Smoothness Prior Information in Principal Component Analysis of Dynamic Image Data

Václav Šmídl[1], Miroslav Kárný[1], Martin Šámal[2], Werner Backfrieder[3], and Zsolt Szabo[4]

[1] Institute of Information Theory and Automation, Academy of Sciences of the Czech Republic, POB 18, CZ-182 08 Prague 8, Czech Republic
smidl@utia.cas.cz, school@utia.cas.cz
[2] Charles University Prague, Czech Republic
samal@cesnet.cz
[3] Institute of Biomedical Engineering and Physics, AKH Vienna, Austria
werner@bmtp.akh-wien.ac.at
[4] Johns Hopkins University, Baltimore, MD, USA
zszabo@welchlink.welch.jhu.edu

Abstract. Principal component analysis is a well developed and understood method of multivariate data processing. Its optimal performance requires knowledge of noise covariance that is not available in most applications. We suggest a method for estimation of noise covariance based on assumed smoothness of the estimated dynamics.

1 Introduction

In medical image processing, principal component analysis (PCA) is used for data compression, noise reduction, and feature extraction purposes. Its usefulness and many advantages are well known. Performance of PCA depends on the amount and characteristics of noise in observed data. In data with a low signal-to-noise ratio (SNR), inhomogeneous, or correlated noise, the performance of PCA can be poor.

The problem has been addressed theoretically in several papers [1,2,3,4,5] with respect to the properties of noise and an optimal scaling of data for PCA was defined. The authors concluded that the optimal metric can be derived directly from the known covariance matrix of the noise, and suggested particular solutions for specific data. With simulated data and known noise, we have found [6] that the methods proposed in [3,4,7] are efficient but their applicability restricted by requirements (e.g. knowledge of the distribution or the covariance matrix of the noise) that are not easily satisfied in practice. That was the motivation for searching for a more practical approach.

We suggest that the covariance matrix of the noise and thus the optimal metric for PCA can be estimated using a rather general prior information on the assumed smoothness of dynamic processes recorded in image sequences. This prior was originally developed for PCA of dynamic scintigraphic data where the assumption on smoothness of time-activity curves and of scintillation spectra is

M.F. Insana, R.M. Leahy (Eds.): IPMI 2001, LNCS 2082, pp. 225–231, 2001.

fully substantiated. However, the same prior information can be applied to a wider class of image sequences.

The prior information is embraced via the Bayesian paradigm [8] and an iterative search for maximum a posteriori probability (MAP) estimation of the parameters is proposed. The performance of the method is demonstrated in the context of simulated and clinical dynamic image data.

2 Problem Description and Solution

The aim is to improve the performance of PCA when the data SNR is low and/or the noise covariance is unknown. This requires a joint estimation of low-rank mean value of data and the covariance matrix of the noise.

Model of Observed Data. The observed image sequence consists of T images having N pixels each, stored column-wise. The images are assumed to be linear combinations of $r \ll \min(N,T)$ underlying images, P $(N \times r)$, weighted by coefficients, Q $(r \times T)$. The observed data O consist of this combination corrupted by an additive zero mean noise E

$$O = \mu + E = PQ + E.$$

The noise is assumed to contain no outlying realizations so that its distribution can be considered normal. Properties of the noise are thus fully characterized by the covariance $\mathcal{C} = \mathcal{E}(E_{it}E_{j\tau})$ where $\mathcal{E}$ denotes mathematical expectation. Its generic entries describe correlations and variances of the noise at pixels i and j $(i,j = 1 \dots N)$ at images t and τ $(t,\tau = 1 \dots T)$. Hence, the observed data O are normal with mean μ and covariance $\mathcal{C}$, symbolically $O \sim \mathcal{N}(\mu, \mathcal{C})$.

