Chapter 2
Multidimensional Image Models and Processing

Victor Krasheninnikov and Konstantin Vasil'ev

Abstract The problems of developing mathematical models and statistical algorithms for processing of multidimensional images and their sequences are presented in this chapter. Different types of random fields are taken for the basic mathematical image model. This implies two main problems associated with image modeling, namely, model analysis and synthesis. The main attention is paid to the correlation aspect, i.e. evaluation of the correlation function of a random field generated by a given model and, vice versa, development of a model generating a random field with a predetermined correlation function. For this purpose, new models (tensor and wave) and new versions of autoregressive models (with multiple roots) are suggested. The problems of image simulation on the curved surfaces are considered. The suggested models are used to synthesize the algorithms of multidimensional image processing and their sequences. The tensor filtration of imaging sequences and recursive filtration of multidimensional images, as well as the asymptotic characteristics of efficiency of random field filtration on grids of arbitrary dimension are suggested. The problem of object and anomaly detection on the background of interfering images is considered for the images of any dimension, e.g. for multi-zone data. It is shown that four equivalent forms of the optimal decision rule, which reflect various aspects of detection procedure, exist. Potential efficiency of anomaly detection is analyzed. The problems of alignment and estimation of parameters for interframe geometric image transformations are considered for multidimensional image sequences. A tensor procedure of simultaneous filtration of multidimensional image sequence and their interframe displacements are suggested. A method based on a fixed point of a complex geometric image transformation was investigated in order to evaluate large interframe displacements. Options for adaptive image processing algorithms are also discussed in this chapter. In this context, pseudo-gradient procedures are taken as a basis, as they do not require

V. Krasheninnikov (✉) · K. Vasil'ev
Ulyanovsk State Technical University, 32 Severni Venets Street,
Ulyanovsk 432027, Russian Federation
e-mail: kvrulstu@mail.ru

K. Vasil'ev
e-mail: 1948vasiliev@gmail.com

M.N. Favorskaya and L.C. Jain (eds.), *Computer Vision in Control Systems-3*,
Intelligent Systems Reference Library 135, https://doi.org/10.1007/978-3-319-67516-9_2

preliminary evaluation of any characteristics of the processed data. This allows to develop the high-performance algorithms that can be implemented in real-time systems.

Keywords Multidimensional image model · Autoregressive model
Tensor model · Wave model · Curved surface · Processing · Potential efficiency
Prediction · Filtration · Anomaly detection · Recognition · Adaptive algorithm
Pseudo-gradient algorithm

2.1 Introduction

Nowadays, the problems of automatic analysis of images and their sequences are becoming more and more important. This is due to the rapid development of the aerospace Earth monitoring systems, radio and sonar systems with spatial arrays, medical devices for early disease diagnosis, and computer vision systems. For some applications, images can be represented as data files set on a multidimensional index grid changing in discrete time. This allows to describe certain 2D or multidimensional images and their temporal sequences with scalar or vector values, e.g. sequence of TV images, multispectral images, spatial field of wind speeds, and others.

Mathematical models are necessary to formalize image processing problems. Random nature of image values determines the use of probability theory and mathematical statistics methods, i.e. image representation in the form of Random Fields (RFs) [1–5]. This implies the problems of the RF model analysis and synthesis. Composition of information and noise RF is usually used as a model of image observations. This model can also include an additional parameter vector, which allows to consider the peculiarities of image registration, e.g. possible mutual spatial displacement of adjacent frames and various abnormalities (object, signal) [6–9]. The reason for processing the sequences of observations can be development of estimations of an informational RF (prediction, filtration, and interpolation). Another important problem is the estimation of vector parameters, e.g. the image model parameters, spatial displacements, parameters of detected objects, and others.

This chapter presents the results obtained by the authors and their colleagues in modeling and statistical processing of 2D and multidimensional images and their sequences. A number of models [2, 3, 10, 11], including multidimensional ones [7, 12], are known to describe the images and their sequences [13, 14]. Section 2.2 covers the problems of the RF description and simulation. Random field is a set of random variables defined on multidimensional spatial grids with rectangular or more complex cells. For this purpose, the Gibbs, autoregressive, Tensor, and wave models are used [3, 12, 15, 16]. It is difficult to solve synthesis problems using autoregressive models, in particular, even modeling of spatially isotropic RF. However, the use of models with multiple roots of a characteristic equation facilitates to obtain approximately isotropic RF [16–19]. For linear transformations of

image frames, it is desirable to use tensor operations that provide the basis for tensor models, in which every recurrent frame is formed from the previous frame and the frame of random variables [16]. The suggested wave models permit to solve the analysis, synthesis, and simulation problems rather efficiently. In these models, the RF is a result of disturbance interaction, which can occur in random places at random times [16, 20]. These models include many well-known models as special cases. Obtaining of the RF with a given type of correlation is achieved by varying a probability distribution of a large-scale disturbance ratio. Section 2.2 also covers different ways to describe the images on arbitrary surfaces. Autoregressive and wave RF models are used for this purpose [16]. For example, the scan of a cylindrical image can be used to represent the speech and other quasi-periodic signals [21], and images on the sphere can be used to describe a planet relief [22].

There are many papers on image filtering. Many methods of algorithm synthesis are developed taking into consideration a type of informative and noise components of observations [6, 15, 23–25]. At the same time, much attention was paid to reduce the computational complexity of filtration procedures, which, in particular, led to the creation of Kalman filters. However, the use of Kalman filtering in image processing causes significant difficulties due to multidimensionality. These difficulties were partially overcome using line-by-line vector filtration, as well as combination of several one-dimensional filters [12]. Section 2.3 introduces a tensor filter of multidimensional image sequences. This filter was developed on the basis of the RF tensor model described in Sect. 2.2. Asymptotic characteristics of efficiency of random field filtration on arbitrary dimension grids are also introduced, as well as a potential accuracy of multidimensional RF filtration.

Many researches are devoted to the problems of object or signal detection and recognition [6, 8, 9, 11, 26, 27]. The problems of synthesis and analysis of optimal decision rules for detecting point and extensive anomalies in the images, including multi-zone data, and their sequences are discussed in Sect. 2.4. Four equivalent forms of the optimal decision rule based on the likelihood ratio are obtained. These four forms are significantly different in computational complexity for large spatial sizes of images [16]. Characteristics of efficiency for anomaly detection in multidimensional images are also derived.

An important task of image sequence processing is their alignment, when due to various reasons there are spatial deformations of adjacent frames. Numerous approaches how to solve this problem, for example, search for characteristic points, light flow analysis, morphological analysis, and correlation-extreme methods, are proposed in [7, 10, 11, 14, 28]. In Sect. 2.5, the problems of alignment and estimation of parameters of image interframe geometrical transformations are considered for sequences of 2D and multidimensional images. On the basis of the tensor model of image sequences, a tensor filter for simultaneous filtration of images and their interframe displacement estimation are synthesized [16]. The authors also present a new method to estimate the parameters of image deformation using a fixed point of a complex geometric transformation of two images [16, 29, 30]. This complex transformation consists of the actual deformation plus an additional

artificially performed geometric transformation. Unknown parameters of actual deformation are determined according to coordinates of the obtained fixed point.

In practice, the image, noise and observation models are usually only partially known, i.e. there is a priory uncertainty. Thus, the synthesis of adaptive processing algorithms is required. For this purpose, many adaptive algorithms were suggested [31, 32]. Section 2.6 outlines the adaptive procedure options, which are included in the algorithms for solving image prediction and alignment [16]. Special attention is paid to the pseudo-gradient adaptation, on the basis of which highly efficient algorithms with comparatively small computational costs are synthesized. These qualities allow their usage in real-time systems, dealing with large images and their sequences. Section 2.7 concluded the chapter.

2.2 Mathematical Models of Images

The mathematical models suitable for multidimensional processing, such as the random fields, tensor models of random fields, autoregressive models of random fields, wave models of random fields, and random fields on surfaces are discussed in Sects. 2.2.1–2.2.5, respectively.

2.2.1 Random Fields

Nowadays, information systems, including spatial sensor devices and digital computing, are widely used. Therefore, the images with discrete spatial and temporal variables will be primarily considered. Without loss of generality, one may assume that the images are the sets of values arranged on multidimensional rectangular grids with a unit step. Two-dimensional and three-dimensional grids are presented in Fig. 2.1a, b, respectively. In general, an image is a set in n-dimensional grid nodes $\Omega = \left\{\bar{j} = (j_1, \ldots, j_n) : j_k = \overline{1, M_k},\ k = \overline{1, n}\right\}$. According to the physical nature, image values may be scalar (e.g. brightness of a monochromatic image), vector (velocity field, color images, displacement field), and more complex (e.g. matrix). If an image value in the node (pixel) $\bar{j}$ is denoted as $x_{\bar{j}}$, then the image is a set of these values on the grid $X = \left\{x_{\bar{j}} : \bar{j} \in \Omega\right\}$.

If the data is a time sequence of images, then a sequence can be assumed as a single image, increasing a grid dimension by one unit. For example, the sequence of planar images (Fig. 2.1a) can be regarded as a single three-dimensional image (Fig. 2.1b).

If it is necessary to specify a time variable, let us set it down at the top $X = \left\{x_{\bar{j}}^i : \bar{j} \in \Omega,\ i \in I\right\}$. This image is a set on the direct product $\Omega \times I$ of grids Ω and I, where I is a set of time index values. The cross section $x^i = \left\{x_{\bar{j}}^i : \bar{j} \in \Omega\right\}$, i.e.

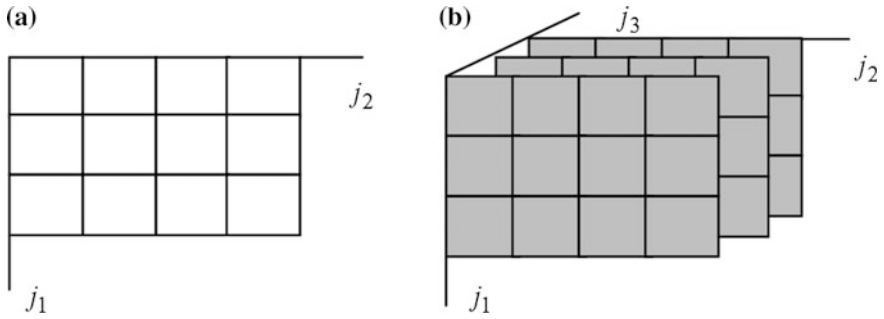

Fig. 2.1 Grids with: **a** two-dimensional image, **b** three-dimensional image

a set of image samples at a fixed value of time index i, is called the ith frame. Each frame is a set on a grid Ω. For example, Fig. 2.1b shows three two-dimensional frames.

Thus, an image can be determined as a function defined on a multi-dimensional grid. The values of the image elements cannot be accurately predicted in advance (otherwise monitoring system would not be necessary). Therefore, it is intrinsic to regard these values as the Random Variables (RVs) using the methods of probability theory and mathematical statistics.

2.2.2 Tensor Models of Random Fields

Consider the RF $X = \left\{x_{\bar{j}}^{i} : i \in I, \bar{j} \in \Omega\right\}$ defined on an $(n + 1)$-dimensional grid $\Omega \times I$, where $\Omega = \{j = (j_1, j_2, j_3, \ldots, j_n)\}$ is an n-dimensional $M_1 \times M_2 \times \cdots \times M_n$-grid, $I = \{i : i = 1, 2, 3, \ldots\}$. Index i may be interpreted as a time index, thus the expression $x^i = \left\{x_{\bar{j}}^{i} : \bar{j} \in \Omega\right\}$ is called a cross section of a field X as in ith frame.

Let the sequence of frames be described by a stochastic difference Eq. 2.1, where $\left\{\xi_{\bar{j}}^{i} : i \in I, \bar{j} \in \Omega\right\}$ is an updating standard Gaussian field, $\xi^i = \left\{\xi_{\bar{j}}^{i} : \bar{j} \in \Omega\right\}$ is the ith field frame, $\varphi^i(x^{i-1}) = \left\{\varphi_{\bar{j}}^{i}(x^{i-1}) : \bar{j} \in \Omega\right\}$ is $M_1 \times M_2 \times \cdots \times M_n$-matrix function, $\vartheta^i(x^{i-1}) = \left\{\vartheta_{\bar{j}\bar{t}}^{i}(x^{i-1}) : \bar{j}, \bar{t} \in \Omega\right\}$ are tensors of rank $2n$ with group indices forming a perturbation component of the ith frame from ξ^i using the rule of tensor multiplication $\vartheta^i \xi^i = \left\{\vartheta_{\bar{j}\bar{t}}^{i}\right\}\left\{\xi_{\bar{t}}^{i}\right\} = \left\{\sum \vartheta_{\bar{j}\bar{t}}^{i} \xi_{\bar{t}}^{i}\right\}$.

$$x^i = \varphi^i(x^{i-1}) + \vartheta^i(x^{i-1})\xi^i \quad i = 1, 2, \ldots \tag{2.1}$$

Transposing this frame supposes a permutation of its group indices $\vartheta^{iT}_{\bar{j}\bar{i}} = \vartheta^{i}_{\bar{i}\bar{j}}$. Note that the superscript i indicates a frame number, i.e. time in our case. Thus, index i is considered to be an umbral index and the summation with respect to it is not extended.

Model provided by Eq. 2.1 allows to describe a very broad class of Markov random frame sequences. In particular, a linear model in Eq. 2.1, where tensors P^i and ϑ^i do not depend on x^{i-1}, describes the Gaussian sequence.

$$x^i = P^i x^{i-1} + \vartheta^i \xi^i \tag{2.2}$$

For this field of Covariance Functions (CFs) there exists a multidimensional covariance matrix defined directly

$$V_x(i,j) = M[(x^i - m^i) \times (x^j - m^j)] \quad V_x(i) = V_x(i,i),$$

where $m^i = M[x^i] = \left\{M[x^i_{\bar{j}}] : \bar{j} \in \Omega\right\}$, and symbol "×" indicates an external matrix multiplication. Thus, $V_x(i)$ and $V^i_{\xi} = M[\xi^i \times \xi^i]$ are symmetrical $M_1 \times M_2 \times \cdots \times M_n \times M_1 \times M_2 \times \cdots \times M_n$-matrices. For complete definition of a random field with a state Eq. 2.2, it is necessary to set the law of first frame x^0 distribution. It is often a Gaussian distribution with mean m^0 and covariance matrix V^0_x.

Model in Eq. 2.2 with constant tensors $P^i = P$ and $\vartheta^i = \vartheta$ is of particular interest that leads to Eq. 2.3, where (k) determines raising to the kth power.

$$V_x(i, i+k) = P^{(k)} V_x(i,i) = P^{(k)} V_x(i) \tag{2.3}$$

If roots of the characteristic equation $\det(\lambda E - P) = 0$ are less than 1 in modulus, than $P^{(k)} \to 0$ as $k \to \infty$, and from Eq. 2.3 we obtain that $V_x(i, i+k) \to 0$ as $k \to \infty$.

Using z-transformation device, let us write Eq. 2.2 in the form

$$x^i = zPx^i + \vartheta\xi^i$$

or

$$(E - zP)x^i = \vartheta\xi^i.$$

Multiplying the congruence on the left by $(E - zP)^{-1}$, the expression x^i through a perturbing field is obtained.

$$x^i = (E - zP)^{-1}\vartheta\xi^i$$

Hence,

$$x^i(x^i)^T = (E - zP)^{-1}\vartheta\xi^i(\xi^i)^T\vartheta^T(E - z^{-1}P^T)^{-1}$$

that after averaging gives the expression of the tensor spectrum for a stationary field

$$V_x(z) = \sum_{k=-\infty}^{\infty} V_x(0,k)z^k = (E - zP)^{-1}\vartheta V_\xi\vartheta^T(E - z^{-1}P^T)^{-1},$$

which is a Laurent series in powers of z with tensor coefficients. The tensor coefficients can be found from Eq. 2.4, where $C = \{z : |z| = 1\}$ is a unit circumference of a complex plane.

$$V_x(0,k) = \frac{1}{2\pi i}\int_{C_1} (E - zP)^{-1}\vartheta\, V_\xi\vartheta^T(E - z^{-1}P^T)^{-1}z^{k-1}dz \tag{2.4}$$

It is sufficient to find $V_x = V_x(0,0)$ from Eq. 2.4, other values are obtained from Eq. 2.3 in a view of $V_x(i,i+k) = P^{(k)}V_x$ for a stationary case. To find V_x it is possible to use limit in Eq. 2.3 as $i \to \infty$ instead of integral in Eq. 2.4

$$V_x = PVP^T + \vartheta\, V_\xi\vartheta^T$$

that is a non-singular system of linear equations relative to tensor components V_x.

Now consider a solution to the model synthesis problem Eq. 2.2, i.e. the problem of finding tensors P^i and ϑ^i, when intraframe $V_x(i,i)$ and interframe $V_x(i-1,i)$ covariance tensors are given. From Eq. 2.3 we obtain

$$V_x(i-1,i) = P^iV_x(i-1,i-1)$$

that is a system of linear equations relative to tensor elements P^i. It is obvious that

$$P^i = V_x(i-1,i)V_x^{-1}(i-1,i-1).$$

Chose a perturbing field with covariance $V_\xi^i = E$. Then the following equation is obtained

$$\vartheta^i\vartheta^{iT} = V_x(i,i) - P^iV_x(i-1,i-1)P^{iT}.$$

For example, for tensor components ϑ^i this equation can be solved on the basis of Gram-Shmidt orthogonalization.

Nonlinear stochastic difference equation may be considered as generalization of the considered tensor model (Eq. 2.2). This equation allows to describe very large class of Markov non-Gaussian RF on n-dimensional grids J_t. Here we have $\left\{\xi_{\bar{l}}^t, \bar{l} \in J_t, t \in T\right\}$, a field of independent, generally speaking, non-Gaussian random variables with known probability density function (PDF) $W(\xi_{\bar{l}}^t)$, $\varphi_{\bar{j}}^t(x_{\bar{l}}^{t-1})$ and $\vartheta_{\bar{j}\bar{l}}^t(x_{\bar{s}}^{t-1})$ are tensors of ranks n and $2n$ correspondingly, which in a general case nonlinearly depend on the values $\left\{\xi_{\bar{j}}\right\}$ of the (t–1)th frame of a multidimensional RF $\left\{x_{\bar{j}}^{t-1}, t \in T, \bar{j} \in J\right\}$.

2.2.3 Autoregressive Models of Random Fields

Tensor representation assumes that for each moment t of discrete time the RF $\left\{x_{\bar{j}}^t, \bar{j} \in J\right\}$ is formed recurrently on the basis of the previous value $\left\{x_{\bar{j}}^{t-1}, \bar{j} \in J\right\}$ and updating RF $\left\{\xi_{\bar{j}}^t, \bar{j} \in J\right\}$ of independent RV. Despite the fact that the calculations are done recurrently while forming each successive frame $\left\{x_{\bar{j}}^t, \bar{j} \in J\right\}$, it is desirable to conduct linear Eq. 2.1 or nonlinear Eq. 2.2 transformation of all elements $\left\{x_{\bar{j}}^{t-1}, \bar{j} \in J\right\}$, $\left\{\xi_{\bar{j}}^t, \bar{j} \in J\right\}$ determined on an n-dimensional spatial grid.

Thus, tensor models make it possible to describe a big class of non-Gaussian and non-homogeneous RF but lead to an overall computational effort during the RF imitation and processing. In this context, there appear questions about the existence of recurrent not only in time but also in space RF representation and the possibility of constructing optimal recurrent algorithms of statistical analysis for such RF.

In their structure, the random fields are much more complex than stochastic processes. First, implementations of the RF are functions of several variables, the theory of which is more complicated than of one variable. Second, the concept of Markov behavior also becomes much more complicated. A random process can develop in course of time. Model from Eq. 2.1 is a mathematical expression of such development. For Markov sequence, the time interval can be broken at any point i for conditionally independent past $\Gamma^- = \left\{x^k : k < i\right\}$ and future $\Gamma^+ = \left\{x^k : k > i\right\}$. However, the RF is defined on an n-dimensional domain Ω. For its geometrical partitioning into two parts Γ^- and Γ^+, at least an $(n - 1)$-dimensional domain Γ is required. Markov RF suggests that for any set Γ (of a certain class) all RV included in Γ^- are conditionally independent from the RV belonging to Γ^+, when values of Γ are known. It is possible to name Γ^-, Γ, and Γ^+ as past, present, and future only roughly. However, Markov property allows to imagine the RF as one developing in time from Γ^- through Γ to Γ^+, in addition Γ moves along Ω with the course of time. For example, if it is assumed that the lines of a two-dimensional grid Ω represent Γ, then the field is formed line by line.

