Learning Multiview Face Subspaces and Facial Pose Estimation Using Independent Component Analysis

Stan Z. Li, XiaoGuang Lu, Xinwen Hou, Xianhua Peng, and Qiansheng Cheng

***Abstract*—An independent component analysis (ICA) based approach is presented for learning view-specific subspace representations of the face object from multiview face examples. ICA, its variants, namely independent subspace analysis (ISA) and topographic independent component analysis (TICA), take into account higher order statistics needed for object view characterization. In contrast, principal component analysis (PCA), which de-correlates the second order moments, can hardly reveal good features for characterizing different views, when the training data comprises a mixture of multiview examples and the learning is done in an unsupervised way with view-unlabeled data. We demonstrate that ICA, TICA, and ISA are able to learn view-specific basis components unsupervisedly from the mixture data. We investigate results learned by ISA in an unsupervised way closely and reveal some surprising findings and thereby explain underlying reasons for the emergent formation of view subspaces. Extensive experimental results are presented.**

***Index Terms*—Appearance-based approach, face analysis, independent component analysis (ICA), independent subspace analysis (ISA), learning by examples, topographic independent component analysis (TICA), view subspaces.**

I. Introduction

APPROXIMATELY 75% of the faces in home photos are nonfrontal [1], and, therefore, it is important for a face recognition system to be able to deal with faces of varying poses. There are two types of pose variations: those due to in-plane rotation and those due to out-of-plane rotation. This paper is concerned with the latter type of variation, which is more difficult to analyze and cope with. We have two objectives: The first is to derive a view-specific subspace (view subspace in brief) representation from a training set of multiview face examples such as those shown in Fig. 1. The second is to design an algorithm for estimating out-of-plane rotations.

Much research has been done in dealing with view and illumination changes [2]–[15]. It has been found that distributions of appearances in linear subspaces such as those based on principal component analysis (PCA) under perceivable variations in viewpoint and illumination are highly nonlinear, nonconvex, complex and perhaps twisted [16]–[20]. The principal component analysis (PCA) based techniques [21], [22], which decorrelate the second order moments, can hardly capture variations due to pose changes. Such variations are related to higher order statistics.

Fig. 1. Multiview face examples.

Talukder and Casasent [23] proposed a maximum discriminating feature (MDF) neural network to extract nonlinear features of high-dimensional data which optimally discriminate between multiple classes. The weights of the neural network are obtained in closed-form, so that the network does not have problems associated with iterative neural network solutions. A comparison of this nonlinear feature technique with other nonlinear techniques that use higher-order statistical information, such as nonlinear PCA, kernel PCA and neural nets, is discussed. Based on the nonlinear MDF features, a modified k-nearest neighbor classifier could be used for facial pose estimation [24], [25].

The use of geometrical features or neural networks for pose estimation has also been investigated for robotics and target recognition. Khotanzad and Liou [26] represent three-dimensional objects by a set of rotation invariant features derived from the complex orthogonal pseudoZernike moments of their two-dimensional (2-D) perspective images, and then obtain the pose parameters, *i.e.,* aspect and elevation angles of the objects, by a two-stage neural network system.

In this paper, we present independent component analysis (ICA) [27], [28] based methods for learning view subspaces from multiview face examples, and thereby performing view-based face classification [29]–[32]. ICA and its variants, namely independent subspace analysis (ISA) [33] and topographic independent component analysis (TICA) [34], take into account higher order statistics required to characterize the view of objects, and are suitable for the learning of view subspaces.

Two types of learning algorithms are presented: supervised and unsupervised. For the unsupervised case where a mixture of multiview face examples are *without* the view labels, we show

Manuscript received October 24, 2002; revised May 16, 2004. This work was carried out at Microsoft Research Asia. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Fernando M. B. Pereira.

S. Z. Li is with Microsoft Research Asia, Beijing 100080, China (e-mail: szli@microsoft.com).

X. Lu is with the Department of Computer Science and Engineering, Michigan State University, East Lansing, MI 48824 USA (e-mail: lvxiaogu@cse.msu.edu).

X. Hou and X. Peng are with the School of Mathematical Sciences, Peking University, Beijing 100871, China (e-mail: xwhou@hci.ia.ac.cn; xhpeng@pku.edu.cn).

Q. Cheng is with the Key Laboratory of Pure and Applied Mathematical Sciences, Peking University, Beijing 100871, China (e-mail: qcheng@pku.edu.cn).

Digital Object Identifier 10.1109/TIP.2005.847295

that applying ICA to view-unlabeled training data yields emergent view-specific basis components of faces; ISA and TICA moreover are able to discover view-based grouping of the basis components, with TICA producing additional view-based ordering between the groups. We then analyze how the above unsupervised ISA learns view subspaces, and thereby present a supervised ISA learning method for more effective estimation of facial poses. The analysis reveals two interesting outcomes: 1) using face examples of a specific view, the ISA actually learns basis components of the complement subspace of that view subspace; 2) using face examples of all but one specific view, the ISA learns basis components of the view subspace corresponding to the excluded view. Using the view label information, a supervised learning algorithm produces sets of basis components which better characterize the view subspaces, and yield higher estimation accuracy for pose estimation. These are supported by extensive experiments.

The rest of the paper is organized as follows. Section II introduces the concepts of ICA, ISA and TICA. Section III presents ICA-based methods for unsupervised learning of view subspaces. Section IV presents the use of learned view-subspace representation for view-based face classification.

