

Chapter 14

Visual Data Recognition and Modeling based on Local Markovian Models

Michal Haindl

Abstract An exceptional 3D wide-sense Markov model which can be completely solved analytically and easily synthesized is presented. The model can be modified to faithfully represent complex local data by adaptive numerically robust recursive estimators of all its statistics. Illumination invariants can be derived from some of its recursive statistics and exploited in content based image retrieval, supervised or unsupervised image recognition. Its modeling efficiency is demonstrated on several analytical and modeling image applications, in particular on unsupervised image or range data segmentation, bidirectional texture function (BTF) synthesis and compression, dynamic texture synthesis and adaptive multispectral and multichannel image and video restoration.

14.1 Introduction

Recognition and processing of multi-dimensional data (or set of spatially related objects) is more accurate and efficient if we take into account all interdependencies between single objects. Objects to be processed like for example multi-spectral pixels in a digitized image, are often mutually dependent (e.g., correlated) with a dependency degree related to a distance between two objects in their corresponding data space. These relations can be incorporated into a pattern recognition process through appropriate multi-dimensional data model. If such a model is probabilistic we can use consistent Bayesian framework for solving many pattern recognition tasks.

Michal Haindl
Institute of Information Theory and Automation of the ASCR, Prague, Czech Republic
e-mail: haindl@utia.cz

Features derived from multi-dimensional data models are information preserving in the sense that they can be used to synthesise data spaces closely resembling original measurement data space as can be illustrated on the recent best visual representation of real material surfaces in the form of bidirectional texture function [220]. Virtual or augmented reality systems require object surfaces covered with realistic nature-like color textures to enhance realism in virtual scenes. Similarly, realistic textures are used in computer games, CAD systems and some other computer graphics applications. Such textures can be either digitized natural textures or textures synthesized from an appropriate mathematical model. However digitized 3D textures are far less convenient alternative, because of extreme virtual system memory demands, visible discontinuities and several other drawbacks [214].

Mathematical multi-dimensional data models are useful for describing many of the multi-dimensional data types provided that we can assume some data homogeneity so some data characteristics are translation invariant. While the 1D models like time series are relatively well researched and they have rich application history in control theory, econometric, medicine and many other recognition applications, multi-dimensional models are much less known and their applications are still limited. The reason is not only unsolved theory difficulties but mainly their huge computing power demands which prevented their wider use until recently.

We introduced in [223] a fast multiresolution Markov random field (MRF) based model and the simultaneous causal autoregressive random field model [222], respectively. Although the former method avoids the time consuming Markov chain Monte Carlo simulation so typical for applications of Markov models it requires several approximations. The later method is very efficient for multispectral image representation not only because it does not suffer with some problems of alternative options (see [214, 217] for details) but it is also easy to analyze as well as to synthesise and last but not least it is still flexible enough to imitate a large set of natural and artificial textures or other spatial data.

It is possible to divide data models applications into two broad categories: synthesis and analysis. Analytical applications include data classifications or unsupervised segmentation, data space directionality analysis, motion detection and some others. Frequent synthesis applications are missing data reconstruction, restoration, image compression and static or dynamic texture synthesis.

In the application Sections 14.4 and further we demonstrate advantages and weak points of the studied Markovian model on several multispectral image recognition and modeling examples.

14.2 3D Causal Simultaneous Autoregressive Model

Modeling visual data requires non-standard multi-dimensional (three-dimensional for static color textures, 4D for videos or even 7D for static BTFs) models. However if such a nD data space can be factorized then these data can be also approximated using a set of lower-dimensional probabilistic models. Although full nD models allow unrestricted spatial-spectral-temporal-angular correlation modeling their main drawback is large amount of parameters to be estimated and in the case of some models (e.g. Markov models) also the necessity to estimate all these parameters simultaneously. The 3D causal simultaneous autoregressive model (3DCAR) is an exceptional model which can be utilized to build much more complex nD data models. For example, the 7D BTF models illustrated in Fig.14.5 are composed from up to one hundred 3DCARs.

A digitized image Y is assumed to be defined on a finite rectangular $N \times M \times d$ lattice I , $r = \{r_1, r_2, r_3\} \in I$ denotes a pixel multiindex with the row, columns and spectral indices, respectively. The notation \bullet has the meaning of all possible values of the corresponding index and $I_r^c \subset I$ is a causal or unilateral neighbourhood of pixel r , i.e.

$$I_r^c = \{s : 1 \leq s_1 \leq r_1, 1 \leq s_2 \leq r_2, s \neq r\} .$$

The 3D causal simultaneous autoregressive model (3DCAR) is the wide-sense Markov model which can be written in the following regression equation form:

$$\tilde{Y}_r = \sum_{s \in I_r^c} A_s \tilde{Y}_{r-s} + e_r \quad \forall r \in I \quad (14.1)$$

where A_s are matrices (14.2) and the zero mean white Gaussian noise vector e_r has uncorrelated components with data indexed from I_r^c but noise vector components can be mutually correlated.

$$A_{s_1, s_2} = \begin{pmatrix} a_{1,1}^{s_1, s_2} & \dots & a_{1,d}^{s_1, s_2} \\ \vdots & \ddots & \vdots \\ a_{d,1}^{s_1, s_2} & \dots & a_{d,d}^{s_1, s_2} \end{pmatrix} \quad (14.2)$$

are $d \times d$ parameter matrices, The model can be expressed in the matrix form (14.51) where

$$X_r = [\tilde{Y}_{r-s}^T : \forall s \in I_r^c] , \quad (14.3)$$

X_r is a $d\eta \times 1$ vector, $\eta = \text{card}(I_r^c)$ and γ

$$\gamma = [A_1, \dots, A_\eta] \quad (14.4)$$

is a $d \times d\eta$ parameter matrix. To simplify notation the multiindexes r, s, \dots have only two components further on in this section.

