Using Regression Techniques for Coping with the One-Sample-Size Problem of Face Recognition

Vitomir Štruc, Rok Gajšek, France Mihelič, Nikola Pavešić

University of Ljubljana, Faculty of Electrical Engineering, Tržaška 25, 1001 Ljubljana, Slovenia
E-mail: vitomir.struc@fe.uni-lj.si, rok.gajsek@fe.uni-lj.si, france.mihelic@fe.uni-lj.si,
nikola.pavesic@fe.uni-lj.si

Abstract. There is a number of face recognition paradigms which ensure good recognition rates with frontal face images. However, the majority of them require an extensive training set and degrade in their performance when an insufficient number of training images is available. This is especially true for applications where only one image per subject is at hand for training. To cope with this *one-sample-size* (OSS) problem, we propose to employ subspace projection based regression techniques rather than modifications of the established face recognition paradigms, such as the principal component or linear discriminant analysis, as it was done in the past. Experiments performed on the XM2VTS and ORL databases show the effectiveness of the proposed approach. Also presented ia a comparative assessment of several regression techniques and some popular face recognition methods.

Key words: Face recognition, feature extraction, regression techniques, subspace projection, one sample size problem.

Uporaba regresijskih metod za samodejno rapoznavanje obrazov

Povzetek. V strokovni literaturi zasledimo kopico pristopov k samodejnemu razpoznavanju obrazov, s katerimi je mogoče zagotoviti razmeroma visoko uspešnost razpoznavanja. Večina teh pristopov pa je učinkovitih zgoj tedaj, ko je na voljo obsežna učna množica slik obrazov, pri čemer mora biti vsaka oseba v učni množici nemalokrat zastopana z vsaj dvema učnima slikama. e je za učenje na voljo le ena slika za vsako izmed oseb v učni množici, se uspešnost razpoznavnja številnih obstoječih pristopov občutno zmanjša. Kot rešitev predstavljenega problema v članku predlagamo uporabo regresijskih metod, ki za osnovo regresije uporabljajo predstavitev slik obrazov v (linearnih in nelinearnih) podprostorih. Učinkovitost regresijskih metod za razpoznavanje obrazov bomo predstavili v seriji identifikacijskih poskusov, izvedenih na dveh javno dostopnih podatkovnih zbirkah - XM2VTS in ORL. Uspešnost razpoznavanja z regresijskimi postopki bomo primerjali še z uspešnostjo razpoznavanja uveljavljenih postopkov samodejnega razpoznavanja obrazov, kot sta postopka *Lastnih* in *Fisherjevih obrazov.*

Ključne besede: razpoznavanje obrazov, izpeljava značilk, regresijske metode, projekcija v podprostor, podatkovni zbirki XM2VTS in ORL

1 Introduction

The existing face recognition techniques have demonstrated good recognition performance on frontal face images when a sufficient number of images is available for training. However, as stated in [1], real-life applications often offer only one training image per subject - a situation that drastically degrades the performance of most face recognition techniques or even worse, renders their employment impossible. We will refer to this situation as the *one-sample-size* (OSS) problem throughout the paper.

To overcome the OSS problem, researchers have presented a number of recognition techniques. In this paper, however, we will focus on the face recognition techniques that have been dominant for years, namely, on the subspace projection techniques. When dealing with subspace projection techniques, one has to distinguish between two kinds of methods: *(i)* unsupervised or expressive techniques, which are applicable regardless of the number of available training images per subject, and *(ii)* supervised or discriminative techniques, which suffer from the OSS problem and are in most cases not feasible when only

one image is at hand for training. Most of the research effort regarding the OSS problem is, therefore, directed at improving the recognition performance of the expressive subspace projection techniques (e.g., principal component analysis - PCA) and modifying the discriminative approaches (e.g., linear discriminant analysis - LDA) to be applicable to one training image per subject.