II. ICA-Based Image Modeling and Subspace Learning

A. ICA

ICA [27], [28] is a linear transform which makes linear mixtures of random variables as statistically independent as possible. It not only decorrelates the second order statistics but also reduces higher-order statistical dependencies [28]. It extracts independent components even if their magnitudes are small whereas PCA extracts components having largest magnitudes. When performed on image patches randomly sampled from natural images, ICA produces some interesting results. Olshausen and Field [35] obtain spatially localized, oriented, bandpass basis functions comparable to those in certain wavelet transforms. Bell and Sejnowski [36] find that independent component of natural scenes are edge-like filters. Lee, Lewicki, and Sejnowski [37] derive an ICA model to represent a mixture of several mutually exclusive classes each of which is described as a linear combination of independent non-Gaussian densities. It is found that the two different class of images have different types of basis functions. In image analysis applications, ICA has also been used for face recognition and texture analysis [38]–[42], as a hopefully better method than PCA. In [42], ICA is used for the unsupervised learning of face representations; it is shown experimentally that the learned ICA representations were superior to representations based on PCA for recognizing faces.

In ICA-based image modeling, a gray-level image $\mathbf{x} = \{x(u,v)\}$, where (u,v) is the pixel location, is represented as a linear combination of m basis functions $\mathbf{b} = \{b_1(u,v), \ldots, b_m(u,v)\}$

$$\mathbf{x}(u,v) = \sum_{i=1}^{m} b_i(u,v) s_i \tag{1}$$

where $\mathbf{s} = (s_1, \ldots, s_m)$ are the combining coefficients. We restrict $\mathbf{b}$ to be an invertible linear system, so that the equation above could be inverted by using the dot-product

$$s_i = \langle \mathbf{w}_i, \mathbf{x} \rangle = \sum_{u,v} w_i(u,v) x(u,v) \tag{2}$$

where the $\mathbf{w} = \mathbf{b}^{-1}$ is the inverse filter.

The crucial assumption made in ICA is that the s_i are nongaussian, and mutually independent random variables. The latter assumption means that the joint distribution of $\mathbf{s}$ can be factorized as

$$p^s(\mathbf{s}) = \prod_{i=1}^{m} p_i^s(s_i) \tag{3}$$

where p_i^s are densities of s_i. The ICA learning problem is to estimate both the basis functions $b_i(u,v)$ and the realizations of the s_i, for all i and (u,v), using a sufficiently large set of training images $\{\mathbf{x}_k(u,v)\}$; so that for any given sample $\mathbf{x}_k(u,v)$ from the training set, information about one of the s_is gives as little information as possible about the others. In other words, the s_is are as independent as possible.

There are several approaches for formulating independence in the ICA model [43] such as minimum mutual information, maximum neg-entropy; a very popular approach is the maximum likelihood [44], [45]. Given an ICA model in (1), and the density of $\mathbf{s}$ in (3), the density of the observation $\mathbf{x}$, or the likelihood of the model, can be formulated as $p(\mathbf{x} \mid \mathbf{w}) = |\det \mathbf{b}|^{-1} p^s(\mathbf{b}^{-1}\mathbf{x}) = |\det \mathbf{w}| p^s(\mathbf{w}\mathbf{x})$. Given N_T training images, $\mathbf{X} = \{\mathbf{x}_k \mid k = 1, \ldots, N_T\}$, the logarithm likelihood can be derived as

$$\log p(\mathbf{X}|\mathbf{w}) = \sum_{k=1}^{N_T} \sum_{i=1}^{m} \log p_i^s(s_{i,k}) + N_T \log |\det \mathbf{w}| \tag{4}$$

where $s_{i,k} = \langle \mathbf{w}_i, \mathbf{x}_k \rangle$ is the coordinate of $\mathbf{x}_k$ in the $\mathbf{w}_i$ axis.

The ICA algorithm leads to sparse coding equivalent to a factorial representation. In other words, the probability distributions of the projection coefficients of a sample on the basis components (feature directions) are sparse, *i.e.,* the density functions are uni-modal and peaked at zero with heavy tails. A sparse distribution leads to super-Gaussianity. Consider a random zero-mean variable x. The fourth cumulant of the distribution of x, also called kurtosis, is defined as

$$\text{kurt}(x) = E\{x^4\} - 3(E\{x^2\})^2. \tag{5}$$

Kurtosis can be considered a measure of the non-Gaussianity of x. Distributions of positive kurtosis are called super-Gaussian, whereas those of negative ones are called sub-Gaussian.

B. ISA

The independent subspace analysis (ISA) is an extension of ICA proposed by Hyvärinen and Hoyer [46]. In ISA, the model is still in the form of (1), but the independence assumption about s_i is relaxed, as compared to ICA. The collection of s_i are divided into a number of L groups. The s_i within a group are dependent on each other, but those in different groups are independent.

Denote the collection of the indices of s_i in group $\ell \in \{1, \ldots, L\}$ by S^ℓ. For each ℓ, the basis components

$\{b_i \mid i \in S^\ell\}$ span the ℓth ISA subspace, and $\sum_{i\in S^\ell} b_i s_i$ is the projection of $\mathbf{x}$ on that subspace. The norm of the projection is given by $\sqrt{\sum_{i\in S^\ell} s_i^2} = \sqrt{\sum_{i\in S^\ell} \langle \mathbf{w}_i, \mathbf{x}\rangle^2}$.

According to the invariant feature subspace theory [47], [48], the norms of the projections on these subspaces represent some higher-order, invariant features. The ISA combines the principle of invariant feature subspace into multidimensional ICA [49] in order to find some invariant features. An invariant feature subspace can be embedded in multidimensional ICA by assuming that for each ℓ, the joint probability distribution of the coefficients s_i $(i \in S^\ell)$ is spherically symmetric, *i.e.,* dependent only on the norm of the s_is. Although the exact nature of the invariance has not been specified in a subspace model, it will emerge from the input data as the maximization is performed in ISA.