An optimal support can be selected as the most probable model given past data, i.e., $\max_j \{p(M_j | Y^{(r-1)})\}$.

$$p(Y^{(r-1)} | M_j) = \int \int p(Y^{(r-1)} | \gamma, \Sigma^{-1}) p(\gamma, \Sigma^{-1} | M_j) d\gamma d\Sigma^{-1} \quad (14.5)$$

and for implemented uniform priors start we get a decision rule [228]:

Theorem 14.1. *The most probable AR model given past data $Y^{(r-1)}$, the normal-Wishart parameter prior and the uniform model prior is the model M_i for which*

$$i = \arg \max_j \{D_j\}$$

$$D_j = -\frac{d}{2} \ln |V_{x(r-1)}| - \frac{\beta(r) - d\eta + d + 1}{2} \ln |\lambda_{(r-1)}| + \frac{d^2 \eta}{2} \ln \pi + \sum_{i=1}^d \left[\ln \Gamma \left(\frac{\beta(r) - d\eta + d + 2 - i}{2} \right) - \ln \Gamma \left(\frac{\beta(0) - d\eta + d + 2 - i}{2} \right) \right] \quad (14.6)$$

where $V_{x(r-1)} = \tilde{V}_{x(r-1)} + V_{x(0)}$ with $\tilde{V}_{x(r-1)}$ defined in (14.12), $V_{x(0)}$ is an appropriate part of V_0 (14.13), $\beta(r)$ is defined in (14.7), (14.8) and $\lambda_{(r-1)}$ is (14.9).

Proof. [213] □

$$\beta(r) = \beta(0) + r - 1 = \beta(r-1) + 1, \quad (14.7)$$

$$\beta(0) > \eta - 2, \quad (14.8)$$

and

$$\lambda_{(r)} = V_{y(r)} - V_{xy(r)}^T V_{x(r)}^{-1} V_{xy(r)}. \quad (14.9)$$

$$V_{r-1} = \tilde{V}_{r-1} + V_0, \quad (14.10)$$

$$\tilde{V}_{r-1} = \begin{pmatrix} \tilde{V}_{y(r-1)} & \tilde{V}_{xy(r-1)}^T \\ r\tilde{V}_{xy(r-1)} & \tilde{V}_{x(r-1)} \end{pmatrix}, \quad (14.11)$$

$$\tilde{V}_{y(r-1)} = \sum_{k=1}^{r-1} Y_k Y_k^T, \quad (14.12)$$

$$\tilde{V}_{xy(r-1)} = \sum_{k=1}^{r-1} X_k Y_k^T, \quad (14.13)$$

$$\tilde{V}_{x(r-1)} = \sum_{k=1}^{r-1} X_k X_k^T. \quad (14.14)$$

Marginal densities $p(\gamma|Y^{(r-1)})$ and $p(\Sigma^{-1}|Y^{(r-1)})$ can be evaluated from (14.15), (14.16), respectively.

$$p(\gamma|Y^{(r-1)}) = \int p(\gamma, \Sigma^{-1}|Y^{(r-1)}) d\Sigma^{-1} \quad (14.15)$$

$$p(\Sigma^{-1}|Y^{(r-1)}) = \int p(\gamma, \Sigma^{-1}|Y^{(r-1)}) d\gamma \quad (14.16)$$

The marginal density $p(\Sigma^{-1}|Y^{(r-1)})$ is the Wishart distribution density [213]

$$p(\Sigma^{-1}|Y^{(r-1)}) = \frac{\pi^{\frac{d(1-d)}{4}} |\Sigma^{-1}|^{\frac{\beta(r)-d\eta}{2}}}{2^{\frac{d(\beta(r)-d\eta+d+1)}{2}} \prod_{i=1}^d \Gamma(\frac{\beta(r)-d\eta+2+d-i}{2})} |\lambda_{(r-1)}|^{\frac{\beta(r)-d\eta+d+1}{2}} \exp\left\{-\frac{1}{2} tr\{\Sigma^{-1}\lambda_{(r-1)}\}\right\} \quad (14.17)$$

with

$$E\left\{\Sigma^{-1}|Y^{(r-1)}\right\} = (\beta(r) - d\eta + d + 1) \lambda_{(r-1)}^{-1} \quad (14.18)$$

$$E\left\{(\Sigma^{-1} - E\{\Sigma^{-1}|Y^{(r-1)}\})^T (\Sigma^{-1} - E\{\Sigma^{-1}|Y^{(r-1)}\}) | Y^{(r-1)}\right\} = \frac{2(\beta(r) - d\eta + 1)}{\lambda_{(r-1)} \lambda_{(r-1)}^T}. \quad (14.19)$$

The marginal density $p(\gamma|Y^{(r-1)})$ is matrix t distribution density [213]:

$$p(\gamma|Y^{(r-1)}) = \frac{\prod_{i=1}^d \Gamma\left(\frac{\beta(r)+d+2-i}{2}\right)}{\prod_{i=1}^d \Gamma\left(\frac{\beta(r)-d\eta+d+2-i}{2}\right)} \pi^{-\frac{d^2\eta}{2}} |\lambda_{(r-1)}|^{-\frac{d\eta}{2}} |V_{x(r-1)}|^{\frac{d}{2}} \\ \left| I + \lambda_{(r-1)}^{-1} (\gamma - \hat{\gamma}_{r-1}) V_{x(r-1)} (\gamma - \hat{\gamma}_{r-1})^T \right|^{-\frac{\beta(r)+d+1}{2}} \quad (14.20)$$

with the mean value

$$E \left\{ \gamma | Y^{(r-1)} \right\} = \hat{\gamma}_{r-1} \quad (14.21)$$

and covariance matrix

$$E \left\{ (\gamma - \hat{\gamma}_{r-1})^T (\gamma - \hat{\gamma}_{r-1}) | Y^{(r-1)} \right\} = \frac{V_{x(r-1)}^{-1} \lambda_{(r-1)}}{\beta(r) - d\eta} . \quad (14.22)$$

Similar statistics can be easily derived [213] for the alternative Jeffreys non-informative parameter prior.

