Proceedings of the
47th IEEE Conference on Decision and Control
Cancun, Mexico, Dec. 9-11, 2008

Some identification techniques in computer vision

Alessandro Chiuso and Giorgio Picci

Abstract— **In this paper we describe, without a pretense of completeness, some modeling and identification techniques which have been proposed recently for applications to computer vision. The emphasis is on methods which, although sometimes still in development, attempt to address *specific* issues of the particular application area.**

Keywords: Computer vision, subspace identification, reciprocal processes, dynamic factor analysis, dynamic textures.

I. Introduction

It is often claimed that one of the main aims of computer vision is to *understand images*. There are many peculiar characteristics of vision as a sensor as compared to other sensing "devices". The image formation process depends on a large number of factors which include geometry (shape of the scene, position of the camera, geometry of the imaging device), photometry (illumination, reflectance properties of the scene etc.) and dynamics (motion of the camera and/or of the objects which compose the scene).

It turns out that, inferring at the same time geometry, photometry and dynamics solely from sequences of images is certainly an ill-posed problem. In our everyday life, however, we manage to make decisions as to what is around us, where it is and how it is moving. These decisions are accurate enough so that we can move around, manipulate and recognized objects and so on. Most often this is accomplished by using prior information concerning the human imaging device (a stereo pair with a fixed baseline), the environment (e.g. we have plenty of prior information concerning the shape of surrounding objects, the material they are made of) as well as on the position and motion of the objects (we would not expect seeing an elephant moving as a bee nor a car being parked in the middle of a lake).

Still, in many cases we are fooled by images or movies; many well known visual illusions have been observed and studied.

For this reason we shall not even attempt to construct ("generative") models which describe geometry, photometry and motion but rather take a "black-box" approach typical of System Identification: we regard images as two dimensional signals having finite support and movies as time-indexed collection of images.

As suggested by Neyman [46], models should be as simple as possible while describing the phenomena we are interested in, at the level of accuracy which is needed to our purposes. There does not exists a "right" model, but rather a model is "good" if it describes well the data to the purpose at hand. Hence, preliminary to model building, one should always have in mind *to which purpose* the model is being built.

For our purpose, this means representing certain features of interest in the image by mathematical models which can be unambiguously interpreted and used, say, for map building or guidance and control of a mobile robot, for manipulating and grasping objects etc. At a first sight this task may look similar to the problem of (say) voice modeling and recognition in speech processing. However while when dealing with one dimensional signals there is a rather standard set of modeling approaches and identification techniques to choose from, one immediately recognizes that in the two-dimensional case there is both an incredible variety of modeling possibilities and at the same time a huge amount of data to be considered and eventually processed. The large variety of possible mathematical models and the need of nonstandard techniques for modeling and processing a large size of data, make the problem of image modeling and understanding very hard. One peculiar characteristic of the modeling problem in computer vision is that there are several (at least three) levels at which it can be approached. The first level is just understanding (modeling) static images. Distinctive features here are that spatial models are intrinsically not causal, contrary to the model classes used for describing phenomena evolving in time, since there is no natural notion of causality which enters in describing spatial correlation. Also real images can almost never be globally described by a unique model class. Different regions of the same image may have very different spatial structures and call for different model classes to realistically capture their structure. What is usually done is to partition the scene into regions where a specifically chosen model class is used to fit the data within a specified accuracy. Separating these regions is the *segmentation problem*, still a very active area of research in the vision community.

The second level concerns *temporal modeling* of a flow of images in time (a "movie"). A time-varying sequence of images carries information on the 3-D structure of the scene which cannot in general be gathered from a single static image. Moreover the scene itself may undergo its own dynamics which is often of interest to capture or estimate. For example the relative motion of the camera with respect to the scene and/or on the relative motion of various objects in the scene.

Dynamic Vision is the discipline which studies the inverse problem of recovering information on (i.e. estimating) the

Alessandro Chiuso is with the Department of Management and Engineering, University of Padova, Vicenza, Italy, chiuso@dei.unipd.it

Giorgio Picci is with the Department of Information Engineering, University of Padova, Italy; email: picci@dei.unipd.it

This work has been partially supported by the PRIN Project "New Methods and Algorithms for Identification and Adaptive Control of Technological Systems".

scene and the relative motion, from a *sequence* of images. Instead of inferring 3-D structure from a single image which is a necessarily incomplete 2-D representation of the scene, dynamic vision attempts to reconstruct the 3-D structure from a sequence of images by exploiting both the spatial structure and the temporal continuity of the scene. The general problem area may be seen as a chapter of nonlinear estimation and/or identification theory but this general classification is hardly of any help in practical solutions of the problem. It is only by exploiting the peculiar geometric and dynamic structure of the sequential inverse-projection problem of dynamic vision that useful and practically implementable solutions can be obtained.

So far temporal modeling and identification of image flows has been approached only for very simple spatial structures, e.g. *dynamic texture* modeling or temporal modeling of a finite set of point features such as for example the joints of moving kynematic chains. Examples of gait modeling will be described in Section VI-A. A definitely more complex task would be to model the spatio-temporal dynamics of an image flow; i.e. the simultaneous evolution of spatial and temporal coordinates (shape and time). This is however a largely unexplored area.

II. Static image modeling

Statistical modeling of images has been the subject of intense research in the past three decades and forms now a vast literature; see for example [25], [10], [2], [45], [65], [67]. Most studied models in the literature are related to the so-called *Gibbs-Markov* (G-M) random fields, borrowed (with some adaptations) from statistical mechanics. Unfortunately these models lead to extremely complicated estimation problems which have to be approached by Monte-Carlo type techniques, such as simulated annealing, MCMC, etc..

In this paper we shall discuss a simple class of stochastic models, known as *reciprocal processes*. These are actually a special class of G-M random fields which have been studied in depth in 1-D; see e.g. [34], [35], [39], [38]. It has been shown that stationary reciprocal processes admit linear *descriptor type* representations with constant parameters which can be seen as a natural non-causal extension of the linear state space models common in time series analysis. Stationary reciprocal processes may naturally live in a finite region of the "time" line (or of the plane) and being described by finitely parametrized linear models, lead (at least in in principle) to much easier identification problems, solvable, say, by "subspace" techniques, similar to those currently widely used in identification of multivariate time series.

One important class of models which we shall not discuss in this survey are *multi-resolution* models [4], [15], [5], [32], [8]. Multi-resolution models are based on the use of the wavelet transform. Taking advantage of the tree representation of the wavelet coefficients, a 2D signal is represented as a stochastic process on a tree. We are not aware of any serious attempt to perform identification of these models from data.

Other approaches include "deterministic" modeling (see e.g. [60]), which however will not be discussed in this paper.

III. Textures

It is natural to think of an image on a finite 2-D lattice of $N \times m$ *pixels* as a *random field*; i.e. a doubly indexed stochastic process $\mathbf{I} = \{\mathbf{I}(k,h)\,;\, k = 1,\dots,N,\ \ h = 1,\dots,m\}$ describing the intensity of the image at each pixel.

For us a **texture** will then be just a *spatially stationary* random field, see, e.g., figure 1.

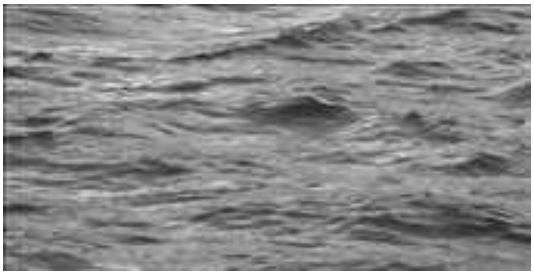

Fig. 1. A stationary image (texture).

Everything will be assumed to be zero mean (for this to hold, one actually may have to compute the average intensity $\bar{\mathbf{I}}$ and consider the logarithms $\log\{\mathbf{I}(k,h)/\bar{\mathbf{I}}\}$). Adopting the usual wide-sense (Gaussian) modeling paradigm of stationary stochastic processes, it appears natural to model textures by linear "2-D stochastic systems". By looking into the early literature in this area, one gets however the impression that the existing 2-D system theory has mostly been driven by an underlying desire of getting an orthodox formal generalization/extension of 1-D systems. This has (in our opinion) laid emphasis on superfluous issues, e.g. the search for extensions to 2-D of the notion of causality, a concept originated in 1-D temporal models which seems to be hardly useful in our context.

In the present setting it seems natural to base the idea of state of a 2-D stochastic model on that of a *Markov Random Field*.

A n-dimensional random field $\{\mathbf{x}(k,h)\}$ on a (finite or infinite) 2-D lattice is *Markovian* if for any closed bounded contour Γ, the random variables in the interior of Γ are conditionally independent of those in the exterior region, given the boundary values $\mathbf{x}_\Gamma := \{\mathbf{x}(k,h)\,;\,(k,h) \in \Gamma\}$. See e.g. [54] for an extensive discussion and precise definitions. The Markov property leads directly to a "local model" of the process. For the best linear estimate of $\mathbf{x}(k,h)$ given all other $\mathbf{x}(k',h')\,;\,(k',h') \neq (k,h)$ must depend only on the value of the process on the pixels immediately surrounding (k,h),

$$\begin{aligned}&\mathbb{E}\left\{\mathbf{x}(k,h) \mid \mathbf{x}(k',h')\,;\,(k',h') \neq (k,h)\right\} = \\ &F_{o+}\mathbf{x}(k+1,h) + F_{o-}\mathbf{x}(k-1,h) + \\ &F_{v+}\mathbf{x}(k,h+1) + F_{v-}\mathbf{x}(k,h-1)\end{aligned}$$

where the F's are $n \times n$ matrices (possibly dependent on (k,h)). Introducing the *conjugate process* ([42])

$$\mathbf{d}(k,h) := \mathbf{x}(k,h) - \mathbb{E}\left\{\mathbf{x}(k,h) \mid \mathbf{x}(k',h')\,;\,(k',h') \neq (k,h)\right\}$$

which by construction is uncorrelated with all random variables $\{\mathbf{x}(k',h')\,;\,(k',h') \neq (k,h)\}$, one readily arrives at the linear model

$$\mathbf{x}(k,h) = F_{o+}\mathbf{x}(k+1,h) + F_{o-}\mathbf{x}(k-1,h) + F_{v+}\mathbf{x}(k,h+1) + F_{v-}\mathbf{x}(k,h-1) + \mathbf{d}(k,h)$$

which should be coupled with suitable boundary conditions. The concept of Markov field reduces in 1-D to that of a *reciprocal process*, which is a generalization of 1-D Markov process [35].