Given an ISA model as in (1), the logarithm of the likelihood of the observations can be formulated as

$$\log p(\mathbf{X} \mid \mathbf{w}) = \sum_{k=1}^{N_T} \sum_{\ell=1}^{L} \log p_\ell^s \left(\sum_{j\in S^\ell} s_{j,k}^2 \right) + N_T \log |\det \mathbf{w}| \quad (6)$$

where $\sum_{j\in S^\ell} s_{j,k}^2$ is the squared norm of the projection of $\mathbf{x}_k$ on the ℓth ISA subspace, and p_ℓ^s are some known density functions (often assumed to be exponential) of the norm. This model specifies the prior information on their independence.

As in ICA, p_ℓ^s in ISA learning is also chosen to be a super-Gaussian distribution [46]. When it is exponential, $p_\ell^s(z) \propto e^{-\sqrt{z}}$, we have $\log p_\ell^s = -\sqrt{\sum_{j\in S^\ell} s_{j,k}^2}$ and the log likelihood as

$$\log p(\mathbf{X} \mid \mathbf{w}) \propto -\sum_{k=1}^{N_T} \sum_{\ell=1}^{L} \sqrt{\sum_{j\in S^\ell} s_{j,k}^2} \quad (7)$$

where we have assumed that $\mathbf{b}$ is an orthogonal matrix so that $\mathbf{w} = \mathbf{b}^T$, and $|\det(\mathbf{w})| = 1$. Maximizing the above likelihood is equivalent to minimizing the following energy with respect to $\mathbf{b}$ (cf. Equation(6) in [46])

$$E(\mathbf{X} \mid \mathbf{b}) = \sum_{k=1}^{N_T} \sum_{\ell=1}^{L} \sqrt{\sum_{j\in S^\ell} \langle \mathbf{b}_j, \mathbf{x}_k\rangle^2}. \quad (8)$$

Learning ISA subspaces can be implemented by using a gradient descent algorithm [46]. Minimizing $E(\mathbf{X} \mid \mathbf{b})$ in (8) with respect to $\mathbf{b}$ results in L groups of ISA basis components.

C. TICA

Topographic independent component analysis (TICA) proposed by Hyvärinen and Hoyer [34] is a further extension to ICA. In TICA, the observed variable $\mathbf{x}$ is also generated as a linear transformation of the components $\mathbf{s} = (s_1, \ldots, s_m)$ as in (1), where m is the dimension of $\mathbf{x}$. In contrast to ICA, the components s_i are no longer independent but mutually energy-correlated according to the generative model $s_i = \sigma_i z_i$ where z_i is a random variable that has the same distribution as s_i given that the energy $\sigma_i^2 = 1$. The z_is are mutually independent and the energy variable σ_i is generated by $\sigma_i = \phi\left(\sum_k h(i,k) u_k\right)$ where u_ks are nonnegative higher-order independent components, ϕ is some nonnegative scalar nonlinearity, $h(i,j)$ is a neighborhood function expressing the proximity between the ith and jth components. The neighborhood function can be defined through a one-dimensional or 2-D neighborhood system, as in self-organizing maps [48]. Thus, components which are close to each other in the 2-D topographic map, *i.e.,* those within a neighborhood, are not assumed to be independent; they are allowed to be correlated in their energies.

Fig. 2. Average faces in 19 views of $0^\circ, 10^\circ, \ldots, 180^\circ$.

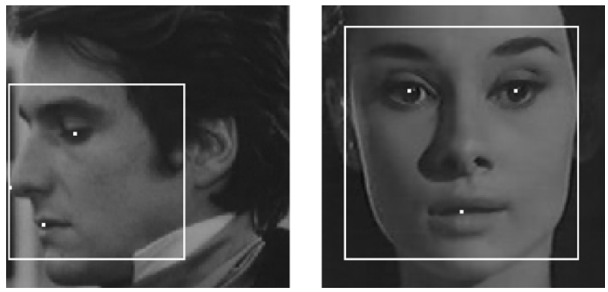


Fig. 3. Feature points and bounding rectangles for two face examples of view around 40° and 90°.

Denote the set of indices of the components neighboring to component i by $\mathcal{N}_i$, the log-likelihood function for the TICA model can be approximated by

$$\log p(\mathbf{X} \mid \mathbf{w}) = \sum_{k=1}^{N_T} \sum_{i=1}^{m} G\left(\sum_{j\in\mathcal{N}_i} h(i,j) s_{j,k}^2 \right) + N_T \log |\det \mathbf{w}| \quad (9)$$

where the function G has a similar role as the log density function of the independent components in classic ICA and could be chosen as many heuristic functions. Learning a TICA model can be achieved by maximizing the log-likelihood. TICA can be considered as the generalization of the model of ISA. The likelihood in (6) can be expressed as a special case of the likelihood (9) with a neighborhood system.

III. Unsupervised Learning of View Subspaces

In this section, we compare and analyze the performance of different unsupervised learning methods, *i.e.,* PCA, ICA, ISA and TICA, in deriving basis components of view subspaces. A multiview face database made at Microsoft Research Asia is used in the following experiments on unsupervised and supervised view-subspace learning. There are a total of about 20 000 face examples, half for training and half for test. The view range is partitioned from 0° (right-side view) to 180° (left-side view) into 19 interval views, each of which spans about 10° as shown in Fig. 2. Due to the symmetry, only one side from 0° to 90° is used, consisting of the ten views. The coordinates of some feature points are manually marked for each face, so the locations of the corresponding points in different view groups are different. The face is then cropped according to the marked points. Fig. 3 illustrates two examples. There are 600 to 2000 original face examples for each view, more frontal view face examples than nonfrontal ones. After these steps, a total of 1000

