Intrinsic Generalization Analysis of Low Dimensional Representations

Xiuwen Liu*, Anuj Srivastava†, DeLiang Wang‡
*Department of Computer Science, Florida State University, Tallahassee, FL 32306-4530 (Email: liux@cs.fsu.edu)
† Department of Statistics, Florida State University, Tallahassee, FL 32306-4330 (Email: anuj@stat.fsu.edu)
‡ Department of Computer & Information Science and Center for Cognitive Science, The Ohio State University
Columbus, OH 43210 (Email: dwang@cis.ohio-state.edu)

***Abstract*— Low dimensional representations of images impose equivalence relations in the image space; the induced equivalence class of an image is named as its intrinsic generalization. The intrinsic generalization of a representation provides a novel way to measure its generalization and leads to more fundamental insights than the commonly used recognition performance, which is heavily influenced by the choice of training and test data. We demonstrate the limitations of linear subspace representations by sampling their intrinsic generalization, and propose a nonlinear representation that overcomes these limitations. The proposed representation projects images nonlinearly into the marginal densities of their filter responses, followed by linear projections of the marginals. We use experiments on large datasets to show that the representations that have better intrinsic generalization also lead to better recognition performance.**

I. Introduction

In recent years, appearance-based methods have become a common choice for building recognition systems based on images. The main motivation is that variations in images of objects can be captured through (a large number of) training images, which avoids the necessity of building complex probability models. Due to the high dimensionality of images, low dimensional representations become necessary in order to develop computationally efficient systems. For example, principal component analysis (PCA) (Hotelling, 1933), also known as Karhunen and Loève transformation (Karhunen, 1947; Loève, 1955), has become a widely used tool for dimension reduction. One of the major limitations of PCA representation is that it is not able to capture statistics higher than the second order. Independent component analysis (ICA) (Comon, 1994) [1] (see Hyvärinen et al. (2001) for a review) has been used to overcome this limitation by imposing statistical independence among the linear bases. By maximizing a discrimination measure among different classes, Fisher discriminant analysis (FDA) offers another popular linear subspace representation (Fisher, 1936). In computer vision, these representations have been applied to face recognition (Sirovich and Kirby, 1987; Turk and Pentland, 1991; Belhumeur et al., 1997).

As the recognition performance of a classifier depends heavily on the choice of training data, it becomes important to study the generalization of a low dimensional representation through the equivalence relation it imposes on the image space. The importance of studying the equivalence classes for generalization is greatly emphasized by Vapnik (2000). In fact, Vapnik named the cardinality of equivalence classes a new concept (to be studied) for learning from small samples (p. 296, Vapnik (2000)). While the cardinality of equivalence classes is important for reducing dimensionality, for recognition performance, it is also important to study the properties of images in a particular equivalence class. Ideally, only images with similar underlying models should be grouped into an equivalence class. We will name this semantics-related aspect of generalization as *intrinsic generalization*. This isolates an intrinsic aspect of a representation that affects the recognition performance. Our study of intrinsic generalization for linear subspaces of images reveals that these representations group images from different models within the same equivalent class and are inherently sensitive to noise and deformations. By analyzing two problems with linear subspace representations, we propose a way to improve their intrinsic generalization, the advantage of which is demonstrated by recognition results. We emphasize that our goal is to characterize an important aspect of a representation through the study of intrinsic generalization, which leads to an important measure to compare different representations.

This paper is organized as follows. Section II gives a definition of generalization and then introduces intrinsic generalization and shows that of linear subspaces. Section III briefly describes the spectral histogram representation of images, and spectral histogram subspace analysis (SHSA), and shows the intrinsic generalization of SHSA through object synthesis. Section IV shows the experimental results on recognition of large datasets. Section V concludes the paper with a discussion on a number of related issues.

II. Intrinsic Generalization

In this paper, an image $\mathbf{I}$ is defined on a finite lattice $\mathcal{L} \subset \mathbf{Z}^2$, the intensity at pixel location $\vec{v} \in \mathcal{L}$ is denoted by $\mathbf{I}(\vec{v}) \in \mathcal{G} = [r_1, r_2]$, where r_1, r_2 bound the dynamic range of the imaging sensor, and Ω the set of all images on $\mathcal{L}$. A representation is a mapping defined as $f : \Omega \to R^K$. For a low dimensional representation, we require $K \ll |\mathcal{L}|$. Before we introduce the notion of intrinsic generalization, we first give a formal definition of generalization of representations.

A. Generalization of Low Dimensional Representations

In learning-based recognition of objects from images, a classifier/recognizer is often trained using some training data and is applied to classify/recognize future inputs in the form of test data. A key issue is to extend good performance on

[1] As pointed out by Simoncelli and Olshausen (2001), ICA is a misnomer as estimated independent components are not guaranteed to be independent.

test inputs using information limited to the training set; it is commonly known as the generalization problem (Bishop, 1995).

There are several ways of formulating the generalization problem and we have chosen the framework laid out by Bishop (1995). Let the observed images be generated from an unknown probability density P on Ω and the underlying true recognition function be denoted by $h : \Omega \mapsto \mathcal{A}$, where $\mathcal{A}$ is the set of all classes. For any classifier function g, its average generalization ability is defined as the probability that $g(\mathbf{I}) = h(\mathbf{I})$, i.e.

$$G(g) = Pr\{\mathbf{I} | \mathbf{I} \in \Omega, g(\mathbf{I}) = h(\mathbf{I})\}. \tag{1}$$

According to (1), obtaining good generalization becomes an optimization over all the possible classifier functions. In practice, since the underlying model $P(\mathbf{I})$ and $h(\mathbf{I})$ are generally unknown, several ways have been proposed. One way is to estimate $G(g)$ directly through a set separate from the training one with known class labels such as cross-validation (Bishop, 1995). Another way is to impose additional constraints on $G(g)$ based on some generic heuristics such as Akaike information criterion (Akaike, 1973) and minimum description length (Rissanen, 1978), where a model with more free parameters is penalized. Yet another approach is to optimize an analytical bound based on statistical analysis such as the worst case performance of all the implementable classifiers of a neural network architecture (Baum and Haussler, 1989). Note that all the existing efforts on generalization have been focused on the generalization of classifiers.