Definition 1: A n-dimensional process $\mathbf{x} := \{\mathbf{x}(k),\, k \in \mathbb{Z}\}$ is *reciprocal* if for all $k \in (k_0,\, k_1)$ and h in the complementary interval $(k_0,\, k_1)^c$ $\mathbf{x}(k)$ and $\mathbf{x}(h)$ are conditionally independent given the boundary values $\mathbf{x}(k_0)$ and $\mathbf{x}(k_1)$.

From now on we shall discuss one-dimensional reciprocal models only. This, on one side, is done to reduce notational complexity and to afford a cleaner exposition. On the other hand, some extensions to 2-D of the procedures exposed below still need to be worked out in full detail.

We may introduce 1-D modeling of an image by simply considering the spatial evolution of the rows (or columns) of the image. By introducing $\mathbf{y}(k) \equiv \mathbf{I}(k,\cdot)$ one can describe $\mathbf{I}$ as an m-dimensional stochastic process $\{\mathbf{y}(k)(k),\, k = 1,2,\dots,N\}$ defined on the finite subinterval $[1,\, N]$ of the integer line. see Fig 2.

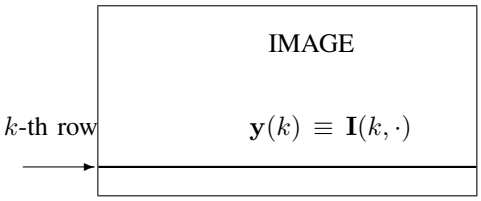

Fig. 2. 1-D modeling of an image

We shall consider processes which are stationary on a finite interval $[1,\, N]$. Write $\mathbf{y} := \{\mathbf{y}(k),\, k = 1,2,\dots,N\}$ as a column vector with N (m-dimensional) components and introduce the covariance matrix $\mathbf{R} := \mathbb{E}\,\mathbf{y}\mathbf{y}^\top$. We shall say that $\mathbf{y}$ is *stationary* if $\mathbf{R}$ has the symmetric block-Toeplitz structure,

$$\mathbf{R} := \mathbb{E}\,\mathbf{y}\mathbf{y}^\top = \begin{bmatrix} R(0) & R(1)^\top & \dots & R(N-1)^\top \\ R(1) & R(0) & R(1)^\top & \dots \\ \dots & & & \dots \\ R(N-1) & \dots & R(1) & R(0) \end{bmatrix}$$

and say that $\mathbf{y}$ is of *full rank* (or *minimal*) if its covariance matrix $\mathbf{R} := \mathbb{E}\,\mathbf{y}\mathbf{y}^\top$ is positive definite.

Assume now that $\mathbf{y}$ is a *periodic* stationary process of period T defined on the integer line: $\mathbf{y}(k+\nu T) := \mathbf{y}(k)$ for arbitrary $\nu \in \mathbb{Z}$. Such a process can equivalently be thought of as being defined on the *discrete group* $\mathbb{Z}_T := \{1,2,\dots,T\}$ with arithmetics mod T. Its covariance matrix, besides being block-Toeplitz, must obey the periodicity constraint

$$R(\tau) = \mathbb{E}\,\mathbf{y}(t+\tau)\mathbf{y}(t+T)^\top = R(\tau - T) = R(T-\tau)^\top$$

which for example implies, $R(1) = R(T-1)^\top$ and so on. Hence a stationary periodic process defined on the discrete group $\mathbb{Z}_T$ has a **symmetric block-circulant** covariance matrix

$$\mathbf{R}{=}\mathbb{E}\,\mathbf{y}\mathbf{y}^\top{=}\begin{bmatrix} R_0 & R_1^\top & \dots & R_\tau^\top & \dots & R_\tau & \dots & R_1 \\ R_1 & R_0 & R_1^\top & \ddots & R_\tau^\top & \dots & \ddots & \vdots \\ \vdots & \ddots & & \ddots & & & \ddots & R_1^\top \\ R_1^\top & \dots & R_\tau^\top & \dots & R_\tau & \dots & R_1 & R_0 \end{bmatrix}$$

which we shall write

$$\mathbf{R}_T = \mathrm{Circ}\{R_0,\, R_1,\, \dots,\, R_\tau, \dots,\, R_\tau^\top,\, \dots,\, R_1^\top\} \quad \text{(III.1)}$$

the subscript denoting the number of blocks.

Note that we can always extend the covariance function of a stationary process defined on $[1,\, N]$ to an enlarged interval of length (at most[1]) $2N$ to make it the covariance function of a periodic process (i.e. making $\mathbf{R}$ symmetric block-circulant). This extension does not require extra information and can always be done on the basis of the available data. Hence a stationary process defined on some finite interval, may without loss of generality be assumed from the ouset to be periodic of period $N \equiv T$.

Next we consider n-dimensional stationary reciprocal processes. Stationary reciprocal processes defined on $[1,\, N]$ may always be assumed to be periodic of period N. For we have the following basic representation result.

Theorem 3.1: Every stationary reciprocal process on $[1,\, N]$ can be represented by a three terms recursion of the following form

$$M\mathbf{x}(k) = F^\top\mathbf{x}(k-1) + F\mathbf{x}(k+1) + \mathbf{e}(k) \quad \text{(III.2)}$$

where M, F are constant matrices and the associated boundary conditions can be taken to be cyclic; i.e.

$$\mathbf{x}(N+1) = \mathbf{x}(1) \qquad \mathbf{x}(0) = \mathbf{x}(N)\,. \quad \text{(III.3)}$$

If $\mathbf{x}$ is of full rank, M is symmetric and positive definite and $\mathbf{e}$ is a locally correlated process; i.e.

$$\mathbb{E}\,\mathbf{e}(k)\mathbf{e}(h)^\top = 0 \qquad |k-h| > 1\,, \quad \text{(III.4)}$$

such that

$$\mathbb{E}\,\mathbf{x}(k)\mathbf{e}(k)^\top = I \qquad \mathbb{E}\,\mathbf{x}(k)\mathbf{e}(h)^\top = 0 \qquad k \neq h\,. \quad \text{(III.5)}$$

For full rank processes, the conjugate process $\mathbf{e}$ is a moving average process of order one; i.e. $\mathbf{e}(k) = \mathbf{w}(k) + B\mathbf{w}(k-1)$ for some white noise $\mathbf{w}$. For non full rank processes this structure needs to be generalized.

The dynamical model (III.2) with boundary conditions (III.3) can be written in matrix notation as

$$\begin{bmatrix} M & -F & 0 & \dots & 0 & -F^\top \\ -F^\top & M & -F & \dots & 0 & 0 \\ 0 & -F^\top & M & -F & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & 0 \\ 0 & \dots & \dots & -F^\top & M & -F \\ -F & \dots & \dots & \dots & -F^\top & M \end{bmatrix} \begin{bmatrix} \mathbf{x}(1) \\ \mathbf{x}(2) \\ \vdots \\ \mathbf{x}(N) \end{bmatrix} = \begin{bmatrix} \mathbf{e}(1) \\ \mathbf{e}(2) \\ \vdots \\ \mathbf{e}(N) \end{bmatrix}$$

[1] see [18].

i.e. as $\Lambda\mathbf{x} = \mathbf{e}$ where the matrix Λ on the left is symmetric block-circulant with a block-tridiagonal structure

$$\Lambda := \text{Circ}\left\{M, -F, 0, \ldots, 0, -F^\top\right\}. \qquad \text{(III.6)}$$

The representation yields a fundamental characterization of the covariance matrix of a full rank reciprocal process.

Theorem 3.2: The covariance of a full-rank reciprocal stationary process $\mathbf{x}$ on $\mathbb{Z}_N$ must be the inverse of a block-tridiagonal circulant matrix.

$$\mathbf{\Sigma} := \mathbb{E}\,\mathbf{x}\mathbf{x}^\top = \Lambda^{-1} = \text{Circ}\left[M, -F, 0, \ldots, 0, -F^\top\right]^{-1}.$$

The proof follows by multiplying from the right $\Lambda\mathbf{x} = \mathbf{e}$ by $\mathbf{x}^\top$, whereby,

$$\Lambda\mathbb{E}\,\mathbf{x}\mathbf{x}^\top = \mathbb{E}\,\mathbf{e}\mathbf{x}^\top = I \qquad \Rightarrow \qquad \mathbb{E}\,\mathbf{x}\mathbf{x}^\top = \Lambda^{-1}.$$

IV. Identification of reciprocal models

An m-dimensional stationary process $\mathbf{y} := \{\mathbf{y}(k), k \in [1, N]\}$ admits a *reciprocal realization*, if there is an n-dimensional reciprocal stationary process $\mathbf{x}$ such that

$$\mathbf{y}(k) = C\mathbf{x}(k) \qquad k \in [1, N]$$

for a suitable constant matrix C. The dynamic equations of a reciprocal realization are of the form

$$\begin{aligned} M\mathbf{x}(k) &= F^\top\mathbf{x}(k-1) + F\mathbf{x}(k+1) + \mathbf{e}(k) \qquad \text{(IV.1)} \\ \mathbf{y}(k) &= C\mathbf{x}(k) \qquad \text{(IV.2)} \end{aligned}$$

J. A. Sand [56] discusses conditions for minimality of a reciprocal realization.

Determining whether such representations exist and computing the parameters of a (minimal) realization, (C, M, F) from the output covariance data $\mathbf{R} \equiv \{R(k), k = 0, 1, \ldots, N-1\}$ is so far an open problem (reciprocal stochastic realization).