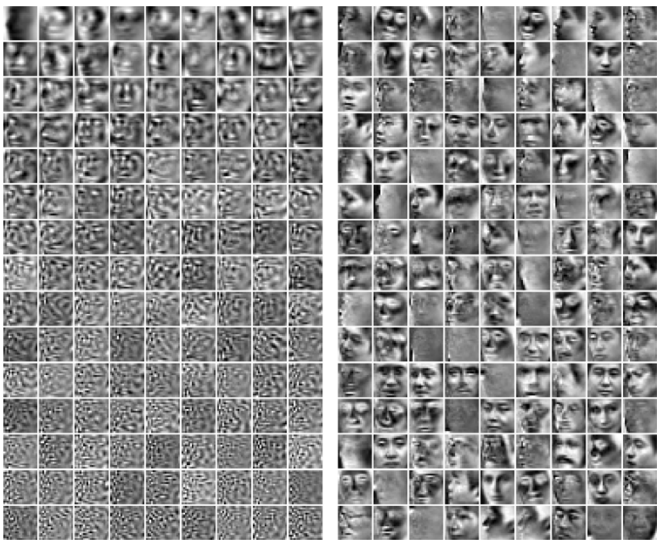


Fig. 4. (Left) Basis components learned by PCA and classic ICA algorithms in an unsupervised way using view-unlabeled multiview examples. PCA components (sorted by descending eigenvalues) present little view-specific information. ICA components are view specific, but there is no ordering between them because they are independent.

labeled face examples are obtained for each of the ten views $\{0°, \ldots, 90°\}$ in training set and test set.

The face images are then preprocessed by illumination correction (by fitting a plane to the image surface and then subtracting it from the image), mean value normalization, and then histogram equalization. Such preprocessing helps but is not crucial to the result. Whitening and dimensionality reduction of the input data is then performed using PCA, as a common practice in ICA, ISA and TICA learning, from 400 to $N = 150$ dimensions. The whitening makes ICA/ISA/TICA computation easier and the dimensionality reduction not only reduces the computational cost but also removes artifacts to some extent to prevent over-fitting. After these, the actually input $\mathbf{x}$ to our view-subspace learning algorithms is a vector of 150 dimensions.

A. ICA

In facial pose estimation, one hopes to find basis functions for each view subspace. Here, ICA, ISA and TICA methods are applied to these datasets to learn view-specific subspaces. Fig. 4 shows the basis components learned by PCA and ICA algorithms in the unsupervised way. The ICA basis components are view specific, whereas the PCA basis components do not present view-specific information. According to the model of ICA, the projections of facial data onto different basis components are independent of each other. This may be interpreted as one component spanning that particular subspace. As such, ICA cannot group basis components with similar views learned to form view subspaces.

B. ISA

The ISA learning method is able to produce groupings of basis component where each group is view specific and constitutes a subspace of that view. Fig. 5 shows the basis components learned by ISA. Indeed, the learned ISA basis components are

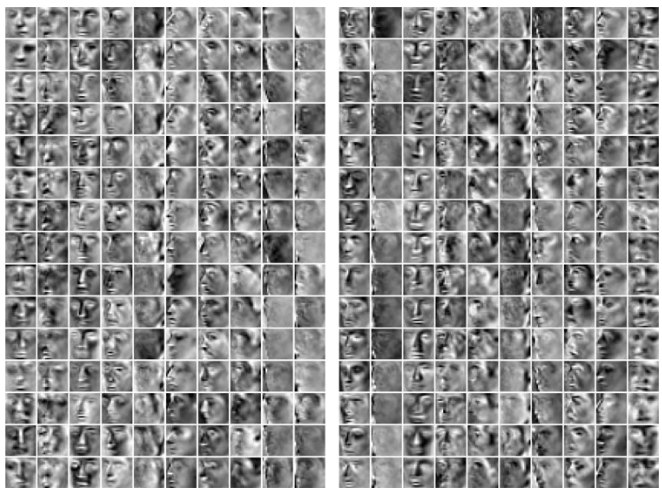


Fig. 5. Two sets of basis components learned by unsupervised ISA, corresponding to different initializations. The components are view specific and each column of the components in an ISA map constitute a view subspace; however, the columns are un-ordered by view because the subspaces are independent.

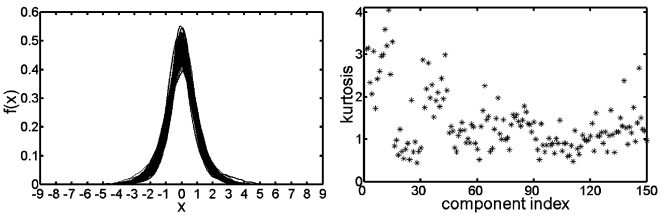


Fig. 6. (Left) Estimated probability density functions and (right) normalized kurtosis of 150 s_is learned by ISA.

view specific, and explicit view-based grouping of these components are formed in the ISA map. Each column constitutes a view subspace.

The unsupervised ISA learning algorithm assumes that the coefficients s_is have sparse distributions. Then the learned basis components should actually have sparse distributions if the ISA model really fits the multiview facial data. Fig. 6 shows the estimated probability density functions and normalized kurtosis $\mathrm{kurt}(s) = E\{s^4\}/(E\{s^2\})^2 - 3$ for the 150 ISA coefficients. All the density functions are uni-modal and peaked at zero with heavy tails; and all the kurtosis are positive. So the probability distribution of all the components are super-Gaussian, which is consistent with the *a priori* assumption used in the derivation of the ISA learning algorithm.

Although the ISA method learns view groupings, different ISA view groups are independent of each other, and, therefore, a view-specific ordering between the groups is not readily available in the the ISA map.

