

Chapter 3

Methods for Detecting of Structural Changes in Computer Vision Systems

Yury S. Radchenko and Aleksey V. Bulygin

Abstract The automation of experimental investigations based on video recording and different artificial vision applications often require that changes in a sequence of frames be detected without the observer's assistance. Variations in brightness, color, and size of an object are easily detectable using energy criteria. Nevertheless, some problems demand the use of algorithms capable of responding to small scale and texture changes of images. These problems can be solved by applying the criteria of Mean Structural Similarity Index Measure (MSSIM) and the developed Mean Nonparametric Structural Similarity Index Measure (MNSSIM), as well as the spectral algorithm for detecting structural changes in a frame, which have been used to good effect in video codec analysis. The profitable features of these criteria are their computational simplicity and their conformance to the human visual system. The criteria have not only a sensitivity for difference of comparing frames, but also have high stability of Gaussian and non-Gaussian (impulse) noises. This chapter describes the MSSIM, the own developed MNSSIM algorithms, and the spectral criterion, which provides the experimental confirmation of operating characteristics and features. The use of these criteria in automatic detection of changes in video captured scientific research scenes, the detection of motion or variable fragments in video frames in the intelligent video systems, and the application in video coding systems are discussed.

Keywords Image quality assessment • Spectral change • Pixel change • Substance phase transitions

Y.S. Radchenko (✉)

Raiophysics Department, Voronezh State University,
Universitetskaya pl. 1, Voronezh 394006, Russian Federation
e-mail: ysradchenko@yandex.ru; deanery@phys.vsu.ru

A.V. Bulygin

Voronezh State University, Universitetskaya pl. 1,
Voronezh 394006, Russian Federation
e-mail: alex987@bk.ru

3.1 Introduction

The automation of scientific researches based on video recording, smart vision, and video coding applications require that changes in a sequence of frames be detected without observer's assistance. Variations in object brightness, color, and size are easily detectable by Mean-Square Error (MSE) and Peak Signal to Noise Ratio (PSNR) energy criteria [1]. However, a number of tasks require algorithms that react on structural (texture) variations of images. Generally, an image region with variable structure is formless.

Another problem of vision systems is a variation of observation conditions and also interference. The problem solution is based on the use of high robust and interference-immunity decision making algorithms. The detection of variations in the image segment structure is based on spectral and correlation analysis of spatial-temporal domain. At present, the quasi-optimum heuristic algorithms applying variations of correlation features exist. However, they are non-invariant to spectrum in various bases in relation to a segment movement and change of texture features.

The structural differences of images can be determined by various techniques. Figure 3.1 shows the criteria and the metrics being a basis to detect these differences. The authors analyzed difference estimation methods and algorithms for images presented by the numbered blocks.

The MSSIM criterion [2–4] and its modification MNSSIM [5] are the tri-criterion functionals that respond to the changes of brightness, contrast, and correlation features of image. Therefore, the MSSIM, the MNSSIM, and other modifications are the energy criteria for detecting image variations. They are sensitive to texture variations as well. The growing popularity of these criteria is proved by their quite appropriate compliance with the human vision system.

The image as a whole or its separate blocks can be expanded in a generalized Fourier series by using the system of orthogonal functions $\varphi_{km}(x, y)$. The following methods can be emphasized among a multitude of orthogonal basis: the Discrete Cosine Transform (DCT) [1] and its integer variant called pseudo-cosine transform [6], Walsh-Hadamard transform [1], wavelet transform [7, 8]. In the researches [9–12], the class of discrete polynomial transforms and easily version discrete Chebyshev transform or the Generalized DCT (GDCT) were proposed. The GDCT has a number of special properties that allow the efficient image processing.

The current research has proved that the spectral algorithms can be a base to implement the image structural variation detectors, which are robust to the change of observation conditions and interference. Let us notice that the MNSSIM algorithms and spectral algorithms are quite simple in computation.

The next Sect. 3.2 covers examples of using the MSSIM and the MNSSIM. Section 3.3 provides the description of spectral criteria of structural image similarity. Spectral field variation detection is discussed in Sect. 3.4. Experimental confirmation of structural similarity criteria is situated in Sect. 3.5. Conclusion is drawn in Sect. 3.6.