Wu and Zhou [2], for example, proposed a modification of the commonly employed PCA-based Eigenface technique called the $(PC)^2A$ method, where, prior to the subspace projection, the face images were combined with their first-order vertical and horizontal projection images with the goal of improving the final recognition performance. Chen et al. [3] presented an extension of the $(PC)^2A$ method called enhanced $(PC)^2A$. In this approach the first-order projection images were replaced with the second-order ones while the other steps of the $(PC)^2A$ method remained the same. Both the $(PC)^2A$ and the enhanced $(PC)^2A$ were reported to outperform the traditional Eiegenface approach for the OSS problem. Wang et al. [4] reported that good recognition rates for the OSS problem can be achieved when subspace projection techniques are trained with the help of a generic database. The authors performed experiments with several established methods within their framework and achieved satisfactory results. Chen et al. [5] described a modification of the commonly used LDA approach tailored towards the OSS problem. They proposed to partition each face image from the training set into multiple non-overlapping sub-images and then use these newly produced samples for training of LDA. With this approach the training set is artificially enlarged, hence, LDA was applicable. The authors reported that their approach outperformed the enhanced $(PC)^2A$ method in their experiments.

From the presented methods we can see that there are two dominant research trends in regard to the OSS problem. Researchers either try to apply a pre-processing technique to the training images to improve the recognition performance of the given face recognition approach or somehow increase the amount of available training data (e.g., with a generic database or sub-sampling of the training images). There is, however, another possibility of how to deal with the OSS problem. One can employ subspace projection-based regression techniques with properly designed response matrices. These techniques are regularly used for classification purposes in the field of chemometrics, but have been largely neglected as a possible solution for the problem of face recognition. As we will show in this paper, regression techniques such as principal component regression (PCR), partial-least-squares regres-

sion (PLSR), kernel principal component regression (KPCR) and kernel partial-least-squares regression (KPLSR) can effectively cope with the OSS problem, while they achieve similar recognition rates as the established expressive and discriminative methods (i.e., the Eigenface technique, the Fisherface approach and generalized discriminant analysis) when more than one image per subject is available for training.

The rest of the paper is organized as follows: in Section 2 the tested regression techniques are briefly reviewed. Section 3 presents the classification rule used, while the experimental setup and the experiments are presented in Sections 4 and 5, respectively. The paper concludes with some final remarks in Section 6.

2   Regression Techniques

In this section we will briefly describe the basic concepts of four regression techniques, i.e., principal component regression (PCR), partial-least-squares regression (PLSR), kernel principal component regression (KPCR) and kernel partial-least-squares regression (KPLSR), and outline how they can be employed for classification, i.e., for face recognition.

Principal Component Regression. PCR is basically a two stage regression technique comprised of the projection of the training data into the principal component subspace followed by a multivariate regression step. Formally, it can be described as follows: let $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n]$ denote a matrix containing in its columns n centered d-dimensional training images from N classes. PCR uses their principal component subspace projections $\mathbf{Z}$, where $\mathbf{Z} = \mathbf{W}^T\mathbf{X}$ and the projection matrix $\mathbf{W}$ is constructed by means of the leading eigenvectors of the covariance matrix of the training images to define a linear regression model, i.e., $\mathbf{Y} = \mathbf{ZB}$ with $\mathbf{Y}$ and $\mathbf{B}$ being the response and regression coefficient matrices, respectively. Here, the matrix $\mathbf{B}$ is computed as $\mathbf{B} = (\mathbf{Z}^T\mathbf{Z})^{-1}\mathbf{ZY}$.

Partial-least-squares Regression. Similar to PCR, PLSR computes a lower dimensional representation of the the training images in form of latent vectors (components, factors) which account for as much as possible of the covariance between the training images in $\mathbf{X}$ and the responses in $\mathbf{Y}$. Thus, it computes latent vectors from $\mathbf{X}$ which are also relevant for $\mathbf{Y}$. Once computed, the latent components are used in the regression step to predict $\mathbf{Y}$. PLSR is commonly performed with the nonlinear iterative partial-least-squares algorithm.