C. TICA

Using TICA learning of view subspaces, we hope to find such a map in which not only the ℓth column of components in the map constitute the bases for the ℓth view, but also the columns are automatically ordered by view. To make basis components for adjacent views correlated with each other, we define a neighborhood in the map such that all components in the $\ell - 1$ and $(\ell + 1)$th columns are neighbors to those in the ℓth column;

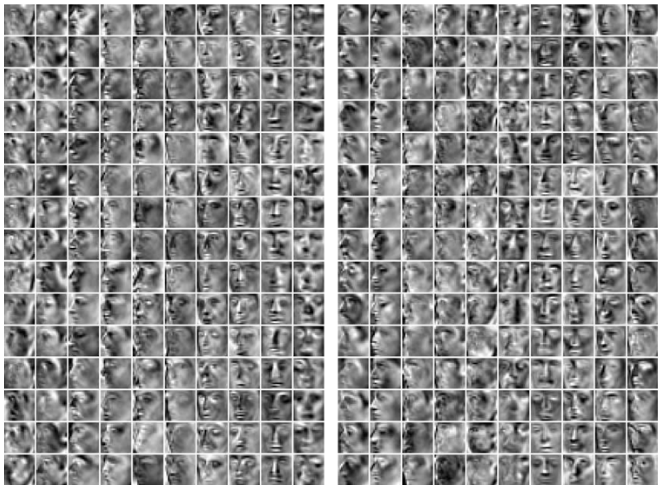


Fig. 7. Two sets of basis components learned by unsupervised TICA algorithms with different initializations. Each column of the components in an TICA map constitute a view subspace, and the columns are view ordered due to the dependencies between neighboring columns.

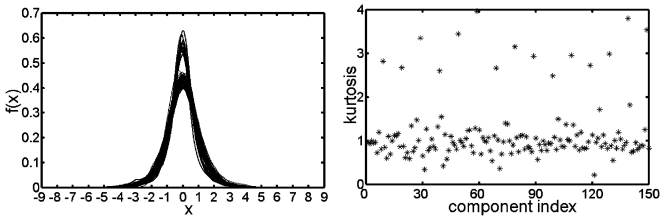


Fig. 8. (Left) Estimated probability density function and (right) normalized kurtosis of each of the 150 s_i learned by TICA.

there should be no neighboring relations beyond that. This is done using the following neighborhood function h

$$h(i,j) = \begin{cases} 1, & \text{if } i \text{ and } j \text{ are in directly adjacent columns} \\ 0, & \text{else.} \end{cases} \tag{10}$$

Fig. 7 shows two maps of basis components learned by unsupervised TICA algorithms with two different initialization and order of training samples. The TICA basis components in one column belong to the same view group as in the ISA case; in addition to the view grouping, an ordering by view are automatically formed in the TICA result due to the modeling of dependency between neighboring view subspaces. In contrast, there is no such ordering in the ISA basis components because the model of ISA assumes different view subspaces are mutually independent.

According to the TICA model, the probability distribution of s_i should also be super-Gaussian, as in the ISA case Fig. 8 shows that the 150 TICA coefficients have positive kurtosis and are also super-Gaussian.

D. How View Subspaces are Learned?

Now, how does the formation of view subspaces emerge in ICA-based learning; in other words, how do they learn "representative" subspaces of facial views from the view unlabeled data? We make a case study using ISA.

Fig. 9. Top row: Basis components of one single-view subspace learned by using a training set consisting of all but the frontal view. Middle block: Basis components of $L = 9$ view subspaces learned from frontal view faces only; they are of all but the frontal view. Bottom row: Basis components of the complement subspace to the subspace spanned by the basis components shown in the middle. They are of the frontal view missing in the middle block.

Two experiments are performed to analyze the underlying mechanisms of ISA learning. First, we use a training set consisting of face examples of all but frontal view, and obtained an ISA result $\{\mathbf{b}_j, j = 1, \ldots, \#(S^1)\}$ by minimizing (8) with $L = 1$. When we visualize each basis component $\mathbf{b}_j$ by image, we find that all basis components $\mathbf{b}_j$ are of frontal view, as shown in the top row of Fig. 9. In other words, the result $\{\mathbf{b}_j, j \in S^1\}$ consists of basis components of frontal view subspace.

In the second experiment, on the other hand, we use a training set consisting of face examples of the frontal view only and set $L = 9$. The learned ISA result $\cup_{\ell=1}^{9}\{\mathbf{b}_j, j \in S^\ell\}$ is shown in the middle block of Fig. 9, the ℓth row of which shows the ℓth set of basis components $\{\mathbf{b}_j, j \in S^\ell\}$, $\ell = 1, \ldots, 9$. As can be seen from the figure, the learned basis components are all but the frontal view. So we can take the learned result $\cup_{\ell=1}^{9}\{\mathbf{b}_j, j \in S^\ell\}$ as basis components of the nine view subspace from 0° to 80°.

To further investigate properties of the ISA subspace spanned by the basis components in the middle block, we compute the orthogonal complement subspace (i.e., its basis components) to the subspace, shown in the bottom row of the figure. We see that the basis components of the complement subspace is exactly of the frontal view.

The answer to the question can be found through analyzing the minimization of the energy function (8). Let $\mathbf{X}^0$ be the PCA subspace of the data points in the original input data space. Because minimizing (8) forces the basis components $\mathbf{b}_i$ ($i \in S^\ell$, $\ell = 1, \ldots, L$) to be as orthogonal as possible to the input data, the L ISA subspaces can be considered as approximately orthogonal to the PCA subspace $\mathbf{X}^0$. A more detailed explanation will be given in the next sub-section.

To summarize, the formation of a view subspace can not be obtained directly by ISA learning algorithm using training data of that view. There are two ways for learning basis components of the kth view subspace: 1) using training face examples of

Fig. 10. Basis components learned by using s-ISA. Each row consists of basis components for one view subspace. These components appear more sensible than those learned by u-ISA.

$L-1$ views (all but view k) to derive $\#(S^k)$ (the number of elements in the set S^k) basis components of the kth view and 2) using training face examples of view k only to derive $L-1$ sets of $\#(S^\ell)$ basis components, which can be considered as the basis components of all but the kth view subspace, and then calculate $\#(S^k)$ basis components orthogonal to all the derived components as the basis of the kth view subspace.