Note however that in our one-dimensional reformulation of the texture modeling problem, the number of pixels in each row (m) will be large and the state vector will generally have a *smaller dimension than the output*: in other words the matrix C will have $n < m$ (independent) columns. Therefore $\mathbf{y}(k) = C\mathbf{x}(k)$ implies that $\mathbf{Y}_k := \text{span}\{\mathbf{y}_i(k)\,;\, i = 1, 2, \ldots, m\} = \text{span}\{\mathbf{x}_i(k)\,;\, i = 1, 2, \ldots, n\} := \mathbf{X}_k$ and hence, under these circumstances, the process $\mathbf{y}$ *will itself be reciprocal*. The (nontrivial and so far unsolved) problem of constructing the state for $\mathbf{y}$ can be bypassed alltogether. Note that in general however $\mathbf{y}$ will be a *singular* (non full-rank) reciprocal process. This last difficulty can be circumvented as it will be explained later. The first step of estimating the C matrix can be accomplished by a SVD decomposition. See [12] for details.

Henceforth we may (and shall) just assume that our measurement data is a sample of the $\mathbf{x}$ process. At first we shall assume that the observed reciprocal process is of full rank. So we are to solve the following problem;

Problem 1: Estimate the parameters (M, F) of a descriptor model (III.2) of an observed **full-rank** reciprocal process $\mathbf{x}$.

This identification problem has been approached from the classical (maximum likelihood) point of view under the assumption of a Gaussian distribution for $\mathbf{x}$. The parametrized density is

$$p_{(M,\,F)}(x) = \frac{1}{\sqrt{(2\pi)^N \det\left(\Lambda^{-1}\right)}} \exp\left(-\frac{1}{2}x^\top\Lambda x\right),$$

where the inverse covariance matrix Λ is parameterized by M and F as in (III.6).

Assuming that T independent sample images of the same texture $\mathbf{x}$ are available and denoting the sample sequence by $\underline{x} := \left(x^{(1)}, .., x^{(T)}\right)$, the log-likelihood function can be written

$$\begin{aligned} L(M, F) &= \log\,\det(\Lambda) - \text{Trace}\left\{MT_0\left(\underline{x}\right)\right\} - \\ &- \text{Trace}\left\{FT_1\left(\underline{x}\right)\right\} \qquad \text{(IV.3)} \end{aligned}$$

which has an exponential class structure [3] with matrix parameters (M, F) and *matrix-valued sufficient statistics* T_0 and T_1 given by:

$$\begin{aligned} T_0\left(\underline{x}\right) &= \frac{1}{T}\sum_{t=1}^{T}\{\sum_{k=0}^{N} x^{(t)}(k)\left[x^{(t)}(k)\right]^\top\} \\ T_1\left(\underline{x}\right) &= \frac{2}{T}\sum_{t=1}^{T}\{\frac{1}{N}\sum_{j=1}^{N} x^{(t)}(k)\left[x^{(t)}(k-1)\right]^\top\} \\ &+ \frac{2}{T}\sum_{t=1}^{T} x^{(t)}(0)\left[x^{(t)}(N)\right]^\top \end{aligned}$$

From exponential class theory it follows that

Proposition 4.1: The statistics T_0 and T_1 are Maximum Likelihood estimators for their expected values, namely

$$\begin{aligned} \frac{1}{N}T_0 &= \hat{\Sigma}(0) = \text{M.L. Estimator of } \mathbb{E}\,\mathbf{x}(k)\mathbf{x}(k)^\top \\ \frac{1}{N}T_1 &= \hat{\Sigma}(1) = \text{M.L. Estimator of } \mathbb{E}\,\mathbf{x}(k+1)\mathbf{x}(k)^\top \end{aligned}$$

In other words, we directly get the M.L. estimates of the main and upper diagonal blocks of the state covariance matrix Σ. However our original problem was to compute M.L. estimates of the parameters M and F. Since there is a one to one relation between Σ and (M, F):

$$\mathbf{\Sigma}^{-1} = \Lambda := \text{Circ}\left[M, -F, 0, \ldots, 0, -F^\top\right] \overset{1:1}{\Leftrightarrow} (M, F)$$

from well-known properties of maximum likelihood, we see that it must be possible to obtain the estimates $\hat{M}$ and $\hat{F}$ uniquely from $\hat{\Sigma}(0)$, $\hat{\Sigma}(1)$. This leads to an instance of the famous,

Covariance Selection Problem (A. P. DEMPSTER, [19]): Determine the (ML) estimates of (M, F) from $\hat{\Sigma}(0)$ and $\hat{\Sigma}(1)$ by solving

$$\hat{\mathbf{\Sigma}} := \hat{\Lambda}^{-1} = \text{Circ}\left[\hat{M}, -\hat{F}, 0, \ldots, -\hat{F}^\top\right]^{-1}$$

in particular impose that the blocks $\hat{\Lambda}_{i,j}$ should be zero exactly where the blocks $\Lambda_{i,j}$ are zero.

From the general theory in [19] it is known that the selection problem has a unique solution. Dempster's original algorithm for solving covariance selection problems is however computationally very intensive; it requires repeated inversion of matrices of size $O(N \times n)$. It does not seem to be useful for real size images. One may resort to approximations, as discussed in the paper [12]. Another way of proceeding is to make connection with *matrix extension problems* studied in linear algebra. See [28, vol II] for a survey. Given the covariance estimates $\hat{\Sigma}(0)$, $\hat{\Sigma}(1)$, one wants to complete the block-Toeplitz matrix

$$\begin{bmatrix} \hat{\Sigma}(0) & \hat{\Sigma}(1)^\top & \dots & ? & ? & ? \\ \hat{\Sigma}(1) & \hat{\Sigma}(0) & \hat{\Sigma}(1)^\top & \dots & ? & ? \\ \vdots & \hat{\Sigma}(1) & \hat{\Sigma}(0) & \hat{\Sigma}(1)^\top & \dots & ? \\ ? & \dots & \ddots & \ddots & \ddots & \\ ? & ? & \dots & \hat{\Sigma}(1) & \hat{\Sigma}(0) & \hat{\Sigma}(1)^\top \\ ? & ? & ? & \dots & \hat{\Sigma}(1) & \hat{\Sigma}(0) \end{bmatrix}$$

in such a way that the inverse $\hat{\Lambda} = \hat{\mathbf{\Sigma}}^{-1}$ has a symmetric *block-tridiagonal-circulant structure*. This can be seen as a particular *band extension problem* of the type dealt in, say [22]. A description of this approach is however outside the scope of this survey and will be found elsewhere [51].

As previously pointed out, the above identification procedure does not generally apply to the one-dimensional texture modeling problem, since the (reciprocal) state process for this example is in general not of full rank. Rank deficiency of the reciprocal state process turns actually out to be rather commonly encountered in applications so we must briefly address the issue. In case of non full rank, it can be shown that the components of $\mathbf{x}(t)$ can all be expressed as delayed versions of a smaller dimensional process $\mathbf{z}(t)$ which is of full rank and admits a general *nearest neighbor representation* of the type

$$\sum_{k=-\nu}^{\nu} M_k \, \mathbf{z}(t+k) + \mathbf{e}_{\mathbf{z}}(t) \qquad t \in \mathbb{Z}_N \tag{IV.4}$$

where M_0 is symmetric positive definite, $M_{-k} = M_k^\top$ and $\mathbf{e}_{\mathbf{z}}$ is the *(normalized) conjugate process* of $\mathbf{z}$, [42] such that

$$\mathbb{E}\, \mathbf{z}(k)\mathbf{e}_{\mathbf{z}}(k)^\top = I\,, \qquad \mathbb{E}\, \mathbf{z}(k)\mathbf{e}_{\mathbf{z}}(h)^\top = 0 \qquad k \neq h$$

In this case $\mathbf{e}_{\mathbf{z}}$ must be a locally correlated process with a correlation window of width ν,

$$\mathbb{E}\, \mathbf{e}_{\mathbf{z}}(k)\mathbf{e}_{\mathbf{z}}(h)^\top = 0 \qquad |k-h| > \nu$$

which implies that $\{\mathbf{e}_{\mathbf{z}}(k)\}$ is an MA type process of order ν,

$$\mathbf{e}_{\mathbf{z}}(k) = \mathbf{w}(k) + B_1\mathbf{w}(k-1) + \ldots + B_\nu \mathbf{w}(k-\nu)$$

for some white noise $\mathbf{w}$.

The model (IV.4) is a natural generalization of the 1-lag descriptor model (III.2). It makes connection with Gaussian Gibbs-Markov models once we interpret $\Lambda = \mathbf{\Sigma}^{-1}$ as the *Potential Matrix*.

In (IV.4) the spatial evolution of each scalar component $\mathbf{z}_j(k)\,; j = 1, \ldots, r$ is influenced by that of its nearest neighbors, in particular by $2\nu_j$ neighboring values of the same component, where ν_j may vary with j. The indices $\nu_j\,, j = 1, \ldots, r$, of which ν is by definition the largest, are structural indices of the model related to the Kronecker (observability) indices of linear system theory. This observation (which needs to be better substantiated) shows the conceptual advantage of the system identification approach as compared to the "physical" a priori modeling approach by Gibbs-Markov models where the structure of the nearest neighbor interaction potential must usually be presumed from the beginning and there is no way to relate it experimentally to the actual data to be described.