Theorem 14.2. *The one-step-ahead predictive posterior density for the normal-Wishart parameter prior has the form of d -dimensional Student's probability density (14.23)*

$$p(Y_r | Y^{(r-1)}) = \frac{\Gamma\left(\frac{\beta(r)-d\eta+d+2}{2}\right)}{\Gamma\left(\frac{\beta(r)-d\eta+2}{2}\right) \pi^{\frac{d}{2}} (1 + X_r^T V_{x(r-1)}^{-1} X_r)^{\frac{d}{2}} |\lambda_{(r-1)}|^{\frac{1}{2}}} \\ \left(1 + \frac{(Y_r - \hat{\gamma}_{r-1} X_r)^T \lambda_{(r-1)}^{-1} (Y_r - \hat{\gamma}_{r-1} X_r)}{1 + X_r^T V_{x(r-1)}^{-1} X_r} \right)^{-\frac{\beta(r)-d\eta+d+2}{2}} , \quad (14.23)$$

with $\beta(r) - d\eta + 2$ degrees of freedom, if $\beta(r) > d\eta$ then the conditional mean value is

$$E \left\{ Y_r | Y^{(r-1)} \right\} = \hat{\gamma}_{r-1} X_r , \quad (14.24)$$

and

$$E \left\{ (Y_r - \hat{Y}_{r-1} X_r)(Y_r - \hat{Y}_{r-1} X_r)^T | Y^{(r-1)} \right\} = \frac{1 + X_r V_{x(r-1)}^{-1} X_r^T}{(\beta(r) - d\eta)} \lambda_{(r-1)} . \quad (14.25)$$

Proof. [213]

□

14.2.1 Adaptivity

The 3DCAR model can be made adaptive if we modify its recursive statistics using exponential forgetting factor, i.e. a constant $\varphi \approx 0.99$. This forgetting factor smaller than 1 is used to weight the influence of older data:

$$\begin{aligned} \hat{Y}_r^T &= \hat{Y}_{r-1}^T + (\varphi^2 + X_r^T V_{x(r-1)}^{-1} X_r)^{-1} V_{x(r-1)}^{-1} X_r (Y_r - \hat{Y}_{r-1} X_r)^T , \\ |V_{x(t)}| &= |V_{x(t-1)}| \varphi^{2\eta} (1 + X_t^T V_{x(t-1)}^{-1} X_t) , \\ \lambda_t &= \lambda_{t-1} (1 + (Y_t - \hat{Y}_{t-1}^T X_t)^T \lambda_{t-1}^{-1} (Y_t - \hat{Y}_{t-1}^T X_t) (\varphi^2 + X_t^T V_{x(t-1)}^{-1} X_t)^{-1}) . \end{aligned}$$

14.2.2 Numerical Stability

The numerical stability of 3DCAR can be quarantined if all its recursive statistics use the square-root factor updating applying either the Cholesky or LDL^T decomposition [216], respectively. Let us denote a lower triangular matrix L_t and a matrix B_t as

$$B_t = V_{i,t}^{-1} = L_t L_t^T , \quad (14.26)$$

$$B_t = (B_{t-1}^{-1} \pm d_t d_t^T)^{-1} , \quad (14.27)$$

$$B_t = (\tilde{B}_{t-1}^{-1} \varphi^{-2} \pm \alpha^2 \tilde{d}_t \tilde{d}_t^T)^{-1} , \quad (14.28)$$

where d_t is an updating vector with exponential forgetting φ and data normalization α . The square-root updating factor of the inversion data gathering matrix B_t can be computed recursively:

$$L_{t,i,j} = \frac{\tilde{\zeta}_{t,j+1}}{\varphi \tilde{\zeta}_{t,j}} \left[L_{t-1,i,j} \mp \frac{\tilde{f}_{t,j}}{\tilde{\zeta}_{t,j+1}^2} \tilde{g}_{t,j+1}^{(i)} \right], \quad (14.29)$$

where $\tilde{f}_t = L_{t-1}^T d_t$ and

$$\tilde{g}_{t,j+1}^{(i)} = \sum_{k=j+1}^i L_{t-1,i,k} \tilde{f}_{t,k}$$

and

$$\zeta_{t,i} = \frac{\alpha}{\varphi} \sqrt{\frac{\varphi^2}{\alpha^2} \pm \sum_{j=i}^n \tilde{f}_{t,j}^2} = \frac{\alpha}{\varphi} \tilde{\zeta}_{t,i} .$$

14.2.3 3DCAR Model Properties

The 3DCAR (analogously also the 2DCAR model) model has advantages in analytical solutions (Bayes, ML, or LS estimates) for I_r , $\hat{\gamma}$, $\hat{\sigma}^2$, \hat{Y}_r statistics. It allows extremely simple fast synthesis, adaptivity and building efficient recursive application algorithms. Its major drawback in some image representation applications might be its mostly artificial causality which can introduce a directional bias into modelled image data.

14.3 Illumination Invariants

Textures are important clues to specify objects present in a visual scene. However, the appearance of natural textures is highly illumination and view angle dependent. As a consequence, most recent realistic texture based classification or segmentation methods require multiple training images [438] captured under all possible illumination and viewing conditions for each class. Such learning is obviously clumsy, probably expensive and very often even impossible if required measurements are not available.

If we assume fixed positions of viewpoint and illumination sources, uniform illumination sources and Lambertian surface reflectance, than two images \tilde{Y}, Y acquired with different illumination spectra can be linearly transformed to each other:

$$\tilde{Y}_r = B Y_r \quad \forall r . \quad (14.30)$$

It is possible to show that assuming (14.30) the following features are illumination invariant:

1. trace: $trace A_m, m = 1, \dots, \eta K$
2. eigenvalues: $v_{m,j}$ of $A_m, m = 1, \dots, \eta K, j = 1, \dots, C$
3. $1 + Z_r^T V_{zz}^{-1} Z_r$,
4. $\sqrt{\sum_r (Y_r - \hat{y}Z_r)^T \lambda^{-1} (Y_r - \hat{y}Z_r)}$,
5. $\sqrt{\sum_r (Y_r - \mu)^T \lambda^{-1} (Y_r - \mu)}$,
 μ is the mean value of vector Y_r ,

Above textural features derived from the 3DCAR model are robust to illumination direction changes, invariant to illumination brightness and spectrum changes, and simultaneously also robust to Gaussian noise degradation. This property was extensively verified on University of Bonn BTF texture measurements [325], where illumination sources are spanned over 75% of possible illumination half-sphere.