Models of Noise Covariance. The array $\mathcal{C}$ is huge with $0.5(NT+1)NT$ distinct elements. It is much larger than the number NT of data O and thus a restricted covariance structure has to be considered. Usually, independence of noise entries in different pixels and images is assumed, all with the same variance $1/\omega > 0$. Then, the model of data O becomes

$$O \sim \mathcal{N}(\mu, I_N \otimes I_T \omega^{-1}) = \left(\frac{\omega}{2\pi}\right)^{NT/2} \exp\left\{-0.5\omega \mathrm{tr}\left[(O-\mu)(O-\mu)'\right]\right\}, \quad (1)$$

where $'$ denotes transposition and tr trace. The covariance is $\mathcal{C} = (I_N \otimes I_T)\omega^{-1}$ where I_N is the identity matrix and $\otimes$ is Kronecker product [9]. The use of the precision ω instead of the variance simplifies formal manipulations.

The maximum likelihood estimate of μ of rank r minimizes the quadratic form in the exponent of (1) and thus coincides with the PCA estimate [10]. The results are poor when the covariance $\mathcal{C}$ does not have the assumed structure and/or the noise level $1/\omega$ is too high compared to the signal values μ. A solution to this problem depends on a more realistic modelling of the noise. Here,

the direct extension $\mathcal{C} = I_N \otimes \Omega^{-1}$ of the classical assumption is considered. The precision matrix Ω models changing covariances of the noise between the individual images. Formally, it is possible to consider $\mathcal{C} = \tilde{\Omega}^{-1} \otimes \Omega^{-1}$ with arbitrary positive definite $N \times N$ matrix $\tilde{\Omega}$. Computational demands are then much higher because the number of pixels N is much larger than the number of images T.

Prior Information. We search for a joint estimator of μ, Ω. It is a non-trivial task as it can be shown that the joint maximum likelihood estimate of μ and Ω does not exist. Thus, it is impossible to separate signal and noise spaces without additional information.

In nuclear medicine, image sequences reflect the changes of pixel values with time or energy. In the former case, the weights Q of images P can be interpreted as time-activity curves, in the latter case as scintillation spectra. In the following text we will use the time interpretation. The weights Q of images P are usually similar so that the observed adjacent images are similar, too. The adjacent observed images are usually similar so that we expect the weights Q of underlying images to be similar. This qualitative information is quantified as follows. The values $Q_{k(t)}$ of k-th curve $k = 1, \ldots, r$ at time $t = 2, \ldots, T$ are related to the preceding values through the simple time-dependent auto-regression

$$Q_{k(t)} \sim \mathcal{N}(a_{t-1} Q_{k(t-1)}, \beta^{-1}), \tag{2}$$

where the precision β and the coefficients $a = [a_1 \ldots a_{T-1}]$, approximating the curve evolution, are assumed to be common to all curves. The arbitrariness of the initial values $Q_{k(1)}$ is modelled by the flat normal probability density function (p.d.f.) $Q_{k(1)} \sim \mathcal{N}(0, 1/\varepsilon)$ with a small precision ε.

These assumptions, applied to $\mu = PQ$ with orthonormal images P, translate into the prior p.d.f. for μ. Its support has to be restricted to μ of the assumed rank $r \ll \min(N, T)$, i.e. to the space of lower dimension. This restriction of the parameter space to a lower dimension modifies the normalization factor [9].

$$\mu \sim K \varepsilon^{0.5r} \beta^{0.5Tr} \exp\left\{-0.5\beta \mathrm{tr}\left(\mu \Delta \Delta' \mu'\right)\right\}, \tag{3}$$

where K is a normalizing constant independent of estimated parameters, Δ is the $(T \times T)$ matrix with the non-zero entries $\Delta_{1,1} = \varepsilon^{0.5}$, $\Delta_{t,t} = 1$, $\Delta_{t-1,t} = -a_{t-1}, t = 2, \ldots, T$ and zero entries otherwise.

The specification of the prior p.d.f. is completed by assuming mutually independent $a_t \sim \mathcal{N}(1, 1/\alpha)$, and $\Omega \sim W(\gamma N, \gamma w I_T)$ where W is the Wishart distribution with parameters γ and ω [9]. These priors assign the highest belief to slowly changing dynamics and diagonal covariance but both are very flat.