Because of the high dimensionality of images, dimension reduction becomes necessary for computational reasons. In case of using a low dimensional representation f, the average generalization ability then becomes the probability that $\hat{g}(f(\mathbf{I})) = h(\mathbf{I})$ for an input $\mathbf{I}$ randomly sampled according to $P(\mathbf{I})$, where $\hat{g}$ is a classifier based on a low dimensional representation f. In other words, we have

$$\begin{array}{l} G(\hat{g}, f) = Pr\{\mathbf{I} | \mathbf{I} \in \Omega, g(f(\mathbf{I})) = h(\mathbf{I})\} \\ = \sum_{f(\mathbf{I})} Pr\{\mathbf{J} | \mathbf{J} \in \Omega, f(\mathbf{J}) = f(\mathbf{I})\} \; 1_{\hat{g}(f(\mathbf{I}))=h(\mathbf{I})}, \end{array} \tag{2}$$

where $1_{x=y}$ is an indicator function, and we use $f(\mathbf{I})$ as the range of f on Ω. From (2), it is clear that f has a significant effect on the generalization of $\hat{g}$. Ideally, we want to group all the images from each class as a single equivalence class. (In this case, the classifier is trivial.) While this is generally not possible for real applications, we want to group images from each class into a small number of equivalence classes, with each class having a large cardinality, as emphasized by Vapnik (2000). This also reduces the number of necessary training images. However, when making each equivalence class as large as possible, we do not want to include images from other classes, as this will make a good classification performance impossible. This leads to the need of analyzing equivalence class structures of low dimensional representations to achieve a good generalization performance.

B. Intrinsic Generalization

The previous analysis shows that the equivalence class structures of low dimensional representations are essential for a good generalization performance. In this paper, we focus on studying the images of a particular equivalence class through statistical sampling.

Definition: Given a representation f, the intrinsic generalization of an image $\mathbf{I}$ under f is defined as

$$S_I(\mathbf{I}) = \{\mathbf{J} | \mathbf{J} \in \Omega, f(\mathbf{J}) = f(\mathbf{I})\} \; \subset \Omega. \tag{3}$$

In other words, intrinsic generalization of image $\mathbf{I}$ includes all the images that cannot be distinguished from $\mathbf{I}$ under representation f. The recognition performance based on f depends critically on the intrinsic generalizations of training images as the images in their intrinsic generalizations are implicitly included in the training set. We define $S_I^0(\mathbf{I})$ as the images sharing the same underlying probability models with $\mathbf{I}$. Ideally, $S_I(\mathbf{I})$ should be as close as possible to $S_I^0(\mathbf{I})$. As $S_I^0(\mathbf{I})$ is generally not available, to explore $S_I(\mathbf{I})$, we employ statistical sampling through the following probability model:

$$q(\mathbf{J}, T) = \frac{1}{Z(T)} \exp\{-D(f(\mathbf{J}), f(\mathbf{I}))/T\}. \tag{4}$$

Here T is a temperature parameter, $D(.,.)$ a Euclidean or other distance measure, and $Z(T)$ is a normalizing function, given as $Z(T) = \sum_{\mathbf{J} \in \Omega} \exp\{-D(f(\mathbf{J}), f(\mathbf{I}))/T\}$. This model has been used for texture synthesis (Zhu et al., 2000) and we generalize it to any representation. It is easy to see that as $T \to 0$, $q(\mathbf{J}, T)$ defines a uniform distribution on $S_I(\mathbf{I})$ (Zhu et al., 2000). The advantage of using a sampler is to be able to generate typical images in $S_I(\mathbf{I})$ so that $S_I(\mathbf{I})$ under f can be examined in a statistical sense.

C. Intrinsic Generalization of Linear Subspaces

Linear subspace representations of images, including PCA, ICA, and FDA, assume that f is a linear map, and $S_I(\mathbf{I})$ forms a linear subspace. While these methods are successful when applied to images belonging to a specific nature, e.g. face images, their generalization seems poor if we consider $S_I(\mathbf{I})$ under these linear subspace methods in Ω.

If $S_I^0(\mathbf{I})$ is available then one can analyze the overlap between the sets $S_I^0(\mathbf{I})$ and $S_I(\mathbf{I})$. If not, then one has to resort to some indirect techniques such a random sampling to compare elements of the two sets. Random sampling seems sufficient in that the typical images in $S_I(\mathbf{I})$ are very different from $\mathbf{I}$. To illustrate these ideas, we have used PCA of the ORL face dataset[2], which consists of 40 subjects with 10 images each; we have obtained similar results using other linear subspaces. We calculate the eigen faces corresponding to the 50 largest eigenvalues. Under PCA, given an image $\mathbf{I}$, $f(\mathbf{I})$ is the projections of $\mathbf{I}$ along eigen faces. We define the reconstructed image[3] of $\mathbf{I}$ as $\pi(\mathbf{I}) = \sum_{i=1}^{K} < \mathbf{I}, \mathbf{V}_i > \mathbf{V}_i$, where $\mathbf{V}_i$ is the ith eigen face and $< .,. >$ is the canonical

[2] http://www.uk.research.att.com/facedatabase.html.