Further development of this idea allows to generalize AutoRegressive (AR) models of random sequences on the RF. If the procedure for scalar sequence formation $x^0, x^1, x^2, \ldots$ usually corresponds to temporally observed values, then the procedure of field $X = \{x_{\bar{j}} : \bar{j} \in \Omega\}$ formation needs a special attention. For this purpose, it is necessary to order grid nodes Ω. Then, it is possible to say about any two elements of the field that one of the elements precedes the other. If $x_{\bar{i}}$ precedes $x_{\bar{j}}$, let us mark it as $(\bar{i}) < (\bar{j})$. There are many variants of such ordering. In a two-dimensional case a saw-toothed scanning (all lines go in one direction, Fig. 2.2a) and a triangular scanning (while passing to the next line the direction is reversed, Fig. 2.2b) are often used. As a result, the RF scanning is converted into a random sequence.

Linear autoregressive models of random fields. The simplest autoregressive model is a linear stochastic Eq. 2.4 with white Gaussian RF $\{\xi_{\bar{j}}\}$, which corresponds to a well-known autoregressive-moving average equation for random sequences (Eq. 2.5).

$$x_{\bar{j}} = \sum_{\bar{l} \in G_j} \alpha_{\bar{j},\bar{l}} x_{\bar{l}} + \sum_{\bar{l} \in Y_j} \beta_{\bar{j},\bar{l}} \xi_{\bar{l}} \quad \bar{j} \in J \tag{2.5}$$

However, unlike its one-dimensional prototype properties of RF generated by Eq. 2.5 are not thoroughly understood, even for constant coefficient models $\alpha_{\bar{j},\bar{l}} = \alpha_{\bar{l}}$, $\beta_{\bar{j},\bar{l}} = \beta_{\bar{l}}$ and unchanging domains $G_{\bar{j}} = G$ and $Y_{\bar{j}} = Y$ (Eq. 2.6).

$$x_{\bar{j}} = \sum_{\bar{l} \in G} \alpha_{\bar{l}} x_{\bar{j}-\bar{l}} + \sum_{\bar{l} \in Y} \beta_{\bar{l}} \xi_{\bar{j}-\bar{l}} \quad \bar{j} \in J \tag{2.6}$$

An important particular case of Eq. 2.6 is represented by Eq. 2.7 of multidimensional AR.

$$x_{\bar{j}} = \sum_{\bar{l} \in G} \alpha_{\bar{l}} x_{\bar{j}-\bar{l}} + \xi_{\bar{j}} \quad \bar{j} \in J \tag{2.7}$$

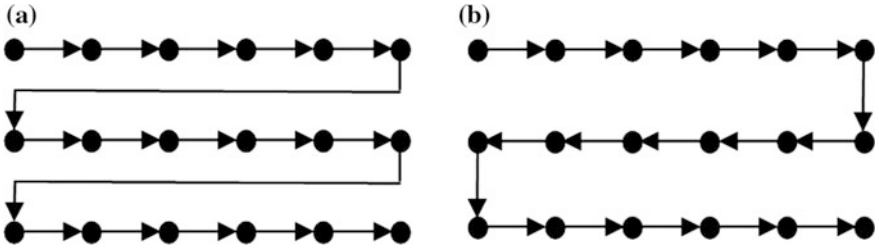

Fig. 2.2 Two-dimensional image scanning: **a** saw-toothed, **b** triangular

The abovementioned equations describe the algorithm of RF formation $\{x_{\bar{j}}\}$ at the point $\bar{j} = (j_1, j_2, \ldots, j_n)$. Besides, it is assumed that at our disposal all values of $x_{\bar{j}-\bar{l}}$, $\bar{l} \in G$, $(\bar{j} - \bar{l}) < (\bar{j})$ and the RF calculated before are given as initial conditions.

Model in Eq. 2.7 has a disadvantage, which complicates its analysis. It is probably a large number of summands on the right side of equation. However, it is possible to reduce their number to a minimum in a following way. The simplest AR equation generating n-dimensional field X, which does not fall into independent fields of less dimension, is represented by Eq. 2.8, where $\bar{e}_k = (0, \ldots, 0, 1, 0, \ldots, 0)$ is a unit vector of the kth coordinate axis.

$$x_{\bar{i}} = \sum_{k=1}^{n} \alpha_k x_{\bar{i}-\bar{e}_k} + \beta \xi_{\bar{i}} \tag{2.8}$$

Any model Eq. 2.7 can be reduced to a model of type Eq. 2.8 with a minimal number of summands. For this purpose, let us use vector autoregressive models, which in their linear form are described by Eq. 2.9, where $\bar{x}_{\bar{i}}$ is a value of a vector field in the node $\bar{i}$, $A_{\bar{j}}$, B are the square matrices, $\{\bar{\xi}_{\bar{i}}\}$ is a renewing standard vector field of independent vectors with independent components.

$$\bar{x}_{\bar{i}} = \sum_{\bar{j} \in D} A_{\bar{j}} \bar{x}_{\bar{i}+\bar{j}} + B \bar{\xi}_{\bar{i}} \tag{2.9}$$

Indeed, consider Habibi model [2] as an example

$$x_{i_1 i_2} = \rho_1 x_{i_1-1, i_2} + \rho_2 x_{i_1, i_2-1} - \rho_1 \rho_2 x_{i_1-1, i_2-1} + \sigma \sqrt{(1-\rho_1^2)(1-\rho_2^2)}\, \xi_{i_1 i_2}. \tag{2.10}$$

Its calculation scheme is shown in Fig. 2.3a.

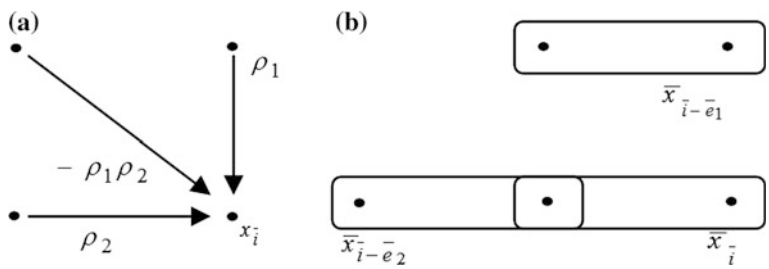

Fig. 2.3 Transition from scalar model to vector model: **a** scalar model, **b** vector model

Let us imagine this model in the form of Eq. 2.9. For this purpose, vectors $\bar{x}_{\bar{i}} = \bar{x}_{ij} = \left(x_{ij}, x_{i,j-1}\right)^{\mathrm{T}}$ and $\bar{\xi}_{\bar{i}} = \left(\xi_{ij}, \varphi_{ij}\right)^{\mathrm{T}}$, $\beta = \sigma\sqrt{(1-\rho_1^2)(1-\rho_2^2)}$, then $\bar{x}_{\bar{i}-\bar{e}_1} = \left(x_{i-1,j}, x_{i-1,j-1}\right)^{\mathrm{T}}$ and $\bar{x}_{\bar{i}-\bar{e}_2} = \left(x_{i,j-1}, x_{i,j-2}\right)^{\mathrm{T}}$ are introduced. Equation 2.10 will be equivalent to the first component of the vector in Eq. 2.11

$$\begin{pmatrix} x_{ij} \\ x_{i,j-1} \end{pmatrix} = \begin{pmatrix} \rho_1 & -\rho_1\rho_2 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} x_{i-1,j} \\ x_{i-1,j-1} \end{pmatrix} + \begin{pmatrix} \rho_2 & 0 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} x_{i,j-1} \\ x_{i,j-2} \end{pmatrix} + \begin{pmatrix} \beta & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \xi_{ij} \\ \varphi_{ij} \end{pmatrix} \tag{2.11}$$

or at obvious notation

$$\bar{x}_{\bar{i}} = A_1\bar{x}_{\bar{i}-\bar{e}_1} + A_2\bar{x}_{\bar{i}-\bar{e}_2} + B\bar{\zeta}_{\bar{i}}$$

that is the minimal vector model of the type Eq. 2.9. Figure 2.3b shows the elements of the field included in the vectors of Eq. 2.11.

At first, let us analyze the main probabilistic characteristics of multidimensional AR models. This class of the RF is generated by linear stochastic difference Eq. 2.12, where $X = \{x_{\bar{i}}, \bar{i} \in \Omega\}$ is a modeled RF defined on an N-dimensional grid $\Omega = \{\bar{i} = (i_1, i_2, \ldots i_N) : i_k = 1 \ldots M_k,\ k = 1 \ldots N\}$, $\{\alpha_{\bar{j}}, \beta, \bar{j} \in D\}$ are model coefficients, $\Xi = \{\xi_{\bar{i}},\ \bar{i} \in \Omega\}$ is a renewing white RF, $D \subset \Omega$ is a casual region.

$$x_{\bar{i}} = \sum_{\bar{j} \in D} \alpha_{\bar{j}} x_{\bar{i}-\bar{j}} + \beta\xi_{\bar{i}} \quad \bar{i} \in \Omega \tag{2.12}$$

Normal RF distribution with independent components is usually chosen as a renewing field Ξ. In this case RF X also has Gaussian distribution.

In its general form, the task of model Eq. 2.12 analysis was described in [16]. A linear spatial filter with transfer function corresponds to the model Eq. 2.12 is given by Eq. 2.13, where $\bar{z}^{-\bar{j}} = z_1^{-j_1} z_2^{-j_2} \ldots z_N^{-j_N}$.

$$H(\bar{z}) = \frac{\beta}{1 - \sum_{\bar{j} \in D} \alpha_{\bar{j}} \bar{z}^{-\bar{j}}} \tag{2.13}$$

Besides, the RF spectral density of X is written as

$$S_x(\bar{z}) = H(\bar{z})\, S_\xi(\bar{z})\, H\left(\bar{z}^{-1}\right) = \sigma_\xi^2 H(\bar{z}) H\left(\bar{z}^{-1}\right),$$

where σ_ξ^2 is the variance of RF Ξ.

The correlation function of X can be calculated by means of backward z-transformation of spectral density

$$R(\bar{r}) = \frac{\beta^2}{(2\pi i)^N} \oint_{C_N} S_x(\bar{z})\, \bar{z}^{\bar{r}-\bar{1}}\, d\bar{z},$$

where $C_N = \{|\bar{z}| = 1\}$ is a unit polycircle in multidimensional complex space.

Analysis of the RF probabilistic properties is simplified if their spectral density can be factorized [12, 16, 33], i.e. $S_x(\bar{z}) = \prod_{k=1}^{N} S_k(z_k)$. Since in such fields, the transfer function of a multidimensional filter $H(\bar{z}) = \prod_{k=1}^{N} H_k(z_k)$ and CF $R(\bar{r}) = \prod_{k=1}^{N} R_k(r_k)$ are also factorized, then it is enough to examine the properties of random sequences generated by one-dimensional AR with the following characteristics $S_k(z_k)$, $H_k(z_k)$ and $R_k(r_k)$, $k = 1 \ldots M_k$.

The drawback of such models is impossibility to describe isotropic RF, e.g. fields with the CF $R(\bar{r}) = R(|\bar{r}|) = R\left(\sqrt{r_1^2 + r_2^2 + \cdots + r_N^2}\right)$. However, the analysis shows that some dimensional models gives the RF close to isotropic ones. In order to obtain close to isotropic RF, the authors [16–18, 33]. Suggested to use one-dimensional filters with multiple roots of characteristic equations $1 - \sum_{j=1}^{n_k} \alpha_{kj} \lambda_k^j = 0$, where n_k, $k = 1 \ldots M$ are the orders of one-dimensional AR.

At first, let us solve these tasks for a one-dimensional model. Consider one-dimensional AR of M length given by Eq. 2.14, where $\{\xi_i\}$ is a Gaussian sequence of independent components with zero mean and variance σ_ξ^2.

$$x_i = \sum_{j=1}^{n} \alpha_j x_{i-j} + \beta\xi_i \quad i = 1 \ldots M \tag{2.14}$$

To solve the synthesis task, it is necessary to determine the AR coefficients $\{\alpha_j, j = 1 \ldots n;\ \beta\}$ using the given root of a characteristic equation $1/\rho$, its multiplicity n and the desired field variance σ_x^2. In the case of multiple roots, this equation can be written in an operator form Eq. 2.15, where z^{-1} is a shift operator.

$$\left(1 - \rho z^{-1}\right)^n x_i = \beta\xi_i \quad i = 1 \ldots M \tag{2.15}$$

Provided that $z^{-k} x_i = x_{k-i}$, let us rewrite Eq. 2.15 in a view of Eq. 2.16.

$$x_i = \sum_{j=1}^{n} (-1)^{j+1} C_n^j \rho^j x_{i-j} + \beta\xi_i \quad i = 1 \ldots M \tag{2.16}$$

Comparison of Eqs. 2.15–2.16 permits to write the expression for coefficients $\alpha_j = \alpha_j(\rho, n)$ as Eq. 2.17.

$$\alpha_j(\rho, n) = (-1)^{j+1} C_n^j \rho^j \quad j = 1 \ldots n \tag{2.17}$$

Value of the unknown parameter β, which is an amplification coefficient in transfer function Eq. 2.13, should be selected so as to make the filter stable. Further, it will be shown how to determine β on the basis of the CF model.

One of the tasks of model statistical analysis is to obtain its CF

$$R(k) = \beta^2(\rho, n)\, \rho^k \sum_{\ell=0}^{n-1} g(n, \ell, k) \frac{\rho^{2(n-\ell-1)}}{(1-\rho^2)^{2n-\ell-1}}, \tag{2.18}$$

where

$$g(n, \ell, k) = \frac{(n+k-1)!(2n-\ell-2)!}{\ell!(n-1)!(n-\ell-1)!(n+k-\ell-1)!} \tag{2.19}$$

and coefficient $\beta = \beta(\rho, n)$ can be obtained from equation $R(0) = 1$

$$\beta^2(\rho, n) = \frac{(1-\rho^2)^{2n-1}}{\sum_{\ell=0}^{n-1} \left(C_{n-1}^{\ell} \rho^{\ell}\right)^2}. \tag{2.20}$$

Equations 2.18–2.20 give a general view of a Normalized Autocorrelation Function (NAF) of a one-dimensional model Eq. 2.14 for given ρ and n. In order to obtain the CF at variances σ_x^2 and σ_ξ^2, which are not equal to 1, it is necessary to multiply the right hand side of Eq. 2.20 by σ_x^2/σ_ξ^2. Thus, we obtain the expression for coefficient β as well

$$\beta = \frac{\sigma_x}{\sigma_\xi} \sqrt{(1-\rho^2)^{2n-1} / \sum_{\ell=0}^{n-1} \left(C_{n-1}^{\ell} \rho^{\ell}\right)^2}. \tag{2.21}$$

Therefore, Eqs. 2.17 and 2.20 completely determine the unknown coefficients of one-dimensional AR model Eq. 2.14 with multiple roots of the characteristic equation.

Now let us consider an N-dimensional case. The RF model for a given variance σ_x^2 is completely determined by a parameter vector $(\rho_1, \rho_2, \ldots, \rho_N)$ and a multiplicity vector $(n_1, n_2, \ldots, n_N)$. Let multidimensional factorable RF be generated by the following AR equations written in an operator form

$$\prod_{k=1}^{N} \left(1 - \rho_k z_k^{-1}\right)^{n_k} x_{\bar{i}} = \left(\prod_{k=1}^{N} \sum_{l=1}^{n_k} \alpha_{kl}\, z_k^{-1}\right) x_{\bar{i}} = \beta \xi_{\bar{i}}, \quad \bar{i} \in \Omega,$$

where N is field dimensionality, ρ_k and n_k are a model parameter and multiplicity root along the kth axis, respectively, Ω is a grid, on which field X is determined.

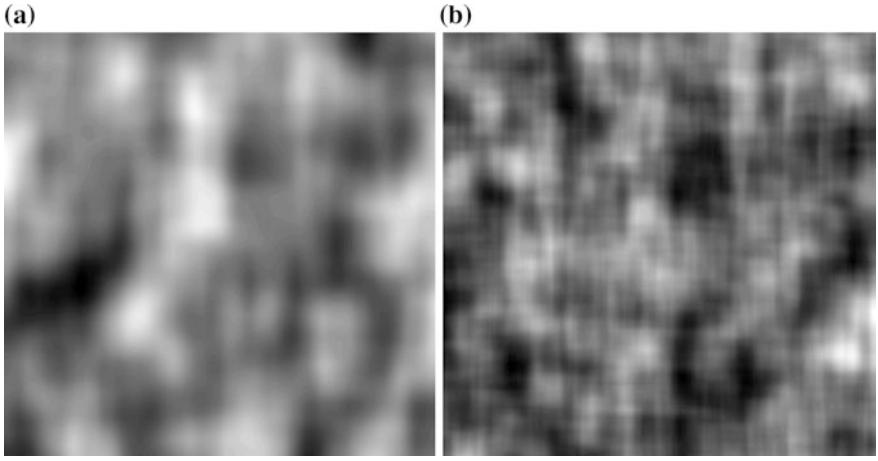

Fig. 2.4 Representation of random fields on the basis of autoregression with multiple roots: **a** multiplicity (2, 2), parameter vector (0.95, 0.95), **b** multiplicity (3, 3), parameter vector (0.95, 0.95)

Coefficients $\alpha_{\bar{j}}$ are the products of corresponding coefficients α_{kj_k} of one-dimensional AR along the kth axis $\alpha_{\bar{j}} = \prod_{k=1}^{N} \alpha_{kj_k}$, where $\bar{j} = (j_1, j_2, \ldots, j_N)$, $j_k = \overline{1 \ldots n_k}$. Coefficient β of a multidimensional model can be found in a similar way $\beta = \frac{\sigma_x}{\sigma_{\bar{\xi}}} \prod_{k=1}^{N} \beta_k$, where β_k is a corresponding standardized one-dimensional coefficient.

An example of image frame (600 × 400 elements) obtained on the basis of the model with multiple roots is depicted in Fig. 2.4. Output analysis shows that the changing parameters and the multiplicity might be obtained a wide range of different patterns, which can help to develop complex models of multi-zone images. At the same time, if a root multiplicity grows, then the modeled RF is approximated in its properties to the isotropic RF. This is also confirmed by the cross section shape of equal level CF shown in Fig. 2.5.

Analysis of the CF models with multiple roots shows that the CF cross-sections of the RF obtained by factorable multi-dimensional models become close to hyperellipsoid with the growth of the characteristic equation root multiplicity. To assess the proximity of such RF to isotropic ones, it is desirable to have quantitative estimation of field anisotropy. For this purpose, let us introduce $\tau(\vec{u}) = \int_0^{+\infty} R(t\vec{u})\, dt$ as the anisotropy of multidimensional RF according to correlation distance τ in $\vec{u}$-direction, where $\vec{u} \in S^{N-1}$ is a point on a hypersphere. In addition, it is possible to suggest the following anisotropy coefficient $A_\tau = \frac{1}{\bar{\tau}} \sqrt{\frac{1}{|S^{N-1}|} \int_{\vec{u} \in S^{N-1}} (\tau(\vec{u}) - \bar{\tau})^2 d\vec{u}}$, $A \geq 0$, where $\bar{\tau} = \frac{1}{|S^{N-1}|} \int_{\vec{u} \in S^{N-1}} \tau(\vec{u}) d\vec{u}$ is a mean angular correlation distance, $\left|S^{N-1}\right|$ is an area of hypersphere surface.

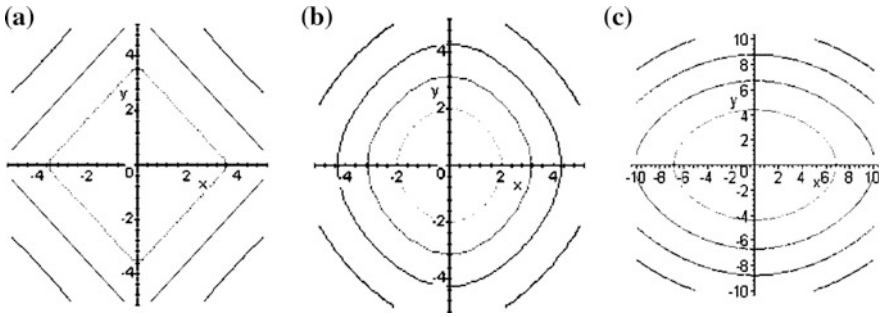

Fig. 2.5 Cross-sections of a correlation function for a two-dimensional model with multiple roots: **a** multiplicity (1, 1), parameters (0.9, 0.9), **b** multiplicity (3, 3), parameters (0.9, 0.9), **c** multiplicity (2, 3), parameters (0.9, 0.9)

The advantage of the proposed anisotropy coefficients is that we should know only field CF because all calculation can be done using standard numerical methods.

2.2.4 Wave Models of Random Fields

Let us consider the RF wave model that generalizes a number of other models and helps to solve the tasks of analysis and synthesis [16, 20, 34] effectively. This model is simple enough and can serve as a basis for simulating images and their sequences with given CF without increasing the number of model parameters.