14.4 Unsupervised Image Recognition

Unsupervised or supervised texture segmentation is the prerequisite for successful content-based image retrieval, scene analysis, automatic acquisition of virtual models, quality control, security, medical applications and many others. Although more than 1000 different methods were already published [479], this problem is still far from being solved. This is among others due to missing reliable performance comparison between different techniques because very limited effort was spent [227] to develop suitable quantitative measures of segmentation quality that can be used to evaluate and compare segmentation algorithms. Spatial interaction models and especially Markov random field-based models are increasingly popular for texture representation [263, 372, 214], etc. Several researchers dealt with the difficult problem of unsupervised segmentation using these models see for example [341, 314, 10, 215], [225, 224, 226], or Chapter 15.

Our unsupervised segmenter is illustrated on a multiscale unsupervised automatic detection of potentially cancerous regions of interest containing fibroglandular tissue in digital screening mammography. The mammogram tissue textures are locally

represented by four causal multispectral random field models recursively evaluated for each pixel and several scales. The segmentation part of the algorithm is based on the underlying Gaussian mixture model and starts with an over segmented initial estimation which is adaptively modified until the optimal number of homogeneous mammogram segments is reached.

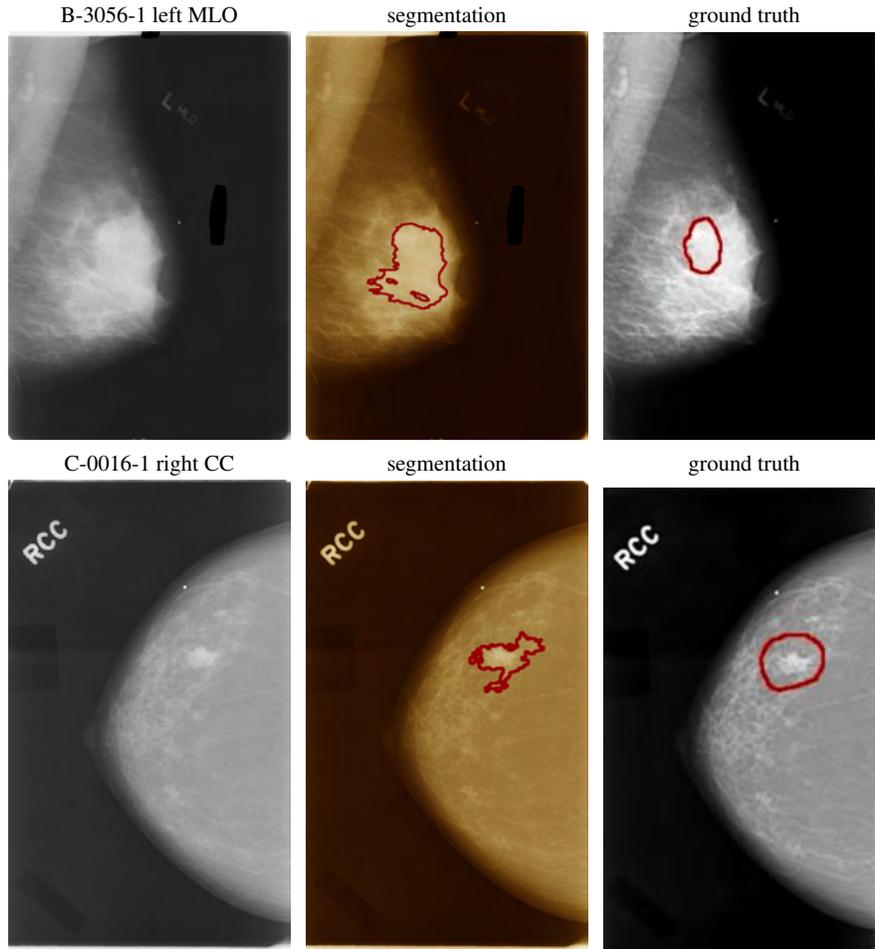


Fig. 14.1 Cancerous mammograms (patients age 58 (top) and 80 (bottom)), radiologist associated ground truth and detected regions of interest using the multiple segmenter approach, respectively.

Our method segments pseudo-color multiresolution mammograms each created from the original grey scale mammogram and its two nonlinear gamma transformations. We assume to down-sample input image Y into $M = 3$ different resolutions $Y^{(m)} = \downarrow^{l_m} Y$ with sampling factors $l_m \quad m = 1, \dots, M$ identical for both directions

and $Y^{(1)} = Y$. Local texture for each pixel $Y_r^{(m)}$ is represented using the 3D CAR model parameter space $\Theta_r^{(m)}$. The concept of decision fusion for high-performance pattern recognition is well known and widely accepted in the area of supervised classification where (often very diverse) classification technologies, each providing complementary sources of information about class membership, can be integrated to provide more accurate, robust and reliable classification decisions than the single classifier applications. The proposed method circumvents the problem of multiple unsupervised segmenters combination [226] by fusing multiple-processed measurements into a single segmenter feature vector.

Smooth pseudo-color mammogram textures require three dimensional models for adequate representation. We assume that single multi spectral texture can be locally modelled using a 3D simultaneous causal autoregressive random field model (3DCAR). This model can be expressed as a stationary causal uncorrelated noise driven 3D autoregressive process [228]:

$$Y_r = \gamma X_r + e_r \quad , \quad (14.31)$$

where $\gamma = [A_1, \dots, A_\eta]$ is the $3 \times 3\eta$ parameter matrix, e_r is a white Gaussian noise vector with zero mean and a constant but unknown variance, X_r is a corresponding vector of the contextual neighbors Y_{r-s} and $r, r-1, \dots$ is a chosen direction of movement on the image index lattice I . The optimal neighborhood (I_r^c) as well as the Bayesian parameters estimation of a 3DCAR model can be found analytically under few additional and acceptable assumptions using the Bayesian approach (14.6). The recursive Bayesian parameter estimation of the 3DCAR model is [228]:

$$\hat{\gamma}_{r-1}^T = \hat{\gamma}_{r-2}^T + \frac{V_{x(r-2)}^{-1} X_{r-1} (Y_{r-1} - \hat{\gamma}_{r-2} X_{r-1})^T}{(1 + X_{r-1}^T V_{x(r-2)}^{-1} X_{r-1})} \quad , \quad (14.32)$$