Kernel Principal Component Regression. Consider a nonlinear mapping Φ of the d-dimensional input variable $\mathbf{x}$ from the original input space $\mathbb{R}^d$ to a

high-dimensional feature space $\mathcal{F}$, i.e., $\Phi : \mathbf{x} \in \mathbb{R}^d \to \Phi(\mathbf{x}) \in \mathcal{F}$. The goal of KPCR is to construct a standard regression model (similar to the one presented in the paragraph on PCR) in the high-dimensional feature space $\mathcal{F}$ rather than in the original input space, thus achieving nonlinear regression. KPCR avoids direct computation of the nonlinear mapping Φ, but rather uses the kernel-trick and performs regression based on the kernel matrix of the training data.

Kernel Partial-least-squares Regression. KPLSR is a non-linear variant of the PCR technique. Like the KPCR method it uses kernel matrices for construction of the regression model in the feature space and consequently achieves nonlinear regression. A detailed description of all the presented regression techniques can be found in [6].

2.1 Using Regression Techniques for Classification

When regression techniques are used for classification, the response matrix used for the construction of the (linear or nonlinear) regression model has to encode the class-membership of the training data. Commonly, the following response matrix is used for training:

$$\mathbf{Y} = \begin{bmatrix} \mathbf{1}_{m_1} & \mathbf{0}_{m_1} & \cdots & \mathbf{0}_{m_1} \\ \mathbf{0}_{m_2} & \mathbf{1}_{m_2} & \cdots & \mathbf{0}_{m_2} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}_{m_N} & \mathbf{0}_{m_N} & \cdots & \mathbf{1}_{m_N} \end{bmatrix}, \quad (1)$$

where N represents the number of classes in the set of n d-dimensional inputs (matrix $\mathbf{X}$), m_i represents the number of inputs in class C_i, $\mathbf{1}_{m_i}$ $(i = 1, 2, ..., M)$ denotes a $m_i \times 1$ vector of all ones and $\mathbf{0}_{m_i}$ $(i = 1, 2, ..., M)$ is a $m_i \times 1$ vector of all zeros. Each of the rows in the matrix $\mathbf{Y}$ represents the desired regression response for the corresponding input training image. The responses computed with the constructed regression model are used for building face templates, where the template for the identity C_i represents the mean vector of the responses corresponding to the training images of the i-th identity.

3 The Classification Rule

The effectiveness of regression techniques and their competitiveness with the established face recognition approaches was tested within a face recognition system operating in the identification mode.

In the identification mode, a feature vector extracted from a given face image is compared to the templates of all subjects enrolled in the system and consequently stored in the systems database. The

identity corresponding to the template which best matches the given feature vector is ultimately assigned to the face image (i.e., to the subject the face image belongs to). A number of classifiers are suitable for this task, for example, the support vector machine (SVM) classifier, the Gaussion mixture model (GMM) classifier or the nearest neighbor (1-NN) classifier[*]. As a compromise between the computational burden required for training the classifier and the recognition performance, the 1-NN classifier is considered in this paper. The 1-NN classifier assigns the identity C_i (for $i \in 1, 2, ..., N$) to the given feature vector $\mathbf{y}$ if the dissimilarity δ between $\mathbf{y}$ and the i-th template $\bar{\mathbf{y}}_i$ is the smallest among all computed dissimilarity scores [11], i.e.,

$$\delta(\mathbf{y}, \bar{\mathbf{y}}_i) = \min_j \delta(\mathbf{y}, \bar{\mathbf{y}}_j) \to \mathbf{y} \in C_i, \quad (2)$$

where $j = 1, 2, ..., N$ and δ denotes the whitened cosine dissimilarity measure which is defined as follows:

$$\delta(\mathbf{y}, \bar{\mathbf{y}}_i) = \frac{(\mathbf{P}^T \mathbf{y})^T (\mathbf{P}^T \bar{\mathbf{y}}_i)}{\|\mathbf{P}^T \mathbf{y}\| \|\mathbf{P}^T \bar{\mathbf{y}}_i\|}. \quad (3)$$

Here, $\mathbf{P}$ stands for the whitening transformation matrix that can be specified by means of the covariance matrix of the templates stored in the systems database and T and $\|\cdot\|$ denote the transpose and the norm operator, respectively. A detailed description of the employed dissimilarity measure can be found in [14].

4 The Databases and Experimental Setup

Two publicly available databases commonly employed for assessing the performance of face recognition algorithms were used in the experiments presented in the remainder of this paper, namely, the XM2VTS and the ORL databases.