In a wave model, the RF is determined by Eq. 2.22, where an $(n + 1)$-dimensional domain $\{(\bar{j}, t)\}$ may be discrete or continuous, $\{(\bar{u}_k, \tau_k)\}$ is a discrete Field of Random Points (FRP) in an $(n + 1)$-dimensional continuous space, t and τ_k are interpreted as time, $\bar{\omega}_k$ is a random vector of function f parameters.

$$x_{\bar{j}}^{t} = \sum_{\{k:\tau_k \le t\}} f((\bar{j}, t), (\bar{u}_k, \tau_k), \bar{\omega}_k) \tag{2.22}$$

This field can be represented as the effect of random disturbances or waves $f((\bar{j}, t), (\bar{u}_k, \tau_k), \bar{\omega}_k)$, appearing in random places $\bar{u}_k$ at random time τ_k and changing according to a given law in time and space.

Selection of function f, the FRP parameters and $\bar{\omega}$ allow us to obtain a vast class of fields, which includes the following models:

1. Poisson fields, when $f((\bar{j}, t),\ (\bar{u}_k, \tau_k), \bar{\omega}_k) = \delta((\bar{j}, t) - (\bar{u}_k, \tau_k))$, δ is a Kronecker delta, and $\{(\bar{u}_k, \tau_k)\}$ is Poisson FRP.
2. Multidimensional filtered Poisson process, when $f((\bar{j}, t),\ (\bar{u}_k, \tau_k), \bar{\omega}_k) = g((\bar{j} - \bar{u}_k t - \tau_k), \xi_k)$, where $\{\xi_k\}$ is a system of scalar RF. This model can

generate only stationary homogeneous fields and the generating waves may differ from each other only in one parameter ξ_k.

3. Weighted sum model, which is obtained at waves $f(\bar{j},t),\ (\bar{u}_k,\tau_k),\bar{\omega}_k) = g(\bar{j},t),(\bar{u}_k,\tau_k))\,\xi_k$, where $\{(\bar{u}_k,\tau_k)\}$ is a set of all grid nodes and g represents corresponding weights of random variables ξ_k. This model corresponds to different RF expansion into a system of basic functions, e.g. [1].
4. Random walk model. The FRP describes a random walk (perhaps with appearance and disappearance) of a set of waves and the choice $\bar{\omega}_k$ determines the dynamics of a wave shape and intensity. Such models can be applied to simulate moving clouds.

Consider a particular case of a wave model, for which correlation tasks of analysis and synthesis can easily be solved (Eq. 2.23, where the FRP is a Poisson one with constant density λ, $\rho_k = |\bar{j} - \bar{u}_k|$ is a distance between $\bar{j}$ and $\bar{u}_k$, $\{R_k\}$ is a system of independent non-negative equally distributed RF with PDF $w(\alpha)$, $\{\xi_k\}$ is a system of independent equally distributed RF).

$$f(\bar{j},t),(\bar{u}_k,\tau_k),\bar{\omega}_k) = g(\rho_k/R_k)\exp(-\mu/|t-\tau_k|)\,\xi_k \tag{2.23}$$

Waves are motionless, independent, have spherical sections in space and exponentially attenuate over time. System $\{\xi_k\}$ determines a wave intensity and $\{R_k\}$ is their spatial scale. Evidently that the generated field X is stationary and homogeneous, has zero mean and isotropic spatial CF. In this case, all summands in Eq. 2.22 are non-correlated and elementary event $\Delta A =$ {in the element $\Delta V = \Delta_{j_1}\Delta_{j_2}\ldots\Delta_{j_n}\Delta\tau$, there appears a point in the FRP, which corresponds to a wave with spatial scale α from the element $\Delta\alpha$} with a probability $P(\Delta A) \approx \lambda\Delta V w(\alpha)\Delta\alpha$. Let us express the CF by an integral with respect to variables $\tau, \alpha, j_1, \ldots, j_n$. After integration with respect to τ and taking into account $g(y) = c\exp(-2y^2)$ we obtain Eq. 2.24.

$$V(\rho,t) = \frac{c^2\pi^{n/2}\lambda}{2^{n+1}}e^{-\mu t}\int_0^\infty \alpha^n \exp\left(-\frac{\rho^2}{\alpha^2}\right)w(\alpha)d\alpha \tag{2.24}$$

When $\rho = t = 0$, a field variance is found from Eq. 2.24

$$\sigma_n^2 = \frac{c^2\pi^{n/2}\lambda}{2^{n+1}\mu}M[R^n].$$

It is proportional to the FRP density λ, efficiency interval $1/\mu$ of wave attenuation and mean value of the nth degree of scale R.

Simulation of a discrete field on an n-dimensional grid $\{\bar{j}\}$ with time quantization Δt can be implemented by the following algorithm. At the initial time $t_0 = 0$, the field values in all nodes are equal to zero. At each subsequent moment $t_m = m\Delta t$, a Poisson FRP with density $\lambda\Delta t$ is formed over continuous space or grid, which

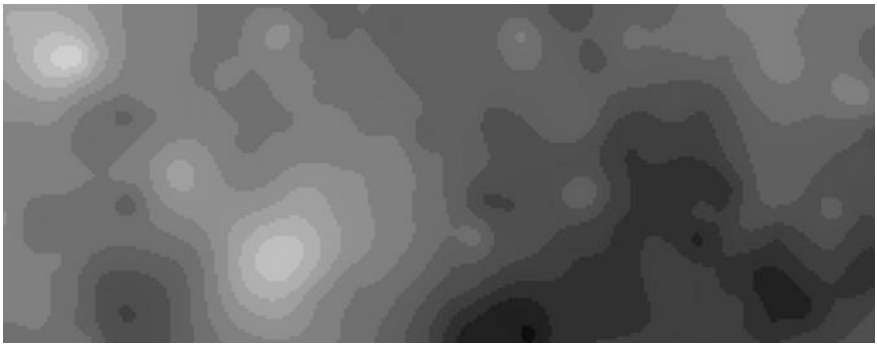

Fig. 2.6 Example of image simulation using a wave model

somehow overlaps $\{\bar{j}\}$. At each generated point $\bar{u}_k$, the RV ξ_k and R_k are formed. After that the following transformation of all field values on grid $\{\bar{j}\}$ is carried out by Eq. 2.25.

$$x_{\bar{j}}^{t_m} = x_{\bar{j}}^{t_{m-1}} \exp(-\mu \cdot \Delta t) + \sum_k g(\rho_k / R_k)\, \xi_k \tag{2.25}$$

In this simulation only large summands (in comparison with a quantization level) can be taken into account. The advantage of this algorithm is its recurrence, as it makes easy to implement a field simulation using computer.

An example of image simulation with the described wave model is represented in Fig. 2.6. It is necessary to underline that Eq. 2.25 actually implements the time-varying images. Therefore, this figure shows only one frame of this process. Each field value is a sums of random numbers of the RV, Thus, generally speaking, the field will not be Gaussian even with Gaussian $\{\xi_k\}$. However, when the model parameter $h = \lambda M[R^n]/\mu$ grows, then the number of summands in Eq. 2.25 with similar distributions increase and the field is normalized.

Now consider the solution of correlation analysis and synthesis task. It follows from Eq. 2.24 that the formed field has an exponential time NAF $e^{-\mu t}$ and space NAF.

$$r(\rho) = \frac{1}{M[R^n]} \int_0^\infty \alpha^n \exp\left(-\rho^2/\alpha^2\right) \omega(\alpha)\, d\alpha \tag{2.26}$$

Thus, solving analysis tasks, when the PDF $\omega(\alpha)$ is given, the required NAF can be found analytically or by numerical integration. Solving synthesis tasks, when the NAF $r(\rho)$ is given, it is necessary to solve integral Eq. 2.26 with respect to unknown $\omega(\alpha)$. As it is not always possible to find an analytical solution of Eq. 2.26, consider a method of its approximate solution. From Eq. 2.26 it follows that at degenerate distribution ($R = \alpha = \text{const}$) we obtain the CF $\exp(-\rho^2/\alpha^2)$. Let

now an arbitrary non-increasing NAF $r(\rho)$ be given. Let us approximate it with adequate accuracy by a sum of Gaussoids with positive coefficients $r(\rho) \approx h(\rho) = \sum_i q_i \exp\left(-\rho^2 / \alpha_i^2\right)$, where $\sum_i q_i = 1$, when $r(0) = 1$. Then, for discrete distribution $P(R = \alpha_i) = k^{-1} q_i / \alpha_i^n$, where $k = \sum_i q_i / \alpha_i^n$, the generated field will have NAF equal to $h(\rho)$. Thus, the generated model allows us approximately to solve a synthesis task by changing only PDF of scale R.

2.2.5 Random Fields on Surfaces

All abovementioned images and the RV were set on rectangular grids of some dimension. In some practical problems, the images can be set on the surface of another type, for example, spherical Earth image or cylindrical image of a rotation shaft. Model representation of such images differs significantly both in the type of a spatial grid and the way of defining correlations. Consider the main peculiarities of presentation of AR images on a cylinder, as well as wave models on almost arbitrary surfaces.

Autoregressive RF model on a cylinder. Consider a cylindrical image, for example, image of a rotation shaft. If it is cut lengthwise and expanded, then it is transformed into a rectangular image. Points along the cross sections are close to each other on the original cylindrical image that is why their values are highly correlated. In the cut image, these points are located at the opposite ends. Thus, the line ends are highly correlated with the new-line characters. However, such images cannot be described by the abovementioned models on the rectangles because the correlations in these models are weakening, when the distances between values increase. Thus, the image points at the line ends do not possess the necessary high correlation. Consider, for example, an image simulated by Habibi model (Fig. 2.7).

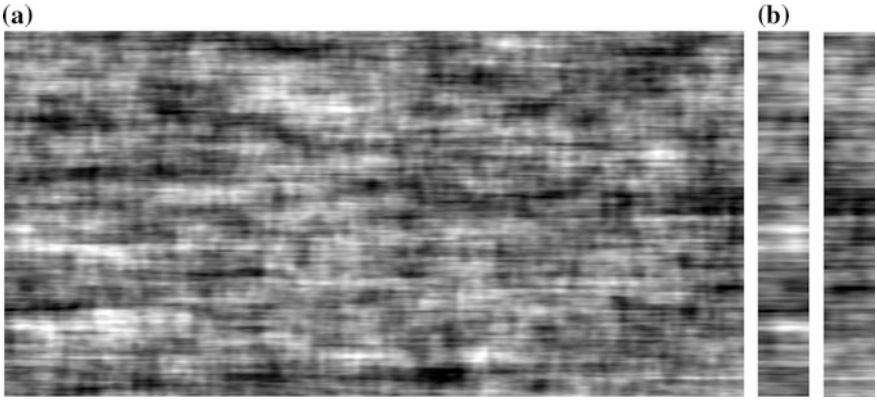

Fig. 2.7 Simulated images: **a** image simulated by Habibi model, **b** the first and the last five columns of image from (**a**)

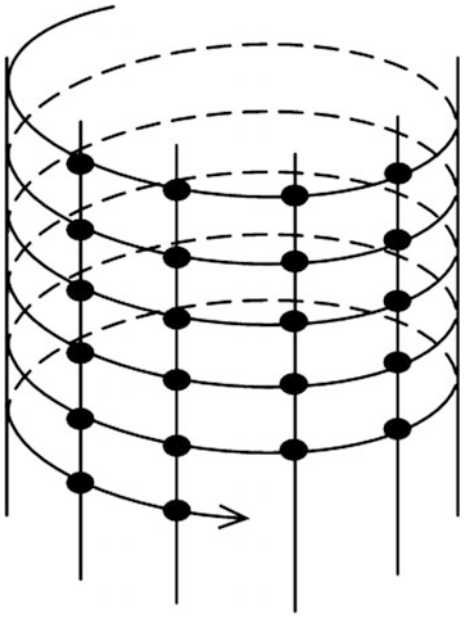

Fig. 2.8 Spiral grid of a cylindrical image

This figure shows that the first and the last columns of the rectangular image are significantly different. This means that, when pasting this image into a cylinder, one will observe great changes in brightness jumps into a cylinder at the junction.

To approximate the model to real cylindrical images, consider a spiral grid on the cylinder shown in Fig. 2.8. Grid lines represent spiral turns. To describe an image set on this grid, let us apply the analog of AR Habibi [2] model Eq. 2.27 where k is a spiral turn number and l is a node number.

$$x_{k,l} = \rho\, x_{k,l-1} + r\, x_{k-1,l} - \rho\, r\, x_{k-1,l-1} + \beta\, \xi_{k,l} \tag{2.27}$$

Here, $l = 0, \ldots, T$, $x_{k,l} = x_{k+1,l-T}$, when $l \geq T$, T is a period, i.e. the number of points in one turn. It should be pointed out that in model Eq. 2.27 the grid can be also regarded a simple cylindrical grid, i.e. as a sequence of circles.

The resulting model of a cylindrical image can be represented [21] in an equivalent form (Eq. 2.28, where $n = kT + l$) as a model of a random process, which is an image scan along the spiral.

$$x_n = \rho\, x_{n-1} + r\, x_{n-T} - \rho\, r\, x_{n-T-1} + \beta\, \xi_n \tag{2.28}$$

Obviously, if r value is close to 1, then the neighboring image lines (spiral turns) will be slightly different from each other. Thus, this model can be used to describe and simulate quasi-periodic signals, e.g. speech signals.

It can be shown that the CF of the model Eq. 2.28 is as following

$$V(n) = \beta^2 \left(\frac{1}{(1-r^2)T} \sum_{k=0}^{T-1} \frac{z_k}{(1-\rho\, z_k)(z_k-\rho)} z_k^n + \frac{s}{(1-\rho^2)(1-rs)(s-r)} \rho^n \right),$$

where $z_k = \sqrt[T]{r} \exp(i2\pi k/T)$ and $s = \rho^T$. In particular, when $n = kT$ we obtain

$$V(kT) = \frac{\beta^2}{(1-\rho^2)(1-r^2)(1-sr)(r-s)} \left((1-s^2) r^{k+1} - (1-r^2) s^{k+1} \right)$$

and find variance, when $k = 0$

$$\sigma^2 = \frac{\beta^2 (1+rs)}{(1-\rho^2)(1-r^2)(1-rs)}$$

and another form of the CF

$$V(kT) = \sigma^2 r^k + \frac{s(r^k - s^k)}{(1-\rho^2)(r-s)}.$$

If $0 \le l \le T$, we have

$$V(l) = \sigma^2 \rho^l + \frac{r(\rho^{T-l} - \rho^{T+l})}{(1-\rho^2)(r-s)(1-rs)}.$$

The scanning of image implementation obtained by use of model Eq. 2.27 is shown in Fig. 2.9a. It is obvious that the values at the line ends are strongly

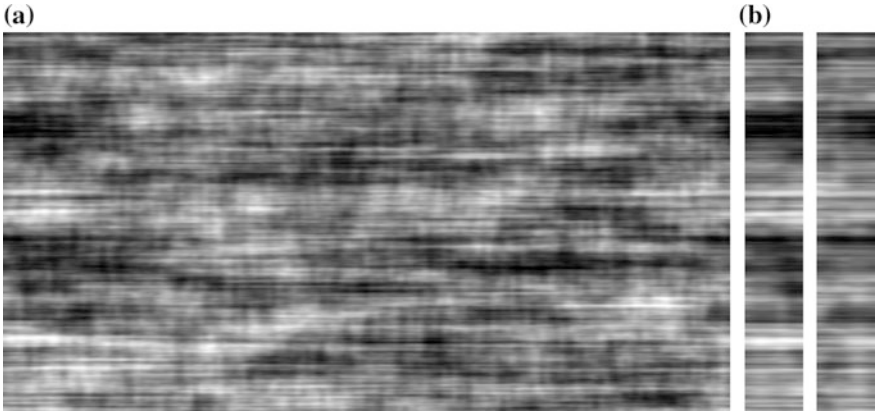

Fig. 2.9 Simulated images: **a** image simulated by model Eq. 2.27, **b** the first and the last five columns of image from (**a**)

correlated (Fig. 2.9b), as it should be for a cylindrical image obtained by image pasting as in Fig. 2.9a.

Wave RF models on an arbitrary surface as cross sections. It was possible to implement the abovementioned autoregressive model on a cylinder and analyze it because there is a simple regular grid on this surface. However, on other surfaces there are no grids with constant configuration and cell size suitable for autoregressive representation. They cannot even be found on a sphere, in spite of its symmetry. Therefore, it is necessary to find other methods of setting the RF on the surfaces. The following approach seems rather promising [16, 20, 34].

Let S be an arbitrary surface. In order to set the RF X on it, one can do a simple thing. Let us take a random field Z in space F containing S and consider X to be values at surface points S of spatial field values Z at these points. In other words, X is a cross-section of field Z with surface S. Thus, it is possible to use any RF model in space. However, one should remember that values must be obtained exactly at the points of surface S. For example, if one uses the AR model set on a rectangular grid, then interpolation into surface points will be necessary. The image on intersection of hyperbolic paraboloid and sphere with three-dimensional autoregressive RF with factorable CF is depicted in Fig. 2.10.

Application of the AR model for the RF construction on surfaces requires calculation of image values at all grid points. Therefore, while simulating images on the surface, a large number of spatial image values will be calculated but not used. Besides, a synthesis of the autoregressive model with predetermined CF is a very complicated process. It prevents imaging on surfaces with a predetermined CF. Wave model is more favorable for imaging on surfaces. This model allows to form values only at desired points (in our case at a set of predetermined points on the surface). The task of correlation synthesis for this model can be solved rather easily.

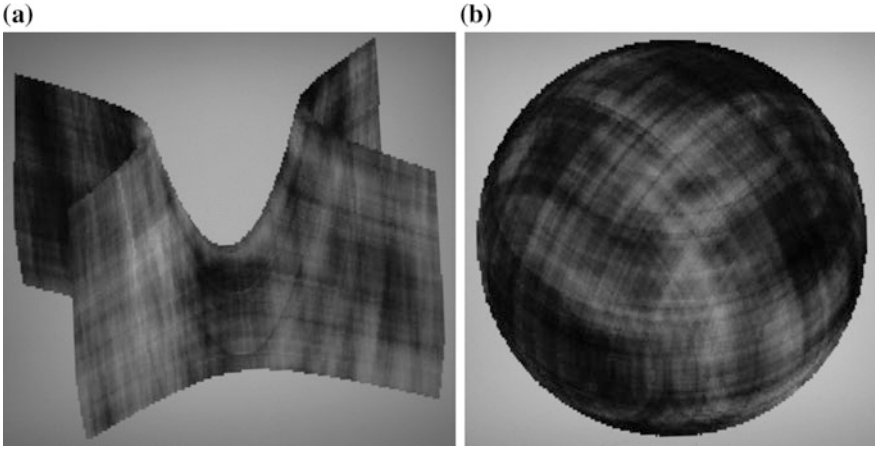

Fig. 2.10 The AR images on: **a** hyperbolic paraboloid, **b** sphere

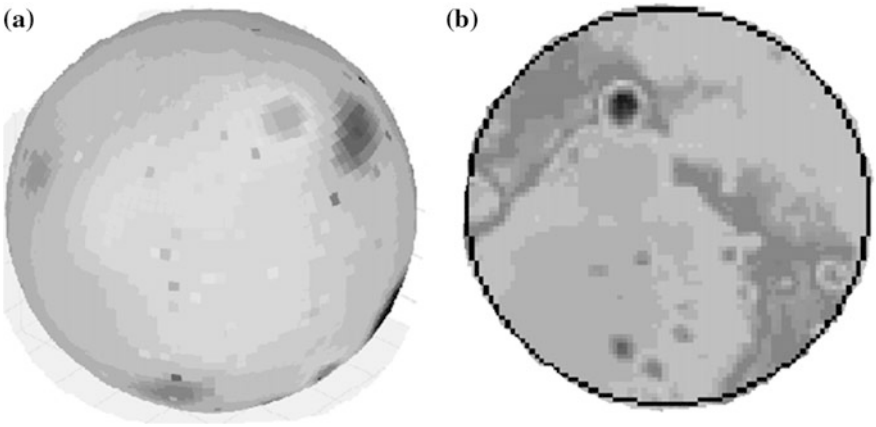

Fig. 2.11 Examples of: **a** relief simulation, **b** real part of lunar surface

Note that the CF of a wave model is a function of the Euclidean distance between spatial points. On the surface, the CF can be determined as a distance function along a geodesic arc. In this case, the CF defined on the surface should be recalculated into the CF according to the Euclidean distance. For example, in the case of a sphere this transformation is easily accomplished.