where $V_{x(r-1)} = \sum_{k=1}^{r-1} X_k X_k^T + V_{x(0)}$. Each matrix contains local estimations of the 3DCAR model parameters. These models have identical contextual neighborhood I_r^c but they differ in their major movement direction (top-down, bottom-up, rightward, leftward). The local texture for each pixel and M resolutions $\alpha_1, \dots, \alpha_M$ is represented by four parametric matrices t, b, r, l e.g. $\hat{\gamma}_r^{i, \alpha_j}$ for $i \in \{t, b, r, l\}$, $j = 1, \dots, M$ which are subsequently compressed using the local PCA (for computational efficiency) into $\tilde{\gamma}_r^{i, \alpha_j}$. Single resolution compressed parameters are composed into M parametric matrices:

$$\tilde{\gamma}_r^{\alpha_j T} = \{ \tilde{\gamma}_r^{t, \alpha_j}, \tilde{\gamma}_r^{b, \alpha_j}, \tilde{\gamma}_r^{r, \alpha_j}, \tilde{\gamma}_r^{l, \alpha_j} \}^T \quad j = 1, \dots, M \quad .$$

The parametric space $\tilde{\gamma}^{\alpha_j}$ is subsequently smooth out, rearranged into a vector and its dimensionality is reduced using the PCA feature extraction ($\tilde{\gamma}^{\alpha_j}$). Finally we add the average local spectral values $\zeta_r^{\alpha_j}$ to the resulting feature vector:

$$\Theta_r = [\bar{\gamma}_r^{\alpha_1}, \zeta_r^{\alpha_1}, \dots, \bar{\gamma}_r^{\alpha_M}, \zeta_r^{\alpha_M}]^T . \quad (14.33)$$

Rough scale pixels parameters are simply mapped to the corresponding fine scale locations.

Multi-spectral, multiresolution texture segmentation is done by clustering in the combined 3DCAR models parameter space Θ defined on the lattice I where Θ_r is the modified parameter vector (14.33) computed for the lattice location r . We assume that this parametric space can be represented using the Gaussian mixture model (GM) with diagonal covariance matrices due to the previous 3DCAR parametric space decorrelation. The Gaussian mixture model for 3DCAR parametric representation is as follows:

$$p(\Theta_r) = \sum_{i=1}^K p_i p(\Theta_r | v_i, \Sigma_i) , \quad (14.34)$$

$$p(\Theta_r | v_i, \Sigma_i) = \frac{|\Sigma_i|^{-\frac{1}{2}}}{(2\pi)^{\frac{d}{2}}} e^{-\frac{(\Theta_r - v_i)^T \Sigma_i^{-1} (\Theta_r - v_i)}{2}} . \quad (14.35)$$

The mixture model equations (14.34),(14.35) are solved using a modified EM algorithm. The algorithm is initialized using v_i, Σ_i statistics estimated from the corresponding regions obtained by regular division of the input detected breast area. An alternative initialisation can be random choice of these statistics. For each possible couple of regions the Kullback Leibler divergence

$$D(p(\Theta_r | v_i, \Sigma_i) || p(\Theta_r | v_j, \Sigma_j)) = \int_{\Omega} p(\Theta_r | v_i, \Sigma_i) \log \left(\frac{p(\Theta_r | v_i, \Sigma_i)}{p(\Theta_r | v_j, \Sigma_j)} \right) d\Theta_r \quad (14.36)$$

is evaluated and the most similar regions, i.e.,

$$\{i, j\} = \arg \min_{k,l} D(p(\Theta_r | v_l, \Sigma_l) || p(\Theta_r | v_k, \Sigma_k))$$

are merged together in each step. This initialisation results in K_{ini} subimages and recomputed statistics v_i, Σ_i . $K_{ini} > K$ where K is the optimal number of textured segments to be found by the algorithm. Two steps of the EM algorithm are repeating after the initialisation. The components with smaller weights than a fixed threshold ($p_j < \frac{0.01}{K_{ini}}$) are eliminated. For every pair of components we estimate their Kullback Leibler divergence (14.36). From the most similar couple, the component with the weight smaller than the threshold is merged to its stronger partner and all statistics are actualized using the EM algorithm. The algorithm stops when

either the likelihood function has negligible increase ($\mathcal{L}_t - \mathcal{L}_{t-1} < 0.01$) or the maximum iteration number threshold is reached.

The parametric vectors representing texture mosaic pixels are assigned to the clusters according to the highest component probabilities, i.e., Y_r is assigned to the cluster ω_{j^*} if

$$\pi_{r,j^*} = \max_j \sum_{s \in I_r} w_s p(\Theta_{r-s} | \mathbf{v}_j, \Sigma_j) ,$$

where w_s are fixed distance-based weights, I_r is a rectangular neighbourhood and $\pi_{r,j^*} > \pi_{thre}$ (otherwise the pixel is unclassified). The area of single cluster blobs is evaluated in the post-processing thematic map filtration step. Regions with similar statistics are merged. Thematic map blobs with area smaller than a given threshold are attached to its neighbour with the highest similarity value. Finally, regions which have grey level mean value difference from the median mean value (over the same type of digitized mammograms) of cancerous ground truth regions larger than a specified threshold are eliminated.

14.5 Multispectral Image Restoration

Physical imaging systems and a recording medium are imperfect and thus a recorded image represents a degraded version of the original scene. Similarly an image is usually further corrupted during its processing, transmission or storage. The image restoration task is to recover an unobservable image given the observed corrupted image with respect to some statistical criterion. Image restoration is the busy research area for already several decades and many restoration algorithms have been proposed [9, 192, 3], see also Chapters 16, 2, 3 and 4.



Fig. 14.2 Original, corrupted, and reconstructed Cymbidium image.