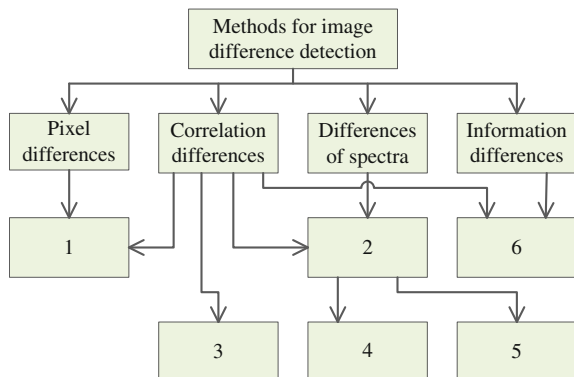


Fig. 3.1 Methods and algorithms for structural image difference detection, where 1 are algorithms based on MSSIM, MNSSIM1, and MNSSIM2 structural criteria, 2 are algorithms based on difference of generalized orthogonal basis spectra, 3 are algorithms based on difference of correlation features of moving video sequence fragments, 4 are spectral algorithms for texture anisotropy detection and estimation, 5 is a spectral algorithm for moving object detection, 6 are algorithms based on the Kullback and the Bhattacharyya information metrics

3.2 Pixel Structural Similarity Criteria

The task of detection and estimation for the structural similarity of two images having uncertain structures is a crucial issue in computer vision. The random images or frames of video sequence can be analyzed. This task cannot be formalized in full and has not yet been solved unambiguously. Objective structural similarity criteria may be classified in the following manner: single factor criteria, multi-factor criteria, and integral criteria being a combination of single factor and multi-factor criteria.

One of the simplest single factor criteria is a deviation of MSE. The MSE is determined by Eq. 3.1 for a separate image brightness or color component, where X, Y are images under a comparison ($X = (x_{ij}), Y = (y_{ij}), i = 1..n, j = 1..m$).

$$MSE(X, Y) = \frac{\sum_{i=1}^n \sum_{j=1}^m (x_{ij} - y_{ij})^2}{n \cdot m} \quad (3.1)$$

This criterion cannot be applied to human vision system. According to the MSE criterion, the images differ from each other, if the brightness reduces by 5 % only (the human vision system does not recognize this while the brightness options of different computer screens vary much more). At the same time, the images with the pronounced color variation of separate point, mild stripes, or frequency distortion resulting in a sharpness loss will be recognized as “almost unchanged”.

The widely used nowadays PSNR for a separate brightness or color component of an image is expressed by Eq. 3.2, where MAX_I is the maximum value assumed by the image component element.

$$PSNR(X, Y) = 10 \cdot \log \left(\frac{MAX_I^2}{MSE} \right) = 10 \cdot \log_{10} \frac{255^2 \cdot n \cdot m}{\sum_{i=1, j=1}^{n, m} (x_{ij} - y_{ij})^2} \quad (3.2)$$

For an RGB image, each component R, G, or B occupies 8 bits and, hence, $MAX_I = 2^8 - 1 = 255$ for this image. This measure is appropriate due to the logarithmic scale. It has the same drawbacks that root-mean-square deviation does [1, 13].

In a number of cases, the criterion should assume a variation of all image colors. The total PSNR for a full color RGB image $PSNR_{RGB}$ is calculated based on the summed squared error of the components provided by Eq. 3.3, where the maximum value is $\max S_{\Sigma}^2 = 3 \cdot 255^2 \cdot n^2 \cdot v$ ($n \times n$ is a number of pixels in a block, for the sake of simplicity the blocks are assumed to be square), v is a number of blocks, $\eta = 1 \dots v$ is a number of the current block, $(x_{i,j}^{(\eta)}, y_{i,j}^{(\eta)})_{R,G,B}$ are brightness of R, G, B components of pixels of blocks of images under comparison.

$$S_{\Sigma}^2 = \sum_{\eta=1}^v \left\{ \sum_{i,j=1}^n (x_{i,j}^{(\eta)} - y_{i,j}^{(\eta)})_R^2 + \sum_{i,j=1}^n (x_{i,j}^{(\eta)} - y_{i,j}^{(\eta)})_G^2 + \sum_{i,j=1}^n (x_{i,j}^{(\eta)} - y_{i,j}^{(\eta)})_B^2 \right\} \quad (3.3)$$

Hence, Eq. 3.3 can be re-written as Eq. 3.4.

$$PSNR_{RGB} = 10 \cdot \lg \left(\frac{\max S_{\Sigma}^2}{S_{\Sigma}^2} \right) = 10 \cdot \lg \left(\frac{3 \cdot 255^2 \cdot n^2 \cdot v}{S_{\Sigma}^2} \right) \quad (3.4)$$

The PSNR can be calculated for images in the YUV color format and other formats by equations similar to Eqs. 3.3–3.4. The combined criterion $PSNR_{RGB}$ provides the image similarity performance, which is more relevant for human vision. However, Eqs. 3.1–3.4 are slightly sensitive to texture image changes.

