SIRFACE vs. FISHERFACE: RECOGNITION USING CLASS SPECIFIC LINEAR PROJECTION

Yangrong Ling[*] *Xiangrong Yin*[+] *Suchendra M. Bhandarkar*[*]

[*]Department of Computer Science
[+]Department of Statistics
The University of Georgia
Athens, GA 30602, USA

ABSTRACT

Using a novel data dimension reduction method proposed in statistics, we develop an appearance-based face recognition algorithm which is insensitive to large variation in lighting direction and facial expression. Taking a pattern classification approach, we consider each pixel in an image as coordinate in a high-dimensional space. However, since faces are not truly Lambertian surfaces and indeed produce self-shadowing, images will deviate from this linear subspace. Rather than explicitly modeling this deviation, we linearly project the image into a subspace in a manner which discounts those regions of the face with large deviation using Sliced Inverse Regression (SIR) [9]. Our face recognition algorithm termed as *Sirface* produces well-separated classes in a low-dimensional subspace, even under severe variation in lighting and facial expression. *Sirface* can be shown to be equivalent to the well known *Fisherface* algorithm [1] in the subspace sense. However, *Sirface* is shown to produce the optimal reduced subspace (with the fewest dimensions) resulting in a lower error rate and reduced computational expense. Experimental results comparing *Sirface* to *Fisherface* on the Yale face database are presented.

1. INTRODUCTION

During the past few years, numerous algorithms, both appearance-based and feature-based, have been proposed for face recognition. While much progress has been made toward recognizing faces under small variations in lighting, facial expression and pose, reliable techniques for recognition under more extreme variations have proven elusive. In this paper, we outline a new approach to appearance-based face recognition; one that is insensitive to large variations in lighting and facial expressions. Note that lighting variability includes not only changes in light source intensity, but also changes in the direction(s) and number of light sources.

Our approach to face recognition exploits two observations:
(1) All of the images of a Lambertian surface, taken from a fixed viewpoint, but under varying illumination, lie in a 3D linear subspace of the high-dimensional image space [1].
(2) Due to shadowing, specularities, and variations in facial expression, the above observation does not always exactly hold. In practice, certain regions of the face may exhibit significant deviation from the linear subspace, and, consequently, are less reliable for the purpose of recognition [1].

We have used these observations to find a linear projection of the faces from the high-dimensional image space to a significantly lower dimensional feature space such that the projection is insensitive to variation in both, lighting direction and facial expression. Thus, the use of proper data dimension reduction techniques is critical. We have applied the concepts of data dimension reduction originating in statistics to the problem of appearance-based face recognition. Note that appearance-based approaches to face recognition preclude the use of an a priori model. In contrast to model-based approaches to face recognition where an explicit geometric and/or photogrammetric representation is needed, appearance-based approaches rely on the learning of an implicit model via selection of sample images of the face under varying conditions of illumination, pose, viewpoint and facial expression.

We have developed an appearance-based face recognition algorithm based on Sliced Inverse Regression (SIR) [9] which we have termed as *Sirface.* The reduced dimensional subspace arrived at by *Sirface* is equivalent to the one obtained from the *Fisherface* algorithm [1] in that it maximizes the ratio of the between-class scatter to the within-class scatter. But *Sirface* is shown to further reduce the dimensionality of the reduced subspace determined by *Fisherface* thus resulting in a reduced subspace of much smaller dimensionality. Thus, *Sirface* is more efficient than *Fisherface* in terms of both, error rate and computational expense.

While there are other methods for data dimensionality reduction such as *correlation*, *Eigenface* and *linear subspace projection*, a comparison of these methods to the *Fisherface* method can be found in the paper by Belhumeur et al. [1]. The *Fisherface* method was shown to be superior to *correlation*, *Eigenface* and *linear subspace projection* methods. Thus, the main purpose of this paper is to compare the *Sirface* algorithm to the Fisherface algorithm as described in Belhumeur et al. [1].

We should point out that Fisher's linear discriminant analysis (LDA), on which the *Fisherface* algorithm is based, is a classical statistical technique, especially in the areas of pattern classification and discriminant analysis. Sliced Inverse Regression (SIR) [9], on the other hand, is a relatively novel technique with origins in statistical regression. The connection between the areas of statistical regression and discriminant analysis was recently established by Kent [8] and by Cook and Yin [4]. A detailed description of the *Fisherface* and *Sirface* algorithms is provided in the following section.