The image degradation is supposed to be approximated by the linear degradation model:

$$X_r = \sum_{s \in I_r} f_s Y_{r-s} + e_r \quad (14.37)$$

where f is a discrete representation of the unknown point-spread function. The point-spread function can be non-homogeneous but we assume its slow changes relative to the size of an image. I_r is some contextual support set, and the degradation noise e is uncorrelated with the unobservable image, i.e.,

$$E\{Y e\} = 0 \quad (14.38)$$

The point-spread function is unknown but such that we can assume the unobservable image Y to be reasonably well approximated by the expectation of the corrupted image

$$\hat{Y} = E\{X\} \quad (14.39)$$

in regions with gradual pixel value changes. Pixels with steep step discontinuities are left unrestored to avoid excessive blurring, i.e.,

$$\hat{Y}_r = \begin{cases} E\{X_r\} & \text{if } |E\{X_r\} - X_r| < \frac{1}{n_s} \sum_s |E\{X_{r-s}\} - X_{r-s}| \\ X_r & \text{otherwise} \end{cases} \quad (14.40)$$

The expectation (14.39) can be expressed as follows:

$$E\{X\} = \int X p(X) dX = \int \begin{pmatrix} X_1 & X_2 & \dots & X_M \\ X_{M+1} & X_{M+2} & \dots & X_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ X_{NM-M+1} & X_{NM-M+2} & \dots & X_{NM} \end{pmatrix} \prod_{r=1}^{NM} p(X_r | X^{(r-1)}) dX_1 \dots dX_{NM} \quad (14.41)$$

where

$$X^{(r-1)} = \{X_{r-1}, \dots, X_1\} \quad (14.42)$$

is a set of noisy pixels in some chosen but fixed ordering. For single matrix elements in (14.41) it holds

$$\begin{aligned}
E\{X_j\} &= \int X_j \prod_{r=1}^{NM} p(X_r | X^{(r-1)}) dX_1 \dots dX_{NM} \\
&= \int X_j \prod_{r=1}^j p(X_r | X^{(r-1)}) dX_1 \dots dX_j \\
&= \int E\{X_j | X^{(j-1)}\} \prod_{r=1}^{j-1} p(X_r | X^{(r-1)}) dX_1 \dots dX_{j-1} \\
&= E\{E\{X_j | X^{(j-1)}\}\} \tag{14.43}
\end{aligned}$$

Let us approximate after having observed $X^{(j-1)}$ the mean value $\hat{Y}_j = E\{X_j\}$ by the $E\{X_j | X^{(j-1)} = x^{(j-1)}\}$ where $x^{(j-1)}$ are known past realisation for j . Thus we suppose that all other possible realisations $x^{(j-1)}$ than the true past pixel values have negligible probabilities. This assumption implies conditional expectations approximately equal to unconditional ones, i.e.,

$$E\{X_j\} \approx E\{X_j | X^{(j-1)}\} , \tag{14.44}$$

and

$$\hat{Y} = E\{X\} \approx$$

$$\begin{pmatrix} E\{X_1|X^{(0)}\} & E\{X_2|X^{(1)}\} & \dots & E\{X_M|X^{(M-1)}\} \\ E\{X_{M+1}|X^{(M)}\} & E\{X_{M+2}|X^{(M+1)}\} & \dots & E\{X_{2M}|X^{(2M-1)}\} \\ \vdots & \vdots & \ddots & \vdots \\ E\{X_{NM-M+1}|X^{(NM-M)}\} & E\{X_{NM-M+2}|X^{(NM-M+1)}\} & \dots & E\{X_{NM}|X^{(NM-1)}\} \end{pmatrix}.$$

Suppose also that the noisy image X can be represented by a causal simultaneous autoregressive model (14.1), then the conditional mean (14.44) values needed for the estimation \hat{Y} is (14.52) if we replace in above equations $X \rightarrow Y, Z \rightarrow X$. The estimator (14.44) can be efficiently computed using the following recursion

$$\hat{\gamma}_r^T = \hat{\gamma}_{r-1}^T + (1 + Z_r^T V_{z(r-1)}^{-1} Z_r)^{-1} V_{z(r-1)}^{-1} Z_r (X_r - \hat{\gamma}_{r-1} Z_r)^T. \quad (14.45)$$

The selection of an appropriate model support (I_r^c) is important to obtain good restoration results. The optimal Bayesian decision rule for this selection is either (14.6) or analogous statistics depending on the parameter prior.

14.5.0.1 Local Estimation of the Point-Spread Function

If we assume a non-homogeneous slowly changing point-spread function, we can estimate its local value using the local least square estimate

$$\hat{\psi}_r = \min_{\psi_r} \left\{ \sum_{\forall r \in J_r} (X_r - \psi_r \hat{W}_r)^2 \right\}. \quad (14.46)$$

The locally optimal estimate is

$$\hat{\psi}_r^T = \tilde{V}_{\hat{W}(r)}^{-1} \tilde{V}_{\hat{W}X(r)} \quad (14.47)$$

where $\tilde{V}_{\hat{W}(r)}, \tilde{V}_{\hat{W}X(r)}$ are corresponding local data gathering matrices analogous to (14.14), (14.12), but using only data from local sub-lattice $J_r \subset I, r \in J_r$. This estimator can be efficiently evaluated using the fast recursive square-root filter introduced in section 14.2.2. If the point-spread function is constant for all lattice position both PSF estimators (local and global) are equivalent.

14.6 Multichannel Image Restoration

The major degradation of a ground-based telescope is caused by random fluctuations originating mostly in the Earth's atmosphere (*seeing*) along the optical path between the object space and the image formation device. The image degradation by seeing is a very complicated process due to blurring, motion, and distortion. The image degradation is described by the changing complex point-spread-function (PSF) of the telescope, which embodies all the important behaviour of the optical image formation system. For the restoration we assume one unknown degradation function involving all degradation aspects.