This suggests a supervised way of ISA learning (s-ISA) in which the training face examples of a view are used to derive the basis components of that view subspace directly without the need for the calculation of the complement subspace. The s-ISA assumes that view label is known for every training example. One view subspace is learned by using the training examples of that view only. The components, thus, learned appear to be clearer and more sensible than those learned by using the unsupervised method (Fig. 10). It is shown in [32] that when the basis is orthonormal, the s-ISA method is equivalent to the view-based PCA method of [50].

IV. Pose Classification in View-Subspaces

The learned view subspaces $\{\mathcal{S}^\ell\}$ provide a basis for pose estimation. The activity of an input image $\mathbf{x}$ in view subspace ℓ is defined as the norm of the projection of $\mathbf{x}$ onto $\mathcal{S}^\ell$ (the ℓth view subspace)

$$F^\ell(\mathbf{x}) = \sum_{i=1}^{M} \langle \mathbf{u}_i^{(\ell)}, \mathbf{x} \rangle^2 \tag{11}$$

where $\mathbf{u}_i^{(\ell)}$ (for $i = 1, \ldots, M$) are the orthogonal basis components of the view subspace $\mathcal{S}^\ell$. The activity corresponds to the response of a complex cell in mammalian primary visual cortex (V1) [46].

The pose estimation is performed by classifying the input $\mathbf{x}$ into one of the view groups according to a principle called maximum view subspace activity (MVSA). This is done as follows: An image is projected onto each view subspace and the subspace activity defined in (11) is then computed. This gives

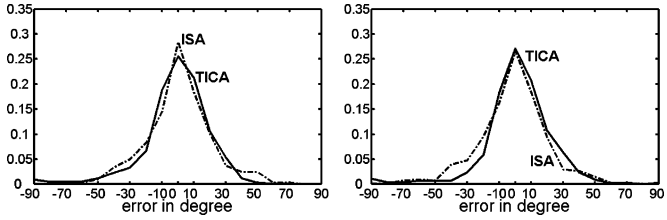


Fig. 11. (Solid lines) Pose estimation error distributions of the unsupervised TICA and (dashed lines) unsupervised ISA for (left) the training and (right) test sets.

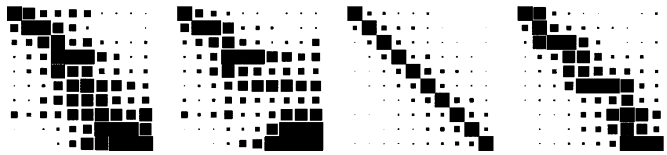


Fig. 12. Hinton diagrams showing the accuracies of pose estimates on the training and test data sets for u-ISA (two on the left) and s-ISA.

$F^0, F^1, \ldots, F^{L-1}$. Then a sample is classified using the MVSA criterion: it belongs to the ℓ'th view if $\ell' = \arg\max_\ell F^\ell$. The MVSA criterion assumes that the training data of a specific view has larger activity in its own view subspace than in other view subspaces.

Now, we give a comparison between unsupervised ISA and TICA in facial pose estimation. A good representation of view subspaces should take into account the intrinsic correlation among view subspaces with similar view. ISA assumes that different view subspaces are independent of each other; while TICA can model this kind of correlation between view subspaces by introducing independency between neighboring components in different view subspaces. In this sense, TICA seems to be a more advantageous model than ISA in unsupervised learning of view subspaces. The distributions of pose estimation errors of the two methods for training and test data sets is shown in Fig. 11. The results show that these two methods have similar performance for pose estimation, which is reasonable since neither of the two unsupervised algorithms takes full advantage of the view label information of the training data.

The pose estimation accuracies of u-ISA and s-ISA are demonstrated in Fig. 12 through Hinton diagrams of confusion matrices (c-matrices). The block size of an entry (i, ℓ) in a c-matrix represents the (normalized) number of samples whose ground truth view label is ℓ and classified into the ith view subspace. The left-most column corresponds to the frontal view for the ground truth, and right-most to the side view. The top row corresponds to the frontal view subspace, and the bottom row is for the side view subspace. The ideal case should be such that the "diagonal" elements of the c-matrix are all ones whereas other elements are all zeros.

Although the applications of ICA-based methods here are for view subspace learning, it would be interesting to consider whether the ICA-based methods would apply to the more general problem of unsupervised object categorization (e.g., [51]).

V. Conclusion and Discussion

The contributions of the paper are the following. First, we presented an ICA-based approach for learning view subspaces. Second, we provided explanations for the emergent formation of view subspaces in the unsupervised ISA (u-ISA) learning. Third, in the probe of the reasons, we found a surprising phenomenon that u-ISA actually derived basis components which were approximately orthogonal to the PCA space determined by the training data, in the sense that the basis components pointed toward regions where the data points were sparse.

References

[1] A. Kuchinsky, C. Pering, M. L. Creech, D. Freeze, B. Serra, and J. Gwizdka, "Fotofile: A consumer multimedia organization and retrieval system," presented at the ACM SIG CHI Conf., Pittsburg, PA, May 1999.

[2] R. Brunelli, "Estimation of pose and illuminant direction for face processing," Massachusetts Inst. Technol., Cambridge, MA, A. I. Memo 1499, 1994.

[3] P. W. Hallinan, "A low-dimensional representation of human faces for arbitrary lighting conditions," in *Proc. IEEE Computer Society Conf. Computer Vision and Pattern Recognition*, 1994, pp. 995–999.

[4] R. Epstein, P. Hallinan, and A. Yuille, "5 ± 2 eigenimages suffice: An empirical investigation of low-dimensional lighting models," in *Proc. IEEE Workshop Physics-Based Vision*, 1995, pp. 108–116.

[5] Y. Adini, Y. Moses, and S. Ullman, "Face recognition: The problem of compensating for changes in illumination direction," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 721–732, Jul. 1997.

[6] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 711–720, Jul. 1997.

[7] K. Etemad and R. Chellapa, *Face Recognition Using Discriminant Eigenvectors*, 1996.