Estimation Algorithm. The observation and noise models, together with the chosen prior distribution on unknown parameters $\Theta = (\mu, \Omega, a_1, \ldots, a_{T-1}, \beta, \varepsilon) = (\mu, \theta)$ determine the posterior p.d.f. of parameters given by the observations O. Its Θ-dependent part reads

$$\begin{aligned}\mathcal{L}(\Theta) = |\Omega|^{0.5N(1+\gamma)} \beta^{0.5Tr} \varepsilon^{0.5r} \times \exp\left\{-0.5\mathrm{tr}\left[(O-\mu)\,\Omega\,(O-\mu)'\right]\right\} \times \\ \times \exp\left\{-0.5\beta \mathrm{tr}\left(\mu \Delta \Delta' \mu'\right) + \gamma w\, \mathrm{tr}(\Omega) + \alpha(a+1)(a+1)'\right\}.\end{aligned} \tag{4}$$

The MAP estimate of Θ maximizes the function (4). Maximization complexity stems mainly from the restricted rank of the mean value μ. This makes an iterative search inevitable. Splitting of the estimated parameter $\Theta = [\mu, \theta]$ simplifies the description of the proposed algorithm.

Algorithm SPCA: Smoothed PCA

1. Choose small values of tuning knobs α, γ, w, select the upper bound $\bar{n} > 0$ on the number of iterations and set the iteration counter $n = 0$.
2. Choose initial estimates θ_n of θ as follows $\Omega_n = I_T,\ \beta_n = \varepsilon_n = a_{nt} = 0,\ t = 1, \ldots, T-1$.
3. Do while μ_n, θ_n are changing and $n < \bar{n}$
   (a) Complete the squares in exponent (4) with respect to μ so that you get
   $$\mathrm{tr}\left[(O\mathcal{A}_n - \mu\mathcal{B}_n)(O\mathcal{A}_n - \mu\mathcal{B}_n)'\right] + \Lambda_n,$$
   where $\mathcal{A}_n = \Omega_n \left(H_n^{-1}\right)'$, $\mathcal{B}_n = H_n^{-1}$ are regular matrices determined by the latest estimates θ_n through the identity $H_n H_n' = \Omega_n + \beta_n \Delta_n' \Delta_n$. The unique matrix remainder Λ_n collects the terms independent of μ.
   (b) Find the estimate $(\mu\mathcal{B}_n)_n$ of $(\mu\mathcal{B}_n)$ by applying standard PCA to the scaled data $(O\mathcal{A}_n\ )$ and compute the estimate $\mu_n = (\mu\mathcal{B}_n)_n {\mathcal{B}_n}^{-1}$ of μ.
   (c) Substitute μ_n into (4), find θ_{n+1} as the maximizer of the obtained expression (it can be mostly done analytically) and increase the iteration counter n.

3 Experiments

SPCA was implemented in Matlab [11] and its performance evaluated in experiments with simulated and clinical data of dynamic scintigraphy. Two illustrative examples are presented: a simple mathematical phantom and a dynamic PET study of the brain with ^{11}C labelled radioligand to serotonin transporters [12].

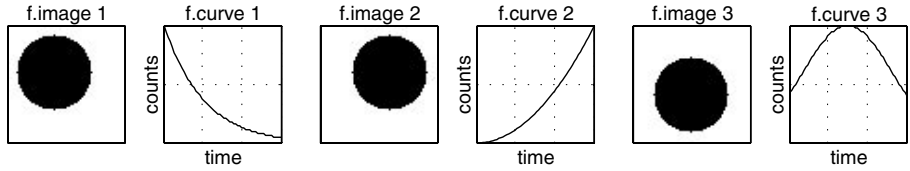

Fig. 1. Factor images and curves used for simulation of dynamic scintigraphic data.