The dynamical equation (IV.4) can be written in matrix form as

$$\begin{bmatrix} M_0 & M_1^\top & \dots & M_\nu^\top & 0 & (*)^\top \\ M_1 & M_0 & M_1 & \dots & M_\nu^\top & 0 \\ \vdots & M_1 & M_0 & M_1^\top & \dots & M_\nu^\top \\ M_\nu & \dots & \ddots & \ddots & \ddots & \\ 0 & M_\nu & \dots & M_1 & M_0 & M_1^\top \\ (*) & 0 & M_\nu & \dots & M_1 & M_0 \end{bmatrix} \begin{bmatrix} \mathbf{z}(1) \\ \mathbf{z}(2) \\ \vdots \\ \mathbf{z}(N) \end{bmatrix} = \begin{bmatrix} \mathbf{e}_{\mathbf{z}}(1) \\ \mathbf{e}_{\mathbf{z}}(2) \\ \vdots \\ \mathbf{e}_{\mathbf{z}}(N) \end{bmatrix}$$

where the asteriscs on the upper and lower corners denote the circulant completion of the matrix. See e.g. [30] for many examples with scalar entries. This matrix, denoted $\Lambda := \mathrm{Circ}\left[M_0, M_1, \ldots, M_\nu, 0, \ldots, 0\,, M_\nu^\top, M_{\nu-1}^\top, \ldots, M_1^\top\right]$ is a *symmetric banded block-circulant matrix.*

Estimation of these models can be done by generalizing what was done for the 1-lag descriptor model for a full rank process. Now the log-likelihood depends on $\nu + 1$ matrix parameters $\{M_k\}$; i.e.

$$L(M_0, \ldots, M_\nu) = \log\det(\Lambda) - \sum_{k=0}^{\nu} \mathrm{Trace}\left\{M_k T_k\left(\underline{x}\right)\right\}$$

where the matrix-valued statistics $T_k\,, k = 0, \ldots, \nu$ are generalized sample covariances. By exponential class theory they are in fact the ML estimates of the true state covariances

$$\begin{array}{rcl} \frac{1}{N}T_0 & = & \hat{\Sigma}(0) = \text{M.L. Estimator of } \mathbb{E}\, \mathbf{z}(k)\mathbf{z}(k)^\top \\ & \ldots & \\ \frac{1}{N}T_\nu & = & \hat{\Sigma}(\nu) = \text{M.L. Estimator of } \mathbb{E}\, \mathbf{z}(k+\nu)\mathbf{z}(k)^\top \end{array}$$

The covariance selection problem now becomes: To compute the (unique) ML estimates of $(M_0,\, M_1, \ldots, M_\nu)$ from $\hat{\Sigma}(0)\, \ldots, \hat{\Sigma}(\nu)$, by solving

$$\hat{\mathbf{\Sigma}}^{-1} = \mathrm{Circ}\left[\hat{M}_0, \ldots, \hat{M}_\nu, 0, \ldots, 0, \hat{M}_\nu^\top, \ldots, \hat{M}_1^\top\right]$$

i.e by imposing that the inverse $\hat{\Lambda} = \hat{\mathbf{\Sigma}}^{-1}$ should have a *symmetric banded block -circulant structure of bandwith ν*. This again can be rephrased as a particular band extension problem for block circulant matrices: *Complete the estimated covariances $\hat{\Sigma}(0), \ldots, \hat{\Sigma}(\nu)$ with a sequence $\hat{\Sigma}(\nu+1), \ldots \hat{\Sigma}(N-1)$ in such a way that the inverse $\hat{\Lambda} = \hat{\mathbf{\Sigma}}^{-1}$ has a symmetric banded block-circulant structure of bandwidth ν.*

Figures 3 and 4 give an idea of what can be achieved by these modeling techniques. Both images are identified with an index one ($\nu = 1$) reciprocal model.

Fig. 3. Trees image: original texture (bottom) and synthesized (top).

Fig. 4. Steam image: original texture (bottom) and synthesized (top).

V. Dynamic models for visual processes

As explained in the introduction we regard a movie (a sequence of images) as a discrete-time signal $\mathbf{I}(x,t) : \mathcal{X} \times \mathbb{N} \to \mathbb{R}^+$. For digital images (finite number of pixels) the set $\mathcal{X}$ is a finite lattice of the form $\mathcal{X} := \{(i,j) : i = 1,..,r,\ j = 1,..,c\}$ where r and c are the number of rows an columns in the image[2]. For simplicity we consider graylevel movies, so that $\mathbf{I}(x,t)$, for each fixed t, is a matrix whose entries take values in $\mathbb{R}^+$. Of course it would be straightforward to extend the results in this paper to color images by considering vector valued signals.

In this section we shall be mainly interested in describing the dynamical properties of the sequence, i.e. our models should capture how the image changes over time. Examples of visual phenomena we are interested in modeling are dynamic textures [20] and gaits (e.g. walking, running etc.). In fact, dynamics has proven to be of paramount importance in extracting information from videos, see for instance the Johansson experiment [36].

At a given time t, the image $\mathbf{I}(\cdot,t)$ can be thought of as a vector process whose dimension is equal to the number of pixel in the image; for instance, if we are interested in modeling a "dynamic texture" (see figure 5 for an example) we shall consider the signal $\mathbf{y}(t) := \mathrm{vec}(\mathbf{I}(\cdot,t))$, obtaining by vectorizing the image I. Sometimes, the specific problem may suggest that only certain functions of the image intensity are of interest. For instance, when modeling a person walking, the signal of interest $\mathbf{y}(t)$ shall contain a description (which is extracted from the image $\mathbf{I}(\cdot,t)$) of the posture like, e.g., the positions of markers from a motion capture system or angles of the person's joints (knees, ankles, shoulders, elbows etc.) [6]. We regard the extraction of $\mathbf{y}$ from the rough image $\mathbf{I}$ as a pre-processing phase which shall not address in this paper.

In any case, it is typical that, regardless of the specific modeling task, the signal extracted from the image intensities $\mathbf{I}(\cdot,t)$, which describes the phenomena of interest, forms a vector say, $\mathbf{y}(t) \in \mathbb{R}^m$ with a "large" number (from tens to thousands), of components.

We are interested, therefore, in classes of models (and identification methodologies thereof) which are suited for high dimensional data. Note also that the number N of samples (i.e. $t = 1,..,N$) is very often of the same order (and sometimes smaller) than the data dimensionality ($N < m$). For instance, in dynamic textures modeling, the number N of images in the sequences is of the order of a few hundreds while m (which is equal to the number of pixels of the image) is certainly of the order of a few tens of thousands. It is therefore apparent that some sort of dimensionality reduction is absolutely needed in this context. Motivated by this requirement, we shall look for an r dimensional ($r \ll m$) white noise input $\mathbf{w} \in \mathbb{R}^r$ with unit covariance matrix $\mathbb{E}\mathbf{w}(t)\mathbf{w}^\top(t) = I$, and a rational transfer function $H(z)$, possibly of "low" McMillan degree n, which describe the output process according to the scheme,

$$\begin{aligned} \mathbf{y}(t) &= H(z)\mathbf{w}(t) + \mathbf{v}(t) \\ &= \mathbf{f}(t) + \mathbf{v}(t). \end{aligned} \tag{V.1}$$

called *Dynamic Factor Models* (DFM). The process $\mathbf{v}(t)$ (sometimes called *idiosyncratic noise*) represents the mismatch between the observed data $\mathbf{y}(t)$ and the "ideal model" $H(z)\mathbf{w}(t)$. It is assumed to be independent of $\mathbf{w}(t)$, zero mean and with uncorrelated components (diagonal covariance matrix).

This latter assumption hinges on the fact that all the "common dynamics" (cross-correlation) in the components of the output process $\mathbf{y}$ should be captured by the factor $\mathbf{f}(t) := H(z)\mathbf{w}(t)$. Dynamic Factor Models models can be though of as dynamic versions of the factor analysis model in statistics. Usually, the dimension m of $\mathbf{y}(t)$ is called the *cross-sectional* dimension, $\mathbf{w}(t)$ is called the *latent variable* and $\mathbf{f}(t) = H(z)\mathbf{w}(t)$ is called the *common factor*.

Sometimes the noise process $\mathbf{v}$ is also assumed to be temporally white. This we shall assume also in this paper.

Factor analysis models have been first developed by psychologists [59], [9]; dynamic version of factor models have attracted a remarkable attention in the statistical and econometric literature, see e.g. [27], [57], [48], [33] and references therein. It is also fair to say that, with a few notable exceptions [37], [62], [50], [52], [16], less attention has been payed to these models in the control engineering community. Recently, a generalization of these models allowing the cross-sectional dimension to go to infinity have been studied; these models are called *Generalized* DFM (GDFM)(see e.g. [24]

[2]In Section III these were denoted N and m; here we shall reserve these symbols to denote other quantities.

and references therein). This class of models could turn out to be useful in modeling visual signals since, as discussed above, often the cross-sectional dimension can be very large (tens of thousands or larger).

Identification of these models has been addressed by several authors, we refer the reader to [17], [49] for recent surveys.

In this paper we shall adopt a slightly different approach, using ideas from stochastic realization theory [40], [52] and subspace methods [61], [41], [14].

The identification procedure used in [20] can be seen as a simplified version of this approach.

VI. Identification

From now on we shall refer to a model class of the form (V.1) where we shall always assume that $\mathbf{v}(t)$ has uncorrelated and temporally white components. Furthermore, we shall assume that the transfer function $H(z)$ admits a factorization of the form

$$H(z) = L \cdot G(z) \quad L \in \mathbb{R}^{m \times p} \tag{VI.1}$$

where $p \ll m$. This just says that that a first dimensionality reduction can be achieved by a static transformation. Without loss of generality we can assume[3] $p \geq r$ and $L^\top L = I$. We postpone to future work an in-depth study of this condition as well as how it relates to similar ones encountered in the literature, see e.g. [49], [24], [48], [33], [17] and references therein.

Condition (VI.1) implies that the common dynamic factor $\mathbf{f}(t) = H(z)\mathbf{w}(t)$ can be reduced, by a static transformation, to a lower (p) dimensional process[4] $\mathbf{f}_r(t) := L^\dagger H(z)\mathbf{w}(t) = G(z)\mathbf{w}(t)$. The multiplicity of the "reduced" factor will however not change and still be $r \leq p$. It will be part of the identification procedure to estimate the matrix L as well as the dimensions p and r.