[3] We assume the mean is taken care of properly.

inner product. Fig. 1(a) shows a face image in the dataset and Fig. 1(b) shows the reconstructed image with $K = 50$. We then use a Gibbs sampler to generate samples of $S_I(\mathbf{I})$ by sampling from $q(\mathbf{J}, T)$ given by (4). Fig. 1(c)-(f) show four samples of $S_I(\mathbf{I})$. (For Fig. 1(f), the object in the middle is used as boundary condition, i.e., pixels on the object are not updated.) In other words, all these images have the same 50 eigen decompositions. Note that $S_I(\mathbf{I})$ is defined on Ω and these images are far from each other in Ω. As expected, the corresponding reconstructed images are identical to Fig. 1(b).

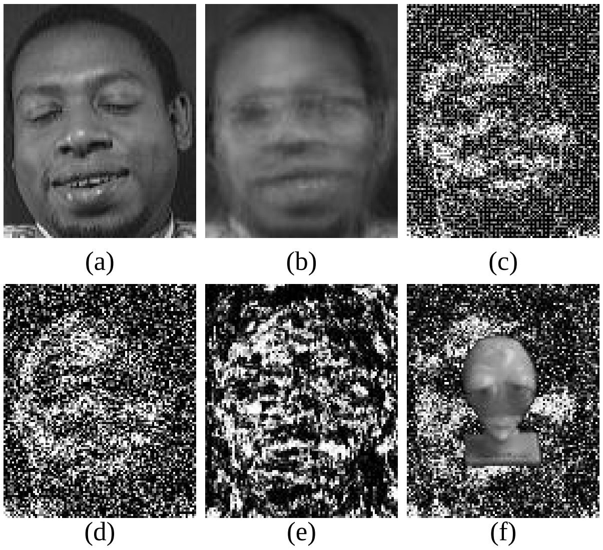


Fig. 1. (a) A face image. (b) Reconstructed image using $K = 50$ principal components. (c)-(f) Four random samples from the set $S_I(\mathbf{I})$, with $\pi(\mathbf{I})$ identical to the one shown in (b).

Because $S_I(\mathbf{I})$ consists of images from various underlying probability models, the linear subspace representations can make the subsequent classification intrinsically sensitive to noise and other deformations. To show that, Fig. 2(a) gives three different face images which share exactly the same eigen representation (bottom row). On the other hand, Fig. 2(b) shows three similar images whose eigen representations correspond to three different faces.

We emphasize here that the sampling is very different from reconstruction. The sampling is to draw a typical sample from the set of all the images with a particular low dimensional representation while reconstruction gives one in the set whose coefficients are zero along the dimensions complement to the given subspace. To illustrate this, Fig. 3 shows an example of a one-dimensional subspace in a two-dimensional space. In this case, the reconstructed "image" of "x" is the point given by "+" in Fig. 3(a) while the sampling can return any point with equal probability along the solid line shown in Fig. 3(b). This shows clearly that the reconstructed image may not provide much information about all the other images having the same low dimensional representation. This has an important implication for building low dimensional generative models. For recognition purpose, it is not sufficient to show that the synthesized image based on a generative model has certain desirable properties as in PCA; to be a good model, all the images in the intrinsic generalization should also exhibit the desirable properties. As generative models become popular in modeling, their intrinsic generalizations should also be studied and analyzed.

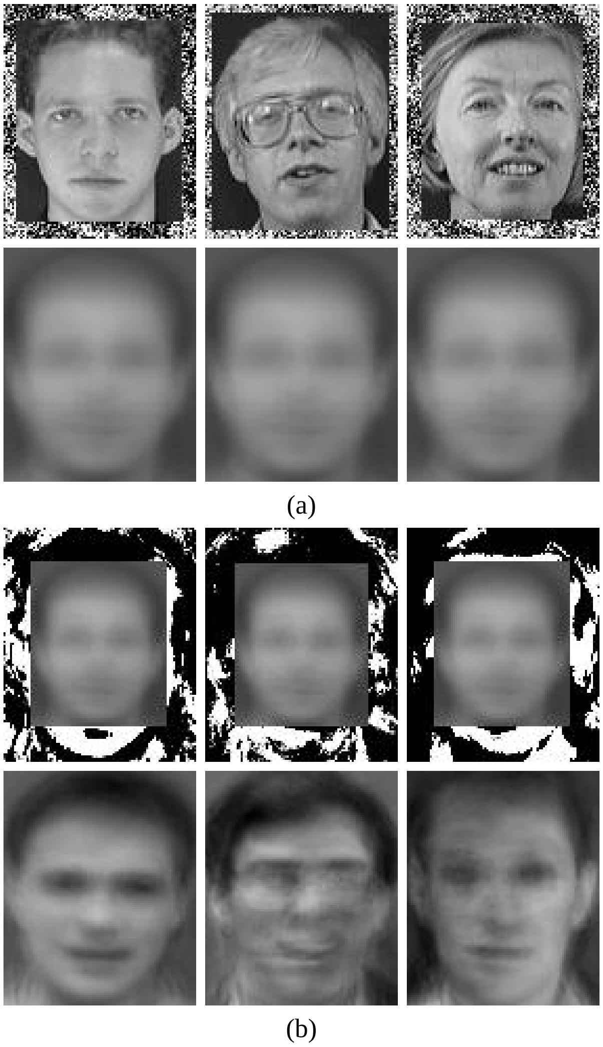


Fig. 2. Examples of different images with identical eigen decompositions and similar images with different eigen decompositions. The top row shows the images and the bottom reconstructed. (a) Three different images with the same eigen representations. (b) Three similar images with different eigen representations.