2. METHODS

The appearance-based face recognition problem can be simply stated as follows: Given a set of face images labeled with the person's identity (*the learning set*) and an unlabeled set of face images from the same group of people (*the test set*), identify each person in the test images.

The typical recognition procedure in appearance-based computer vision is to use the learning set to establish some classification rules for the underlying patterns and then apply these rules to classify the patterns in the test set into the appropriate classes. Formally, let us consider a set of n sample images $\boldsymbol{X}_1,\ldots,\boldsymbol{X}_n$ taking values in a p-dimensional image space, and assume that each image belongs to one of C classes 1,..., c. Thus we need to establish the appropriate classification rules for the p-dimensional input image space and classify each of the $\boldsymbol{X}_i$'s to its right class. This is, in fact, a classical pattern classification problem. Since p is large ($p = l\times k$ for an image of size $l\times k$ pixels), we would like to reduce the p dimensional image space to the smallest q dimensional image subspace (more specifically, find a $p\times q$ matrix $\boldsymbol{B}$ where $\boldsymbol{B}^T\boldsymbol{X}$ is the smallest q dimensional image subspace) which retains all the necessary classification information. This means that we will classify each input pattern $\boldsymbol{X}_i$ to the same class regardless of whether we use the original p-dimensional image space or the reduced q-dimensional image subspace. More formally, the subspace spanned by the columns of $\boldsymbol{B}$ is a central discriminant subspace [4]. Using the reduced q-dimensional image space has several potential advantages. For example, if $q \leq 3$, one can more easily visualize the sample data. A reduced dimensional subspace serves to filter out the noisy or irrelevant portions of the input image space thus reducing both, the classification error rate and the computational expense. Although the input data is in its original $\boldsymbol{X}$-scale, we can always transform the input data to an equivalent $\boldsymbol{Z}$-scale where

$$Z = \sum\nolimits_X^{-\frac{1}{2}}(X - \mu_X)$$

and Σ_X and μ_X are, respectively, the covariance matrix and the mean vector of $\boldsymbol{X}$. Here we assume that Σ_X is nonsingular, else we can first reduce the dimensionality of the original $\boldsymbol{X}$ using principal components analysis (PCA) to the point where the resulting Σ_X becomes nonsingular. The use of the $\boldsymbol{Z}$-scale allows for easy comparison of the various dimensionality reduction techniques.

In the next section, we examine two pattern classification techniques for the appearance-based face recognition problem, namely, the *Fisherface* technique and the *Sirface* technique. We approach this problem within the pattern classification paradigm, considering each of the pixel values in the sample images as a coordinate in a high-dimensional input space (i.e., the *image space*).

2.1. The Fisherface Technique

Let μ_i and Σ_i be the mean vector and covariance matrix for class i. Define

$$S_B = \sum_{i=1}^{c} n_i(\mu_i - \mu)(\mu_i - \mu)^T = \sum_{i=1}^{c} n_i(\mu_i \mu_i^T)$$

and

$$S_w = \sum_{i=1}^{c} \sum_{Z_k \in i} (Z_k - \mu_i)(Z_k - \mu_i)^T$$

Note that in the $\boldsymbol{Z}$-scale, $\mu = 0$. We can always assume that $\boldsymbol{S}_W$ is nonsingular else use PCA to reduce its rank so that the resulting reduced dimensional subspace has nonsingular $\boldsymbol{S}_W$. Belhumeur et al. [1] developed an algorithm called *Fisherface* to determine a matrix $\Gamma = (\gamma_1, \ldots, \gamma_q)$ that maximizes the ratio of the between class scatter to the within class scatter given by:

$$\frac{\gamma^T S_B \gamma}{\gamma^T S_W \gamma}$$

under the constraint $\Gamma^T S_W \Gamma = I$. When $c = 2$, this technique is traditionally called Fisher's linear discriminant analysis (LDA). When $c>2$, it is traditionally called canonical covariate analysis [10].

The *Fisherface* method is a well-known technique in classification and discriminant analysis. The optimal situation, from a pattern classification standpoint, is encountered when the X_i's are normally distributed for each class i with each class having the same covariance matrix, i.e. $\Sigma_i = \Sigma$ for $i = 1,...,c$. When not all $\Sigma_i = \Sigma$, important classification information could be lost, hence the difference among the Σ_i's needs to be considered. While the *Fisherface* method does not need the assumption of normal class distributions, its absence could result in suboptimal classification.