As an example, consider a relief simulation according to Unified Lunar Control Network (ULCN 2005), consisting of 272,931 points on the lunar surface. These points were obtained from space missions and observations from the Earth. On the basis of these data, an approximating ellipsoid was constructed using the least square method. Deviations from the catalog of the relief heights of this ellipsoid with mean zero samples were taken for image values on an ellipsoid. The CF dependant on the Euclidean distance between two points on an ellipsoid was evaluated according to the data obtained. This CF was approximated by the sum of Gaussoids. The approximation is used to obtain a probability distribution scale R parameter while simulating image on the ellipsoid, i.e. relief deviations from this ellipsoid [22]. Figure 2.11a shows an example of a relief simulation on the ellipsoid segment (greater heights correspond to greater brightness). By comparison, Fig. 2.11b shows some part of the real lunar surface. Visually, these figures are similar. It indicates that the suggested model is rather sufficient.

2.3 Image Filtering

Extraction of useful image component based on their noise observations is an important task, since it gives the opportunity to improve image on the background of noise [4, 12, 15, 23–25]. Furthermore, as it will be shown Sect. 2.4 the anomaly detection problem also makes it necessary to determine the covariance matrices of

filtering errors. However, there exist many different approaches to solve the problem of determining the potential accuracy of estimation and construction of multidimensional RF filtering algorithms. In the first case, in order to obtain the relatively simple analytical relations, it is advisable to refer to asymptotic formulae of the Wiener filter. During constructing multidimensional image estimation on the background noise, it is desirable to find the optimal algorithm structures (or structures close to optimal) with low computational cost. It can be done using recursive estimation methods based on the abovementioned image models [16].

The issue of efficiency of optimal image filtering is considered in Sect. 2.3.1, while tensor representation of the Kalman filter is discussed in Sect. 2.3.2.

2.3.1 *Efficiency of Optimal Image Filtering*

To find an optimal filtering error variance σ_ε^2, consider homogeneous information RF $x_{\bar{j}}$, set on an infinite n-dimensional grid J. Suppose that for use of observations $z_{\bar{j}} = x_{\bar{j}} + \theta_{\bar{j}}$ it is necessary to give the best (with minimal error variance) linear estimation $\hat{x}_{\bar{0}} = \sum_{\bar{j} \in J} h_{\bar{j}} z_{\bar{j}}$ of information RF element $x_{\bar{0}}$. Condition for minimum σ_ε^2 can be written as a system of linear Eq. 2.29 (as an n-dimensional analogue of the Wiener-Hopf equations), where $R(\bar{q}) = M\{x_{\bar{j}},\ x_{\bar{j}+\bar{q}}\}$ is a covariation function. In addition, the smallest error variance is $\sigma_\varepsilon^2 = \sigma_\theta^2 h_{\bar{0}}$.

$$h_{\bar{q}} \sigma_\theta^2 + \sum_{\bar{j} \in J} h_{\bar{j}} R(\bar{q} - \bar{j}) = R(\bar{q}), \quad \bar{q} \in J \tag{2.29}$$

Unfortunately, the analytical solution of Eq. 2.29 can be found only for a very small class of "separable" exponential CF $R(\bar{q}) = \sigma_x^2 \prod_{i=1}^n \rho_i^{q_i}$ [12]. However, supposing that the cells of spatial grid J are small in comparison with the RF, it is possible to replace the system Eq. 2.29 with one integral equation. Then

$$\sigma_\varepsilon^2 = \sigma_\theta^2 h_{\bar{0}} \cong \sigma_\theta^2 \frac{1}{(2\pi)^n} \int_{-\infty}^{\infty} (f(\bar{\lambda})\, d\bar{\lambda}) \,/\, (\sigma_\theta^2 + f(\bar{\lambda})), \tag{2.30}$$

where

$$f(\bar{\lambda}) = \int_{R^n} R(\bar{u}) \exp\left(-j(\bar{\lambda}, \bar{u})\right) d\bar{u} \tag{2.31}$$

is the RF spectral density of RF $x(\bar{u})$, $\bar{u}$, $\bar{\lambda} \in R^n$, $(\bar{\lambda}, \bar{u}) = \sum_{k=1}^n \lambda_k u_k$, $d\bar{\lambda} = d\lambda_1 d\lambda_2 \ldots d\lambda_n$.

Thus, for analysis of filtering efficiency it is sufficient to find a spectral density in Eq. 2.31 of information RF and to make calculations using Eq. 2.30. The greatest challenge is usually connected with n-multiple integration in Eq. 2.31 and, particularly, in Eq. 2.30. These problems may be significantly lower, when the RF $x(\bar{u})$ is isotropic throughout space R^n or any other subspaces $\Omega_m \subset R^n$. Indeed, after entering spherical coordinates in R^n Eqs. 2.30 and 2.31 for isotropic RF can be written in the forms of Eqs. 2.32 and 2.33, where $J_\nu(\cdot)$ is Bessel functions of ν order, $\Gamma(\cdot)$ is a complete gamma function, $k = |\lambda|,\ \rho = |\bar{u}|$.

$$\frac{\sigma_\varepsilon^2}{\sigma_\theta^2} = \frac{1}{(2\pi)^{n-1}\Gamma(0.5n)} \int_0^\infty \frac{k^{n-1} f(k)}{\sigma_\theta^2 + f(k)} dk \tag{2.32}$$

$$f(k) = (2\pi)^{n/2} \int_0^\infty R(\rho)\rho^{n-1} \frac{J_{0.5n-1}}{(k\rho)^{0.5n-1}} d\rho \tag{2.33}$$

For isotropic RF on spaces with an odd number dimensions, the Bessel functions in Eq. 2.32 can be expressed by means of elementary functions. In these cases, the calculations, using Eqs. 2.32 and 2.33, are rather easy. For example, for isotropic RF with

$$R(\rho) = \sigma_x^2 \exp(-\alpha\rho), \quad \rho = \sqrt{\sum_{k=1}^{n} u_k^2},$$

we obtain Eq. 2.34, where $\gamma = \alpha^n/q$, $q = \sigma_x^2/\sigma_\theta^2$ is information and interference RF variance ratio, $\Phi_1(\nu) = 2/(1+\nu^2)$, $\Phi_3(\nu) = 8\pi/(1+\nu^2)^2$, $\Phi_5(\nu) = 64\pi^2/(1+\nu^2)^3$, $\Phi_7(\nu) = 96\pi^3/(1+\nu^2)^4$, etc.

$$\frac{\sigma_\varepsilon^2}{\sigma_\theta^2} = \frac{\gamma q}{(2\pi)^{n-1}\Gamma(0.5n)} \int_0^\infty \frac{\nu^{n-1}\Phi_n(\nu)}{\gamma + \Phi_n(\nu)} d\nu \tag{2.34}$$

For isotropic exponentially correlated RF that is set on spaces with even number of dimensions $n = 2N$, there is a simple formula for integral Eq. 2.33

$$f(k) = (2\pi)^M \sigma_x^2 k^{-2(M-1)} (-1)^M \frac{d^M}{d\alpha^M} \left(\frac{(\sqrt{\alpha^2+k^2} - \alpha)^{M-1}}{\sqrt{\alpha^2+k^2}} \right).$$

In addition, a minimal variance of filtering error can also be presented as Eq. 2.34, where $\Phi_2(\nu) = 2\pi/\sqrt{(1+\nu^2)^3}$, $\Phi_4(\nu) = 6\pi/\sqrt{(1+\nu^2)^5}$, etc. Unfortunately, for large dimension n of homogeneous RF rather complicated formulae can be obtained and it is advisable to use numerical methods while

calculating Eq. 2.34. At the same time, according to the analysis, if the generalized parameter $\gamma = \alpha^n/q$ increases, then variance of optimal filtering error converges to

$$\sigma_\varepsilon^2/\sigma_x^2 = \beta_n \sqrt[n+1]{\gamma},$$

where $\beta_1 \cong 0.707,\ \beta_2 \cong 0.66,\ \beta_3 \cong 0.63,\ \beta_4 \cong 0.606, \ldots \beta_\infty \cong 0.58$.

Thus, the considered relations give a possibility to obtain rather simple estimations of potential RF filtering accuracy on the background of noise. In addition, the obtained limit values of filtering variance allow to construct the characteristics of optimal algorithms to detect the spatial anomalies on the background of multidimensional interfering images.

2.3.2 *Tensor Kalman Filter*

Analyzed Wiener filtering procedures can be used in applications with a limited number of k frames and small grids G and J_t. However, there is a significant variety of problems, when observations are carried out continuously and the number of elements in the domain $J_t \otimes T$ is arbitrary large. Nevertheless, a prediction using the weighted summation of all previous to the certain frame observations $\left\{z_{\bar{j}}^t\right\}$ can be unreasonably time-consuming. In such situation, it is desirable to impose the additional restrictions to the considered image models and use the effective recurrent procedures to construct some optimal prediction.

One of the least overloaded restrictions allowing to find a recurrent solution to the problem is a description of interfering image frame sequence by means of nonlinear tensor stochastic differential Eq. 2.1. Based on observations of frame sequence

$$z_{\bar{j}}^t = x_{\bar{j}}^t + \theta_{\bar{j}}^t \quad \bar{j} \in J_t \quad t = 1, 2, \ldots,$$

representing an additive mixture of information RF $x_{\bar{l}}^t$ and the RF of white noise $\theta_{\bar{l}}^t$, it is necessary to find the best estimate $x_{\bar{l}}^t$ of a recurrent frame of information RF. To find such estimation, let us use the criterion of average maximal gain [16] and an invariant imbedding method. As a result, the recursive estimation rule (Eq. 2.35) can obtained, where $\hat{x}_{\ni\bar{j}}^t = \varphi_{\bar{j}}^t(\hat{x}_{\bar{l}}^{t-1}),\ \bar{j} \in J_{t-1}$ is an optimal RF prediction, $P_{\ni}^t = (\varphi^t(\hat{x}_{\bar{l}}^{t-1}))'\, P^{t-1}(\varphi^t(\hat{x}_{\bar{l}}^{t-1}))' + \upsilon^t(\hat{x}_{\ni\bar{l}}^t)\upsilon^t(\hat{x}_{\ni\bar{l}}^t)$ is a covariance matrix of prediction errors, $x_{\ni\bar{j}}^1 = 0;\ P_{\ni\bar{j}\bar{l}}^1 = M\left\{x_{\bar{j}}^1 x_{\bar{l}}^1\right\}$.

$$\begin{gathered}\hat{x}_{\bar{j}}^t = \hat{x}_{\ni\bar{j}}^t + P_{\bar{j}\bar{l}}^t V_\theta^{-1}(z_{\bar{l}}^t - \hat{x}_{\ni\bar{l}}^t),\ \bar{l},\ \bar{j} \in J_t \\ P_{\bar{j}\bar{l}}^t = P_{\ni\bar{j}\bar{l}}^t (E + V_\theta^{-1} P_{\ni\bar{j}\bar{l}}^t)^{-1}\end{gathered} \tag{2.35}$$

Equations 2.35 allow to find extrapolated estimates $\widehat{x}^t_{\ni\bar{j}}$ and covariance matrices of extrapolation errors $P^t_{\ni\bar{j}\bar{l}}$ recursively as new observations $\left\{z^i_{\bar{j}},\ \bar{j}\in J\right\}$ of successive RF frames become available. For Gaussian RF defined by linear stochastic equations, when $\varphi^t_{\bar{j}}(x^{t-1}_{\bar{l}}) = P^t_{\bar{j}\bar{l}}x^{t-1}_{\bar{l}}$ and $\upsilon^t_{\bar{j}\bar{q}}(x^{t-1}_{\bar{l}}) = \upsilon^t_{\bar{j}\bar{q}}(x^{t-1}_{\bar{l}})$, the procedure of filtration-interpolation together with algorithm Eq. 2.35 gives a strictly optimal solution how to detect the multidimensional signals in image sequence. In this case, domain G coincides with J_t, Eq. 2.35 determines the sequence of interframe observation processing and, while generating the log-likelihood ratio, the weighted summation of observations $z^t_{\bar{j}},\ \bar{j}\in J_t$ and predictions $\widehat{x}^t_{\ni\bar{j}},\bar{j}\in J_t$ of a successive frame of a multidimensional image is used.

Let now X be the field with multiplicative correlation function $M\left\{x^t_{\bar{j}}\times x^s_{\bar{l}}\right\} = \sigma^2_x\rho^{|t-s|}\prod_{k=1}^n r_k^{|j_k-t_k|}$, where ρ is a time correlation coefficient, r_k is a coefficient along the kth spatial axis, σ^2_x is a field variance. Then, when $t = s$, an intraframe covariance is obtained and, when $t = s - 1$, an interframe covariance $V^{t|t-1}_x = \rho\, V^{t|t}_x$ is yielded, where $R = R_1\times R_2\times\cdots\times R_m$, R_k is a correlation tensor of the kth row.

In this case $\rho^t = \rho\, E, \nu^{t-1}(\nu^{t-1})^T = V^t_x = V_x$, and Eq. 2.35 can be written as

$$\widehat{x}^t = \rho\widehat{x}^{t-1} + P^t\left(z^t - \rho\widehat{x}^{t-1}\right),\ P^t = P^t_{\bar{j}}\left(E + P^t_{\ni}\right)^{-1},$$
$$P^t_{\ni} = \rho^2P^t + (1-\rho^2)\,qR,\ P^1_{\ni} = qR$$

where $q = \sigma^2_x/V_\theta$ is a signal/noise ratio, tensors $P^t_{\ni}$ and P^t are normalized by noise variance and represent relative error covariance of the extrapolated and current estimates expressed in noise variance.

As an example, consider a field with 3×2 grid J_t, i.e. the case, when frames x^t consist of six points

$$x^t = \begin{pmatrix} x^t_{11} & x^t_{21} & x^t_{31}\\ x^t_{12} & x^t_{22} & x^t_{32}\end{pmatrix}.$$

Thus, correlation matrices of the first and second row are as follows

$$R_1 = \begin{pmatrix} 1 & r_1 & r_1^2\\ r_1 & 1 & r_1\\ r_1^2 & r_1 & 1\end{pmatrix}\quad R_2 = \begin{pmatrix} 1 & r_2\\ r_2 & 1\end{pmatrix}.$$

Consequently,

$$V_x = \sigma_x^2 R_1 \times R_2 = \begin{pmatrix} 1 & r_1 & r_1^2 & r_2 & r_2 r_1 & r_2 r_1^2 \\ r_1 & 1 & r_1 & r_2 r_1 & r_2 & r_2 r_1 \\ r_1^2 & r_1 & 1 & r_2 r_1^2 & r_2 r_1 & r_2 \\ r_2 & r_2 r_1 & r_2 r_1^2 & 1 & r_1 & r_1^2 \\ r_2 r_1 & r_2 & r_2 r_1 & r_1 & 1 & r_1 \\ r_2 r_1^2 & r_2 r_1 & r_2 & r_1^2 & r_1 & 1 \end{pmatrix}$$

Note that error covariance matrices $P^t_{\ni}$ and P^t have the same form.

Tensor P^t elements are filtering error covariance x^k, which in this case depend only on four parameters: the correlation coefficients r_1, r_2, and ρ and also signal/noise ratio $q = \sigma_x^2 / V_\theta$. When $q \gg 1$ and $t \to \infty$, the coefficients $P^t_{j_1 j_2}$ converge to limit $\left\{P^t_{j_1 j_2}\right\} = P_{j_1 j_2}$ rather quickly. Thus, it is often possible to apply limit values at once. It will worsen the filtering results only at the beginning but the amount of computation will be significantly reduced (or the storage space if coefficients $\left\{P^t_{j_1 j_2}\right\}$ are calculated beforehand).

It is rather important that the deduced tensor filtering Eq. 2.35 can be easily generalized in case of almost arbitrary interaction

$$z^t_{\bar{j}} = S^t_{\bar{j}}(x^t_{\bar{l}}, \theta^t_{\bar{l}})\ \bar{j},\ \bar{l} \in J_t$$

of information RF and noise [16]. Besides, using a modified method of invariant imbedding and the abovementioned models it is possible to synthesize recurrent procedures to test multialternative hypotheses such as

$$H_v : \left\{z^t = S^t_v(x^t_v, \theta^t)\right\}.$$

2.4 Anomalies Detection in Noisy Background

In many applications, it is often desirable to detect anomalies that may appear in the signal, separate image, or recurrent image of frame sequence [6, 8, 14, 26, 35, 36]. For example, these anomalies may be forest fires, pathological changes in medical images, new objects in a security area, etc. Optimal algorithms for signal detection are suggested in Sect. 2.4.1. Efficiency of anomaly detection is represented in Sect. 2.4.2.

Fig. 2.12 Domain G, where and only where a signal becomes apparent, Q is a complimentary domain

2.4.1 Optimal Algorithms for Signal Detection

Let observations of the RF $x_{\bar{j}}^t$ with the space-time correlation and additive noise RF $\theta_{\bar{j}}^t$, which consists of independent random variables with zero means and variances V_θ, be made by Eq. 2.36, where parameter vector $\bar{\chi}_t$ describes, for example, possible mutual spatial displacements and rotations of nearest-neighbor image frames.

$$z_{\bar{j}}^t = x_{\bar{j}}^t(\bar{\chi}_t) + \theta_{\bar{j}}^t \quad \bar{j} \in J_t \quad t = 1 \ldots k \tag{2.36}$$

Let appearance of a deterministic signal cause changes in the model Eq. 2.36 only in $G \subset J$ (Fig. 2.12). In particular, the domain G can be a part of the last frame observed (Eq. 2.37), where $\left\{s_{\bar{j}}, \quad \bar{j} \in G\right\}$ are the values of a useful (detectable) signal.

$$z_{\bar{j}}^k = s_{\bar{j}} + x_{\bar{j}}^k(\bar{x}_t) + \theta_{\bar{j}}^k \quad \bar{j} \in G \tag{2.37}$$

Under considered circumstances, it is necessary to find out a rule to test the hypothesis H_0 on the absence of anomalies in the domain G against alternative assumption H_1 on the validity of the model Eq. 2.37.

At the predetermined probabilistic characteristics of model components Eqs. 2.36 and 2.37, it is possible to estimate corresponding conditional PDF of observations $W(Z|H_0)$ and $W(Z|H_1)$. Therefore, to solve the detection problem one should compare the results with the threshold level Λ_0 of Likelihood Ratio (LR).

$$\Lambda = \frac{W(Z|H_1)}{W(Z|H_0)} \begin{cases} \geq \Lambda_0 - \textit{signal}, \\ < \Lambda_0 - \textit{no signal} \end{cases} \tag{2.38}$$

To simplify the calculations, let us represent the conditional PDF in a product form $W(Z|H_{0,1}) = W(Z_0|H_{0,1})\, W(Z_G|Z_0, H_{0,1})$, where Z_G is a set of observations in domain G, Z_0 is a set of all observations, which do not belong to the domain of the intended signal. Since $W(Z_0|H_0) = W(Z_0|H_1)$, the LR in Eq. 2.38 can be rewritten in the form

$$\Lambda = \frac{W(Z_G|Z_0, H_1)}{W(Z_G|Z_0, H_0)}. \tag{2.39}$$

Let us approximate the conditional PDF, included in the LR Eq. 2.39, using Gaussian distribution (Eq. 2.40), where $m_0 = \{m_{0\bar{j}}\}$, $V_0 = \{V_{0\bar{l}\bar{j}}\}$, and $m_1 = \{m_{1\bar{j}}\}$, $V_0 = \{V_{0\bar{l}\bar{j}}\}$, $\bar{i},\bar{j} \in G$ are conditional means and spatial covariance matrices of observations Z_G in the absence and presence of a useful signal, respectively.

$$W(Z_G|Z_0, H_{0,1}) = \frac{1}{(2\pi)^{N/2}\sqrt{\det V_{0,1}}} \exp\left(-0.5||z^k - m_{0,1}||^2_{V^{-1}_{0,1}}\right) \tag{2.40}$$

Using the observation models Eqs. 2.36 and 2.37, the following formulae for the conditional mean values is obtained $m_{0\bar{j}} = \hat{x}^k_{\ni\bar{j}}$, $m_{1\bar{j}} = s_{\bar{j}} + x^k_{\ni\bar{j}}$, $\bar{j} \in G$, where $\hat{x}^k_{\ni\bar{j}} = M\left\{x^k_{\bar{j}}|Z_0\right\}$ is the optimal prediction of the RF values $x^k_{\bar{j}}$, $\bar{j} \in G$ obtained on the basis of all available observations Z_0, which do not belong to a signal domain. Matrices V_0 and V_1 are the same $V_0 = V_1 = V = P_\ni + V_\theta$, where $P_\ni$ is a covariance matrix of optimal prediction errors. Substituting the obtained formulae in Eqs. 2.39 and 2.40 and taking logarithms, it is possible to find the following algorithm for signal detection (Eq. 2.41), where $\lambda = \ln\Lambda_0 + 0,5\left\|s_{\bar{j}}\right\|^2_{V^{-1}}$ is a detection threshold.

$$L = \sum_{\bar{l},\bar{j}\in G} s_{\bar{j}} V^{-1}_{\bar{l}\bar{j}}(z^k_{\bar{j}} - \hat{x}^k_{\ni\bar{j}}) \begin{cases} \geq \Lambda_0 - \textit{signal}, \\ < \Lambda_0 - \textit{no signal} \end{cases} \tag{2.41}$$

Thus, the procedure of anomaly detection includes a compensation of interfering images $z^k_{\bar{j}} - \hat{x}^k_{\ni\bar{i}}$ by subtracting optimal prediction $\hat{x}^k_{\ni\bar{j}}$ from observation $z^k_{\bar{j}}$. The prediction is found on the basis of all observations, which do not belong to domain G. Let us call this type of prediction as "domain prediction" (Fig. 2.13a). After compensation of interference, the RF linear weighted summation of all residuals is carried out.