The first database, i.e., the XM2VTS database, contains 2360 (color) facial images that correspond to 295 subjects. Two images of each subject were captured during four recording sessions. Hence, a total of eight facial images per subject is available for training and performance assessment of ones face recognition algorithms. Furthermore, as the sessions were distributed over a period of five months, different images of the same subject exhibit variations in terms of hairstyle, pose, facial expression, etc. The images are stored in the portable pixel map format at a resolution of $720 \times 576 \times 3$ pixels [7]. The second

---

[*]Of course, there are several other classifiers; however, the listed ones are among the most commonly used in the field of face recognition.

XM2VTS							
No. of training samples	Grey-scale images				Gabor feature vectors			
	EF	FF	PCR	PLSR	EF	FF	PCR	PLSR
1	48.9	N/A	60.4	70.2	54.8	N/A	68.4	92.2
2	62.0	67.9	85.4	82.9	77.5	84.4	87.7	98.4
3	65.6	81.4	87.5	87.9	80.3	98.6	96.1	99.1
4	71.6	91.5	95.5	88.7	86.5	99.0	97.9	99.6

ORL							
No. of training samples	Grey-scale images				Gabor feature vectors			
	EF	FF	PCR	PLSR	EF	FF	PCR	PLSR
1	50.7	N/A	65.8	71.7	58.1	N/A	73.9	75.8
2	66.2	69.8	86.9	88.1	73.2	77.4	89.7	93.4
3	72.0	90.2	93.2	93.9	82.4	95.1	96.4	96.8
4	76.1	93.4	92.9	95.0	86.6	98.0	99.1	99.2
5	78.8	94.7	96.5	97.5	91.2	98.9	99.5	99.6

Table 1. Rank-one and average rank-one recognition rates in % for the identification experiments performed on the XM2VTS and ORL databases
Tabela 1. Uspešnost in povprečna uspešnost identifikacije (v %) pri rangu ena določeni v poizkusih na podatkovnih zbirkah XM2VTS in ORL

database, i.e., the ORL database, used in our experiments was acquired at the Olliveti Research Laboratory in Cambridge, U.K. [8]. It contains 400 images of 40 distinct subjects, i.e., 10 facial images per subject, which are stored at a resolution of 112×92 pixels and 256 grey levels in the portable grey map format. The images display diversity across illumination, pose and facial expression.

Prior to the experiments, images from both databases were subjected to a pre-processing procedure which comprised: *(i)* a conversion of the original color images to grey-scale intensity images (only for the XM2VTS database), *(ii)* a geometric normalization procedure that (based on manually determined eye coordinates) rotated and scaled the images in such a way that the eye-centers were located at pre-defined positions and finally cropped the face region to a standard size of 128×128 pixels for the XM2VTS and 64×64 pixels for the ORL database and *(iii)* a photometric normalization procedure which featured a conversion of the pixel intensity distribution of the images to $\mathcal{N}(0,1)$.

A similar experimental setup was chosen for both databases. In the first step, images from both databases were partitioned into two groups: *(i)* the group of training images and *(ii)* the group of test images. The former was employed for training the regression as well as all other techniques implemented in the experiments, while the latter served solely for the final performance assessment.

For the XM2VTS database four sets of identification experiments were performed. In the first set, one training image per subject was used for training, while the remaining images were left for the performance assessment. In the second set of experiments, the number of training images was increased to two, in the third set to three and in the last set, four images were employed for training. In all four sets of experiments the training images were selected randomly amongst the eight images of each subject.

For the ORL database five sets of face recognition experiments were performed. Again, the number of (randomly chosen) training images was increased from one to five, while the left over images were employed for testing. However, as the database contains only images of 40 subjects, the experiments were repeated five times. Hence, the results for the ORL database are given in terms of the qtextitaverage rank-one recognition rate, as opposed to the XM2VTS database, where the results are presented in terms of the rank-one recognition rate.