The mathematical phantom consisted of 60 images of size 64×64. Each image was a linear combination of three factor images with circular structures. They are shown in Figure 1 which includes also the curves simulating intensity changes with time. A flat background and uncorrelated Gaussian noise (1) with a high variance was added to the simulated images. Figure 2 demonstrates six

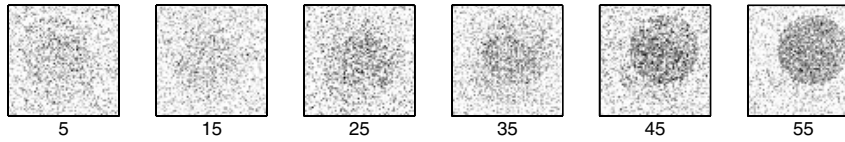

Fig. 2. Six samples from the analyzed series of 60 noisy images.

of 60 images (no. 5, 15, 25, 35, 45, and 55) in the resulting image series. PCA of the simulated data should recognize three underlying dynamic components. The first three most significant principal components (PCs) produced by PCA are demonstrated in Figure 3, those produced by SPCA in Figure 4.

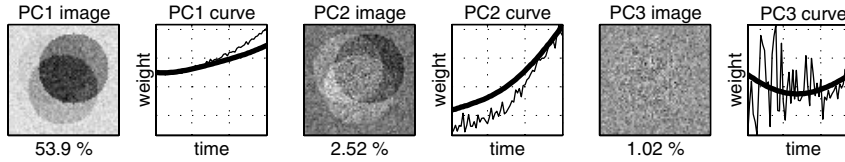

Fig. 3. The first three most significant PCs produced by PCA of simulated data. Numbers in % are relative contributions of PCs to original data. In noiseless data, true contributions of the first three PCs are 95.0, 4.5, and 0.5 %. The curves show the weights of respective PCs in original images. Thick lines show true weights of PCs extracted from noiseless data.

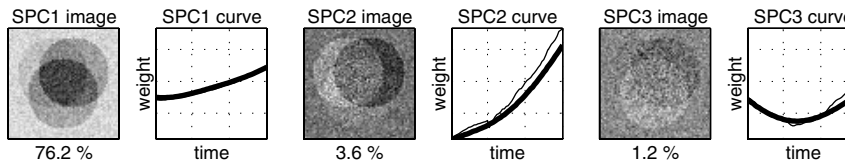

Fig. 4. The first three most significant PCs produced by SPCA of simulated data. The third PC is well defined and the curve reflects well the corresponding dynamics (polarity of PCs is arbitrary). Unlike PCs in Figure 3, PCs in Figure 4 can be successfuly rotated in order to recover the images and curves of underlying dynamic structures shown in Figure 1.

A dynamic PET study of the brain with ^{11}C labelled radioligand to serotonin transporter sites consisted of 18 images recorded in progressively extended time intervals in order to compensate for a very fast decay of ^{11}C and to obtain an acceptable contrast between the specific and non-specific binding of the radioligand that increases with time. Figure 5 demonstrates six of 18 images in the recorded image series. PCA was expected to recognize two underlying dynamic components (the signal of specific and non-specific binding). The first two most significant PCs produced by PCA are demonstrated in Figure 6, those produced by SPCA in Figure 7.

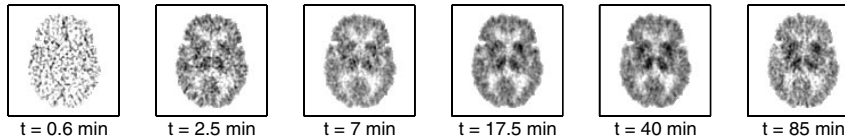

Fig. 5. Six samples from the analyzed series of 18 dynamic PET images.

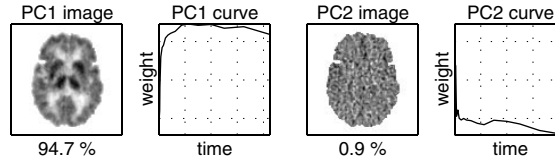

Fig. 6. The first two most significant PCs produced by PCA of dynamic PET brain study. Only the first PC shows the brain structure, the second PC reflects mostly noise.