[8] A. Shashua, "On photometric issues in 3d visual recognition from a single 2d image," *Int. J. Comput. Vis.*, vol. 21, pp. 99–122, 1997.

[9] S. Baker, S. Nayar, and H. Murase, "Parametric feature detection," *Int. J. Comput. Vis.*, vol. 27, no. 1, pp. 27–50, Mar. 1998.

[10] P. N. Belhumeur and D. J. Kriegman, "What is the set of images of an object under all possible illumination conditions," *Int. J. Comput. Vis.*, vol. 28, no. 3, pp. 245–260, Jul. 1998.

[11] A. S. Georghiades, D. J. Kriegman, and P. N. Belhumeur, "Illumination cones for recognition under variable lighting: Faces," in *Proc. IEEE Computer Society Conf. Computer Vision and Pattern Recognition*, 1998, pp. 52–59.

[12] H. F. Chen, P. N. Belhumeur, and D. W. Jacobs, "In search of illumination invariants," in *Proc. IEEE Computer Society Conf. Computer Vision and Pattern Recognition*, 2000, pp. 254–261.

[13] A. Yilmaz and M. Gokmen, "Eigenhill vs. eigenface and eigenedge," in *Proc. Int. Conf. Pattern Recognition*, Barcelona, Spain, 2000, pp. 827–830.

[14] J. Hornegger, H. Niemann, and R. Risack, "Appearance-based object recognition using optimal feature transforms," *Pattern Recognit.*, vol. 33, no. 2, pp. 209–224, Feb. 2000.

[15] A. Shashua and T. R. Raviv, "The quotient image: Class based re-rendering and recognition with varying illuminations," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 2, pp. 129–139, May 2001.

[16] M. Bichsel and A. P. Pentland, "Human face recognition and the face image set's topology," *CVGIP: Image Understanding*, vol. 59, pp. 254–261, 1994.

[17] H. Murase and S. K. Nayar, "Visual learning and recognition of 3-D objects from appearance," *Int. J. Comput. Vis.*, vol. 14, pp. 5–24, 1995.

[18] S. Gong, S. McKenna, and J. Collins, "An investigation into face pose distribution," presented at the IEEE Int. Conf. Face and Gesture Recognition, 1996.

[19] D. Graham and N. Allinson, "Face recognition from unfamiliar views: Subspace methods and pose dependency," in *Proc. 3rd Int. Conf. Automatic Face and Gesture Recognition*, Nara, Japan, Apr. 1998, pp. 348–353.

[20] H. Borotschnig, L. Paletta, M. Prantl, and A. Pinz, "Active aobject recognition in parametric eigenspace," in *Proc. 9th Brit. Machine Vision Conf.*, Southampton, UK, 1998, pp. 63–72.

[21] M. Kirby and L. Sirovich, "Application of the karhunen-loeve procedure for the characterization of human faces," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 12, no. 1, pp. 103–108, Jan. 1990.

[22] M. A. Turk and A. P. Pentland, "Face recognition using eigenfaces," in *Proc. IEEE Computer Society Conf. Computer Vision and Pattern Recognition*, Jun. 1991, pp. 586–591.

[23] A. Talukder and D. Casasent, "A closed-form neural network for discriminatory feature extraction from high-dimensional data," *Neural Netw.*, vol. 14, no. 9, pp. 1201–1218, 2001.

[24] ——, "Pose invariant recognition of faces at unknown aspect views," presented at the Int. Joint Conf. Neural Networks (IJCNN), 1999.

[25] ——, "Classification and pose estimation of objects using nonlinear features," in *Proc. SPIE: Applications and Science of Computational Intelligence*, vol. 3390, Apr. 1998, pp. 12–23.

[26] A. Khotanzad and J. J.-H. Liou, "Recognition and pose estimation of unoccluded three-dimensional objects from a two-dimensional perspective view by banks of neural networks," *IEEE Trans. Neural Netw.*, vol. 7, no. 4, pp. 897–906, Aug. 1996.

[27] C. Jutten and J. Herault, "Blind separation of sources, part I: An adaptive algorithm based on neuromimetic architecture," *Signal Process.*, vol. 24, pp. 1–10, 1991.

[28] P. Comon, "Independent component analysis—A new concept?," *Signal Process.*, vol. 36, pp. 287–314, 1994.

[29] S. Z. Li, X. G. Lu, H. J. Zhang, Q. Fu, and Y. Cheng, "Learning topographic representation for multi-view object appearances," in *Proc. ICASSP*, vol. 2, Salt Lake City, UT, May 8–11, 2001, pp. 1329–1332.

[30] S. Z. Li, X. G. Lu, and H. J. Zhang, "View based clustering of object appearances based on independent subspace analysis," in *Proc. IEEE Int. Conf. Computer Vision*, vol. 2, Vancouver, BC, Canada, Jul. 2001, pp. 295–300.

[31] ——, "View-subspace analysis of multi-view face patterns," in *Proc. IEEE ICCV Workshop on Recognition, Analysis, and Tracking of Faces and Gestures in Real-time Systems*, Vancouve, BCr, Canada, Jul. 13, 2001, pp. 125–132.

[32] S. Z. Li, X. H. Peng, X. W. Hou, H. J. Zhang, and Q. S. Cheng, "Multi-view face pose estimation based on supervised learning," presented at the IEEE Int. Conf. Automatic Face and Gesture Recognition, Washington, DC, May 2002.

[33] A. Hyvärinen, P. Hoyer, and M. Inki, "Topographic independent component analysis," Neural Comput., vol. 13, to be published.

[34] A. Hyvärinen and P. Hoyer, "Emergence of topography and complex cell properties from natural images using extensions of ica," *Adv. Neural Inf. Process. Syst.*, vol. 12, pp. 827–833, 2000.