For future use let us observe that the covariance function of the process $\mathbf{y}(t)$ takes the form:

$$\Sigma_{\mathbf{y}}(0) = L\Sigma_{\mathbf{f}_r}(0)L^\top + \Sigma_{\mathbf{v}} \tag{VI.2}$$

and

$$\Sigma_{\mathbf{y}}(k) = L\Sigma_{\mathbf{f}_r}(k)L^\top \quad k \neq 0 \tag{VI.3}$$

The identification algorithm used in [20], [23] can be summarized as follows:

- Compute the SVD of $\Sigma_{\mathbf{y}}(0)$. An orthonormal basis $\hat{L}$ for the principal subspace (of dimension p) of this matrix is used as estimator of L.
  This corresponds to computing a Karhunen-Loeve expansion $\hat{\mathbf{y}}(t)$ of $\mathbf{y}(t)$ using p principal components:
  $$\hat{\mathbf{y}}(t) = \hat{L}\hat{L}^\top \mathbf{y}(t)$$
- The coefficients of the PCA expansion are computed as
  $$\mathbf{z}(t) := \hat{L}^\top \mathbf{y}(t) = \mathbf{f}_r(t) + \hat{L}^\top \mathbf{v}(t) \tag{VI.4}$$
- An AR(1) model for the PCA coefficients is estimated solving the linear equation
  $$\mathbf{z}(t) = A\mathbf{z}(t-1) + \mathbf{e}(t) \tag{VI.5}$$
  in the least squares sense.

In this last step an ARMA model could be estimated instead, and subspace techniques can be used to this purpose [61], [11]; this is done, for instance, in gait modeling [6]. The simple model (VI.5) has been used in dynamic textures since, experimentally, it provides good performance in both synthesis and recognition [20], [21] while being simple enough. Sample images from both the original and synthesized sequence using this method are reported in figure 5.

As we shall discuss in the next section, some applications such as gaits modeling call for dynamical models which include also periodic modes (simple unreachable eigenvalues on the unit circle). Subspace methods can indeed be extended to this purpose, see [6] and Section VI-A.

A few remarks are now in order concerning this procedure:

a) the first PCA step provides a consistent estimator of L only under the assumption $\Sigma_{\mathbf{v}} = \sigma^2 I$. Similarly to the approach suggested in [49], instead, L could be estimated from Singular Value Decomposition of $\Sigma_{\mathbf{y}}(k)$, $k > 0$. This is consistent regardless of the matrix $\Sigma_{\mathbf{v}}$ and only relies on whiteness of $\mathbf{v}(t)$.

b) Even under consistent estimation of L, the PCA coefficients (VI.4) are the sum of the common (reduced) factor $\mathbf{f}_r(t)$ and the noise component $\hat{L}^\top \mathbf{v}(t)$. Therefore, even under assumption (VI.1), $\mathbf{z}(t)$ is a full rank process which could be modeled, with some care, using a standard DFM of the form (V.1) itself.

c) The first dimensionality reduction using PCA is necessary because in computer vision problems the cross-sectional dimension m is often much larger than the number of data itself. Still, even after this reduction, $\mathbf{z}$ can be high dimensional. For instance, in the case of dynamic texture a reasonable number of principal components ranges in the interval $30 - 50$ for images whose size is of the order of 400×300 pixels, see e.g. [20]. Unfortunately the last step in the algorithm above does not seek for a low dimensional input (latent variable) as prescribed by DFM; in fact, generically, $\mathbf{e}(t)$ in (VI.5) is a full rank process.

We shall now attempt to outline a procedure which overcomes the shortcomings mentioned above and fully exploits the structure of DFM (V.1) under the additional assumption (VI.1). We shall not enter into the fine details required for the actual implementation; this is actually subject of ongoing research.

The main conceptual steps could be enumerated as follows:

1) Perform a first dimensionality reduction as in (VI.1). As suggested in [33], [49] the matrix L can be estimated from PCA of the covariance matrix (VI.3), for some $k > 0$. This provides a consistent estimator of

[3] If this was not the case the model (V.1) could be written with a smaller number of latent variables $\mathbf{w}(t)$.

[4] The superscript † denotes Moore-Penrose pseudoinverse.

(the column space of) L under the condition that $\mathbf{v}(t)$ is white noise.

2) Compute the reduced output $\mathbf{z}(t)$ as in (VI.4).
3) The reduced output component satisfies

$$\mathbf{z}(t) = \hat{L}^\top L G(z)\mathbf{w}(t) + \hat{L}^\top \mathbf{v}(t).$$

Let (A, B, C, D) be a state space realization of $\hat{L}^\top L G(z)$, i.e. $\hat{L}^\top L G(z) = C(zI - A)^{-1}B + D$. Then $\mathbf{z}(t)$ can be given a realization of the form

$$\begin{aligned} \mathbf{x}(t+1) &= A\mathbf{x}(t) + B\mathbf{w}(t) \\ \mathbf{z}(t) &= C\mathbf{x}(t) + D\mathbf{w}(t) + \hat{L}^\top \mathbf{v}(t) \end{aligned} \quad \text{(VI.6)}$$

where $\mathbf{w}(t)$ and $\hat{L}^\top \mathbf{v}(t)$ are uncorrelated white noise processes of dimension r and p respectively.
Using subspace techniques [61], [41], [11] we can recover the matrices $A, C, \bar{C}, \Sigma_{\mathbf{z}}(0)$ where $\bar{C} := \mathbb{E}\mathbf{z}(t-1)\mathbf{x}^\top(t)$ and $\Sigma_{\mathbf{z}}(0) := \mathbb{E}\mathbf{z}(t)\mathbf{z}^\top(t)$.
It is well known that all minimal (stochastic) realizations of the process $\mathbf{z}(t)$ can be parametrized by the symmetric positive definite solutions $P = P^\top > 0$ of the LMI (positive real lemma)

$$\begin{bmatrix} P - APA^\top & \bar{C}^\top - APC^\top \\ \bar{C} - CPA^\top & \Sigma_{\mathbf{z}}(0) - CPC^\top \end{bmatrix} \geq 0 \quad \text{(VI.7)}$$

Usually one is interested in the stable, minimum phase model (spectral factor) for $\mathbf{z}(t)$ which corresponds to the minimal solution P^- of the LMI (VI.7). Instead, here, we are interested in models for which the dimension r of white input in the state equation is as small as possible (this is called multiplicity of $\mathbf{x}(t)$ in the literature of stochastic processes [55]). This is achieved by finding solutions P which minimize the rank of the matrix $Q := P - APA^\top$. Under suitable assumptions the "model noise" $B\mathbf{w}(t)$ and the "measurement noise" $\hat{L}^\top \mathbf{v}(t)$ are uncorrelated, hence there exist solutions P which make $\bar{C}^\top - APC^\top = 0$. This we call a "diagonalizing" P. In particular this certainly holds if $G(\infty) = D = 0$.
This is in general a difficult problem and further investigation is needed. We shall now give for granted that such a P has been found.

4) The matrices B, D and the output noise variance $\hat{L}^\top \Sigma_{\mathbf{v}} \hat{L}$ can be found as follows.
Since P is solution of (VI.7) there exist matrices $\bar{B}$ and $\bar{D}$ satisfying

$$\begin{aligned} \bar{B}\bar{B}^\top &= P - APA^\top \\ \bar{B}\bar{D}^\top &= \bar{C}^\top - APC^\top \\ \bar{D}\bar{D}^\top &= \Sigma_{\mathbf{z}}(0) - CPC^\top \end{aligned}$$

If $P - APA^\top$ is rank deficient, w.l.o.g. $\bar{B}$ can be chosen in the form $\bar{B} = [B\ 0]$. Let also $\bar{D} = [D\ \tilde{D}]$. It thus follows that

$$\begin{aligned} BB^\top &= P - APA^\top \\ BD^\top &= \bar{C}^\top - APC^\top \\ DD^\top + \tilde{D}\tilde{D}^\top &= \Sigma_{\mathbf{z}}(0) - CPC^\top \end{aligned}$$

Hence $\hat{L}^\top \Sigma_{\mathbf{v}} \hat{L} = \tilde{D}\tilde{D}^\top$.

There are a number of (statistical) issues related to the procedure outlined above, among which the decision concerning the rank of $P - APA^\top$, which are not discussed here but are of paramount importance for the actual implementation when using sample moments and estimators $\hat{A}$, $\hat{C}$, $\hat{\bar{C}}$. We postpone discussion of these aspects to future work.

Also the relation of this procedure with state-of-the art techniques for identification of DFM, see e.g. [17], [33], [49], [24], is outside the scope of this tutorial paper and will be addressed elsewhere.

A. Modeling with purely deterministic components

In certain vision application such as textures and gaits it is desirable to model signals with periodic components. People in vision have sometimes used ad-hoc remedies to enforce periodic modes in dynamic textures models, see e.g. [66]. Other work (see e.g. [1]) have postulated a strictly periodic structure of the data by assuming that the eigenvalues of A are uniformly spaced on the unit circle. This corresponds exactly to performing Discrete Fourier Transform of the sequence to be modeled, yielding a completely equivalent representation of the data itself.

In [6] subspace methods have been extended in order to model signals which contains both periodic and purely "stochastic" components. We shall briefly overview this approach here. In this Section we shall assume that the first (PCA based) dimensionality reduction has been performed and we therefore work with the reduced signal $\mathbf{z}(t)$. We use the notation $\hat{\mathbf{z}}(t|t-1)$ to denote the best (minimum variance) linear predictor of $\mathbf{z}(t)$ given its past history $\{\mathbf{z}(t-1), \mathbf{z}(t-2), ...\}$. It is well known (first part of Wold's decomposition theorem [55], [47]) that every stationary random process $\mathbf{z}(t)$ can be decomposed into two parts

$$\mathbf{z}(t) = \mathbf{z}_d(t) + \mathbf{z}_s(t) \quad \text{(VI.8)}$$

where $\mathbf{z}_d(t)$ is a *purely deterministic* (PD) process which can be predicted exactly as a linear combination of its past (i.e. $\mathbf{z}_d(t) = \hat{\mathbf{z}}_d(t|t-1)$) , and $\mathbf{z}_s(t)$ is a *purely non-deterministic* (PND) process (or "purely stochastic," hence the choice of subscript s), uncorrelated from $\mathbf{z}_d(t)$, for which the one step ahead prediction error $\mathbf{z}_s(t) - \hat{\mathbf{z}}_s(t|t-1)$ has positive definite variance. From Wold's decomposition theorem [55], the PND part can be given an infinite moving average representation of the form $\mathbf{z}_s(t) = \sum_{\tau=0}^{\infty} W(\tau)\mathbf{e}(t-\tau)$, where $W(\tau)$ is a sequence of matrices such that $W(0) = I$, $\sum_{\tau=0}^{\infty} \|W(\tau)\|^2 < \infty$ and $\mathbf{e}(t)$ is the *innovation process* $\mathbf{e}(t) := \mathbf{z}_s(t) - \hat{\mathbf{z}}_s(t|t-1)$.