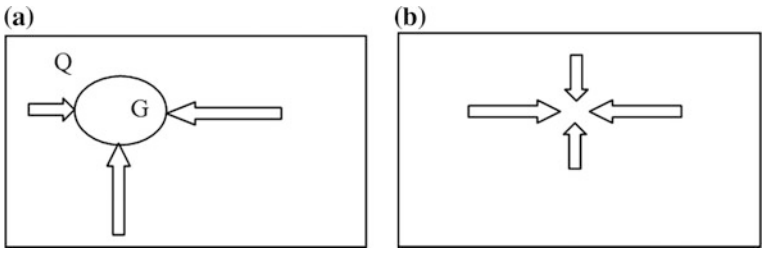

Fig. 2.13 Anomaly detection: **a** domain prediction, **b** point prediction

Another form of optimal detection procedure can be obtained assuming that any useful signal can occupy all image frames, i.e. domain G includes all multi-dimensional grids $J_1, J_2, \ldots, J_t$. Then, the best prediction $x^k_{\ni \bar{j}} \equiv 0$, $\bar{j} \in G$, and algorithm Eq. 2.41 can be written as Eq. 2.42, where $V_{\bar{l}\bar{j}} = V_{x\bar{l}\bar{j}} + V_\theta E_{\bar{l}\bar{j}}$, $V_{x\bar{l}\bar{j}}$, $\bar{l}, \bar{j} \in G$ is a covariance matrix of an interfering image.

$$L = \sum_{\bar{l},\bar{j} \in G} s_{\bar{l}} V^{-1}_{\bar{l}\bar{j}} z^k_{\bar{j}} \begin{cases} \geq \Lambda_0 - \textit{signal} \\ < \Lambda_0 - \textit{no signal} \end{cases} \tag{2.42}$$

Direct implementation of procedure Eq. 2.42 is difficult because of a large number of computations. However, an expanding the spatial symmetric matrix into the product of two triangle ones $V^{-1}_{\bar{l}\bar{j}} = A_{\bar{l}\bar{v}} A_{\bar{v}\bar{l}}$, $\bar{l}, \bar{v}, \bar{j} \in G$, it is possible to present Eq. 2.42 in the form of Eq. 2.43, which corresponds to the preliminary decorrelation of image sequences $z^t_{\bar{j}}$, $\bar{j} \in G$, and subsequent weighted summation with weights $s_{\bar{l}} A_{\bar{l}\bar{v}}$.

$$L = \sum_{\bar{l},\bar{j},\bar{v} \in G} s_{\bar{l}} A_{\bar{l}\bar{v}} A_{\bar{v}\bar{j}} z^t_{\bar{j}} \begin{cases} \geq \Lambda_0 - \textit{signal}, \\ < \Lambda_0 - \textit{no signal} \end{cases} \tag{2.43}$$

In many cases, such approach helps to find acceptable in practice quasi-optimal algorithms of decorrelation. At the same time an analysis of adaptive recursive filters, which properties are similar to decorrelation ones, is an important area to look for relatively simple technical or software implementation of system for image sequence processing. The new quality of algorithms Eqs. 2.42 and 2.43 in comparison to Eq. 2.41 is the division of time-consuming operation of optimal prediction or decorrelation, which is not connected with a signal form, and relatively simple weighted summation, which takes into account the type of a useful signal. It allows to solve both problems of anomaly detection with unknown location and more complex problems of multialternative detection (recognition) of several signal types rather easily.

There is one more form (Eq. 2.44) of decision rule, which gives the same result [16]. Let us represent $V^{-1}_{\bar{l}\bar{j}} z^k_{\bar{j}}$ in Eq. 2.42 in the form $V^{-1}\bar{z} = \begin{pmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{pmatrix}^{-1} \begin{pmatrix} \bar{z}_1 \\ \bar{z}_2 \end{pmatrix}$, where $\bar{z} = \begin{pmatrix} \bar{z}_1 \\ \bar{z}_2 \end{pmatrix}$ is divided into two part like $\bar{z}_1 = (z^k_1, \ldots, z^k_m)^T$ and $\bar{z}_2 = (z^k_{m+1}, \ldots z^k_n)^T$, $V_{ij} = M[\bar{z}_i \bar{z}^T_j]$. Using Frobenious formula we obtain

$$V^{-1}\bar{z} = V^{-1} \begin{pmatrix} \bar{z}_1 \\ \bar{z}_2 \end{pmatrix} = \begin{pmatrix} T^{-1} & -T^{-1} V_{12} \\ -V^{-1}_{22} V_{21} T^{-1} & V_{22} + V^{-1}_{22} V_{21} T^{-1} V_{12} V^{-1}_{22} \end{pmatrix} \begin{pmatrix} \bar{z}_1 \\ \bar{z}_2 \end{pmatrix},$$

where $T = V_{11} - V_{12}V_{22}^{-1}V_{21}$. Then, the first m components of $V^{-1}\bar{z}$ are $T^{-1}(\bar{z}_1 - V_{12}V_{22}^{-1}\bar{z}_2)$, where $V_{12}V_{22}^{-1}\bar{z}_2 = M[\bar{z}_1|\bar{z}_2] = \tilde{\bar{z}}_1$ is the optimal prediction of $\bar{z}_1$ according observation $\bar{z}_2$ and T is covariance matrix of its errors $\Delta = \bar{z}_1 - V_{12}V_{22}^{-1}\bar{z}_2 = \bar{z}_1 - \tilde{\bar{z}}_1$. If $\bar{z}_1$ consists of one element (i.e. $\bar{z}_1 = (z_1)$), then $T^{-1}(\bar{z}_1 - V_{12}V_{22}^{-1}\bar{z}_2) = (z_1 - \tilde{z}_1)/\sigma_1$, where $\tilde{z}_1 = M[z_1|(\bar{z}\backslash z_1)]$ is the optimal prediction of z_1 according to all other elements of $\bar{z}$, σ_1^2 is a variance of error $\Delta_1 = z_1 - \tilde{z}_1$. But any element of $\bar{z}$ can be taken as z_1. Thus,

$$V^{-1}\bar{z} = C^{-1}(\bar{z} - \tilde{\bar{z}}) = C^{-1}\Delta,$$

where $\tilde{\bar{z}}$ is the set of predictions of all elements $z_{\bar{j}}^k$ and every element is predicted according to all other elements of $\bar{z}$, i.e. $\tilde{z}_{\bar{j}} = M[z_{\bar{j}}|(\bar{z}\backslash z_{\bar{j}})]$, diagonal matrix C consists of variances $\sigma_{\bar{j}}^2 = M[(\bar{z} - \tilde{\bar{z}})^2]$ of errors $\Delta = \bar{z} - \tilde{\bar{z}}$. Thus $C^{-1}\Delta = \left\{(z_{\bar{j}}^k - \tilde{z}_{\bar{j}}^k)/\sigma_{\bar{j}}^2 : \bar{j} \in G\right\}$ and we obtain Eq. 2.44 which is equivalent to Eqs. 2.41–2.43.

$$L = \sum_{\bar{j}\in G} s_{\bar{l}}\delta_{\bar{j}} = \sum_{\bar{j}\in G} s_{\bar{l}}(z_{\bar{j}}^k - \tilde{z}_{\bar{j}}^k)/\sigma_{\bar{j}}^2 \begin{cases} \geq \Lambda_0 - \textit{signal}, \\ < \Lambda_0 - \textit{no signal} \end{cases} \tag{2.44}$$

It is based on the "point prediction" $\tilde{z}_{\bar{j}}^k = M[z_{\bar{j}}^k|(Z\backslash z_{\bar{j}}^k)]$ (Fig. 2.13b), which is made taking into account all other points of Z and assuming that there is no signal.

Despite the equality of statistics Eqs. 2.42 and 2.44, there is a fundamental difference between them. In Eq. 2.42, prediction and compensation are made according to observations that do not have a signal. Thus, if there is a signal in G, then it will be distorted only by prediction errors. If these errors are small, the compensation residuals will be close to the values of detectable S_G (it is possible to see the signal with little distortion). In Eq. 2.44 while constructing a point prediction, all other observations are used including those with a signal. Therefore, in the residuals of this compensation every signal value will be distorted not only by prediction errors of an interfering image but by other signal elements as well. Even if prediction errors are small, more distorted signal will be observed.

A significant drawback of all considered algorithms is quite a complex analysis of signal detection efficiency. While searching the ways to simplify this analysis, we managed to obtain one more procedure to detect signals [16]. For this purpose, it is enough to substitute a known connection between the tensor estimations in Eq. 2.41

$$\hat{x}_j^t = \hat{x}_{\ni j}^t + P_{lj}^t V_\theta^{-1}(z_l^t - \hat{x}_{\ni j}^t) \quad (l,j) \in G_0^t \quad t = 1, \ldots, N,$$

where $\hat{x}_j^t$ is an optimal RF estimation in domain $j \in G_0^t$, $\quad t = 1, \ldots, N$, made on the basis of all observations z_l^t, $\quad l \in G_0^t$, $\quad t = 1, \ldots, N$ and P_{lj}^t is a covariance matrix of filtering errors. After elementary transformations and considering that

$P(E+V_\theta^{-1}P_\ni)=P_\ni$, $z_l-\hat{x}_{\ni j}=(E+V_\theta^{-1}P_\ni)(z-\hat{x})$, the decision rule based on the optimal estimates $\hat{x}_j^t$ (Eq. 2.45) can be found.

$$L=s_l^tV_{\theta lj}^{-1}(z_j^t-\hat{x}_j^t)\begin{cases}\geq\Lambda_0-signal\\<\Lambda_0-no\ signal\end{cases} \tag{2.45}$$

Thus, at Gaussian approximation the obtained detection procedure suggests the optimal RF filtration, calculation of covariance matrix of filtering errors, and weighted summation using Eq. 2.45. Since algorithm Eq. 2.45, in contrast to known detectors does not require the time-consuming calculation of covariance matrices of prediction errors, it can be used not only in image processing system but also for probability analysis of detection efficiency. It is important to note that a total amount of computation according to rule Eq. 2.45 is almost the same as in Eq. 2.44 and can be used as another variant to construct a detector at low computational costs and at an unknown spatial position of anomalies. It is especially evident if the relationship between the estimation and point prediction is used

$$x_j^t=\tilde{z}_j^t+P_j^t(1/\sigma_j^2)(z_j^t-\tilde{z}_j^t),$$

where all formula components are scalar and there is no summation over the same lower indices. Substituting this equation in Eq. 2.45, Eq. 2.44 is obtained.

These results allow to specify the conditions, under which a proposed replacement of conditional PDF by normal distribution is valid. First of all, it is a wide class of the Gaussian models Eqs. 2.36 and 2.37. In these cases, the procedures Eqs. 2.41–2.45 are optimal. When model Eqs. 2.36 and 2.37 components are non-Gaussian, the sufficient optimum condition is the possibility to approximate the posterior PDF prediction $x_{\ni\bar{j}}^k$ using normal distribution. Note that the last condition is fulfilled in many applied tasks of the RF processing with significant space-time correlative relationship and is usually equivalent to the high posteriori prediction accuracy.

2.4.2 *Efficiency of Anomaly Detection*

On the basis of synthesized optimal algorithms for anomaly detection, the relatively simple quasi-decision rules can be built. These rules use only some part of available observations, different prediction methods or decorrelation. In these cases, as well as while studying the potential possibilities of real detection systems there arises a problem of calculating optimal algorithm characteristics.

In the Gaussian approximation, the conditional statistics distributions Eqs. 2.41–2.44 will also be Gaussian. Therefore, for calculation of the efficiency of anomaly detection it is sufficient to find conditional moments of statistics in Eq. 2.41

$$M\{L/H_0\} = 0 \quad M\{L/H_1\} = S_{\bar{e}} V_{\bar{j}\bar{e}}^{-1} S_{\bar{j}} \quad \sigma_L^2 = S_{\bar{e}} V_{\bar{j}\bar{e}}^{-1} S_{\bar{j}}.$$

Thus, at a given quantile x_F of level P_F of a normal distribution a detection threshold $\lambda = x_F \sigma_L$ and probability of correct detection can be determined by Eq. 2.46, where $\Delta_D = \sqrt{S_{\bar{e}} V_{\bar{j}\bar{e}}^{-1} S_{\bar{j}}}, \quad \bar{j},\ \bar{e} \in G, \quad \Phi_\circ(x) = (1/\sqrt{2\pi}) \int_0^x \exp(-z^2/2)\, dz.$

$$P_D = 0,5 + \Phi_0(\Delta_D - x_F) \tag{2.46}$$

It can be shown that there is a simple relationship between the spatial covariance matrices $P_{\ni \bar{j}\bar{e}}$ of optimal prediction errors and $P_{\bar{j}\bar{e}} = M\left\{(x_{\bar{j}}^t - \hat{x}_{\bar{j}}^t)(x_{\bar{e}}^t - \hat{x}_{\bar{e}}^t)\right\}$ of optimal linear estimation errors $\hat{x}_{\bar{j}}^t = \hat{x}_{\bar{j}}^t(Z)$. Indeed, analyzing Wiener-Hopf equation for multidimensional discrete RF, it is possible to obtain the equation

$$P_{\ni \bar{j}\bar{e}} = (E_{\bar{j}\bar{q}} - P_{\bar{j}\bar{v}} V_{\theta\bar{v}\bar{q}}^{-1})^{-1} P_{\bar{q}\bar{e}},$$

where $V_{\theta\bar{v}\bar{q}} = \sigma_\theta^2 E_{\bar{v}\bar{q}}$. Substituting the obtained link between matrices in Eq. 2.46, Eq. 2.47 provides efficacy calculation of signal detection.

$$\Delta_D = \sqrt{S_{\bar{e}} V_\theta^{-1} (E_{\bar{e}\bar{j}} - V_\theta^{-1} P_{\bar{e}\bar{j}}) S_{\bar{j}}} \tag{2.47}$$

If there are no interfering images $x_{\bar{j}}$ or if they are estimated precisely $P_{\bar{e}\bar{j}} = 0$, the deterministic signal detection performance on the background of white noise is determined by a well-known signal/noise ratio $\Delta_D = \sqrt{\sum_{\bar{j} \in G} S_{\bar{j}}^2 / V_\theta}$. Errors $x_{\bar{j}}^t$ with time-space correlation decrease the detection accuracy, which is measured by non-zero covariance matrix of optimal RF estimation $x_{\bar{j}}^t$ on the background of white noise $\theta_{\bar{j}}^t$, $\bar{j} \in J$.

Consider two important examples how it is possible to calculate the potential efficiency of point anomaly detection. Suppose that a point signal $s_{\bar{j}} = s_0$ is being detected. It occupies one element J_t of grid G. In this case

$$\Delta_D = \sqrt{Q(1 - \sigma_\varepsilon^2 / \sigma_\theta^2)},$$

where $Q = s_0^2 / \sigma_\theta^2$ is a signal/noise ratio, σ_ε^2 is an error variance at optimal estimation of one RF element $x_{\bar{j}}^t$, $\bar{j} \in G$, on the basis of all available observations $z_{\bar{j}}^t$, $\bar{j} \in J_t$, $t \in T$. Magnitude of error variance σ_ε^2 can be obtained by known methods of the RF optimal estimation theory, which are observed on the background of noise [16].

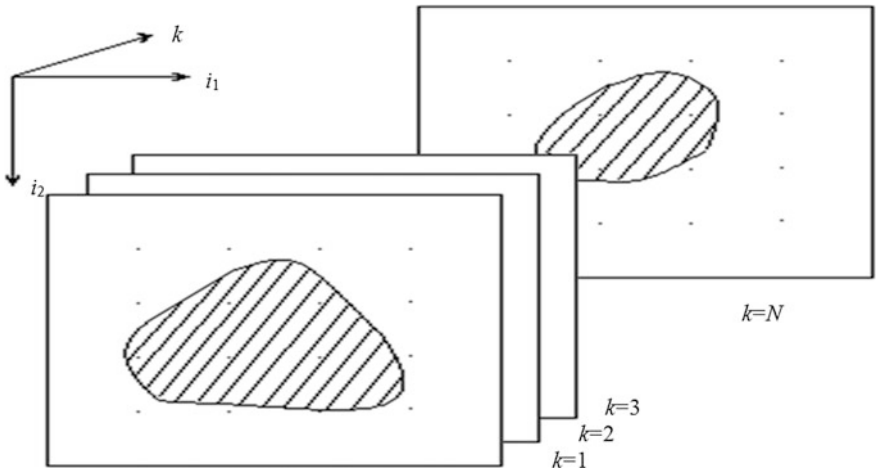

Fig. 2.14 Frames of a multi-zone image

Consider the task of finding a point object based on the results of observations of imagery M_3 of $M_1 \times M_2$ size (Fig. 2.14). Such a problem can arise, for example, during multi-zone observations of one and the same object using different x-systems functioning in different spectral range [7, 11, 12, 37]. In this case, an observation model Eq. 2.36 can be written as following

$$z_{\bar{j}} = x_{\bar{j}} + \theta_{\bar{j}},\ \bar{j} = (j_1, j_2, j_3);\ j_k = 1 \ldots M_k.$$

Assume that data sets $z_{\bar{j}}, j_3 = 1 \ldots M_3$ are spatially shifted. Therefore, the appearance of the desired signal increase the RF level by s_0 in one and the same element each image frame, of which is numbered $\left(j_1^0, j_2^0\right)$ (Fig. 2.14). Covariance functions of interfering images can be written as follows

$$R_x(m_1, m_2, m_3) = M\left\{x_{j_1,j_2,j_3} x_{(j_1+m_1),(j_3+m_3),(j_3+m_3)}\right\} = R_1(m_1, m_2)\rho^{1(m_3)},$$

where $R_1(m_1, m_2) = R_x(m_1, m_2, m_3 = 0)$ is one and the same covariance function from M_3 image frames, $1(m_3) = 1$ if $m_3 \neq 0$, and $1(m_3 = 0) = 0$. Note that such representation means equality of correlation distances between any pair from M_3 analyzed image frames. After simple calculations it is possible to say that Eq. 2.47 for a given task can be transformed into

$$\Delta_D = \sqrt{M_3} Q \sqrt{1 - F_R(\bar{0}) \frac{(1 + (M_3 - 1)\rho) G_1(\omega_1, \omega_2)}{\sigma_\theta^2 + (1 + (M_3 - 1)\rho) G_1(\omega_1, \omega_2)}},$$

where $Q = s_0 / \sigma_\theta^2;\ G_1(\omega_1, \omega_2) = F\{R_1(\omega_1, \omega_2)\};\ F(\omega_1, \omega_2)\{\cdot\}$ and $F_R(m_1, m_2)\{\cdot\}$ are direct and inverse two-dimensional discrete Fourier transformations, $F_R(\bar{0})\{\cdot\} = F_R(m_1 = 0,\ m_2 = 0)\{\cdot\}$.

The analysis of the obtained equation leads to the following important conclusions about the effectiveness of the point anomaly optimal detection on an arbitrary number of simultaneously processed images. In the absence of interframe correlation ($\rho = 0$), the co-processing of M_3 frames yields to gain a threshold signal in $\sqrt{M_3}$ times in comparison to signal detection on one image frame notwithstanding the covariance function $R_1(m_1, m_2)$ and size $M_1 \times M_2$ of a frame. The interframe correlation ($\rho \neq 0$) leads to the losses in detection efficiency, which correspond to additional increase in variance σ_x^2 of interfering images in $(1 + (M_3 - 1)\,\rho)$ times.

The abovementioned properties make it easy to recalculate the characteristics of point anomaly detection in a single image frame in the case of co-processing of arbitrary number of mutually-correlated image frames.

2.5 Image Alignment

While synthesizing algorithms for solving different tasks of frame sequence processing $z^1, z^2, \ldots$, it is usually assumed that observations $z_{\bar{j}}^1, z_{\bar{j}}^2, \ldots$ in the node $\bar{j}$ of a grid Ω correspond to the same space point. In the real world, the receiver motion, its imperfection, and other factors lead to the fact that $z_{\bar{j}}^1, z_{\bar{j}}^2, \ldots$ correspond to the different space points. As a result, the sequence $z^1, z^2, \ldots$ will demonstrate interframe displacement from frame to frame, i.e. Interframe Geometric Transformation (IGT), such as shifts, rotations, etc. If such distortions are not taken into account, the efficiency of processing algorithms can dramatically decrease. In particular, the IGT evaluation is used in video stabilization [28].