These results, while generated using PCA, are valid for an arbitrary linear subspace since the sampling tries to match the representation. The main problems of linear subspace representations, as revealed here, are that these representations can not account for the fact that most images in the image space are white noise images. Additionally, they can not incorporate important spatial constraints in images.

III. Spectral Histogram Subspace Analysis

A. Spectral Histogram Representation of Images

As discussed earlier, an ideal representation f for $\mathbf{I}$ will be such that $S_I(\mathbf{I}) = S_I^0(\mathbf{I})$. There are two important limitations of the linear methods that need to be addressed: (i) As the vast majority images in Ω are white noise images, a

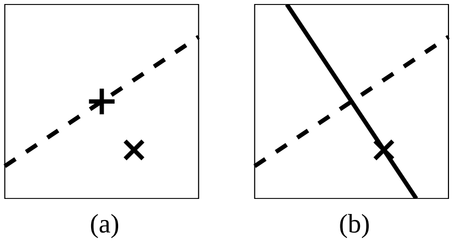


Fig. 3. An illustration example of the difference between sampling and reconstruction. Here the dashed line represents a one-dimensional subspace in a two-dimensional space. For a training example (marked as 'x'), the sampling is to draw a random point along the solid line in (b) while the reconstructed image is a single point given by '+' in (a).

good approximation of $S_I^0(\mathbf{I})$ for an image of object(s) must handle white noise images effectively; otherwise, $S_I(\mathbf{I})$ will concentrate on white noise images. Experiments show that linear representations suffer from this problem. (ii) Another issue is the linear superposition assumption, where each basis contributes linearly independently to the image. In contrast, pixels on objects are dependent and efficient models should exploit this dependency.

The issue of white noise images can be dealt with effectively through the method of types (Csiszar and Korner, 1981) as the white noise images are grouped together under types. However, the direct use of types does not provide enough constraints as only the histogram of images is used. We generalize the type definition by including marginals of filter responses (of the input image) with respect to a set of filters, which also incorporates local pixel dependence through filtering.

The representation of using marginals of filtered images can be justified in many ways: (i) by assuming that *small disjoint regions in the frequency domain are statistically independent*. That is, partition the frequency domain into small disjoint regions and model each region by its marginal distribution. The partitioning of the frequency also leads to spatial filters. (ii) Wavelet decompositions of images are local in both space and frequency, and hence, provide attractive representations for objects in the images. We convolve an image with the filters and compute the histograms of the resulting images. Each image is then represented by a vector consisting of all the marginal distributions. We shall call this representation *spectral histogram representation* (Liu and Wang, 2002), each of these vectors a *spectral histogram*, and the set of all valid vectors the *spectral histogram space*. Elements of a spectral histogram relate to the image pixels in a nonlinear fashion, and hence, avoid the linearity issue mentioned earlier.

This representation has also been suggested through psychophysical studies on texture modeling (Chubb et al., 1994), and has been used in texture modeling and synthesis (Heeger and Bergen, 1995; Zhu et al., 1997; Liu and Wang, 2002), texture classification (Liu and Wang, 2003), and face recognition (Liu and Cheng, 2003). Both the histogram of input images (Swain and Ballard, 1991) and joint histograms of local fields (Schiele and Crowley, 2000) have been used for object recognition.

B. Spectral Histogram Subspace Analysis

In this method, the strategy is to first represent each image in the spectral histogram space and then apply a linear subspace method, such as PCA, ICA or FDA, in the spectral histogram space[4]. Name these corresponding methods as SHPCA, SHICA, and SHFDA, and call them collectively as spectral histogram subspace analysis (SHSA).

To demonstrate the effectiveness of SHSA representations, we explore their intrinsic generalization through sampling. As in the linear subspace case, we use SHPCA for experiments; similar results have been obtained using other linear spectral histogram subspaces.

First, bases in the spectral histogram space are computed based on training images. Given an image, its spectral histogram representation is computed and then projected onto a spectral histogram subspace. We use a Gibbs sampling procedure to generate images that share the same spectral histogram representation. Fig. 4 shows two sets of examples; Fig. 4(a) shows three texture images and Fig. 4(b) shows three objects. These examples show that the spectral histogram subspace representation captures photometric features as well as topological structures, which are important to characterize and recognize images.