Note however that, as in the previous section, we are not necessarily interested in the minimum-phase model for $\mathbf{z}_s$. Hence we consider the decomposition:

$$\mathbf{z}_s(t) = \sum_{\tau=0}^{\infty} G(\tau)\mathbf{w}(t-\tau) + \hat{L}^\top \mathbf{v}(t) \quad \text{(VI.9)}$$

where $G(0) = D_s$, $\sum_{\tau=0}^{\infty} \|G(\tau)\|^2 < \infty$ and $\mathbf{w}(t)$ is a normalized white noise process with uncorrelated components. In general $W(\tau)$ needs not be equal to $G(\tau)$, as we

shall discuss in the next section. In fact this freedom can be used to match higher order statistics which might in turn be useful both to the purpose of synthesis and recognition.

It is possible to show that the PD component $\mathbf{z}_d(t)$ can be represented as the superposition of (possibly infinitely many) sinusoidal signals. However, from a practical standpoint, we can assume that $\mathbf{z}_d(t)$ is the superposition of a finite number of sinusoids and hence can be represented, for $t > t_0$, as the output of an autonomous system of state dimension n_d

$$\left\{ \begin{array}{rcl} \mathbf{x}_d(t+1) & = & A_d \mathbf{x}_d(t) \\ \mathbf{z}_d(t) & = & C_d \mathbf{x}_d(t) \end{array} \right. \quad \text{(VI.10)}$$

with the constraint that A_d has eigenvalues on the unit circle and is diagonalizable. Without loss of generality the pair (A_d, C_d) can be taken to be observable. From stationarity of $\mathbf{z}_d$, $\mathbf{x}_d(t)$ is also stationary and $P_d = \text{Var}\{\mathbf{x}_d(t)\}$ satisfies the homogeneous Lyapunov equation $P_d = A_d P_d A_d^\top$. Since the choice of basis in the state space is arbitrary, one can choose it so that $P_d = I$; with this canonical choice, we have

$$A_d A_d^\top = I \quad \text{(VI.11)}$$

showing that, in this particular basis, A_d needs to be orthogonal.

Similarly, it is possible to give a state space realization to the representation (VI.9) in the form

$$\left\{ \begin{array}{rcl} \mathbf{x}_s(t+1) & = & A_s \mathbf{x}_s(t) + B_s \mathbf{w}(t) \\ \mathbf{x}_s(t) & = & C_s \mathbf{x}_s(t) + D_s \mathbf{w}(t) \end{array} \right. \quad \text{(VI.12)}$$

where $\mathbf{x}_s(t) \in \mathbb{R}^{n_s}$. Defining the aggregate state $\mathbf{x}(t) = [\mathbf{x}_d^\top(t)\ \mathbf{x}_s^\top(t)]^\top$, $\mathbf{x}(t) \in \mathbb{R}^n$, we obtain a generative model of the the stationary process $\mathbf{z}(t) \in \mathbb{R}^m$ in state-space form $(VI.6)$ with

$$\begin{array}{l} A = \left[\begin{array}{cc} A_d & 0 \\ 0 & A_s \end{array} \right] \quad B = \left[\begin{array}{c} 0 \\ B_s \end{array} \right] \\ C = \left[\begin{array}{cc} C_d & C_s \end{array} \right] \quad D = D_s \\ |\lambda_i(A_d)| = 1,\ i = 1, .., n_d \ \ |\lambda_j(A_s)| < 1,\ \ j = 1, .., n_s \\ \mathbf{x}(t) = \left[\begin{array}{c} \mathbf{x}_d(t) \\ \mathbf{x}_s(t) \end{array} \right],\ \mathbf{x}_d(t) \in \mathbb{R}^{n_d},\ \mathbf{x}_s(t) \in \mathbb{R}^{n_s} \\ E[\mathbf{w}(t)] = 0 \quad , \quad E\left[\mathbf{w}(t)\mathbf{w}(t)^\top\right] = I. \end{array} \quad \text{(VI.13)}$$

where $\mathbf{x}_d(t)$ and $\mathbf{x}_s(t)$ are the deterministic and stochastic components of the state corresponding to (VI.8) and $[\mathbf{x}_0 = \left[\begin{array}{cc} \mathbf{x}_{0d}^\top & \mathbf{x}_{0s}^\top \end{array} \right]^\top$ is the initial condition.

Estimation of sinusoidal components (corrupted by white noise) is addressed by standard algorithms such as MUSIC [58] and ESPRIT [53]; however these algorithms disregard non-periodic components which are instead of considerable importance for tasks such as synthesis and/or recognition. In the paper [6] we have discussed a modification of subspace methods which allows to handle critically stable systems as (VI.13); further work is indeed needed to investigate the statistical properties of the algorithm proposed. We shall just give an outline of the algorithm and refer the reader to [6] for the details.

1) From $\mathbf{z}(t)$ estimate the state space $\hat{\mathbf{x}}(t)$ and the matrices $\hat{A}$, $\hat{K}$ and $\hat{C}$ of the innovation model
$$\begin{array}{rcl} \mathbf{x}(t+1) & = & A\mathbf{x}(t) + K\mathbf{e}(t) \\ \mathbf{z}(t) & = & C\mathbf{x}(t) + \mathbf{e}(t) \end{array}$$
using the subspace method in [61], [11].
2) Change the basis in the estimated state space according to the eigenvalue decomposition of $\hat{A}$. Let $\hat{T}\hat{\mathbf{x}}(t) := [\hat{\mathbf{x}}_d^\top(t)\ \hat{\mathbf{x}}_s^\top(t)]^\top$ be the state basis which (block) diagonalizes the state transition matrix
$$\hat{T}\hat{A}\hat{T}^{-1} = \left[\begin{array}{cc} \hat{A}_1 & 0 \\ 0 & \hat{A}_2 \end{array} \right]$$
where $\hat{A}_1$ has eigenvalues "close" to the unit circle and its principal subspace is "almost unreachable". The decision as to whether an eigenvalue is close to the unit circle and/or its corresponding eigenspace is almost unreachable ought to be based on statistical properties of the estimators $\hat{A}$, $\hat{K}$ which we shall not discuss here. Statistical properties [13] of subspace methods may turn out to be useful to this purpose.
Without loss of generality the state transformation $\hat{T}$ can be chosen so that the sample covariance matrix of $\hat{\mathbf{x}}_d$ is the identity. With this choice of basis it is known (see eq. (VI.11)) that the matrix A_d should be orthogonal. Then $\hat{A}_d$ can be estimated by solving[5]
$$\hat{A}_d = \min_{Ad \in O(n_d)} \|\hat{\mathbf{x}}_d(t+1) - A_d \hat{\mathbf{x}}_d(t)\|_F^2.$$
This is called "matrix Procrustes problem" [64], [29] and its solution can be obtained from Singular Value Decomposition $USV = \hat{\Sigma}_{\hat{\mathbf{x}}_d' \hat{\mathbf{x}}_d}$ of the sample covariance between $\hat{\mathbf{x}}_d(t+1)$ and $\hat{\mathbf{x}}_d(t)$ as $\hat{A}_d = UV^\top$.
3) Define $[\hat{C}_d\ \hat{C}_2] := C\hat{T}^{-1}$. Using the estimators $\hat{A}_d$ and $\hat{C}_d$ estimate the initial condition $\hat{x}_d(0)$ solving
$$\hat{x}_d(0) := \arg\min \| \sum_{t=0}^{N} z(t) - \hat{C}_d \hat{A}_d^t x_d(0) \|_F^2$$
4) Estimate the stochastic component $\hat{z}_s(t) := z(t) - \hat{C}_d \hat{A}_d^t \hat{x}_d(0)$. Apply the subspace algorithm in [61], [11] to $\hat{z}_s(t)$ and compute estimates $\hat{A}_s$, $\hat{C}_s$, $\hat{K}_s$ of the innovation model
$$\begin{array}{rcl} \mathbf{x}_s(t+1) & = & A_s \mathbf{x}_s(t) + K_s \mathbf{e}(t) \\ \mathbf{z}_s(t) & = & C_s \mathbf{x}(t) + \mathbf{e}(t) \end{array}$$

B. Matching high order statistics

Experience with texture and gait modeling [20], [21], [6], [7] shows that, often, linear models are not rich enough to the purpose of synthesis and recognition. However, tests in [6] suggest that for human gait data the linearity assumption cannot be ruled out if also higher order statistics are considered.

In fact, while the linear (minimum phase) model captures the second order statistics of the data, proper choices of the

[5] $O(n_d)$ is the orthogonal group of $n_d \times n_d$ matrices and the subscript $_F$ denotes Frobenius norm.

input distribution as well as of the particular spectral factor (parametrized by the solutions P of the LMI (VI.7)) may be used to model also higher order statistics of the data. For reasons of space, in this short tutorial, we shall not enter into the details of how this can be done and intead refer the reader to [6] and references therein. Suffices here to say that, indeed, these richer models which consider also non-Gaussian inputs and non-minimum phase models do improve linear models described in the previous Section. Along these lines, we also refer the reader to the work [43] for a non-linear, Monte Carlo based, approach to texture synthesis.

VII. Classification

Measuring the "distance" between data sequences is a prerequisite for performing classification (and recognition) tasks. A straightforward approach would be, perhaps, to compare the sequences of data themselves, e.g. measuring their L_2 distance. Of course there are several problems with this approach; for instance it might become hard if the data sequences have different lengths, if some data are missing, if two sequences of data are measurements of the same phenomena using different "sensors", like, e.g. two movies of the same person walking taken from two different viewpoints.

A common approach in statistics is to assume that the data have been generated according to a probabilistic model in a given class. Then estimation of the model (within the chosen class) which best describes the data and classification are intimately related. The recent paper [26] discusses how distances can be measured when one is only interested in second order statistical properties, i.e. power spectra.