The presented experimental setup was chosen for the following two reasons: *(i)* it allows us to assess the performance of the regression techniques with respect to the OSS problem and *(ii)* it enables a comparative assessment of the recognition performance of the regression techniques and other established face recognition methods when a different amount of training

XM2VTS							
No. of training samples	Grey-scale images				Gabor feature vectors			
	KPCA	GDA	KPCR	KPLSR	KPCA	GDA	KPCR	KPLSR
1	53.6	N/A	67.4	69.0	85.3	N/A	86.9	90.8
2	65.1	78.0	70.1	72.4	92.5	97.1	94.2	97.0
3	71.9	94.0	78.4	81.5	97.8	99.3	98.6	99.1
4	79.5	95.8	84.7	87.5	99.1	99.7	98.9	99.8

ORL							
No. of training samples	Grey-scale images				Gabor feature vectors			
	KPCA	GDA	KPCR	KPSLR	KPCA	GDA	KPCR	KPSLR
1	51.9	N/A	64.1	65.2	68.4	N/A	69.1	75.9
2	67.6	82.0	80.3	82.1	84.5	92.0	91.6	92.4
3	75.6	91.1	86.8	91.5	89.7	95.7	93.6	97.7
4	79.3	94.4	91.3	95.4	92.2	98.5	99.1	99.2
5	82.4	95.2	92.5	95.0	95.3	99.4	99.3	99.3

Table 2. Rank-one and average rank-one recognition rates in % for the identification experiments performed on the XM2VTS and ORL databases
Tabela 2. Uspešnost in povprečna uspešnost identifikacije (v %) pri rangu ena določeni v poizkusih na podatkovnih zbirkah XM2VTS in ORL

data is available.

5   The Experiments

The first series of our face recognition experiments aimed at assessing the performance of the linear regression techniques PCR and PSLR and compare it to that of two established linear face recognition techniques, namely, the Eigenface [10] and the Fisherface [9] approaches - denoted as EF and FF in Table 1. The experiments were performed with optimized parameters, i.e., for each face recognition technique the number of features was chosen in such a way that the technique resulted in the best recognition performance, using the classification rule and similarity measure presented in Section 3. All techniques were applied to the preprocessed grey-scale images of both databases and to the augmented Gabor feature vectors which were computed following the work presented in [11]. It has to be noted that a detailed description of the Gabor wavelet-based methods is beyond the scope of this paper. The Gabor representation of face images is used in our experiments only to show the recognition performance achievable with regression techniques when only one image per subject is available for training. The reader is referred to [11] for details on the Gabor wavelet-based methods.

The results of the experiments for the XM2VTS and ORL databases are presented in Table 1. Here,

the expression *N/A* denotes that the technique is not applicable considering the available number of training images.

From the results we can see that for the OSS problem the regression techniques performed best amongst all the tested methods with the PLSR method achieving higher recognition rates than the PCR technique. Furthermore, when more than one face image per subject was used in the training stage, the regression techniques resulted in similar and, in some cases, even better recognition rates than the Fisherface method which again performed better than the Eigenface approach. Generally, the regression techniques offer an appealing alternative to the commonly employed subspace projection techniques.

In our second series of face recognition experiments we assessed the performance of two kernel (nonlinear) regression techniques, i.e., KPCR and KPLSR, and the two kernel (nonlinear) counterparts of the Eigenface and Fisherface methods, i.e., the kernel principal component analysis (KPCA)[12] and the generalized discriminant analysis (GDA)[13]. As in the first series of experiments, the nearest neighbor classification rule in conjunction with the whitened cosine similarity measure was used for all the tested methods. Again all the methods were optimized to yield the best possible recognition rate. The results of the experiments in terms of the rank-one and average rank-one recognition rates are presented in Table

2.

Similar to the first series of experiments, the regression techniques again performed best among all the methods for the OSS problem. Considering the overall performance of the subspace projection techniques, i.e., the recognition rates obtained for different numbers of training images, we can see that the kernel methods outperformed their linear counterparts. The kernel regression techniques, on the other hand, exhibited only small recognition improvements or resulted in worse recognition rates as the linear ones, which is quite unexpected. The overall conclusion with respect to the suitability of regression techniques, be it either linear or non-linear (kernel), for face recognition still holds: they provide effective means to tackle the OSS problem and also achieve good recognition performance when more than one image per subject is available for training.