Suppose Y represents a true but unobservable monospectral image defined on the finite rectangular $N \times M$ underlying lattice I . Suppose further that we have a set of d observable images \mathcal{X} where each $X_{\bullet,i} \in \mathcal{X}$ is the i -th version of Y distorted by the unknown PSF and noise independent of the signal. The notation \bullet designates of all possible values of the corresponding multiindex (e.g. the multiindex $r = \{r_1, r_2\}$ which has the row and columns indices, respectively). We assume knowledge of all pixels from the reconstructed scene. For the treatment of the more difficult problem when some data are missing see [229], [230]. The image degradation is supposed to be approximated by the linear discrete spatial domain degradation model

$$X_{r,\bullet} = \sum_{s \in I_r} H_s Y_{r-s} + \varepsilon_{r,\bullet} \quad (14.48)$$

where H is a discrete representation of the unknown point-spread function, $X_{r,\bullet}$ is the $d \times 1$ vector of the r -th pixel in different distortions and Y_{r-s} are ideal (unobservable) image pixels. The point-spread function is assumed to be either homogeneous or it can be non-homogeneous but in this case we assume it slowly changes relative to the size of an image. I_r is some contextual support set, and a noise vector ε is uncorrelated with the true image, i.e., $E\{Y \varepsilon_{\bullet,i}\} = 0$. The point-spread function is unknown but such that we can assume the unobservable image Y to be reasonably well approximated by the expectation of the corrupted image

$$\hat{Y} = E\{X_{\bullet,i}\} \quad (14.49)$$

in regions with gradual pixel value changes, and the i -th degraded image $X_{\bullet,i} \in \mathcal{X}$ is the least degraded image from the set \mathcal{X} . The index i of the least degraded image is excluded from the following equations (14.50)-(14.52) to simplify the corresponding notation. The above method (14.49) changes all pixels in the restored image and thus blurs discontinuities present in the scene although to much less extent than the classical restoration methods due to our restoration model (14.51) adaptivity. This excessive blurring can be avoided if pixels with steep step discontinuities are left unrestored, i.e.,

$$\hat{Y}_r = \begin{cases} E\{X_r\} & \text{if } p(X_r|X^{(r-1)}) > \kappa \\ X_r & \text{otherwise} \end{cases}, \quad (14.50)$$

where κ is a probabilistic threshold based on the prediction density. Single matrix elements in the expectation $E\{X\}$ are approximated [231] by the conditional expectation $E\{X_j|X^{(j-1)} = x^{(j-1)}\}$ where $x^{(j-1)}$ are known past realisation for j . Thus we suppose that all other possible realisation $x^{(j-1)}$ than the true past pixel values have negligible probabilities. This assumption implies conditional expectations approximately equal to unconditional ones, i.e., then the expectation is $E\{X_j\} \approx E\{X_j|X^{(j-1)}\}$.

Suppose further that a noisy image can be represented by an adaptive 2.5D causal simultaneous autoregressive model

$$X_{r,i} = \gamma Z_r + \varepsilon_r, \quad (14.51)$$

where $\gamma = [A_1, \dots, A_\eta]$, $\eta = \text{card}(I_r^c)$ is a $1 \times d\eta$ parameter matrix, Z_r is a corresponding vector of X_{r-s} , ε_r is a white Gaussian noise vector with zero mean, and a constant but unknown covariance matrix Σ . The noise vector is uncorrelated with data from a causal neighbourhood I_r^c . $A_s = [a_{s,1}, \dots, a_{s,d}] \quad \forall s$ are parameter vectors. The model adaptivity is introduced using the exponential forgetting factor technique in parameter learning part of the algorithm. The conditional mean value can be derived under few acceptable conditions [231] in the following form:

$$E\{X_r|X^{(r-1)}\} = V_{zx(r-1)}^T V_{zz(r-1)}^{-T} Z_r \quad (14.52)$$

where $V_{r-1} = \tilde{V}_{r-1} + I$ and

$$\tilde{V}_{r-1} = \begin{pmatrix} \sum_{k=1}^{r-1} X_k X_k^T & \sum_{k=1}^{r-1} X_k Z_k^T \\ \sum_{k=1}^{r-1} Z_k X_k^T & \sum_{k=1}^{r-1} Z_k Z_k^T \end{pmatrix} = \begin{pmatrix} \tilde{V}_{xx(r-1)} & \tilde{V}_{zx(r-1)}^T \\ \tilde{V}_{zx(r-1)} & \tilde{V}_{zz(r-1)} \end{pmatrix}.$$

An appropriate model support (I_r^c) can be found using the Bayesian decision rule (cf. [231]).

The proposed recursive multitemporal blur minimizing reconstruction method is very fast (approximately five times faster than the median filter) robust and its reconstruction results surpasses some standard reconstruction methods, which we were able to implement for the verification. Our causal model has the advantage to have the analytical solution for all needed model statistics. Possible artifacts introduced by this type of models are diminished by introducing **adaptivity** into the model. This novel formulation allow us to obtain extremely fast adaptive multichannel /

multitemporal restoration and it can be easily paralleled as well as generalized for multispectral (e.g. color, multispectral satellite images) or registered images which is seldom the case for alternative methods.

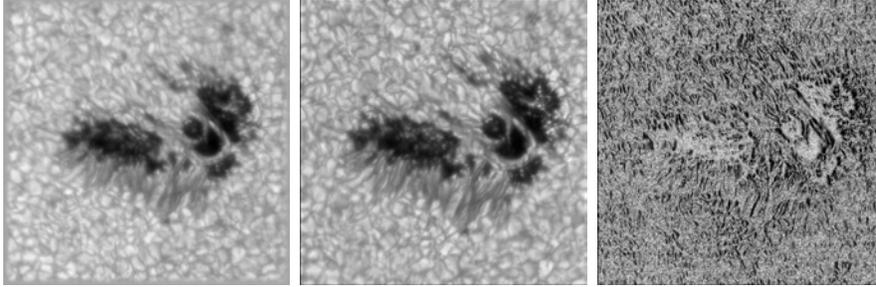


Fig. 14.3 The measured degraded (left), reconstructed sunspot image using our method, and its corresponding prediction probability image.

14.7 Video Restoration

Every movie deteriorates with usage and time irrespective of any care it gets. Movies (on both optical and magnetic materials) suffer with blotches, dirt, sparkles and noise, scratches, missing or heavily corrupted frames, mold, flickering, jittering, image vibrations and some other problems. For each kind of the defect usually different kind of restoration algorithm is needed. The scratch notion in this section means every coherent region with missing data (simultaneously in all spectral bands) in a color movie frame. Our method [218] reconstructs missing multispectral (e.g., color) pixels from available data in neighboring frames and pixels from the corrupted frame as well. A digitized color movie is supposed to be represented with the 3.5D causal AR model (see section 14.2):

$$Y_{r_1, r_2, \bullet, r_4} = \gamma X_{r_1, r_2, \bullet, r_4} + e_{r_1, r_2, \bullet, r_4} \quad \forall r \in I . \quad (14.53)$$

The missing scratch data are reconstructed from the topologically nearest known data in the lattice I using temporal and spatial correlation in the neighbourhood. Scratch pixels are computed from the set of one-step-ahead predictions using the conditional mean predictor

$$\tilde{Y}_r = E \left\{ Y_{r_1, r_2, \bullet, r_4} | Y^{(r-1)} \right\} = \hat{\gamma}_{r-1} X_{r_1, r_2, \bullet, r_4} , \quad (14.54)$$

where

$$Y^{(r-1)} = \{Y_{r-1}, Y_{r-2}, \dots, Y_1\}$$

is the known process history and $\hat{\gamma}_{r-1}$ is the estimator of unknown model parameter matrix γ (see section 14.2).