IV. Experimental Results for Recognition

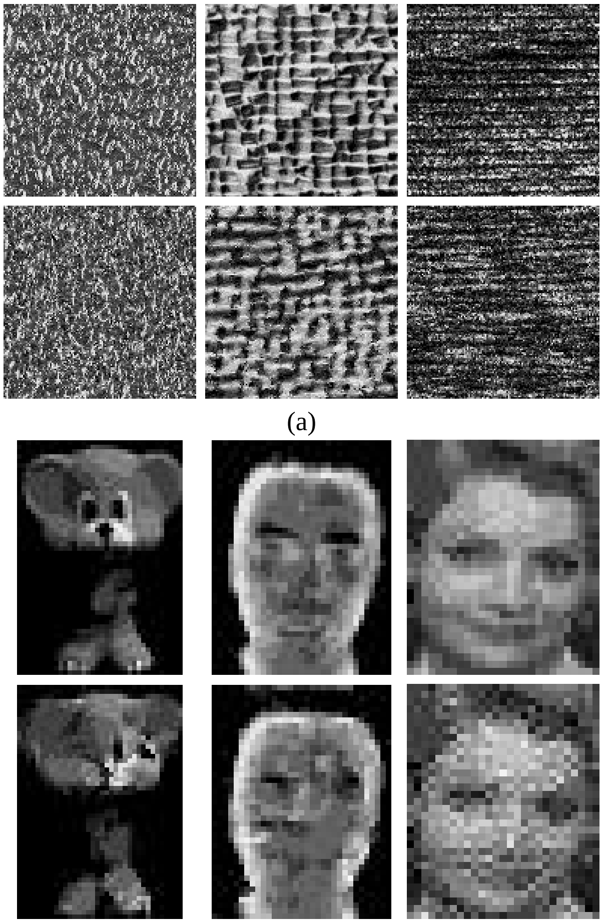
To demonstrate the effectiveness of SHSA representations, we use several data sets and compare their performance with that of linear subspace representations. In our experiments, the number of principal components is determined by thresholding the ratio of a component's eigenvalue and the largest eigenvalue. (If the same threshold is applied to PCA, it tends to keep more components as the dimension of the input space here is much larger). We have used the same number of components for ICA, FDA, SHICA, and SHFDA as PCA and SHPCA respectively. Here ICA is calculated using the FastICA algorithm (Hyvärinen, 1999) and FDA based on an algorithm by Belhumeur et al. (1997). We use the nearest neighbor classifier for recognition. To calculate the spectral histogram, we use a fixed set of 21 filters. These filters were chosen automatically from a larger set of Gabor and Laplacian of Gaussian filters using a filter selection algorithm (Liu and Wang, 2001) for the ORL face dataset.

A classifier's performance in a low dimensional space depends on intrinsic generalization of the representation to the test data. The result for a new image of a classifier is determined by the decision region partitions in the feature space and thus in the image space. Given a training set B, we define the extrinsic generalization set of $\mathbf{I} \in B$ as:

$$\begin{aligned} S_E(\mathbf{I}) \quad = \quad & \{\mathbf{J} | \mathbf{J} \in \Omega, L(\mathbf{J}|B) = L(\mathbf{I})\} \ \setminus \\ & \{\mathbf{J} | \mathbf{J} \in \Omega, L(\mathbf{J}|B \setminus \mathbf{I}) = L(\mathbf{I})\}. \end{aligned} \quad (5)$$

Here $L(\mathbf{I})$ is the label of $\mathbf{I}$ and $L(\mathbf{J}|B)$ is the label of image $\mathbf{J}$ assigned by the classifier trained on the set B. To separate

[4] Note a reconstructed spectral histogram may be outside the spectral histogram space and here we ignore this complication.



(a)

(b)

Fig. 4. Samples from SHPCA intrinsic generalization. In each panel the top row shows the input image and the bottom a typical sample from its intrinsic generalization. (a) Three textures. (b) One object and one face image. Boundary conditions need to be taken with care when sampling from $S_I(\mathbf{I})$.

the effectiveness of a representation from that of the choice of training and test data, we have also used (uniformly) randomly generated bases, which we call random component analysis (RCA) and spectral histogram random component analysis (SHRCA).

First we use the Columbia Object Image Library (COIL-100)[5] dataset, which consists of images of 100 3-D objects with varying pose, texture, shape and size, 21 of which are shown Fig. 5. Pontil and Verri (1998) applied SVM (Support Vector Machines) method to 3D object recognition. Yang et al. (2000) proposed a learning algorithm, named SNoW (Sparse Network of Winnows) for appearance based recognition and applied their learning algorithm to the full COIL dataset and compared with SVM methods. They tested their method by varying the number of training views. For a fair comparison with the results in Yang et al. (2000), these color images were converted to gray level images and downsampled to size 32×32, which are used in all the experiments described here.

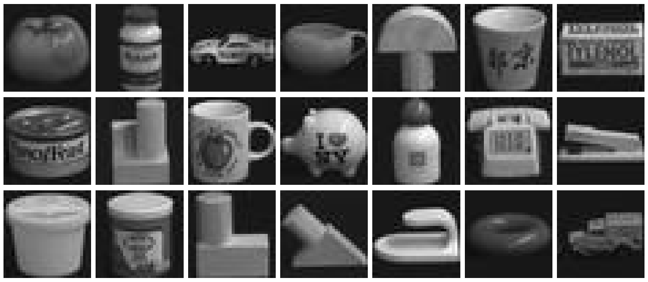


Fig. 5. 21 selected objects from the 100 objects in the database.

As in Yang et al. (2000), we vary the number of training views per object. Tab. I shows the recognition rates on the dataset using PCA, ICA, FDA, SNoW (Yang et al., 2000) and Linear SVM (Yang et al., 2000) as well as SHSA methods. While the COIL-100 dataset is considered to be easy with enough training data, Tab. I reveals clearly the generalization capability of different representations. Under all the conditions, SHSA methods outperform the other methods. Among the SHSA methods, SHFDA gives the best performance. However, FDA in the image space does not outperform other linear methods; this is because these images are not linearly separable in the image space and linear FDA bases are not effective. This result is consistent with that of Martinez and Kak (2001), which showed that FDA may not outperform PCA in general. Another interesting point of Tab. I is that RCA and SHRCA give comparable results to those of other bases, suggesting that the choice of commonly used different bases within a space may not be that critical for recognition as none of them is considerably better than a random one in term of recognition performance.