2.2. The Sirface Technique

Sliced Inverse Regression (SIR) [9] was originally developed for data dimensionality reduction in statistical regression problems. Let (Y_i, X_i) $i = 1,...,n$ be an input sample, where Y is a response variable and X is a predictor vector. Li [9] considered the inverse mean of $\mathbf{E}(X|Y)$ in the Z-scale, by constructing the matrix $\text{Var}(\mathbf{E}(Z|Y))$. Singular value decomposition (SVD) is used to find the minimum dimension of this matrix. As in the case of the *Fisherface* algorithm, we assume that S_W is nonsingular, else we first reduce its rank using PCA to make it nonsingular as is done in [1]. Kent [8] mentioned that SIR is equivalent to LDA when Y is a categorical variable. Cook and Yin [4] further developed this connection. Yin and Cook [11] recently showed a direct link between SIR and LDA. For a slightly different matrix whose columns span the same subspace as SIR, Geisser [5] proved a similar result.

In fact, for a categorical Y, the SIR matrix is given by $M_{SIR} = \frac{1}{n} S_B$. Thus, we only need to apply SVD to $\frac{1}{n} S_B$, i.e., find its d non-zero eigenvalues and their corresponding eigenvectors. We refer to this method as *Sirface*.

2.3. Comparison with The Fisherface Technique

In *Sirface*, the subspace spanned by the d non-zero eigenvectors of M_{SIR} can be shown to be the same as the subspace spanned by the $c-1$ eigenvectors in *Fisherface*. Hence using these d vectors does not result in any loss of classification information. However, *Fisherface* uses the m eigenvectors corresponding to the m largest eigenvalues for a prespecified value of m. If $m<d$, then *Fisherface* may lose important classification information. If $m>d$, then *Fisherface* can be seen to use redundant predictor vectors. In the case of *Sirface,* there exists a formal method for determining the optimal dimensionality d of the reduced dimensional subspace based on input data. Thus *Sirface* is the optimal method in this sense.

Sirface can result in dimensionality reduction beyond that possible with *Fisherface*. In the absence of any further information, the *Fisherface* technique is constrained to choose $c-1$ (where c is the number of classes) as the dimensionality of the reduced subspace. Any further reduction in dimensionality is possible only via exhaustive search. *Sirface*, on the other hand, yields a reduced dimensionality of d which is often $< c-1$. While LDA (and hence *Fisherface*) is, in general, not robust to non-normal data [7], *Sirface* can be shown to find outliers in the input data and hence is much more robust [11].

2.4. Test for determining *d*

Let

$$M_{SIR} = \Gamma^T \begin{pmatrix} D & 0 \\ 0 & 0 \end{pmatrix} \Gamma$$

where Γ is a $p\times p$ orthogonal matrix. Furthermore, $\Gamma^T=(\Gamma_1, \Gamma_0)$, where Γ_0 is $p\times(p\text{-}d)$ matrix. D is a $d\times d$ diagonal matrix whose elements $\lambda_1 \geq \cdots \geq \lambda_d$ are the eigenvalues of M_{SIR}. If d is known, we can compute the estimated eigenvalues $\hat{\lambda}_1 \geq \cdots \geq \hat{\lambda}_d$ of the sample matrix $\hat{M}_{SIR}$, and use the corresponding eigenvectors $\hat{l}_1 \geq \cdots \geq \hat{l}_d$. If d is unknown, an inference procedure about d is required. The relevant test statistic [9] is given by

$$\hat{\Lambda}_d = n \sum_{j=d+1}^{p} \hat{\lambda}_j$$

where $\hat{\lambda}_j$'s are the eigenvalues of the sample matrix $\hat{M}_{SIR}$.

The general asymptotic distribution for the SIR test statistic is a linear combination of independent chi-square random variables each with one degree of freedom [2]. For some special cases, such as normal predictors the test statistic has a central chi-square distribution [3, 9].

3. EXPERIMENTAL RESULTS

Initial results on the Yale Face Database B [6] tabulated in Table 1 have shown that, whereas in the case of *Fisherface*, a reduced dimensionality of $c-1$ (where c is the number of classes), is adequate, *Sirface* is capable of determining a reduced dimensionality much lower than $c-1$ without the need for exhaustive search and without compromising the classification accuracy. In the case of both, *Sirface* and *Fisherface*, the training images and the test images were projected onto the computed subspace. A simple nearest-neighbor classifier was used to classify the test images. *Sirface* was seen to achieve a classification accuracy of over 90% with as low as 5

Table 1. Comparison of the Sirface and Fisherface methods on the Yale Face Database B [6]

