representation. *Neural Systems*, 7:477–500, 1996.
47. D. I. Perret, E. T. Rolls, and W. Caan. Visual neurons responsive to faces in the monkey temporal cortex. *Experimental Brain Research*, (47):329–342, 1982.

48. Y. Petrov and Z. Li. Local correlations, information redundancy, and the sufficient pixel depth in natural images. *Journal Optical Society of America A*, 20(1):56–66, 2003.
49. E. Rolls, G. Baylis, and M. Hasselmo. The responses of neurons in the cortex in the superior temporal sulcus of the monkey to band-pass spatial frequency filtered faces. *Vision Research*, 27(3):311–326, 1987.
50. E. T. Rolls and A. Treves. The relative advantages of sparse versus distributed encoding for associative neural networks in the brain. *Network* 1:407–421, 1990.
51. J. W. Tanaka and M. J. Farah. Parts and wholes in face recognition. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, 46A:225–245, 1993.
52. K. Tanaka, C. Saito, Y. Fukada, and M. Moriya. Integration of form, texture, and color information in the inferotemporal cortex of the macaque. In *Vision, Memory and the Temporal Lobe*, pages 101–109, 1990.
53. A. Tewes, R. P. Würtz, and C. von der Malsburg. A flexible object model for recognizing and synthesizing facial expressions. In *Proc. of the Int. Conf. on Audio-and Video-based Biometric Person Authentication*, pages 81–90, 2005.
54. P. Thompson. Margaret Thatcher: a new illusion. *Perception*, 9:483–484, Pion Limited, London, 1980.
55. M. Turk and A. Pentland. Eigenfaces for recognition. *Cognitive Neuroscience*, 3(1):71–86, 1991.
56. T. Valentine. A unified account of the effects of distinctiveness, inversion, and race in face recognition. *Quarterly Journal of Experimental Psychology*, 43 A:161–204, 1991.
57. D. Valentin and H. Abdi. Can a linear auto-associator recognize faces from new orientations? *Journal of the Optical Society of America A*, 13:717–724, 1996.
58. E. Wachsmuth, M. W. Oram, and D. I. Perrett. Recognition of objects and their component parts: responses of single units in the temporal cortex of the macaque. *Cerebral Cortex*, 4(5):509–522, 1994.
59. L. Wiskott, J.-M. Fellous, N. Krüger, and C. von der Malsburg. Face recognition by elastic bunch graph matching. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 19(7):775–779, 1997.
60. R. P. Würtz. Object recognition robust under translations, deformations and changes in background. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 19(7):769–775, 1997.
61. R. P. Würtz. Organic Computing methods for face recognition. *it – Information Technology*, 47(4):207–211, 2005.
62. P. C. Yuen, and J. H. Lai. Face representation using independent component analysis. *Pattern Recognition*, 35(6):1247–1257, 2002.
63. J. Yang, D. Zhang, and J. Yang. Is ICA Significantly Better than PCA for Face Recognition? *10^{th} IEEE International Conference on Computer Vision*, 2005.
64. Z. Zhang, M. Lyons, M. Schuster, and S. Akamatsu. Comparison between geometry-based and Gabor-wavelets-based facial expression recognition using multi-layer perceptron. In *Proc. of Third IEEE Int. Conf. Automatic Face and Gesture Recognition, April 14-16, 1998, Nara, Japan*, pages 454–459, 1998.