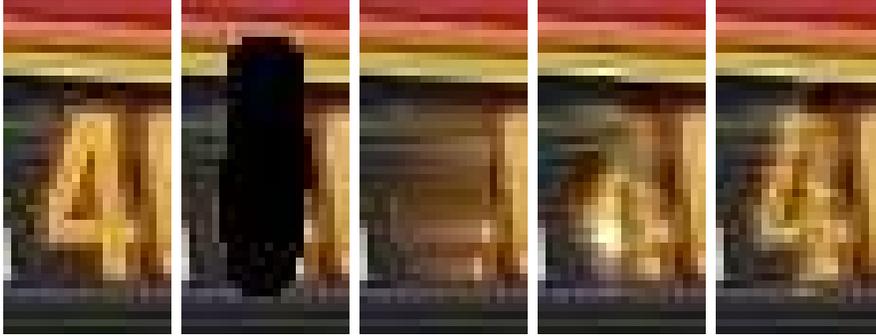


Fig. 14.4 A car frame restoration (original, scratch, quadratic interpolation method, 3D CAR, 3.5D CAR).

A model movement towards the scratch is assumed. When the model reaches the scratch, the corrupted pixel prediction is evaluated. This is performed for each line in the scratch from top and bottom edge of the scratch using two symmetrical downwards and upwards moving models and their results are averaged. This helps to counterbalance artificial restriction on the contextual neighbourhood which has to be causal. Similarly another couple of models is moving in the opposite direction. Two computed predictions for each missing pixel have to be combined. Simple averaging is not appropriate, because each of both predictors has different distance from the last known original data and consequently it has also a different precision. Hence the exponential interleaving was used to weight the data influence from each side of the scratch as a function of the horizontal position of the predicted pixel on a scratch line to be reconstructed.

14.8 Texture Synthesis and Compression

Texture synthesis methods may be divided primarily into intelligent sampling and model-based methods. Sampling approaches [96, 130, 129, 233, 472, 109, 474] rely on sophisticated sampling from real texture measurements while the model-based techniques [32, 33, 196, 214, 223, 222, 339, 480] describe texture data using mul-

tidimensional mathematical models and their synthesis is based on the estimated model parameters only.

There are several texture modeling approaches published [262, 223, 222] and some survey articles are also available [214, 217]. Most published texture models are restricted only to monospectral textures for few models developed for multispectral (mostly color) textures refer [32, 33, 223, 222].

The Bidirectional Texture Function (BTF) [219, 221, 220, 151] is the most advanced representation of visual properties for realistic real-world materials. BTF describes rough texture appearance for varying illumination and viewing conditions. Such a function can be represented by thousands of measurements (images) per material sample. The resulting BTF size excludes its direct rendering in graphical applications and some compression of these huge BTF data spaces is obviously inevitable. The BTF modeling ultimate aim is to create a visual impression of the same material without a pixel-wise correspondence to the original measurements. The cornerstone of our BTF compression and modeling method is the replacement of a huge number of original BTF measurements by their efficient parametric estimates derived from an underlying set of 2DCAR or 3DCAR spatial probabilistic models.



Fig. 14.5 BTF measurements mapped on part of a car gearbox. Original BTF data (enlarged using image tiling) (left) compared with synthesized BTF (right) for four distinct materials: *wood01*, *foil01*, *wood02*, *foil02* (3D model courtesy of DaimlerChrysler).

The off-line part of the algorithm [220] starts with the BTF illumination / view $(\theta_i, \phi_i / \theta_v, \phi_v)$ space segmentation into several subspace images using the K-means algorithm on color cumulative histograms features. Thus we trade off between an extreme compression ratio and the visual quality by using several probabilistic BTF subspace dedicated models. The overall roughness of a textured surface significantly influences the BTF texture appearance. Such a surface can be specified using its

range map, which is estimated by the photometric stereo approach. The subspace 2DCAR texture model starts with a spectral PCA-based decorrelation of subspace image (result of BTF segmentation) into mono-spectral factors. The 3DCAR model does not need this decorrelation step. Each of these factors is subsequently decomposed into sub-band components using the multi-resolution Gaussian-Laplacian pyramid. This allows to use simpler 2D/3D CAR models to model wide range of textures. Each such sub-band component is analyzed by a dedicated CAR factor model to obtain a compact set of model parameters.

The 3DCAR model [220, 151] offers a huge BTF compression ratio unattainable by any alternative sampling-based BTF synthesis method. Simultaneously this model can be used to reconstruct missing parts of the BTF measurement space.

14.9 Conclusion

The 3DCAR models are among rare exceptions in the Markovian model family that allow to derive extremely efficient and fast data processing algorithms. All their statistics can be evaluated recursively and they do not need any Monte Carlo sampling typical for other Markovian models. The 3DCAR models have the advantage over non causal (3DAR) ones that they can be treated analytically. It is possible to find analytical solution of model parameters, optimal model support, model predictor, etc. Similarly the 3DCAR model synthesis is very simple and a causal SAR RF can be directly generated from model equation. The disadvantage on the other hand is the causality which is usually rather artificially imposed on image data for algorithmic reasons and it is seldom supported by real image data. A causal model introduces an arbitrary directional bias, which depends on the orientation of a causal neighbourhood used. The CAR model can represent only wide-sense stationary data and data which have linear mutual relationships.

Acknowledgements This research was supported by the projects GAČR 102/08/0593, 102/07/1594 and partially by the MŠMT grants 1M0572 DAR, 2C06019.