TABLE I

RECOGNITION RESULTS FOR THE COIL-100 DATASET

	Training / test per object			
Methods	36 / 36	18 / 54	8 / 64	4 / 68
PCA	98.6%	96.7%	87.2%	75.8%
ICA	98.6%	96.5%	87.9%	76.0%
RCA	98.6%	96.3%	87.0%	75.4%
FDA	97.6%	92.6%	82.1%	56.8%
SNoW (Yang et al., 2000)	95.8%	92.3%	85.1%	81.5%
Linear SVM (Yang et al., 2000)	96.0%	91.3%	84.8%	78.5%
SHPCA	99.4%	97.1%	89.3%	82.9%
SHICA	99.4%	97.0%	89.2%	82.7%
SHRCA	99.4%	96.9%	89.1%	83.0%
SHFDA	99.9%	98.9%	94.4%	87.4%

While SHSA representations are translation invariant by definition, the linear subspaces of images are not as translation will change the random variables associated with pixels in the images. Theoretically speaking, SHSA representations are not rotation invariant in general as filters are orientation sensitive. It has been shown that Gabor filters can tolerate a certain amount of rotation (Lades et al., 1993). Tables II and III show the recognition results using PCA, ICA, SHPCA, and SHICA representations with respect to 2-D translations and rotations; similar results have been obtained using other linear subspaces.

[5] Available at http://www.cs.columbia.edu/CAVE.

Here 12 views were used for training and the remaining 60 views were used for testing for each object and borders with background color were added to images so that translations and rotations can be done without clipping the objects. As Tab. II shows, PCA and ICA are very sensitive to translation. While neither PCA, ICA, SHPCA, or SHICA is rotation invariant, SHPCA and SHICA are less sensitive to rotation as shown in Tab. III. In addition, for SHPCA and SHICA the correct recognition rate to the closest three does not decrease much, suggesting that SHPCA and SHICA can be used as an effective means to propose candidates for more accurate models.

TABLE II

RECOGNITION RESULTS FOR THE COIL-100 DATASET WITH RESPECT TO TRANSLATION

	Horizontal / vertical translation				
Methods	0 / 0	2 / 0	0 / 2	2 / 2	4 / 4
PCA	93.3%	81.7%	46.4%	32.3%	4.9%
ICA	93.2%	81.9%	46.8%	32.3%	5.1%
SHPCA	94.7%	94.7%	94.7%	94.7%	94.7%
SHICA	94.6%	94.6%	94.6%	94.6%	94.6%

TABLE III

RECOGNITION RESULTS FOR THE COIL-100 DATASET WITH RESPECT TO 2D ROTATION

	Recognition rate with the correct to be the first				
	Rotation angle				
Methods	0°	5°	10°	15°	20°
PCA	93.3%	91.6%	83.2%	60.8%	38.1%
ICA	93.2%	91.7%	82.6%	61.8%	39.0%
SHPCA	94.7%	93.4%	89.5%	79.4%	66.3%
SHICA	94.6%	93.2%	89.2%	79.0%	66.1%
	Recognition rate with the correct to be among the first three				
PCA	96.9%	96.5%	93.7%	80.8%	60.6%
ICA	96.9%	96.5%	93.6%	81.3%	60.9%
SHPCA	97.1%	96.9%	95.9%	92.5%	83.0%
SHICA	97.0%	96.9%	95.9%	92.3%	82.8%

Eigen faces have been used widely for face recognition applications (see e.g. Sirovich and Kirby (1987); Kirby and Sirovich (1990); Turk and Pentland (1991)). We have also applied our method to ORL[6], a standard face dataset. The dataset consists of faces of 40 subjects with 10 images for each. The images were taken at different times with different lighting conditions on a dark background. While only limited side movement and tilt were allowed, there was no restriction on facial expression.

The procedure is the same as that for 3D object recognition. We use the same 21 filters and vary the number of the training images per subject and the remaining ones are used for testing. We randomly choose the training images from the dataset to avoid potential bias due to the choice of training faces. The average and worst results of 100 trials from linear subspace representations as well as SHSA ones are shown in Table IV. SHSA performance is significantly better than the corresponding linear subspace in the image space essentially because different lighting conditions and facial expressions make the pixel-wise representation not reliable for recognition (Zhang et al., 1997). The results obtained here are also significantly better than those obtained in Zhang et al. (1997) on the same dataset.

TABLE IV

RECOGNITION RESULTS FOR THE ORL FACE DATASET OF 100 TRIALS WITH DIFFERENT NUMBER OF TRAINING/TEST FACES

	Training/test per subject			
	Average rate		Worst rate	
Methods	5 / 5	3 / 7	5 / 5	3 / 7
PCA	94.5%	88.0%	89.0%	80.7%
ICA	94.0%	86.0%	89.0%	75.4%
FDA	98.9%	96.3%	95.5%	91.1%
RCA	85.6%	73.7%	79.5%	65.7%
SHPCA	98.5%	94.6%	95.0%	87.9%
SHICA	98.15%	94.10%	94.00%	87.86%
SHFDA	99.3%	98.5%	98.0%	95.7%
SHRCA	96.2%	90.2%	92.5%	84.6%

We have also applied our method to a dataset of 40 real texture images[7] with size of 256×256. This dataset contains different kinds of natural textures. Also some of textures are similar to others in the dataset, making the classification very challenging.

To do the texture classification, we partition each texture image into non-overlapping patches with size 32×32, resulting in a total of 64 patches per texture type. We then choose a specified number of patches as the training set and the rest are used for testing. As in the face recognition experiment, we randomly choose a given number of patches as training and run our method 100 times. Table V shows the average, best, and worst classification results of 100 trials with 32 training patches and 32 test patches per texture type. It is clear that the SHSA representations give significantly better results. Not surprisingly, linear subspaces of images do not give satisfactory results as the pixel-wise difference between textures is not a meaningful distance for textures.