Dataset	N_{class}	$N_{image/class}$	N_{train}	N_{test}	d_{Fisher}	d_{Sir}	ε_{Fisher}	ε_{Sir}
1	20	5	100	140	19	9	5.0%	5.0%
2	20	5	100	120	19	9	7.5%	7.5%
3	20	3	60	140	19	6	20.0%	21.4%
4	10	5	50	70	9	5	4.3%	4.3%
5	10	7	70	70	9	7	1.5%	1.5%

N_{class}: number of classes, $N_{image/class}$: number of training images in each class, N_{train}: number of training images, N_{test}: number of test images, d_{Fisher}: reduced dimensionality resulting from the Fisherface method, d_{Sir}: reduced dimensionality resulting from the Sirface method, ε_{Fisher}: classification error of the Fisherface method, ε_{Sir}: classification error of the Sirface method

Figure 1. Sample images of a single person under different illumination conditions and with differing poses (from the Yale Face Database B [6])

training images per class (Table 1, Datasets 1, 2, 4, and 5). The classification accuracy of *Sirface* drops to 78.6% only when the number of training images per class is reduced to 3 (Table 1, Dataset 3).

4. CONCLUSIONS AND FUTURE WORK

In this paper we proposed a novel technique for data dimensionality reduction in the context of appearance-based face recognition. This technique is based on Sliced Inverse Regression (SIR) and is termed as *Sirface*. Initial experiments on the Yale Face Database B show that *Sirface* can yield classification accuracy comparable to the well-known *Fisherface* technique while resulting in dimensionality reduction beyond that possible with *Fisherface*. Whereas in the case of *Fisherface*, the optimum reduced dimensionality can be determined only via exhaustive reach, *Sirface* has an associated formal hypothesis testing procedure for determining the optimum reduced dimensionality. Further testing of *Sirface* on a wider set of human faces is currently in progress. Data dimensionality reduction techniques based on higher-order moments such as the Sliced Average Variance Estimation (SAVE) technique are also being investigated. Higher-order moment-based techniques such as SAVE are expected to be capable of distinguishing more subtle features, such as changes in facial expression, which *Sirface* and *Fisherface* are found to be incapable of.

Acknowledgment: This research was supported in part by the University of Georgia President's Venture Fund through the generous gifts of the University of Georgia Partners.

REFERENCES

[1] P.N. Belhumeur, J.P. Hespanha and D.J. Kriegman, Eigenfaces vs. Fisherfaces: recognition using class-specific linear projection, *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 19, No. 7, pp. 711-720, 1997.

[2] E. Bura and R.D. Cook, Estimating the structural dimension of regressions via parametric inverse regression, *Journal of Royal Statistical Society, B*., 63, No. 2, pp. 393-410, 2001.

[3] R.D. Cook, *Regression Graphics: Ideas for Studying Regressions through Graphics,* New York: Wiley, 1998.

[4] R.D. Cook, and X. Yin, Dimension reduction and visualization in discriminant analysis, invited paper with discussion in the *New Australia and New Zealand Journal of Statistics*, Vol. 43, No. 2, pp. 147-199, 2001.

[5] S. Geisser, Discrimination, allocatory and separatory, linear aspects, In: J. Van Ryzin, Ed., *Classification and Clustering*, Academic Press, New York, pp. 301-330, 1977.

[6] A.S. Georghiades, P.N. Belhumeur, and D.J. Kriegman, From Few to Many: Illumination Cone Models for Face Recognition under Variable Lighting and Pose, *IEEE Trans. Pattern Anal. Mach. Intelligence*, Vol. 23, No. 6, pp. 643-660, 2001

[7] W. J. Krzanowski, The performance of Fisher's Linear Discriminant Function Under Non-Optimal Conditions, *Technometrics*, Vol. 19, No. 2, pp 191-200, 1997.

[8] J.T. Kent, Discussion of Li (1991), *Journal of the American Statistical Association*, Vol. 86, pp. 336-337, 1991.

[9] K.C. Li, Sliced inverse regression for dimension reduction (with discussion), *Journal of the American Statistical Association*, Vol. 86, pp. 316-342, 1991.

[10] G.J. McLachlan, *Discriminant Analysis and Statistical Pattern Recognition*, John Wiley, New York, 1992.

[11] X. Yin, and R.D. Cook, Dimension reduction for the conditional *k*-th moment in regression, *Journal of the Royal Statistical Society* Ser B, 64, pp. 159-175, 2002.