Other approaches which have been recently proposed at the interface between of control and computer vision include deterministic approaches (model validation based), such as [44], [1].

In this short tutorial we shall discuss a possible methodology which is tailored to the model considered in the previous sections, i.e. linear state space models with given initial conditions and, possibly, non-Gaussian inputs. The modeling stage aims at capturing the invariant properties of the sequence, while describing its variability. These invariant properties are encoded in the dynamical model, but also in the initial condition and in the input distribution. In [6], the Kernel based distance introduced in [63], has been extended to account for dynamical properties, initial conditions and input distribution.

A. Kernel-based distance

In this section we shall refer to a model M as a collection $M := \{A, B, C, D, x_0, p_{\mathbf{u}}\}$, which include the system parameters (A, B, C, D), the initial condition x_0 and the input distribution $p_{\mathbf{w}}$. These will be outcomes of the identification experiment. The space of models M (say of all M's giving rise to stationary output processes $\mathbf{y}$) can be endowed with an inner product as follows. Let $\mathbf{y}(t)$ and $\mathbf{y}^{'}(t)$ be output process of M and M'. Let us assume, for a moment, that models M and M' have the same input[6] $\mathbf{w}$ with probability density function $p_{\mathbf{w}}$. Following [63], for a suitable positive definite matrix W,

$$k(M, M') := \max_{\tau} \mathbb{E}_{p_{\mathbf{w}}} \left(\sum_{i=0}^{\infty} e^{-\lambda t} \mathbf{y}^{\top}(t) W \mathbf{y}^{'}(t-\tau) \right)$$

Using the Binet-Cauchy theorem, it has been shown in [63] that this is indeed a "valid" inner product. The maximization over τ is necessary in order to "align" sequences which differ only up to a phase shift. Computational aspects and details can be found in [6], where also the extension to the case in which the two models M and M' have different input distribution is discussed.

The inner product $k(M, M')$ induces a distance between models defined as

$$d(M, M') := k(M, M) + k(M', M') - 2k(M, M')$$

This distance can be effectively used to compare video sequences encoded through the models M and M'. Some results are reported in figure 6 (taken from [6]).

VIII. Conclusions and future work

In this paper we have discussed some modeling and identification techniques which have been proposed recently for applications to computer vision. In most cases there are a number of open issues which we have been listed along the way; our future work will address these open questions. We would also like to remark that, besides the few cases dealt with in this paper, Computer Vision poses challenging problems which can be cast in the framework of system identification and data analysis; many of these cannot be solved using off-the-shelf tools thus providing inspiration for future work.

IX. Acknowledgements

Much of the material presented in this work is the outcome of many years of collaboration and fruitful discussions with Stefano Soatto, Alessandro Bissacco, Gianfranco Doretto, Payam Saisan, which are gratefully acknowledged.

References

[1] S. Al-Takrouri and A. Savkin, "A model validation approach to texture recognition," in *Proc. of ECC*, 2007, pp. 1537–1544.

[2] N. Balram and J. Moura, "Noncausal Gauss Markov random fields: parameter structure and estimation," *IEEE Trans. Information Theory*, vol. IT-39, no. 4, pp. 1333–1355, 1993.

[3] O. Barndorff-Nielsen, *Information and Exponential Families in Statistical Theory*. Chichester: Wiley, 1978.

[4] M. Baseville, A. Benveniste, K. Chou, S. Golden, R. Nikoukhah, and A. Willsky, "Modeling and estimation of multiresolution stochastic processes," *IEEE Trans. on Inf.Theory*, vol. 38, no. 2, pp. 766–785, 1992.

[5] A. Benveniste, R. Nikoukhah, and A. Willsky, "Multiscale system theory," in *Proceedings of the 29th CDC*, 1990, pp. 2484–2489.

[6] A. Bissacco, A. Chiuso, and S. Soatto, "Classification and recognition of dynamical models: The role of phase, independent components, kernels and optimal transport," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 29, no. 11, pp. 1958–1972, 2007.

[6]This is reasonable as long as the inputs of M and M' have the same statistical distribution.

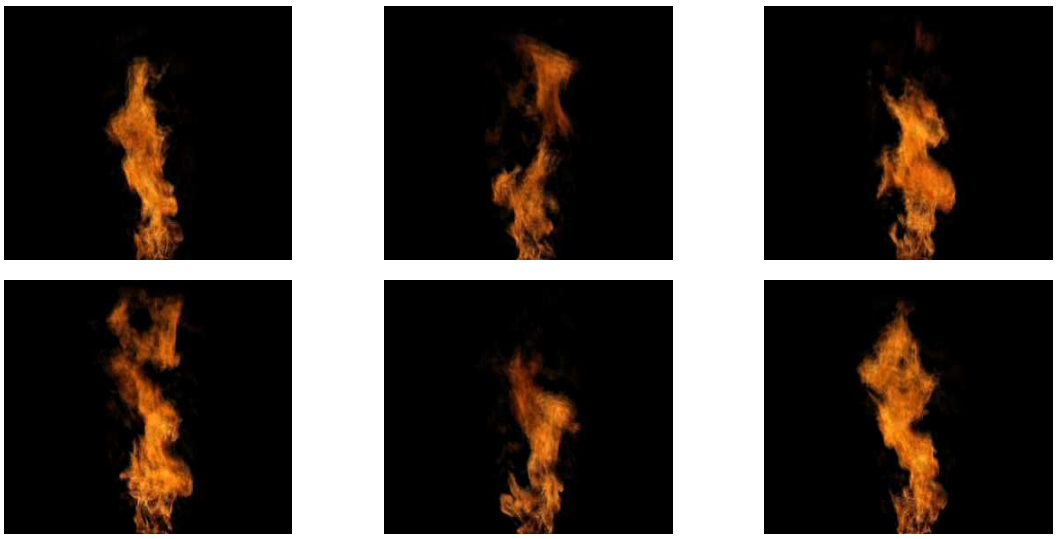

Fig. 5. Original (3 snapshots) dynamic texture: flame sequence. Top: original, bottom synthesized.

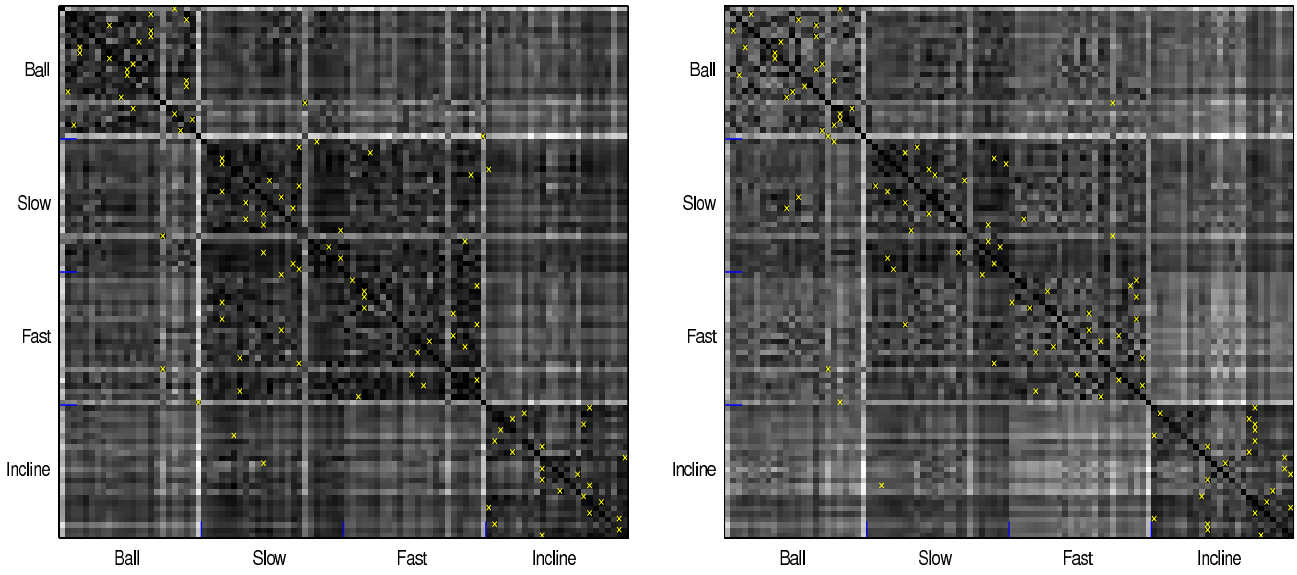

Fig. 6. State and input kernel distances. We show the confusion matrices representing trace kernel distances between non-Gaussian linear models learned from walking sequences in the Mobo dataset [31]. There are 4 motion classes and 24 individuals performing these motions, for a total of 96 sequences. For each sequence we learn a linear model and then measure distance between models by the trace kernels. On the left we show results using kernels on initial states only, on the right we display the confusion matrix obtained from the trace kernels that include the effect of the input. For each row a cross indicates the nearest neighbor. It is clear how the additional information provided by the input statistics results in improved gait classification performances: we have 17 (17.7%) nearest neighbors mismatches (i.e. closest models that do not belong to the same gait class) using the state-only distance, while only 9 (9.3%) with the complete trace kernel distance.

[7] A. Bissacco, P. Saisan, and S. Soatto, "Dynamic modeling of human gaits," in *Proc. of the IFAC Symposium on System Identification*, 2003.

[8] J. D. Bonet, "Multiresolution sampling procedure for analysis and synthesis of texture images," *Computer Graphics. ACM SIGGRAPH*, pp. 361–368, 1997.

[9] C. Burt, "Experimental test of general intelligence," *British J. of Psycology*, vol. 3, pp. 94–177, 1909.

[10] R. Chellapa and R. L. Kashyap, "Digital image restoration using spatial interacton models," *IEEE Trans. Automatic Control*, vol. AC-35, no. 9, pp. 1013–1023, 1990.

[11] A. Chiuso, "The role of Vector AutoRegressive modeling in predictor based subspace identification," *Automatica*, vol. 43, no. 6, pp. 1034–1048, June 2007.