4 Discussion and Conclusions

Preliminary experiments with simulated and clinical data have shown that in comparison with PCA, the SPCA is able to improve the separation of the signal from noise, and to enhance contrast in the images of principal components. We believe that the method proposed in this paper may improve the results of PCA applied to dynamic scintigraphic data recorded with varying acquisition intervals, in several energy windows, and studies with short-lived radionuclides. All those data are occasionally corrupted by potentially strong, correlated, and variable noise that may result in suboptimal performance of PCA. The prior information used in the proposed method is rather general and not necessarily restricted to scintigraphic data. In addition, alternative prior information - better suited to a specific problem - can be chosen and the methodology proposed in this paper still used with benefit. The method can be further developed to support the estimation of the number of significant factors and to benefit from similar prior information applied also to the images of principal components. Formally, these extensions are relatively straightforward. However, the increase in complexity of calculations is significant and approximations have to be found in order to make the solution feasible.

Acknowledgements

The work has been partially supported by the following grants: Austro-Czech project Kontakt II-16 (ME-228), GACR 102/99/1564, IGA MZCR NN5382-3/99, NIH no.AA11653, and NIH no.AG14400.

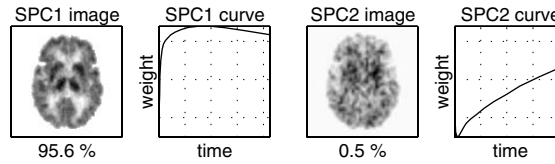

Fig. 7. The first two most significant PCs produced by SPCA of the dynamic PET brain study. The second PC is weak but well differentiated from noise. Unlike the PCs in Figure 6, the first two PCs in Figure 7 can be successfully rotated to the realistic images and curves of underlying specific and non-specific binding maps.

References

1. Anderson TW. Estimating linear statistical relationships. Ann Statist 1984; 12:1-45.
2. Fine J, Pousse A. Asymptotic study of the multivariate functional model. Application to the metric choice in principal component analysis. Statistics 1992; 23: 63-83.
3. Benali H, Buvat I, Frouin F, Bazin JP, DiPaola R. A statistical model for the determination of the optimal metric in factor analysis of medical image sequences. Phys Med Biol 1993; 38:1065-1080.
4. Pedersen F, Bergstroem M, Bengtsson E, Langstroem B. Principal component analysis of dynamic positron emission tomography studies. Eur J Nucl Med 1994; 21:1285-1292.
5. Hermansen F, Lammertsma AA. Linear dimension reduction of sequences of medical images: I. Optimal inner products. Phys Med Biol 1995; 40:1909-1920.
6. Šámal M, Kárný M, Benali H, Backfrieder W, Todd-Pokropek A, Bergmann H. Experimental comparison of data transformation procedures for analysis of principal components. Phys Med Biol 1999; 44:2821-2834.
7. Hermansen F, Ashburner J, Spinks TJ, Kooner JS, Camici PG, Lammertsma AA. Generation of myocardial factor images directly from the dynamic oxygen-15-water scan without use of an oxygen-15-carbon monoxide blood-pool scan. J Nucl Med 1998; 39:1696-1702.
8. Berger JO. Statistical Decision Theory and Bayesian Analysis. New York, Springer, 1985.
9. Rao CR. Linear Statistical Inference and its Application. New York, Wiley, 1973.
10. Golub GH, VanLoan CF. Matrix Computations. Baltimore, J Hopkins Univ Press, 1989.
11. Matlab v. 5.3.1 (R11.1), The MathWorks Inc., Natick, MA 01760-1500, USA, `http://www.mathworks.com`.
12. Parsey RV, Kegeles LS, Hwang D-R, Simpson N, Abi-Dargham A, Mawlawi O, Slifstein M, Van Heertum RL, Mann J, Laruelle M. In vivo quantification of brain serotonin transporters in humans using [^{11}C] McN 5652. J Nucl Med 2000; 41(9):1465-1477.