6 Conclusion

In this paper regression techniques were introduced for coping with the one-sample-size problem of face recognition. Four regression techniques, namely, principal component regression, partial-least-squares regression, kernel principal component regression and kernel partial-least-squares regression, were tested for their recognition performance in a scenario where only one face image per subject was at hand for training. The experimental results obtained on the XM2VTS and ORL databases suggest that regression techniques successfully handle the one-sample-size problem and ensure recognition rates comparable or even better than those of the established face recognition techniques, such as the Eigenface approach, the Fisherface approach, kernel principal component analysis and generalized discriminant analysis, when more than one training image is available.

7 Acknowledgements

8 References

[1] W. Deng, J. Hu, J. Gao, *Robust face recognition from one training sample per person.* In: L. Wang, K. Chen, Y.S. Ong (Eds.): LNCS 3610, pp. 915–924, Springer, (2005).

[2] J. Wu, Z.H. Zhou, Face recognition with one training image per person. *Pattern Recogntion Letters* 23(14), 1711–1719, (2002).

[3] S.C. Chen, D.Q. Zhang, Z.H. Zhou, Enhanced $(PCA)^2$A for face recogniton with one training image per person. *Pattern Recognition Letters* 25(10), 1173–1181, (2004).

[4] J. Wang, K.N. Plataniotis, J. Lu, A.N. Venetsanopoulos, On solving the FR problem with one training sample per subject. *Patteern Recognition* 39(9), 1746–1762, (2006).

[5] S. Chen, J. Liu, Z.H. Zhou, Making FLDA apllicable to face recognition with one sample per person. *Pattern recognition* 37(7), 1553–1555, (2004).

[6] R. Rosipal, L.J. Trejo, Kernel partial least squares regression in reproducing kernel Hilbert space. *Journal of Machine Learning Research* 2, 97–123, (2001).

[7] K. Messer, J. Matas, J. Kittler, J. Luettin, J., G. Maitre, *XM2VTSDB: the extended M2VTS database.* In: Proceedings of AVBPA'99, pp. 72–77, USA, (1999).

[8] F. Samaria, A. Harter, *Parameterisation of a stochastic model for human face identification.* In: Proceedings of ACV'94, pp. 138–142, USA, (1994).

[9] P.N. Belhumeur, J.P. Hespanha, D.J. Kriegman, *Eigenfaces vs. Fisherfaces.* In: Proceedings of the 4th ECCV, pp. 45–58, UK, (1996).

[10] M. Turk, A. Pentland: Eigenfaces for Recognition. *Journal of Cognitive Neurosicence* 3(1), 71–86, (1991).

[11] C. Liu, H. Wechsler, Gabor feature based classification using the enhanced Fisher linear discriminant model for face recognition. *IEEE TPAMI* 11(4), 467–476, (2002).

[12] B. Schölkopf, A.J. Smola, K.R. Müller, Nonlinear component analysis as a kernel eigenvalue problem. *Neural Computation* 10, 1299–1319, (1998).

[13] G. Baudat, F.E. Anouar, Generalized discriminant analysis using a kernel approach. *Neural Copmutation* 12(10), 2385–2404 (2000).

[14] C. Liu, The Bayes decision rule induced similarity measures. *IEEE TPAMI* 29(6), 1086–1090, (2007).

Vitomir Štruc is currently working as a researcher at the Laboratory of Artificial Perception, Systems and Cybernetics (LUKS) at the Faculty of Electrical Engineering in Ljubljana. His research interests include pattern recognition, machine learning and biometrics.

Rok Gajšek is currently working as a researcher at the Laboratory LUKS at the Faculty of Electrical Engineering of the University in Ljubljana. His research interests include speech recognition, signal processing and pattern recognition.

France Mihelič is a full professor at the Faculty of Electrical Engineering in Ljubljana. His research interests include pattern recognition, speech recognition and understanding, speech synthesis and signal processing.

Nikola Pavešić is currently head of the Laboratory LUKS at the Faculty of Electrical Engineering in Ljubljana. His research interests include pattern recognition, neural networks, image processing, speech processing, and information theory.