[12] A. Chiuso, A. Ferrante, and G. Picci, "Reciprocal realization and modeling of textured images," in *Proceedings of the 44rd IEEE Conference on Decision and Control*, Seville, Spain, December 2005.

[13] A. Chiuso and G. Picci, "The asymptotic variance of subspace estimates," *Journal of Econometrics*, vol. 118, no. 1-2, pp. 257–291, 2004.

[14] ——, "On the ill-conditioning of subspace identification with inputs," *Automatica*, vol. 40, no. 4, pp. 575–589, 2004.

[15] K. Chou, A. Willsky, and A. Benveniste, "Multiscale recursive estimation, data fusion, and regularization," *IEEE Trans. on Automatic Control*, vol. 39, no. 3, pp. 464–478, 1994.

[16] M. Deistler and W. Scherrer, "Structure theory for linear dynamic errors-in-variables models," *SIAM J. Control and Optimization*, vol. 36, pp. 2148–2175, 1998.

[17] M. Deistler and C. Zinner, "Modeling high-dimensional time series by

generalized linear dynamic factor models: and introductory survey," *Communications in Information Systems*, vol. 2, pp. 153–166, 2007.

[18] A. Dembo, C. L. Mallos, and L. Shepp, "Embedding nonnegative definite toeplitz matrices in nonnegative definite circulant matrices with application to covariance estimation," *IEEE Trans. Information Theory*, vol. IT-35, pp. 1206–1212, 1989.

[19] A. Dempster, "Covariance selection," *Biometrics*, vol. 28, pp. 157–175, 1972.

[20] G. Doretto, A. Chiuso, S. Soatto, and Y. Wu, "Dynamic textures," *International Journal of Computer Vision*, vol. 51, no. 2, pp. 91–109, February 2003.

[21] G. Doretto, D. Cremers, P. Favaro, and S. Soatto, "Dynamic texture segmentation," in *Proceedings of the International Conference on Computer Vision*, vol. 2, Nice, France, October 2003, pp. 1236–1242.

[22] H. Dym and I. Gohberg, "Extension of band matrices with band inverses," *Linear Algebra and Applications*, vol. 36, pp. 1–24, 1981.

[23] A. Fitzgibbon, "Stochastic rigidity: Image registration for nowhere-static scenes," in *Proc. IEEE International Conf. Computer Vision (ICCV)*, vol. 1, Vancouver, Canada, 2001, pp. 662–670.

[24] M. Forni, M. Hallin, M. Lippi, and L. Reichlin, "The generalized dynamic-factor model: identification and estimation," *The Review of Economics and Statistics*, vol. 82, pp. 540–554, 2004.

[25] S. Geman and D. Geman, "Stochastic relaxation, gibbs distribution and bayesian restoration of images," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. PAMI-6, pp. 721–741, 1984.

[26] T. Georgiou, "Distances and Riemannian metrics for spectral density functions," *IEEE Trans. on Signal Processing*, vol. 55, no. 8, pp. 3995–4003, 2007.

[27] J. Geweke, "The dynamic factor analysis of economic time series," in *Latent Variables in Socio-Economics Models*, D. Aigner and A. Goldberger, Eds. Amsterdam: North-Holland, 1977.

[28] I. Gohberg, Goldberg, and M. Kaashoek, *Classes of Linear Operators vol II*. Boston: Birkhauser, 1994.

[29] G. Golub and C. Van Loan, *Matrix Computation*, 2nd ed. The Johns Hopkins Univ. Press., 1989.

[30] R. M. Gray, *Toeplitz and Circulant Matrices:*. John Wiley & Sons, 1979.

[31] R. Gross and J. Shi, "The cmu motion of body (mobo) database," Robotics Institute, Carnegie Mellon University, Tech. Rep., 2001.

[32] D. Heeger and J. Bergen, "Pyramid-based texture analysis/synthesis," in *SIGGRAPH 95 Conf. Proc.*, 1995.

[33] Y. Hu and R. Chou, "On the Peña-Box model," *Journal of Time Series Analysis*, vol. 25, pp. 811–830, 2004.

[34] B. Jamison, "Reciprocal processes: the stationary Gaussian case," *Ann. Math Stat.*, vol. 41, pp. 1624–1630, 1970.

[35] ——, "Reciprocal processes," *Zeit. Warschelinchkeitstheorie und Verw. Gebiete*, vol. 30, pp. 65–86, 1974.

[36] G. Johansson, "Visual motion perception," *Scientific American*, vol. 232, pp. 76–88, 1975.

[37] R. Kalman, "Identifability and problems of model selction in econometrics," in *Advances in econometrics*, W. Hildebrandt, Ed. Cambridge: Cambridge University Press, 1983.

[38] B. C. Levy and A. Ferrante, "Characterization of stationary discrete-time Gaussian reciprocal processes over a finite interval," *SIAM J. Matrix Anal. Appl.*, vol. 24, pp. 334–355, 2002.

[39] B. C. Levy, R. Frezza, and A. Krener, "Modeling and estimation of discrete-time Gaussian reciprocal processes," *IEEE Trans. Automatic Control*, vol. AC-35, no. 9, pp. 1013–1023, 1990.

[40] A. Lindquist and G. Picci, "A geometric approach to modelling and estimation of linear stochastic systems," *Journal of Mathematical Systems, Estimation and Control*, vol. 1, pp. 241–333, 1991.

[41] ——, "Canonical correlation analysis, approximate covariance extension and identification of stationary time series," *Automatica*, vol. 32, pp. 709–733, 1996.

[42] P. Masani, "The prediction theory of multivariate stochastic proceses, iii," *Acta Mathematica*, vol. 104, pp. 141–162, 1960.

[43] A. Masiero and A. Chiuso, "Non-linear temporal texture synthesis: a Monte Carlo approach," in *Computer Vision - Proc. of ECCV*, A. Leonardis, H. Bischof, and A. Prinz, Eds. Spinger Verlag, 2006.

[44] M. Mazzaro, M. Sznaier, and O. Camps, "A model (in)validation approach to gait classification," *IEEE Trans. on Pattern Analysis and MAchine Intelligence*, vol. 27, no. 11, pp. 1820–1825, 2005.

[45] J. Moura and N. Balram, "Recursive structure of noncausal Gauss Markov random fields," *IEEE Trans. Information Theory*, vol. IT-38, no. 2, pp. 334–354, 1992.

[46] J. Neymann, "University of california publications in statistics," Vol. I, University of California Press, 1954.

[47] A. Papoulis, "Predictable processes and Wold's decomposition: an review," *IEEE Transactions on Acoustic, Speech and Signal Processing*, vol. 33, no. 4, pp. 933–938, 1985.

[48] D. Peña and G. Box, "Identifying a simplifying structure in time series," *J. Amer. Stat. Ass.*, vol. 82, pp. 836–843, 1987.

[49] D. Peña and P. Poncela, "Nonstationary dynamic factor analysis," *Journal of Statistical Planning and Inference*, vol. 136, pp. 1237–1257, 2006.

[50] G. Picci, "Parametrization od factor analysis models," *J. of Econometrics*, vol. 41, pp. 17–38, 1987.

[51] G. Picci and F. Carli, "Band extension of block-circulant matrices," in *Proceedings of the 18th MTNS symposium*, Blacksbourg VA, July 2008.

[52] G. Picci and S. Pinzoni, "Dynamic factor-analysis models for stationary processes," *IMA Journal of MAthematical Control & Information*, vol. 3, pp. 185–210, 1986.

[53] R. H. Roy, "Esprit - a subspace rotation approach to estimation of parameters of cisoids in noise," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. Vol. ASSP-34, pp. pp. 1340–1342, Oct. 1986.

[54] Y. A. Rozanov, *Markov Random Fields*. New York: Springer Verlag, 1982.

[55] Y. Rozanov, *Stationary Random Processes*. San Francisco: Holden-Day, 1967.

[56] J. A. Sand, "Reciprocal realizations on the circle," *SIAM J. Control and Optimization*, vol. 34, pp. 507–520, 1996.

[57] T. Sargent and C. Sims, "Business cycle modeling without pretending to have too much a priori economic theory," in *New Methods in Business Research*, C. Sims, Ed. Minneapolis: Federal Reserve Bank of Minneapolis, 1977.

[58] R. O. Schmidt, "A signal subspace approach to multiple emitter location and spectral estimation," Ph.D. dissertation, Stanford University, Nov. 1981.

[59] C. Sperman, "General intelligence, objectively determined and measured," *American J. of Psycology*, vol. 15, pp. 201–203, 1904.

[60] M. Sznaier, O. Camps, and C. Mazzaro, "Finite horizon model reduction of a class of neutrally stable systems with applications to texture synthesis and recognition," in *Proc. of the IEEE Conf. on Decision and Control*, 2004.

[61] P. Van Overschee and B. De Moor, "Subspace algorithms for the stochastic identification problem," *Automatica*, vol. 29, pp. 649–660, 1993.

[62] J. van Schuppen, "Stochastic realization problems motivated by econometric modeling," in *Modeling Identification and robust control*, C. Byrnes and A. Lindquist, Eds. Amsterdam: North-Holland, 1986.

[63] S. Vishwanathan, R. Vidal, and A. J. Smola., "Binet-cauchy kernels on dynamical systems and its application to the analysis of dynamic scenes." *International Journal of Computer Vision*, vol. 73, no. 1, pp. 95 – 119, 2007.

[64] G. Wahba, "A least squares estimate of satellite attitude," *SIAM Review, Problems and Solutions section*, vol. 8, no. 3, 1966.

[65] J. W. Woods, "Two-dimensional discrete Markovian fields," *IEEE Transaction Information Theory*, vol. IT-18, no. 2, pp. 232–240, March 1972.

[66] L. Yuan, F. Wen, C. Liu, and H. Shum, "Synthsizing dynamic texture with closed-loop linear dynamic system," in *Proc. of European Conf. on Computer Vision (ECCV)*, 2004, pp. 603–616.

[67] S. Zhu, Y. Wu, and D. Mumford, "Minimax entropy principle and its application to texture modeling," *Neural Computation*, vol. 9, no. 8, 1997.