However, the IGT can be not only disturbing factors but also contain useful information. For example, analyzing the IGT of frames obtained at different time intervals one can track the aircraft or submarine course under limited visibility.

Further, tensor shift filtering is yielded in Sect. 2.5.1. Random field alignment of images with interframe geometric transformation is described in Sect. 2.5.2. Section 2.5.3 provides a discussion about alignment of two frames of Gaussian random field. Method of fixed point at frame alignment is represented in Sect. 2.5.4.

2.5.1 Tensor Shift Filtering

Tensor Kalman filter described in Sect. 2.3.2, which estimates frame sequence defined by tensor Model, can be extended in order to joint estimation of frames and interframe shift of these frames. Consider this extension. Let a sequence of m-dimensional frames be defined by a linear tensor stochastic equation

$$x^t = \rho^t x^{t-1} + \vartheta^t \xi^t \quad t = 1, 2, \ldots$$

and their observations look like Eq. 2.48, where $y^t = (y_1^t, \ldots, y_m^t)$ is the IGT parameter vector of the tth frame, $x^t(y^t)$ is a proper observation of a frame x^t with parameters y^t, θ^t, is a white Gaussian RF of observation noise.

$$z^t = x^t(y^t) + \theta^t \tag{2.48}$$

Let the sequence of shift vectors be also described by a linear stochastic equation

$$y^t = \Im^t y^{t-1} + \Phi^t \eta^t \quad t = 1, 2, \ldots,$$

with $(m \times n)$-matrices $\Im^t, \Phi^t$ and a white Gaussian generating vector $\eta^t = (\eta_1^t, \ldots, \eta_m^t)$ of a shift model.

Using observations Eq. 2.48, it is necessary to find an estimation of the successive RF frame x^t and estimation y^t, when obtaining a recurrent observation z^t. To find these estimations, let us use the tensor filtering equations as in Eq. 2.35. In these equations, both the frame itself x^t and its parameters y^t are included in the estimated frame x^t. This integration is a combined tensor, so filtering equations become a bit more complicated. Omitting the intermediate calculations, the resulting algorithm (Eq. 2.49) of recurrent estimation of the RF and displacement [16] are given, where $\widehat{x}_\ni^t = x^t(\widehat{y}_\ni^t), \widehat{y}_\ni^t = \Im^t \widehat{y}^{t-1}$.

$$\begin{aligned} \widehat{x}^t &= \widehat{x}_\ni^t + P_x^t (V_\theta^t)^{-1} (z^t - \widehat{x}_\ni^t) + P_B^t (V_\theta^t)^{-1} \frac{d\widehat{x}_\ni^t}{d\alpha^t} (z^t - \widehat{x}_\ni^t) \\ \widehat{y}^t &= \widehat{y}_\ni^t + P_y^t (V_\theta^t)^{-1} \frac{d\widehat{x}_\ni^t}{d\alpha^t} (z^t - \widehat{x}_\ni^t) + (P_B^t)^T (V_\theta^t)^{-1} (z^t - \widehat{x}_\ni^t) \end{aligned} \tag{2.49}$$

The recurrent relations between tensor coefficients of Eq. 2.49 are given in Eqs. 2.50–2.51. Note that although filter in Eqs. 2.50–2.51 can solve this task it is rather difficult to use it in real situations. Besides the computational problems, it is connected with model specification Eq. 2.48, namely, defining function $x^t(y^t)$, i.e. how the frame x^t looks like at the IGT parameters y^t. However, if the IGT type is given (for example, shift or rotation), it is much easier to define the function $x^t(y^t)$.

$$\begin{cases} P_x^t A_x^t + P_B^t A_{yx}^t = P_{\ni x}^t, \ (P_B^t)^T A_x^t + P_y^t A_{yx}^t = (P_{\ni B}^t)^T \\ P_x^t A_{xy}^t + P_B^t A_y^t = P_{\ni B}^t, \ (P_B^t)^T A_{xy}^t + P_y^t A_{yx}^t = P_{\ni B}^t \end{cases} \tag{2.50}$$

$$\begin{cases} A_x^t = E + (V_\theta^t)^{-1} P_{\ni x}^t + (V_\theta^t)^{-1} \dfrac{dx^t}{dy^t} (P_{\ni B}^t)^T \\ A_y^t = E - (V_\theta^t)^{-1} \left(\dfrac{dx^t}{dy^t}\right)^2 P_{\ni y}^t + (V_\theta^t)^{-1} \dfrac{dx^t}{dy^t} P_{\ni B}^t \\ A_{xy}^t = -(V_\theta^t)^{-1} \dfrac{dx^t}{dy^t} P_{\ni y}^t + (V_\theta^t)^{-1} P_{\ni B}^t \\ A_{yx}^t = -(V_\theta^t)^{-1} \dfrac{dx^t}{dy^t} P_{\ni x}^t + (V_\theta^t)^{-1} \left(\dfrac{dx^t}{dy^t}\right)^2 (P_{\ni B}^t)^T \\ P_{\ni x}^t = \Re^t P_x^{t-1} (\Re^t)^T + V_x^t,\ P_{\ni B}^t = \Re^t P_B^{t-1} (\Im^t)^T \\ P_{\ni B}^t = \Im^t P_B^{t-1} (\Im^t)^T + V_y^t,\ V_y^t = \Phi^t (\Phi^t)^T \end{cases} \tag{2.51}$$

2.5.2 Random Field Alignment of Images with Interframe Geometric Transformation

Let us examine the task of the RF alignment, when a parameter distribution for the IGT is not given. For this purpose, consider the IGT model at first.

Let the RF $\dot{X} = \left\{x_{\bar{u}}^i : \bar{u} \in U,\ i = 1, 2, \ldots\right\}$ be set on some continuous domain U at any specific time i. Each frame $x^i = \left\{x_{\bar{j}}^i : \bar{j} \in \Omega_i\right\}$ of a grid field X is a system of values $x^i = \left\{x_{\bar{u}}^i : \bar{u} \in U\right\}$ on the grid $\Omega_i = \left\{\bar{j} : (j_1, \ldots j_n) : j_k = \overline{1, M_k}\right\}$. Besides, the position and shape of grids Ω_i can vary with time, while an index size $M_1 \times M_2 \times \ldots \times M_n$ remains constant. Some possible positions of two-dimensional grids Ω_{i-1} (continuous line) and Ω_i (dotted line) are depicted in Fig. 2.15.

Each of grids Ω_i may be considered as a system of coordinates. Thus, the task of the RF alignment can be formulated as a task of finding of coordinates' transformation of nodes of grid Ω_i nodes in the system of coordinates Ω_u of domain U. Sometimes, (e.g. to compensate the noise) it is easier to find the transformation of coordinate Ω_i into Ω_{i-1}, i.e. to align each recurrent frame x^i with proceeding frame x^{i-1}.

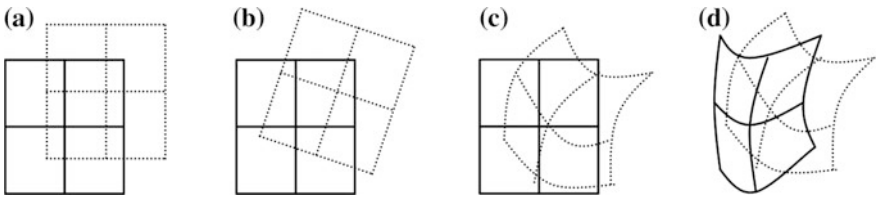

Fig. 2.15 Positions of grids Ω_{i-1} (*continuous line*) and Ω_i (*dotted line*) in two sequenced frames: **a** grid Ω_i is obtained from a rectangular grid Ω_{i-1} by a parallel shift, **b** grid Ω_i is obtained from grid Ω_{i-1} by the shift and rotation, **c** grid Ω_{i-1} is curvilinear, **d** both grids are curvilinear

Consider the task of aligning two frames x^{i-1} and x^i. In general (Fig. 2.15d), it is necessary to estimate a grid form Ω_{i-1} and find transformation of coordinates Ω_i into Ω_{i-1}. Even for stationary fields $\dot{X}$, an estimation of a grid form Ω_{i-1}, i.e. estimation of mutual value ordering x_j^{i-1}, is rather complex and of low accuracy. Let us restrict ourselves to the case, when Ω_{i-1} is a rectangular grid with a unit step (Fig. 2.15a–c). This task simplification is not a rough approximation of a real situation, as value grids are usually close to rectangular.

Let an observation model of field X be Eq. 2.52, where $\theta = \left\{\theta_j^i\right\}$ is a field of independent RV.

$$z^i = x^i + \theta^i \quad i = 1, 2, \ldots \tag{2.52}$$

If the RF $\dot{X}$ is stationary, then a conditional mutual PDF $w\left(z^{i-1}, z^i | f\right)$ can be obtained, where f is a transformation of coordinates Ω_i in (rectangular) coordinate system Ω_{i-1}. This makes it possible to apply various statistical estimations to combine a pair of frames x^{i-1} and x^i, for example, maximum likelihood estimate by Eq. 2.53.

$$f = \arg \max_f w(z^{i-1}, z^i | f) \tag{2.53}$$

In the general case of transformations f, all coordinates $f(\bar{j})$ of all frame values x^i should be estimated in the system Ω_{i-1}. Therefore, function f in Eq. 2.53 contains a large number of parameters and calculation becomes rather complicated.

Alignment task is sufficiently simplified if the type of transformation f is considered. Then, it is desirable only to determine its parameters $\bar{\alpha}$. In such cases, an estimation Eq. 2.53 takes the following form

$$\hat{\bar{\alpha}} = \arg \max_{\bar{\alpha}} w(z^{i-1}, z^i | \bar{\alpha}) \tag{2.54}$$

and contains only a few parameters. Estimation Eq. 2.54 can be used in the case of the RF a priori parametric uncertainty $\dot{X}$ and observation model Eq. 2.52. For this parameter, vector $\bar{\alpha}$ must be supplemented by the RF unknown parameters $\dot{X}$ and an observation model.

2.5.3 *Alignment of Two Frames of Gaussian Random Field*

Let Gaussian stationary field $\dot{X}$ has a zero mean and the CF as Eq. 2.55, where $\rho(i, |u_1|, \ldots |u_n|)$ is a field $\dot{X}$ correlation coefficient at a distance $|i|$ along the time axis and at a distance $|u_k|$ along the kth spatial axis.

$$V(i,\bar{u}) = M[\dot{x}^{j}_{\bar{v}}\dot{x}^{j+i}_{\bar{v}+\bar{u}}] = \sigma_x^2\rho(|i|,|u_1|,\ldots|u_n|) \tag{2.55}$$

In observation model Eq. 2.52, noises θ will also be assumed as Gaussian with zero mean and constant value σ_θ^2.

Let there exist some observations z^{i-1} of a frame x^{i-1} at the nodes of a rectangular grid Ω_{i-1} with a unit step and some observations z^i of a frame x^i at the grid Ω_i nodes. It is necessary to estimate the parameters $\bar{\alpha}$ of IGT x^{i-1} and x^i, i.e. to find the estimation $\hat{\bar{\alpha}}$ of parameters in $f(\bar{j},\bar{\alpha})$. If we assume that $i = 1$ and chose the grid Ω_1 axis coinciding with the coordinate axis, in which the CF as Eq. 2.55 is set, then the Gaussian joint PDF of a frame x^1 and its observations can easily be calculated from Eqs. 2.52 and 2.55. If parameter vector $\bar{\alpha}$ is given, then position of grid Ω_2 relative to grid Ω_1 becomes definite. Therefore, it is possible to find a joint conditional PDF of observations $Z = (z^1, z^2)$ at given $\bar{\alpha}$ as Eq. 2.56, where $V(\bar{\alpha})$ is a covariance matrix of observations Z, k is a number of elements in Z.

$$w(Z|\bar{\alpha}) = \frac{1}{(2\pi)^{k/2}\det^{1/2}(V(\bar{\alpha}))}\exp\left(-\frac{1}{2}ZV^{-1}(\bar{\alpha})Z\right) \tag{2.56}$$

It should be noted that calculation of Eq. 2.56 maximum is rather a complicated task which is absolutely unrealizable in real-time systems. In order to simplify the Task, consider the estimation, which can be obtained as a result of only exponent maximization (in Eq. 2.56) provided by Eq. 2.57.

$$\hat{\bar{\alpha}} = \arg\min_{\bar{\alpha}} ZV^{-1}(\bar{\alpha})Z = \arg\min_{\bar{\alpha}} J(Z,\bar{\alpha}) \tag{2.57}$$

Functional $J(Z,\bar{\alpha})$ in Eq. 2.57 is the Mahalanobis distance of sample Z from the origin for covariance matrix $V(\bar{\alpha})$. Thus, the estimation in Eq. 2.57 minimizes the Mahalanobis distance of observed Z from the origin. Consider this distance in detail. From Eq. 2.44, it follows that the observation Eq. 2.58 takes place, where $Z^*(\bar{\alpha})$ is an optimal (in our case linear) prediction of observations Z into a point, $\Delta^*_Z(\bar{\alpha})$ are prediction errors, i.e. compensation residues into a point, $C(\bar{\alpha})$ is a diagonal matrix of error variance $\Delta^*_Z(\bar{\alpha})$.

$$J(Z,\bar{\alpha}) = ZC^{-1}(\bar{\alpha})(Z - Z^*(\bar{\alpha})) = ZC^{-1}(\bar{\alpha})\Delta^*_Z(\bar{\alpha}) \tag{2.58}$$

If observations z^1 and z^2 are used twice ($z^1 \to z^2$ and $z^2 \to z^1$), then Eq. 2.59 can be obtained, where $\hat{z}^1(\bar{\alpha})$ is an observation prediction of z^1 by z^2, i.e. prediction into domain, $\hat{\Delta}_1(\bar{\alpha})$ are the prediction errors, $N_1(\bar{\alpha})$ is an error covariance $\hat{\Delta}_1(\bar{\alpha})$, $\hat{z}^2(\bar{\alpha})$ is a prediction of z^2 along z^1, $N_2(\bar{\alpha})$ is an error covariance $\hat{\Delta}_2(\bar{\alpha})$ of this prediction.

$$J(z^1, z^2, \bar{\alpha}) = z^1 N_1^{-1}(\bar{\alpha})(z^1 - \hat{z}^1(\bar{\alpha})) + z^2 N_2^{-1}(\bar{\alpha})(z^2 - \hat{z}^2(\bar{\alpha})) \\ = z^1 N_1^{-1}(\bar{\alpha})\hat{\Delta}_1(\bar{\alpha}) + z^2 N_2^{-1}(\bar{\alpha})\hat{\Delta}_2(\bar{\alpha}) \tag{2.59}$$

The evaluation can be somewhat modified by introducing the conditional PDF as the following product

$$w(z^1, z^2|\bar{\alpha}) = w(z^1|\bar{\alpha})\, w(z^2|z^1, \bar{\alpha}).$$

As $w(z^1|\bar{\alpha}) = w(z^1)$ does not depend on $\bar{\alpha}$, it is enough to maximize conditional PDF $w(z^2|z^1, \bar{\alpha}) = \frac{1}{(2\pi)^{k/4}\det^{1/2}(N_2(\bar{\alpha}))}\exp\left(-\frac{1}{2}(z^2 - \hat{z}^2(\bar{\alpha}))ZN_2^{-1}(\bar{\alpha})(z^2 - \hat{z}^2(\bar{\alpha}))\right)$, where $\hat{z}^2(\bar{\alpha})$ and $N_2(\bar{\alpha})$ are the same as in Eq. 2.59. From here, it is also possible to obtain a simplified evaluation by the functional in Eq. 2.60, which is the Mahalanobis distance between observations z^2 and their predictions $\hat{z}^2(\bar{\alpha})$ according to z^1.

$$J(z^1, z^2, \bar{\alpha}) = [z^2 - \hat{z}^2(\bar{\alpha})]N_2^{-1}(\bar{\alpha})[z^2 - \hat{z}^2(\bar{\alpha})] \tag{2.60}$$

Experimental studies show a poor estimation quality obtained using the Mahalanobis distance. It can be explained by the fact that in Eqs. 2.58–2.60 the compensation residues and matrices inversed to $C(\bar{\alpha})$ or to $N_i(\bar{\alpha})$ are multiplied. In addition, the compensation residues can decrease along $\bar{\alpha}$ only to a certain limit. Thus, minimization of these expressions can occur mainly due to increase of matrix elements $C(\bar{\alpha})$ or $N_i(\bar{\alpha})$. This does not happen while the MLM estimates, as the PDF Eq. 2.56, are inversely proportional to the square root of a covariance matrix determinant $V(\bar{\alpha})$.

Much better IGT estimates can be obtained by minimizing the compensation residues $\Delta_Z^*(\bar{\alpha})$, $\hat{\Delta}_1(\bar{\alpha})$, or $\hat{\Delta}_2(\bar{\alpha})$. It is easier to minimize residues

$$\hat{\Delta}_2(\bar{\alpha}) = z^2 - \hat{z}^2(\bar{\alpha}),$$

as the optimal compensation of the recurrent frame z^2 according to the previous frame z^1 is determined. It can often be the main alignment task. Using this approach, it is possible to obtain the estimates in Eq. 2.61 and other estimates, which depend on the used metric. Further, variety of estimates can be obtained using various predictions $\hat{z}_j^2(\bar{\alpha})$, for example, optimal prediction or various interpolations of observation z^1, initially set only on the grid Ω_1.

$$\hat{\bar{a}} = \arg\min_{\bar{\alpha}} M\left[\sum_j \left|z_j^2 - \hat{z}_j^2(\bar{\alpha})\right|\right] \quad \hat{\bar{a}} = \arg\min_{\bar{\alpha}} M\left[\sum_j (z_j^2 - \hat{z}_j^2(\bar{\alpha}))^2\right] \tag{2.61}$$

Generally speaking, it should be mentioned that the compensation estimates Eq. 2.61 are biased. They do not estimate the IGT parameters $\bar{\alpha}$ but only optimize

some compensation in terms of a certain metric. Thus, frames z^1 and z^2 alignment with the use of estimated parameters $\bar{\alpha}$ is a pseudo-alignment in the sense of the best compensation of the selected type. However, in the case of a good choice of prediction function rather effective alignment is often provided.

2.5.4 Method of Fixed Point at Frame Alignment

Many methods of image alignment and the IGT parameter estimation have small operating areas, i.e. they are reasonable only if the estimated parameters are small. Section 2.5.4 describes a method of estimating the IGT parameters with a large operating area. However, the accuracy of this method is low. It can be used to obtain initial parameter approximation, which is then used in a more precise estimation method with a small operating area.

Let mapping $F : W \rightarrow W$ of set W into itself be given. The Fixed Point (FP) of mapping F is any element $w \in W$, for which $F(w) = w$. That is, while mapping F such a point does not change and transforms into itself. Let us apply the concept of the FP to estimate the IGT parameters of images [16, 29, 30].

Consider two images $x(W)$ and $y(J)$ given on integer m-dimensional grids $W = \{w\} = \{(w_1, \ldots, w_m)\}$ and $J = \{j\} = \{(j_1, \ldots, j_m)\}$. Let the IGT type be known and connect positions of these images

$$w = F(j; \bar{\alpha}),$$

where $\bar{\alpha}$ are the unknown IGT parameters, which are subject to estimation. After completing an auxiliary transformation IGT P of image $y(J)$, the image $z(J)$ connected with $x(W)$ by means of a complex transformation is obtained.

$$w = P(F(j; \bar{\alpha})) = H(j; \bar{\alpha})$$

Suppose that this transformation has only one FP v (Eq. 2.62).

$$v = H(v; \bar{\alpha}) \tag{2.62}$$

Then Eq. 2.62 is transformed into a system of m equations relative to parameters $\bar{\alpha}$. If the number of IGT parameters is equal to image dimension, then it is possible to define the parameters from Eq. 2.62. If the IGT has a larger number of parameters, then it is possible to perform K auxiliary transformation P_k, find the FP v_k of each complex transformation H_k, and obtain a system of K equations

$$v_k = H_k(v_k; \bar{\text{alpha}}), \quad k = 1, \ldots, K,$$

from which the estimates for all transformation F parameters $\bar{\alpha}$ can be obtained. The place of the FP after auxiliary transformation is illustrated by Fig. 2.16. The

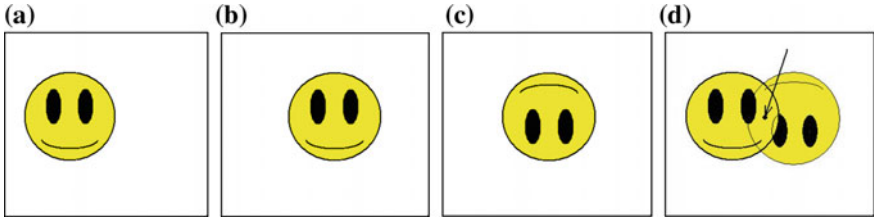

Fig. 2.16 Fixed point of a complex transformation shift-rotation: **a** original image, **b** shift of image **a**, **c** rotation of image (**b**), **d** combination of images (**a**) and (**c**)

original image shows a Smiley (Fig. 2.16a). If this image is shifted right on α_1, the Smiley will shift on α_1 as well (Fig. 2.16b). It is obvious that there is no the FP at a parallel shift. Let us perform an auxiliary rotation of Fig. 2.16b for angle π center-relative. Then, an image in Fig. 2.16c is obtained. Figure 2.16d shows images in Fig. 2.16a, c together. It is evident, that the point marked by an arrow is the same for both Smileys. This is the FP of image transformation of Fig. 2.16a into Fig. 2.16c.