TABLE V

AVERAGE RECOGNITION RESULTS FOR THE TEXTURE DATASET

Methods	Average rate	Best rate	Worst rate
PCA	22.8%	24.2%	21.1%
ICA	23.2%	25.4%	21.6%
FDA	57.5%	60.8%	17.6%
RCA	17.8%	19.5%	16.5%
SHPCA	94.2%	95.3%	92.8%
SHICA	93.8%	94.9%	92.7%
SHFDA	97.9%	98.5%	97.1%
SHRCA	88.3%	89.8%	87.0%

V. DISCUSSION

One of the major obstacles of developing a generic vision system is the generalization of the adopted underlying

[6] http://www.uk.research.att.com/facedatabase.html

[7] Available at http://www-dbv.cs.uni-bonn.de/image/texture.tar.gz.

representation. By studying the intrinsic generalization of a representation, we can better understand and predict its performance under different conditions. To our knowledge, this is the first attempt to provide a quantitative generalization measure intrinsic to a representation; in contrast, generalization is commonly tied to recognition performance, which depends on the choice of the classifier and the choice of training and test data, as shown by the experiments here. Our study on the intrinsic generalization of linear subspace representations in the image space shows that they cannot generalize well as images from different models tend to be grouped into one equivalence class; we emphasize that this result holds for any low dimensional linear subspace in the image space. We have suggested a way to improve the intrinsic generalization by implementing linear subspaces in the spectral histogram space. We have demonstrated substantial improvement in recognition on large datasets.

However, our goal is not to show that SHSA representation is optimal in general. In fact, if classes consist of white noise like images, SHSA representations would be very ineffective. Rather our emphasis is on the importance of the underlying representation for object images. An ideal representation of image $\mathbf{I}$ is $S_I^0(\mathbf{I})$, which can be implemented only when the true underlying object models and the physical imaging process are available; this leads to the analysis-by-synthesis paradigm (Grenander, 1993). When $S_I^0(\mathbf{I})$ is not available explicitly, one needs to approximate it. There is a trivial solution for a good approximation by forcing $S_I(\mathbf{I}) = \{\mathbf{I}\}$. However, the generalization is very poor and it requires literally all possible images in the training set. A good representation should approximate $S_I^0(\mathbf{I})$ well and $|S_I(\mathbf{I})|$ should be as large as possible. These two constraints provide the axes of forming a continuous spectrum of different representations and allow us to study and compare them. For example, only marginal distributions are used in the spectral histogram representation; one can describe and synthesize image $\mathbf{I}$ better by incorporating joint statistics (Portilla and Simoncelli, 2000); however, this obviously decreases $|S_I(\mathbf{I})|$. Within linear subspace methods, one can also decrease $|S_I(\mathbf{I})|$ by imposing additional constraints on bases and coefficients, such as the non-negative constraints (Lee and Seung, 1999)[8]. Due to the complexity of $S_I^0(\mathbf{I})$, a very close approximation using some low dimensional representations may not be feasible. An alternative is to combine the analysis-by-synthesis paradigm (Grenander, 1993) and a low dimensional representation based approach. The hypothesis pruning by Srivastava et al. (2002) provides such an example, where a low dimensional representation selects plausible hypotheses for a analysis-by-synthesis model. In this framework, the difference among low dimensional representations is their effectiveness of selecting good hypotheses rather than providing a final answer.

Within the spectral histogram representation, in addition to linear subspace methods, the dimension can also be reduced by choosing a subset of filters. Filter selection for performance optimization has been studied by Liu and Wang (2003). Filters can also be learned by optimizing a heuristic-based measure such as statistical independence (Liu and Cheng, 2003) or by maximizing the recognition performance (Liu and Srivastava, 2003). We have also developed a two-parameter analytical form for the marginals of filter responses (Srivastava et al., 2002), which reduces the dimension of spectral histogram representation significantly.

An important question for recognition applications, which is not addressed here, is which representation is optimal given the choice of the representation space. The experimental results here suggest that the SHFDA in general gives the best performance; however, no optimality can be established theoretically as SHFDA's optimality requires that the underlying probability distributions are Gaussian and linear discrimination functions are used (Liu et al., 2003). To find optimal representations, computationally efficient algorithms have recently been developed (Liu et al., 2003; Liu and Srivastava, 2003) and their effectiveness has been demonstrated in the image space and in the spectral histogram space.

Acknowledgments This research was supported in part by grants NIMA NMA201-01-2010, NSF DMS-0101429, and ARO DAAD19-99-1-0267. D.L.W. was supported in part by an NSF grant (IIS-0081058) and an AFOSR grant (F49620-01-1-0027). The authors would like to thank the producers of the COIL, ORL, and texture datasets for making them publicly available.

References

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. N. Petrov and F. Csaki (eds.), *2nd International Symposium on Information Theory*, pp. 267–281

Baum, E. B. and Haussler, D. (1989). What size net gives valid generalization? Neural Computation, 1, 151–160.

Belhumeur, P. N., Hepanha, J. P., and Kriegman, D. J. (1997). Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 19, 711–720.

Bishop, C. M. Neural Networks for Pattern Recognition. Oxford, UK: Oxford University Press.

Chubb, C., Econopouly, J., and Landy, M. S. (1994). Histogram contrast analysis and the visual segregation of IID textures. Journal of the Optical Society of America A, 11, 2350–2374.

Comon, P. (1994). Independent component analysis, A new concept? Signal Processing, 36, 287–314.