[35] B. A. Olshausen and D. J. Field, "Natural image statistics and efficient coding," *Network*, vol. 7, pp. 333–339, 1996.

[36] A. J. Bell and T. J. Sejnowski, "The independent components of natural scenes are edge filters," *Vis. Res.*, vol. 37, pp. 3327–3338, 1997.

[37] T. Lee, M. Lewicki, and T. Sejnowski, "ICA mixture models for unsupervised classification of nongaussian classes and automatic context switching in blind separation," *Pattern Anal. Mach. Intell.*, vol. 22, no. 10, Oct. 2000.

[38] M. S. Bartlett, H. M. Lades, and T. J. Sejnowski, "Independent component representations for face recognition," in *Proc. SPIE Conf. Human Vision and Electronic Imaging III*, vol. 3299, 1998, pp. 528–539.

[39] B. Moghaddam, "Principal manifolds and Bayesian subspaces for visual recognition," in *Proc. Int. Conf. Computer Vision*, Sep. 1999, citeseer.nj.nec.com/moghaddam99principal.html, pp. 1131–1136.

[40] C. Liu and H. Wechsler, "Comparative assessment of independent component analysis (ICA) for face recognition," presented at the 2nd Int. Conf. Audio- and Video-based Biometric Person Authentication, Washington, DC, Mar. 1999.

[41] R. Manduchi and J. Portilla, "Independent component analysis of textures," presented at the IEEE Int. Conf. Computer Vision, Corfu, Greece, 1999.

[42] J. M. M. S. Bartlett and T. Sejnowski, "Face recognition by independent component analysis," *IEEE Trans. Neural Netw.*, vol. 13, no. 5, pp. 1450–1464, Oct. 2002.

[43] A. Hyvärinen and E. Oja, "Independent component analysis: Algorithms and applications," *Neural Netw.*, vol. 13, no. 4, pp. 411–430, 2000.

[44] D.-T. Pham, P. Garrat, and C. Jutten, "Separation of a mixture of independent sources through a maximum likelihood approach," in *Proc. EUSIPCO*, 1992, pp. 771–774.

[45] J.-F. Cardoso, "Blind signal separation: Statistical principles," *Proc. IEEE*, vol. 90, no. 10, pp. 2009–2025, Oct. 1998.

[46] A. Hyvärinen and P. Hoyer, "Emergence of phase and shift invariant features by decomposition of natural images into independent feature subspaces," *Neural Comput.*, vol. 12, no. 7, pp. 1705–1720, 2000.

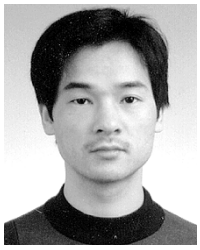
[47] T. Kohonen, "Emergence of invariant-feature detectors in the adaptive-subspace self-organizing maps," *Biol. Cybern.*, vol. 75, pp. 281–291, 1996.
[48] ——, *Self-Organizing Maps*, 2nd ed, ser. Information Sciences. Heidelberg, Germany: Springer, 1997.
[49] J.-F. Cardoso, "Multidimensional independent component anlysis," presented at the Int. Conf. Acoustic, Speech and Signal Processing, Seattle, WA, 1998.
[50] A. P. Pentland, B. Moghaddam, and T. Starner, "View-based and modular eigenspaces for face recognition," in *Proc. IEEE Computer Society Conf. Computer Vision and Pattern Recognition*, 1994, pp. 84–91.
[51] R. Fergus, P. Perona, and A. Zisserman, "Object class recognition by unsupervised scale-invariant learning," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, Madison, WI, Jun. 2003.

Stan Z. Li received the B.Eng. degree from Hunan University, Hunan, China, the M.Eng. degree from the National University of Defense Technology, and the Ph.D. degree from Surrey University, Surrey, U.K., where he was also a Research Fellow.

He is a Researcher at Microsoft Research Asia, Beijing, China. He joined Microsoft Research Asia in May 2000 from his post as an Associate Professor with Nanyang Technological University, Singapore. He is the author of *Markov Random Field Modeling in Image Analysis* (New York: Springer-Verlag, 2nd edition, 2001), and he co-edited, with Anil K. Jain, the *Handbook of Face Recognition* (New York: Springer-Verlag, 2004). His current research interest is in pattern recognition and machine learning, image analysis, face recognition technologies, and biometrics, and he has published over 160 refereed papers and book chapters in these areas.

XiaoGuang Lu received the B.S. and M.S. degrees from the Department of Automation, Tsinghua University, China, in 1997 and 2000, respectively. He is currently pursuing the Ph.D. degree in pattern recognition at the Image Processing Laboratory, Department of Computer Science and Engineering, Michigan State University, East Lansing.

He spent one year at Microsoft Research Asia, Beijing, China, as a visiting student from 2000 to 2001.



Xinwen Hou received the B.S. degree in physics from Zhengzhou University, China, in 1995, the M.S. degree in electronic engineering from the University of Science and Technology of China in 1998, and the Ph.D. degree in mathematics from Peking University, Beijing, China, in 2001.

From 2001 to 2003, he was a Postdoctorate with Nankai University, China. He is currently a Researcher with the Automation Institute, Chinese Academy of Science. His work is centered on independent component analysis, shape tracking, and face recognition.

Xianhua Peng received the B.S. and M.S. degrees in applied mathematics from the School of Mathematical Sciences, Peking University, Beijing, China, in 2000 and 2003, respectively.

His research interests include pattern recognition, digital signal processing, stochastic models, simulations, and applications to telecommunications.

Qiansheng Cheng received the B.S. degree in mathematics from Peking University, Beijing, China, in 1963.

Since May 1989, he has been a Professor with the School of Mathematical Sciences, Peking University. His current research interests include signal processing, nonlinear time series analysis, and pattern recognition.

Prof. Cheng is the Vice Chairman of the Chinese Signal Processing Society and he won the Chinese National Natural Science Award.