It is essential that the value of the estimated parameters $\bar{\alpha}$ has no effect on the structure of the proposed algorithm and we need to find the FP of a complex transformation that matters. For this purpose, it is necessary that the common part of the images $x(W)$ and $z(J)$ ought not to be too small and contain the FP of the complex IGT, i.e. the restrictions for parameter $\bar{\alpha}$ value are not very strong.

It is obvious that the accuracy of parameter $\bar{\alpha}$ estimation depends on estimation error of the FP position, which has a pixel order that leads to errors of parameter $\bar{\alpha}$ estimation for analyzed IGT. In order to make estimation $\bar{\alpha}$ more accurate, it is possible to use the estimate, which is obtained with the help of a FP in a more precise algorithm as initial approximation. The algorithm can be less time-consuming but with a smaller operating area. Thus, a pair of algorithms allows to improve speed of operation in real-time systems connected with extraction of useful information from image sequences.

Fixed point method for 2D images. Consider the application of the FP method for estimating the IGT parameters of two-dimensional images. The extension of this method on images of larger dimension leads to technical complications only.

Let $x(u, v)$ and $y(i, j)$ be two 2D images with the known IGT (f, g), which relates the positions of these images (Eq. 2.63), where $\bar{\alpha}$ are estimated parameters.

$$u = f(i, j; \bar{\alpha}), \; v = g(i, j; \bar{\alpha}) \tag{2.63}$$

Conducting auxiliary transformation (p, q) of an image $y(i, j)$, an image $z(i, j)$ connected with $x(u, v)$ by means of a complex transformation is obtained.

$$k = f(p(i, j), q(i, j); \bar{\alpha}) = F(i, j; \bar{\alpha})$$
$$l = g(p(i, j), q(i, j); \bar{\alpha}) = G(i, j; \bar{\alpha})$$

Suppose that this transformation has only one FP (u,v) satisfying Eq. 2.64.

$$u = F(u, v; \bar{\alpha}) \quad v = G(u, v; \bar{\alpha}) \tag{2.64}$$

If one manages to find this FP (at least approximately), then Eq. 2.64 is transformed into a set of equations relative to parameters $\bar{\alpha}$. If the IGT has only two parameters (for example, in the case of a parallel shift), then these parameters can be obtained from two Eq. 2.64. If the IGT has more than two parameters, then it is possible to conduct K auxiliary transformations, find their FP and obtain a set of equations

$$u_k = F_k(u_k, v_k; \bar{\alpha}) \quad v_k = G_m(u_k, v_k; \bar{\alpha}) \quad k = 1, \ldots, K,$$

from which the estimates of all IGT parameters can be obtained.

The most difficult aspect of this method is to find the FP. The necessary criterion for the set of equations Eq. 2.63, i.e. (u, v) immobility, is

$$x(F(u, v; \bar{\alpha}),\ G(u, v; \bar{\alpha})) = z(u, v).$$

However, this condition is not sufficient as there may be other points with values $z(i,j)$ on the image $x(u, v)$ but the FP is only one of them. Therefore, it is necessary to chose an auxiliary transformation for each specific type of the IGT and find the sufficient characteristics of point immobility.

Consider a common type of the IGT: turn on the angle α around the image centre, scale change with coefficient s, and parallel shift on vector (a, b). Both images $x(u, v)$ and $y(i,j)$ are set on an integer grid. Let us place the origin (0, 0) in the central pixel of the grid. Thus, the coordinate transformation (i,j) of an image $y(i,j)$ into coordinates (u, v) of an image $x(u, v)$ has the form of Eq. 2.65.

$$u = a + s(i\cos\alpha - j\sin\alpha) \quad v = b + s(i\sin\alpha + j\cos\alpha) \tag{2.65}$$

Coordinates (u, v) may be fractional, then it may be desirable to calculate values $x(u, v)$ interpolation of the grid image $x(u, v)$. According to the given images $x(u, v)$ and $y(i,j)$, it is necessary to estimate the IGT parameters using Eq. 2.65. As an auxiliary transformation, let us take image rotation $y(i,j)$ around its center pixel (assuming that its coordinates are (0,0)) on the angle π, which gives an image $z(i,j) = y(-i, -j)$. In Eq. 2.65, this rotation is equivalent to increase of α by π. Thus, the set Eq. 2.63 has the following form

$$u = a - s(u\cos\alpha - v\sin\alpha),\ v = b - s(u\sin\alpha + v\cos\alpha).$$

It has the only solution (Eq. 2.66).

$$u = [a(1 + s\cos\alpha) + bs\sin\alpha]/[(1 + s\cos\alpha)^2 + (s\sin\alpha)^2]$$
$$v = [b(1 + s\cos\alpha) - as\sin\alpha]/[(1 + s\cos\alpha)^2 + (s\sin\alpha)^2] \quad (2.66)$$

If $\alpha \approx 0,\ s \approx 1$, then $u \approx a/2, v \approx b/2$ and an approximate Eq. 2.67 is obtained.

$$a \approx 2u,\ b \approx 2v \quad (2.67)$$

Thus, having estimated position (u, v) of the FP of image transformation $x(u, v)$ into image $z(i,j)$, the shift parameters (a, b) can be estimated using Eq. 2.67.

Now let us turn to the FP detection. For this purpose, consider an image $\Delta(i,j) = |z(i,j) - x(i,j)|$. Image values $x(u, v)$ and $z(u, v)$ at the FP (u, v) coincide, therefore $\Delta(u, v) = 0$. However, there may be other points, in which $\Delta(u, v) = 0$ because values $x(i,j)$ and $z(i,j)$ are equal at random. At first, let $\alpha = 0$ and $s = 1$ in Eq. 2.65, then it is obvious that image $\Delta(i,j)$ transforms into itself, when rotated at angle π about the FP. That is, this image has central symmetry relative to the FP (u, v). Therefore,

$$\varepsilon(u, v; m, n) = |\Delta(u + m, v + n) - \Delta(u - m, v - n)| = 0$$

for any m and n. However, there may be other Points, at which $\varepsilon(u, v; m, n) = 0$ for some values m and n. It is unlikely, that $\varepsilon(u, v; m, n) = 0$ for many values m and n at once if (i,j) is not the FP. Therefore, the statistics values (Eq. 2.68) are more likely to be small, when a point (i,j) is located near the FP.

$$\varepsilon(i,j) = \sum_{m=0}^{r} \sum_{n=-r}^{r} \varepsilon(i,j; m, n) \quad (2.68)$$

Thus, a position of the minimum point (i,j) of statistics Eq. 2.68 can be considered as a position estimation of the FP (u, v).

If there is rotation and scale changes (subject to $\alpha \neq 0,\ s \neq 1$), then the central symmetry of image $\Delta(i,j)$ is broken. However, if the rotation angle and scale change are small, the symmetry distortion will be small. That is why, the minimum point of statistics in Eq. 2.68 will slightly deviate from the desired point, i.e. there will appear a small error in defining its position. The mutual brightness image distortions $x(i,j)$ and $z(i,j)$, for example, noise, can lead to the secondary errors. Summation in Eq. 2.68 causes smoothing. Thus, a noise impact is slightly reduced.

Additional noise reduction effect can be obtained by considering the approximate central symmetry of an image $\varepsilon(i,j)$ and using statistics similar to Eq. 2.68

$$\delta(i,j) = \sum_{m=0}^{r} \sum_{n=-r}^{r} |\varepsilon(i + m, j + n) - \varepsilon(i - m, j - n).$$

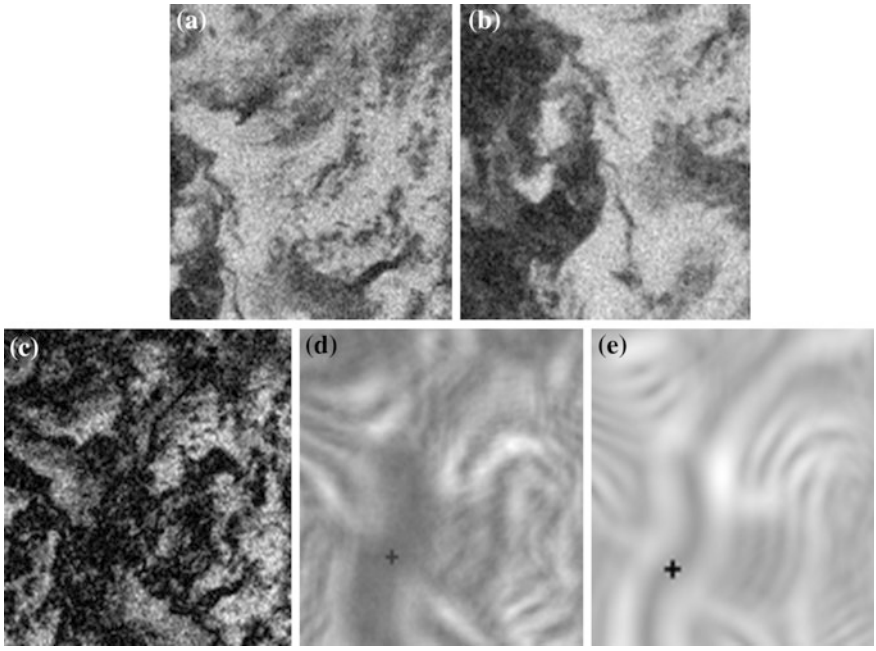

Fig. 2.17 Obtaining of a fixed point at shift, rotation, scale change and noise: **a** and **b** original images, **c** image $\Delta(i,j)$, **d** image $\varepsilon(i,j)$, **e** image $\delta(i,j)$

This method is illustrated by Fig. 2.17. Cloud-cover images in Fig. 2.17a, b have the IGT parameters Eq. 2.65, i.e. shift $a = 43.2$, $b = -38.7$, rotation $\alpha = 0.1$ rad and scale change $s = 0.9$. They are distorted by a strong additive white noise with $\sigma = 30$. Thus, the clouds are slightly visible. Figure 2.17c–e show images $\Delta(i,j)$, $\varepsilon(i,j)$, and $\delta(i,j)$, respectively. Minimum points of images $\varepsilon(i,j)$ and $\delta(i,j)$ are marked with crosses, their coordinates are (3, – 11) and (23, –19). Relative shift estimations obtained according to statistics $\varepsilon(i,j)$ and $\delta(i,j)$ are (6, –22) and (46, –38), respectively. The second one is taken into account. It should be considered as a good result for such rigorous estimation conditions.

The tests of the described algorithm showed that if $|\alpha| \leq 0.1$ rad, $|s-1| \leq 0.1$, and white noise with $\sigma \leq 30$, the estimation error even of large values of shift parameter (tens of pixels) of simulated and real images did not exceed 4-6 pixels.

2.6 Adaptive Algorithms of Image Processing

The image processing algorithms discussed above have been synthesized, mainly, assuming that the image, noise and observation models are thoroughly defined. In practice, such models are usually only partially known, i.e. there exists a priory

uncertainty in defining an image model. Thus, the synthesis of adaptive processing algorithms is required. Many adaptive algorithms, in particular, image processing, were developed [31, 32].

Various types of algorithms, such as the pseudo-gradient adaptive, pseudo-gradient adaptive prediction, and pseudo-gradient algorithms of image alignment, are discussed in Sects. 2.6.1–2.6.3, respectively.

2.6.1 Pseudo-Gradient Adaptive Algorithms

According to the purpose of data processing, the adaptive algorithms can be divided into two types: argument and criterion. The starting point for algorithm synthesis is a minimization of average losses, which is formally expressed by quality functional $R(\bar{\alpha}, Z) = J(\bar{\alpha})$, i.e. the criterion, which value should be minimized according to some parameters $\bar{\alpha}$. However, the requirements for this minimization may be different. The objective of argument tasks is the most accurate determination of the minimum point $\bar{\alpha}^*$ (probably, a variable). This type of tasks includes the measurement, filtration, and prediction problems. The objective of criterion tasks is approximation of $J(\bar{\alpha})$ to its minimal value $J^* = J(\bar{\alpha}^*)$, the parameters themselves $\bar{\alpha}$ are irrelevant and may greatly differ from $\bar{\alpha}^*$.

According to the method of optimal parameter estimation, the adaptive algorithms can be divided into identification and non-identification ones. In identification algorithms, the unknown characteristics γ are estimated according to all data available. Then, the estimates $\hat{\gamma}$ obtained are used as the correct ones. As a result, the parameters for the algorithms in the form $\bar{\alpha} = \bar{\alpha}(\hat{\gamma})$ are obtained. This is the core of multiple modified decision rules. The application of these algorithms is complicated by an additional procedure of estimates $\hat{\gamma}$ and instability of further computations to errors of these estimates. In the algorithms without identification, the criterion $J(\bar{\alpha})$ minimization is conducted by variation of $\bar{\alpha}$ without intermediate estimation of any characteristics of the given data. Moreover, $\bar{\alpha}$ may be chosen iteratively during the current processing according to the observations over current values $J(\bar{\alpha})$.

To implement such algorithms, it is necessary to estimate current values $J(\bar{\alpha})$, i.e. the criterion should be observable and this fact limits the application area of the approach. Sometimes it is possible to replace non-observable $J(\bar{\alpha})$ with the other observable criterion $J_1(\bar{\alpha})$. It is required that minimum points $J(\bar{\alpha})$ and $J_1(\bar{\alpha})$ coincide in argument tasks and $J(\bar{\alpha})$ should approximate to $J^* = J(\bar{\alpha}^*)$ as $J_1(\bar{\alpha})$ approximates to $J_1^* = J_1(\bar{\alpha}_1^*)$ in criterion tasks.

Even if the data is thoroughly described, it is not always possible to find the optimal decision rule because of mathematical difficulties. Even if it is possible to find it, it is often unacceptably time consuming. Besides, the initial data model usually describes the reality only approximately. For these reasons, it is often impossible to find and apply the optimal rule under real conditions. Therefore, it is

necessary to apply the quasi-optimal implementable rules with a small loss in processing quality. To find such rules, it is possible to use the simplified data models, which describe only their fundamental properties. The obtained rules (algorithms) contain some undetermined parameters $\bar{\alpha}$. These rules should be chosen in such manner that the algorithm gave the best result, when the certain processed data is used.

Let the processing structure be defined and the quality criterion for problem solving be formulated in terms of functional minimization $J(\bar{\alpha})$, which characterizes the average loss, when the processing is performed with parameters $\bar{\alpha}$. In a view of the unexpected uncertainty of data description, it is impossible to determine the optimal parameters $\bar{\alpha}^*$ beforehand. Therefore, the adaptive procedure is desirable, which together with the processing procedure composes the adaptive algorithm. These parameters $\bar{\alpha}$ are determined in terms of a certain implementation (observation) Z of the object processed.

Thus, the adaptation problem is formulated as a problem of function minimization $J(\bar{\alpha}) = J(\bar{\alpha}, Z)$ for a particular data Z. Pseudo-Gradient (PG) algorithms showed good results while solving this problem. There are some numerical methods for extremum search. The most common are various modifications of the gradient algorithm (Eq. 2.69), where $\bar{\alpha}_n$ is an approximation to the minimum point of the functional $J(\bar{\alpha})$ following $\bar{\alpha}_{n-1}$, μ_n is a positive numerical sequence determining the step length, $\nabla J(\bar{\alpha})$ is a gradient of $J(\bar{\alpha})$.

$$\bar{\alpha}_n = \bar{\alpha}_{n-1} - \mu_n \nabla J(\bar{\alpha}_{n-1}) \tag{2.69}$$

Each step in Eq. 2.69 is made towards the quickest decrease of $J(\bar{\alpha})$, although under certain conditions convergence $\bar{\alpha}_n \to \bar{\alpha}^*$ takes place, it may be very slow.

Multiple computations $\nabla J(\bar{\alpha}_{n-1}, Z)$ prevent the application of these methods in image processing. Each computation usually includes the whole processing procedure of Z with parameters $\bar{\alpha}_{n-1}$. It is possible to significantly decrease the amount of computations, if instead of $\nabla J(\bar{\alpha}_{n-1}, Z)$ a reduction $\nabla Q(\bar{\alpha}_{n-1}) = \nabla J(\bar{\alpha}_{n-1}, Z_n)$ is taken, i.e. to calculate the gradient not across the whole Z but only across some of its part Z_n, for example, in a sliding window. Then, in Eq. 2.69 instead of an exact gradient value its value with a random error $\bar{\delta}_n$ is used and obtained in Eq. 2.70.

$$\bar{\alpha}_n = \bar{\alpha}_{n-1} - \mu_n(\nabla J(\bar{\alpha}_{n-1}, Z) + \bar{\delta}_n) = \bar{\alpha}_{n-1} - \mu_n \nabla Q(\bar{\alpha}_{n-1}) \tag{2.70}$$

Sequence $\bar{\alpha}_n$ becomes random, thus the very fact of its convergence to $\bar{\alpha}^*$ is random as well. Generally speaking, the random errors $\bar{\delta}_n$ in Eq. 2.70 cannot be a serious obstacle to convergence $\bar{\alpha}_n \to \bar{\alpha}^*$.

The concept of the PG was introduced in [31]. On its basis, a common approach to the analysis and synthesis of algorithms of functional stochastic minimization was developed. The class of the PG algorithms is very wide and it includes all (or nearly all) adaptive and learning algorithms. These algorithms are based on the procedure (Eq. 2.71), where $\bar{\beta}_n$ is a random (particularly, deterministic) direction, generally speaking, depending on previous values $\bar{\alpha}_i$ and n.

$$\bar{\alpha}_n = \bar{\alpha}_{n-1} - \mu_n \bar{\beta}_n \tag{2.71}$$

Vector $\bar{\beta}_n$ is called the PG of $J(\bar{\alpha})$ at a point $\bar{\alpha}_{n-1}$ if the pseudo-gradient condition

$$[\nabla J(\bar{\alpha}_{n-1})]^T M[\vec{\beta}_n] \geq 0$$

is fulfilled, where the left-hand member is a scalar product, i.e. on the average F is directed at an acute angle to the exact gradient. Algorithm Eq. 2.71 is called the PG if $\bar{\beta}_n$ is the PG at each one-step transition. In this case, the transitions in Eq. 2.71 will be made on the average towards reduction of $J(\bar{\alpha})$ and it is possible to hope for convergence $\bar{\alpha}_n \to \bar{\alpha}^*$ as $n \to \infty$, although some transitions can be made towards increase of $J(\bar{\alpha})$. Indeed, the fulfillment of relatively weak conditions [31] is sufficient for the convergence with probability one for any initial approximation $\bar{\alpha}_0$.

Note that algorithm Eq. 2.71 is much more general than Eq. 2.70, as in Eq. 2.71 the possibility of calculating $J(\bar{\alpha})$ or $\nabla J(\bar{\alpha})$ is not assumed even with a random error, i.e. $J(\bar{\alpha})$ can be also non-observable. Still the availability of the observable PG is required. In particular, even noisy gradient value of another functional $J_1(\bar{\alpha})$ may be chosen as $\bar{\beta}_n$ but it should have the same minimum point as $J(\bar{\alpha})$.

If it is possible for $\bar{\beta}_n$ to depend on previous values $\bar{\alpha}_i$, then it allows to use the PG algorithms to process not only one-dimensional but also multidimensional data in the order of their scanning.

Until now it was assumed that the task is to find the minimum point $\bar{\alpha}^*$ of the functional $J(\bar{\alpha}, Z)$, which is unique for all Z. Such a point exists but the processing will be optimal if data Z is homogeneous. The convergence $\bar{\alpha}_n \to \bar{\alpha}^*$ requires the convergence $\mu_n \to 0$. However, if we bound μ_n from below (for example, take constants $\mu_n = \mu$), then the dispersions of estimate errors $\bar{\alpha}_n$ of parameters $\bar{\alpha}^*$ will stop decreasing and will have the order μ^2, while the parameters $\bar{\alpha}_n$ themselves will oscillate about $\bar{\alpha}^*$. Thus, if a homogeneous data processing is conducted simultaneously with the estimation of $\bar{\alpha}^*$ (when $\mu_n = \mu$), then upon reaching a steady state operating conditions only some quasi-optimal processing, i.e. close to optimal, can be conducted.