Csiszar, I. and Korner, J. (1981). Information Theory: Coding Theorems for Discrete Memoryless Systems. New York: Academic Press.

Fisher, R. A. (1936). The use of multiple measures in taxonomic problems. Annals of Eugenics, 7, 179–188.

[8] Rigorously speaking, the bases with non-negative constraints do not form linear subspaces anymore.

Grenander, U. (1993). General Pattern Theory. Oxford, UK: Clarendon Press.

Grenander, U. and Srivastava, A. (2001). Probability models for clutter in natural images. IEEE Transactions on Pattern Analysis and Machine Intelligence, 23, 424–429.

Heeger, D. J. and Bergen, J. R. (1995). Pyramid-based texture analysis/synthesis. In *Proceedings of SIGGRAPHS*, pp. 229–238.

Hotelling, H. (1933). Analysis of a complex of statistical variables in principal components. Journal of Educational Psychology, 24, 417–441, 498–520.

Hyvärinen, A. (1999). Fast and robust fixed-point algorithm for independent component analysis. IEEE Transactions on Neural Networks, 10, 626–634.

Hyvärinen, A., Karhunen, J., and Oja, E. (2001). Independent Component Analysis. New York: Wiley-Interscience.

Karhunen, K. (1947). On linear methods in probability theory. Annales Academiae Scientiarum Fennicae, Ser. A1, 37, 3–79 (English translation, Doc. T-131, The RAND Corp., Santa Monica, CA 1960).

Kirby, M. and Sirovich, L. (1990). Application of the Karhunen-Loève procedure for the characterization of human faces. IEEE Transactions on Pattern Analysis and Machine Intelligence, 12, 103–108.

Loève, M. M. (1955). Probability Theory. Princeton, N.J.: Van Nostrand.

Lades, M., Vorbruggen, J. C., Buhmann, J., Lange, J., von der Malsburg, C., Wurtz, R. P., and Konen, W. (1993). Distortion invariant object recognition in the dynamic link architecture. IEEE Transactions on Computers, 42, 300–311.

Lee, D. D. and Seung, S. (1999). Learning the parts of objects by non-negative matrix factorization. Nature, 401, 788–791.

Liu, X. and Cheng, L. (2003). Independent spectral representations of images for recognition. Journal of the Optical Society of America, A, in press.

Liu, X. and Srivastava, A. (2003). Stochastic geometric search for optimal linear representations of images. In *Proceedings of the Fourth International Workshop on Energy Minimization Methods in Computer Vision and Pattern Recognition*, in press.

Liu, X., Srivastava, A., and Gallivan, K. (2003). Optimal linear representations of images for object recognition. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, in press.

Liu, X. and Wang, D. L. (2001). Appearance-based recognition using perceptual components. In *Proceedings of the International Joint Conference on Neural Networks*, pp. 1943–1948.

Liu, X. and Wang, D. L. (2002). A spectral histogram model for texton modeling and texture discrimination. Vision Research, 42, 2617–2634.

Liu, X. and Wang, D. L. (2003). Texture classification using spectral histograms. IEEE Transactions on Image Processing, in press.

Martinez, A. M. and Kak, A. C. (2001). PCA versus LDA. IEEE Transactions on Pattern Analysis and Machine Intelligence, 23, 228–233.

Murase, S. K. and Nayar, S. K. (1995). Visual learning and recognition of 3-d objects from appearance. International Journal of Computer Vision, 14, 5–24.

Pontil, M. and Verri, A. (1998). Support vector machines for 3D object recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 20, 637–646.

Rissanen, J. (1978). Modeling by shortest data description. Automatica, 14, 465–471.

Portilla, J. and Simoncelli, E. P. (2000). A parametric texture model based on joint statistics of complex wavelets. International Journal of Computer Vision, 40, 49–71.

Schiele, B. and Crowley, J. L. (2000). Recognition without correspondence using multidimensional receptive field histograms. International Journal of Computer Vision, 36, 31–50.

Simoncelli, E. P. and Olshausen, B. A. (2001). Natural image statistics and neural representation. Annual Review of Neuroscience, 24, 1193–1216.

Sirovich, L. and Kirby, M. (1987). Low-dimensional procedure for the characterization of human faces. Journal of the Optical Society of America A, 4, 519–524.

Srivastava, A., Liu, X., and Grenander, U. (2002). Universal analytical forms for modeling image probabilities. IEEE Transactions on Pattern Analysis and Machine Intelligence, 23, 1200–1214.

Swain, M. J. and Ballard, D. H. (1991). Color indexing. International Journal of Computer Vision, 7, 11–32.

Turk, M. and Pentland, A. (1991). Eigenfaces for recognition. Journal of Cognitive Neuroscience, 3, 71–86.

Vapnik, V. N. (2000). The Nature of Statistical Learning Theory, 2nd ed. New York: Springer-Verlag.

Yang, M. H., Roth, D., and Ahuja, N. (2000). Learning to recognize 3D objects with SNoW. In: *Proceedings of the Sixth European Conference on Computer Vision*, pp. 439–454.

Zhang, J., Yan, Y., and Lades, M. (1997). Face recognition: Eigenface, elastic matching, and neural nets. Proceedings of IEEE, 85, 1423–1435.

Zhu, S. C., Liu, X., and Wu, Y. (2000). Exploring texture ensembles by efficient Markov chain Monte Carlo. IEEE Transactions on Pattern Recognition and Machine Intelligence, 22, 554–569.

Zhu, S. C., Wu, Y. N., and Mumford, D. (1997). Minimax entropy principle and its application to texture modeling. Neural Computation, 9, 1627–1660.