If data Z is heterogeneous and there is an abrupt change in its characteristics, then optimal parameter values $\bar{\alpha}^*$ can also change. If a processing proceeds, then it is possible that immediately after such a rapid change the processing quality will get worse and then gradually quasi-optimal results will be achieved once again. When there is only a smooth change in observational characteristics Z (more precisely if optimal parameter values $\bar{\alpha}^*$ change smoothly), which is comparable with the procedure Eq. 2.71 transient rate, it is possible to use the PG algorithms to process the heterogeneous data without their fragmentation into areas of a relatively homogeneous structure.

Thus, the PG algorithms are easy to implement and they can be applied to a wide variety of homogeneous and heterogeneous data. Their adaptation can be performed directly during the processing procedure, thus data delay line is not required. The

indicated positive qualities of the PG adaptive algorithms make them preferable in the processing of images and other large data arrays.

Further, the application of the PG algorithms for the prediction and image alignment as examples is considered.

2.6.2 Pseudo-Gradient Adaptive Prediction Algorithms

While solving some tasks of image processing, there often arises an auxiliary problem of their prediction, i.e. the problem of estimation formation $x_{\bar{i}}^*$ of an image element $x_{\bar{i}}$ according to a certain set (pattern) of observations $Z_{\bar{i}}$, which does not include the predicted element itself. In particular, in the Kalman filters the extrapolated estimate is the prediction. In most real situations, the accuracy of prediction increases with the extension of a pattern but at the same time the computational costs significantly increase.

The prediction is considered to be optimal if the minimum of the mean square prediction error $M[(x_{\bar{i}}^* - x_i)^2] = M[\Delta_i^2]$ is achieved. In this case, a conditional mean $M[x_{\bar{i}}|Z_{\bar{i}}] = f(Z_{\bar{i}})$, i.e. some function of random variables $Z_{\bar{i}}$, is considered to be the optimal prediction. Type of function f depends on the distribution type of image elements, i.e. on its model. Thus, the optimal prediction may be presented as $x_{\bar{i}}^* = f(\bar{\alpha}_{\bar{i}}, Z_{\bar{i}})$, where $\bar{\alpha}_{\bar{i}}$ are the prediction function parameters. For example, in case of the Gaussian images with a zero mean the optimal prediction is linear $x_{\bar{i}}^* = \bar{\alpha}_{\bar{i}}^T Z_{\bar{i}}$.

If a form of the prediction function is determined, the problem is reduced to its optimization, i.e. it is necessary to find out the optimal values $\bar{\alpha}_{\bar{i}}^*$ of parameters $\bar{\alpha}_{\bar{i}}$, whereby a minimum square error is achieved (Eq. 2.72).

$$J(\bar{\alpha}_{\bar{i}}) = M[(f(\bar{\alpha}_{\bar{i}}, Z_{\bar{i}}) - x_{\bar{i}})^2] \tag{2.72}$$

Let us form the PG algorithm, which minimizes this functional. It needs an observable pseudo-gradient $\bar{\beta}_n$. Non-observable value $x_{\bar{i}}$ (the problem under consideration is to estimate $x_{\bar{i}}$) is contained in Eq. 2.72. Therefore, the functional Eq. 2.72 is non-observable, thus it is impossible to obtain the PG as an implementation of its gradient. However, the observations $z_{\bar{i}} = x_{\bar{i}} + \theta_{\bar{i}}$ are observable, their optimal prediction due to independency and centered of noise coincides with the optimal prediction of elements of a true image. Thus, functional Eq. 2.72 can be replaced by an auxiliary functional

$$J_1(\bar{\alpha}_{\bar{i}}) = M\left[(f(\bar{\alpha}_{\bar{i}}, Z_{\bar{i}}) - z_{\bar{i}})^2\right], \tag{2.73}$$

its observable PG may be as Eq. 2.74.

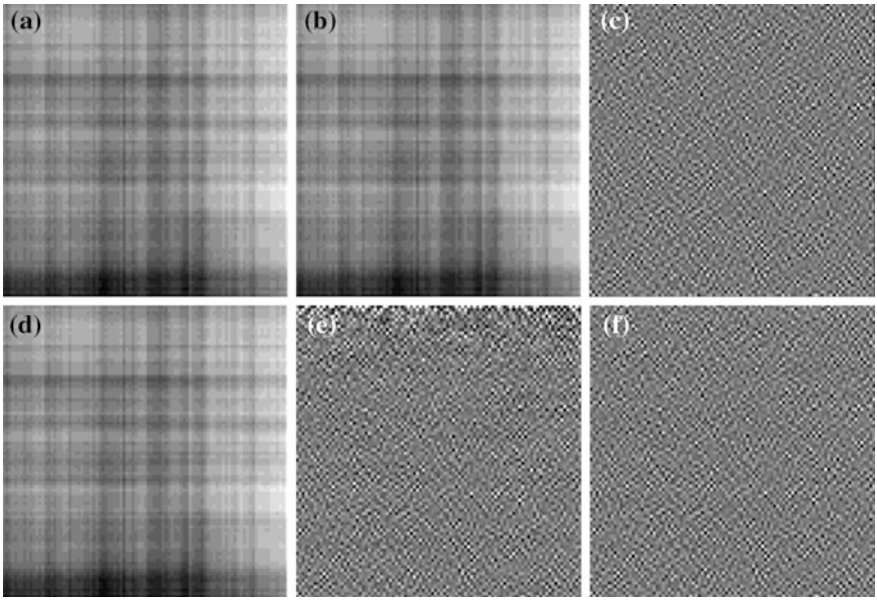

Fig. 2.18 Example of an image prediction: **a** initial image, **b** optimal prediction, **c** optimal prediction errors, **d** the PG prediction, **e** and **f** the PG prediction errors

$$\bar{\beta}_n = \nabla(f(\bar{\alpha}_n, Z_n) - z_n)^2 = 2\Delta_n \frac{\partial f(\bar{\alpha}_n, Z_n)}{\partial \bar{\alpha}_n} \tag{2.74}$$

Finding a gradient in Eq. 2.74 is rather easy, since the prediction f function is given. For example, if the prediction is linear, then

$$\partial f(\bar{\alpha}_n, Z_n) / \partial \bar{\alpha}_n = \partial f\left(\bar{\alpha}_n^T Z_n\right) / \partial \bar{\alpha}_n = Z_n.$$

Note that these algorithms are not associated with any evaluation of image parameters, they are efficient in computation. It enables to implement them in real-time systems.

An example of the PG algorithm application for linear prediction of a homogeneous Gaussian image with a factorable exponential CF is shown in Fig. 2.18. The prediction of an image element is a weighted sum of eight nearest neighbors. The initial image, its optimal prediction, its errors, the adaptive PG prediction, and its errors are represented in Fig. 2.18a–e, respectively. Visually Fig. 2.18b, d are the same, since the adaptive prediction is close to optimal. However, the prediction errors (enlarged for visualization) in Fig. 2.18c, e are different from each other. In the first few (top) lines of the adaptive prediction (Fig. 2.18d) errors are relatively large, as there the process of prediction parameter adjustment takes place. This process takes place rather quickly. Figure 2.18f shows the prediction errors during

the second image scanning, which develops from the values of prediction parameters established by the end of the first pass. Analysis of Fig. 2.18c, f allows to make a conclusion that there is no difference between them. It means that the prediction has nearly converged to an optimal one.

Thus, using the adaptive PG algorithms it is possible to obtain a good quality of image prediction at a low computational cost.

2.6.3 Pseudo-Gradient Algorithms of Image Alignment

As it was mentioned in Sect. 2.5, the compensation approach can be a good basis to synthesize algorithms of the IGT parameter estimation, i.e. finding a rather precise estimate of one frame values according to the observations of another frame.

If the IGT type $f(\bar{j}, \bar{\alpha})$ is known, then the problem of frame z^1 and z^2 alignment is sufficiently simplified. It can be formulated as a task of functional minimization (Eq. 2.75), where $\bar{\alpha}$ are parameters of the IGT model, $\hat{z}^2_{\bar{j}}(\bar{\alpha})$ is a prediction into point $z^2_{\bar{j}}$ according to observations z^1.

$$J(\bar{\alpha}) = M\left[\sum_{\bar{j}} (z^2_{\bar{j}} - \hat{z}^2_{\bar{j}}(\bar{\alpha}))^2\right] \tag{2.75}$$

Different interpolations of a grid frame z^1 are usually used as prediction $\hat{z}^2_{\bar{j}}(\bar{\alpha})$ in alignment problems. A set of parameters $\bar{\alpha}^*$, which contributes minimization (Eq. 2.75), is the IGT parameter estimation. The required alignment of frames z^1 and z^2 is obtained by substituting $\bar{\alpha}^*$ into defined by the IGT model transformation $f(\bar{j}, \bar{\alpha})$ of a coordinate system of grid Ω_1, in which frame z^1 is determined, into a coordinate system of frame z^2.

Thus, the problem of the IGT parameter estimation is reduced to the problem of prediction $\hat{z}^2_{\bar{j}}(\bar{\alpha})$ optimization with respect to parameters $\bar{\alpha}$. This problem can be solved using the PG algorithms, as it was done in Sect. 2.6.2.

However, the criterion (Eq. 2.75) can be used only if the brightness distortions are small enough. Suppose that the brightness distortion of the two frames can be approximated by a linear function. In this case, a correlation coefficient between these images should be high. Therefore, the maximum of a local sample correlation coefficient between the observations of the two frames in shifted windows can be taken as a criterion for alignment quality and the PG algorithm can be used to maximize this function.

As it has already been mentioned, the alignment tasks for unknown IGT types are considered to be more complex. Then, it is possible to find some type of transformation $f(\bar{j}, \bar{\alpha})$, which considers parameters $\bar{\alpha}$ as variables. Then the alignment task can be reformulated as a minimization task of functional (Eq. 2.75) with

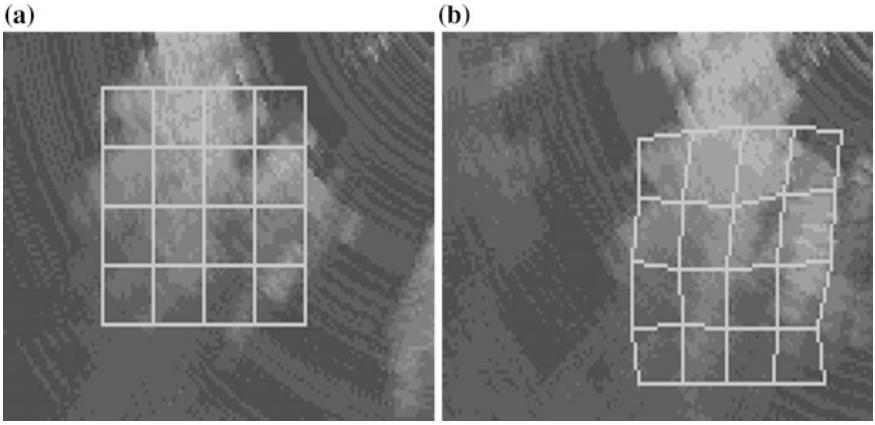

Fig. 2.19 Alignment of radar cloud images: **a** image with rectangle grid, **b** image with distorted grid

variable arguments $\bar{\alpha}$. In such case, the PG algorithms can also be applied, all we need is a parameter μ_n value bounded below. An example of alignment of radar cloud images using the criterion of maximum correlation is depicted in Fig. 2.19.

Figure 2.19b is received by the radar a few minutes after Fig. 2.19a. In this interval, a form of clouds has significantly changed. The clouds did not only move but they changed their shape and density. As a result, their images have the IGT of a general form and significant brightness distortions. To illustrate the alignment result, a conventional rectangular grid is drawn in Fig. 2.19a. Figure 2.19b shows its estimated position (in fact, the estimated position of the elements from Fig. 2.19a in Fig. 2.19b). If the image fragments are compared in the corresponding cells of these both grids, then it is possible to make a conclusion that an alignment is accurate enough. Note that according to the found shifts of image elements it is possible to estimate local speeds of air displacement, build the wind field, which is used, for example, to ensure the flight safety near airports.

Alignment task becomes even more complicated if the brightness distortion is high enough and has of a non-linear character. Then, correlation-extreme criteria are also inapplicable. In this case, the alignment may be performed on the basis of a morphological analysis, i.e. taking into account the image form [10, 11]. In this context, the PG procedures [16] can also be applied.

2.7 Conclusions

New approaches to solve the problems of modeling and statistical analysis of multidimensional data sequences, which can be presented in the form of changing images, are considered in this chapter. A description of images in the form of a random field, which is set on an integer two-dimensional or multidimensional

integer grid, is taken as a basis. A number of known model were analyzed and new probabilistic models were developed. It enabled an adequate description of a large class of static and variable images, including ones set on various surfaces. These models are subsequently used to formalize the image processing tasks and synthesize corresponding algorithms. The chapter touches upon the problems of image filtering, detection of point and lengthy anomalies on the background of noise with spatial correlation, image alignment and estimation of parameters of their interframe geometric transformations. Optimal and quasi-optimal algorithms that help to solve these problems were obtained. Their efficiency and computational complexity were analyzed. Special attention is paid to adaptive algorithms of image processing, when there is uncertainty in the description of the processed data.

Acknowledgements The reported study was funded by the Russian Fund for Basic Researches according to the research projects № 16-41-732041 and № 16-41-732027.

References

1. Shalygin, A.S., Palagin, Y.I.: Applied Methods of Statistical Modeling. Mechanical Engineering Leningrad: Mashinostroenie (1986)
2. Habibi, A.: Two-dimensional Bayesian estimate of images. Proc. IEEE **60**(7), 878–883 (1972)
3. Gimel'farb, G.L.: Image Textures and Gibbs Random Fields. Kluwer Academic Publishers, Dordrecht (1999)
4. Woods, J.W.: Two-dimensional Kalman filtering. In: Huang, T.S. (ed.) Two-Dimensional Digital Signal Processing I: Linear Filters. TAP, vol. 42 pp. 155–205 Springer, Berlin, Heidelberg, New York (1981)
5. Yaroslavsky, L.: Digital Picture Processing. An Introduction. Springer, Berlin, Heidelberg (1985)
6. Duda, R.O., Hart, P.E., Stork, D.G.: Pattern Classification, 2nd edn. Wiley-Interscience, New York (2000)
7. Dudgeon, D.E., Mersereau, R.M.: Multidimensional Digital Signal Processing. Signal Processing Series. Prentice-Hall, Englewood Cliffs, New York (1984)
8. Favorskaya, M.N., Levtin, K.: Early smoke detection in outdoor space by spatio-temporal clustering using a single video camera. In: Tweedale, J.W., Jain, L.C. (eds.) Recent Advances in Knowledge-Based Paradigms and Applications. AISC, vol. 234, pp. 43–56. Springer International Publishing, Switzerland (2014)
9. Gonzalez, R.C., Woods, R.E.: Digital Image Processing, 4th edn. Pearson/Prentice-Hall, New York (2017)
10. Serra, J. (ed.): Image Analysis and Mathematical Morphology. Vol 2: Theoretical Advances. Academic Press, London (1988)
11. Vizilter, Y.V., Pyt'ev, Y.P., Chulichkov, A.I., Mestetskiy, L.M.: Morphological image analysis for computer vision applications. In: Favorskaya, M.N., Jain, L.C. (eds.) Computer Vision in Control Systems-1, ISRL, vol. 73, pp. 9–58. Springer International Publishing, Switzerland (2015)
12. Gruzman, I.C., Kirichuk, V.P., Kosikh, G.I., Peretryagin, G.I., Spector, A.A.: Digital Image Processing in Informative Systems. Novosibirsk State Technical University (2000) (in Russian)
13. Huang, T.S. (ed.): Image Sequence Analysis. Springer, Berlin, Heidelberg, New York (1981)

14. Huang, T.S. (ed.): Image Sequence Processing and Dynamic Scene Analysis. Springer, New York (1983)
15. Soifer, V.A. (ed.): Computer Image Processing. Part I: Basic Concepts and Theory. VDM Verlag Dr. Muller E.K. (2009)
16. Vasil'ev, K.K., Krasheninnikov, V.R.: Statistical Analysis of Images. Ulyanovsk State Technical University (2015) (in Russian)
17. Vasil'ev, K.K., Dement'ev, V.E., Andriyanov, N.A.: Doubly stochastic models of images. Pattern Recognit. Image Anal. **25**(1), 105–110 (2015)
18. Vasil'ev, K.K., Popov, O.V.: Autoregression models of random fields with multiple roots. Pattern Recognit. Image Anal. **9**(2), 327–328 (1999)
19. Vasil'ev, K.K., Dement'ev, V.E., Andriyanov, N.A.: Application of mixed models for solving the problem on restoring and estimating image parameters. Pattern Recognit. Image Anal. **26**(1), 240–247 (2016)
20. Krasheninnikov, V.R.: Correlation analysis and synthesis of random field wave models. Pattern Recognit. Image Anal. **25**(1), 41–46 (2015)
21. Krasheninnikov, V.R., Kalinov, D.V., Pankratov, YuG: Spiral autoregressive model of a quasi-periodic signal. Pattern Recognit. Image Anal. **8**(1), 211–213 (2001)
22. Krasheninnikov, V.R., Mikeev, R.R., Kuzmin, M.V.: The model and algorithm for simulation of planets relief as surfaces image. Radioengineering **175**, 192–194 (2012). (in Russian)
23. Dikshit, S.: A recursive Kalman window approach to image restoration. IEEE Trans Acoust. Speech Signal Process. **30**(2), 125–140 (1982)
24. Jähne, B.: Digital Image Processing, 6th edn. Springer, Berlin, Heidelberg (2005)
25. Pratt, W.K.: Digital Image Processing. PIKS Inside. 3rd ed. Wiley, New York (2001)
26. Prewitt, J.M.S.: Object enhancement and extraction. In: Lipkin, B.S., Rosenfeld, A. (eds.) Picture Processing and Psychopictorics, pp. 75–149. Academic Press, New York (1970)
27. Zhuravlev, Yu.I.: An algebraic approach to recognition or classifications problems. Pattern Recognit. Image Anal. **8**(1), 59–100 (1998)
28. Favorskaya, M., Jain, L.C., Buryachenko, V.: Digital video stabilization in static and dynamic scenes. In: Favorskaya, M.N., Jain, L.C. (eds.) Computer Vision in Control Systems-1, ISRL, vol. 73, pp. 261–309 Springer International Publishing, Switzerland (2015)
29. Krasheninnikov, V.R., Potapov, M.A.: A way to detect the straight line trajectory of an immovable point for estimating parameters of geometrical transformation of 3D images. Pattern Recognit. Image Anal. **21**(2), 280–284 (2011)
30. Krasheninnikov, V.R., Potapov, M.A.: Estimation of parameters of geometric transformation of images by fixed point method. Pattern Recognit. Image Anal. **22**(2), 303–317 (2012)
31. Polyak, B.T., YaZ, Tsypkin: Optimal pseudogradient adaptation procedure. Autom. Remote Control **8**, 74–84 (1980)
32. Widrow, B., Stearns, S.D.: Adaptive Signal Processing. Prentice-Hall Inc., Englewood, Cliffs, NJ (1985)
33. Vasil'ev, K.K.: Statistical analysis of multidimensional images. Pattern Recognit. Image Anal. **9**(4), 732–748 (1999)
34. Krasheninnikov, V.R.: Wave image models on the surfaces. In: 8th Open German-Russian Workshop on Pattern Recognition and Image Understanding Nizhny, Novgorod, pp. 154–157 (2011)
35. Krasheninnikov, V.R., Kuznetsov, V.V., Lebedeva, E.Y., Krasheninnikova, N.A.: Optimization of dictionary and model library for recognition of speech commands based on cross-correlation portraits. Pattern Recognit. Image Anal. **23**(1), 80–86 (2013)
36. Krasheninnikov, V.R., Kopylova, A.S.: Algorithms for automated processing images of blood serum facies. Pattern Recognit. Image Anal. **22**(4), 583–592 (2012)
37. Vasil'ev, K.K., Dement'ev, V.E., Luchkov, N.V.: Analysis of efficiency of detecting extended signals on multidimensional grids. Pattern Recognit. Image Anal. **22**(2), 400–408 (2012)

http://www.springer.com/978-3-319-67515-2

Computer Vision in Control Systems-3
Aerial and Satellite Image Processing
Favorskaya, M.N.; Jain, L.C. (Eds.)
2018, XIV, 343 p. 145 illus., 45 illus. in color., Hardcover
ISBN: 978-3-319-67515-2