At the moment one of the criteria closest to the subjective perception of the recovered image quality, if the MSSIM [2–4, 14] characterizes a similarity of X and Y images by brightness, contrast, and structure, i.e. it is tri-factor. It appears as Eq. 3.5, where X_{η}, Y_{η} are the images compared in a block having number $\eta = 1 \dots v$, v is a number of blocks.

$$MSSIM = \frac{1}{v} \cdot \sum_{\eta=1}^v SSIM(X_{\eta}, Y_{\eta}) \quad (3.5)$$

The SSIM criterion is a block criterion. This means that it is applied not to the whole image at once but to its separate parts—equal blocks of the image, and later this value is averaged by all computed blocks producing the resulting MSSIM value for the whole image. In the general case, the $SSIM(X, Y)$ value for each block is calculated by Eq. 3.6, where $l(x, y)$ is a brightness comparison functional, $c(x, y)$ is a contrast comparison functional, $s(x, y)$ is a structure comparison functional, α, β, γ are the control coefficients.

$$SSIM(X, Y) = l(x, y)^\alpha c(x, y)^\beta s(x, y)^\gamma \quad (3.6)$$

In accordance with control coefficients $\alpha = \beta = \gamma = 1$ [4], the comparison functionals in the blocks are calculated by Eq. 3.7.

$$l(x, y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \quad c(x, y) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \quad s(x, y) = \frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3} \quad (3.7)$$

Here variables have the following meanings:

– μ_x, μ_y are the sample mean for image blocks X_η and Y_η , respectively,

$$\mu_x = \frac{1}{N^2} \cdot \sum_{i,j=0}^{N-1} x_{ij} \quad \mu_y = \frac{1}{N^2} \cdot \sum_{i,j=0}^{N-1} y_{ij}.$$

– σ_x^2, σ_y^2 are sample variance for image blocks X_η and Y_η , respectively,

$$\sigma_x^2 = \frac{1}{N^2} \cdot \sum_{i,j=0}^{N-1} (x_{ij} - \mu_x)^2 \quad \sigma_y^2 = \frac{1}{N^2} \cdot \sum_{i,j=0}^{N-1} (y_{ij} - \mu_y)^2.$$

– σ_{xy} is a moment of correlation between image blocks X_η and Y_η

$$\sigma_{xy} = \frac{1}{N^2} \cdot \sum_{i,j=0}^{N-1} (x_{ij} - \mu_x)(y_{ij} - \mu_y).$$

Within constant C_3 , a functional $s(x, y)$ coincides with the Pearson's sample correlation coefficient. C_1, C_2, C_3 are small constants preventing incorrect behavior of the criterion when the moments are cleared. In accordance with [2–4] one can assume $C_1 = (0.01L)^2, C_2 = (0.03L)^2, C_3 = 0.5C_2, L$ is image bit width.

The MSSIM criterion assumes values from -1 to 1 . The value of 1 is obtained only in the case, when one and the same image is compared. This means that closer an image compared to the original image, closer the criterion value to 1 .

A selection of functional $s(x, y)$ in Eq. 3.7 as a measure of structural difference is mostly justified, when the comparing vectors have the values with multivariate Gaussian distribution. Therefore, the MSSIM criterion perfectly distinguishes textures in a form of Gaussian noise. When the laws of distortion distribution are unknown (non-Gaussian), it is reasonable to apply estimates of the respective structural characteristics based on nonparametric statistics [15]. To estimate mean brightness, it is efficient to use a sample median with higher stability as compared with the sample mean and one of the rank correlation coefficients instead of the Pearson's correlation coefficient. Two modifications of the structural similarity criterion based on the nonparametric MNSSIM1 and MNSSIM2 was proposed in [5, 7].

For MNSSIM1, the functionals below (Eq. 3.8) are suggested instead of Eq. 3.7

$$l(x, y) = \frac{2m_x m_y + C_1}{m_x^2 + m_y^2 + C_1} \quad c(x, y) = \frac{2\sigma m_x \sigma m_y + C_2}{\sigma m_x^2 + \sigma m_y^2 + C_2} \quad s(x, y) = Rs(x, y), \quad (3.8)$$

where $m_x = \text{mediana}(\vec{x})$ and $m_y = \text{mediana}(\vec{y})$ are medians of the brightness vectors in image blocks x and y , respectively, $\sigma m_x^2 = \text{mediana}[(\vec{x} - m_x)^2]$ and $\sigma m_y^2 = \text{mediana}[(\vec{y} - m_y)^2]$ are medians of the squared vector difference of brightness and the median, $Rs(x, y)$ is Spearman's rank correlation coefficient [15]. The Rs values change from -1 to 1 , while $Rs = 0$ means the absence of correlation.

The functionals (Eq. 3.9) are used in the MNSSIM2.

$$l(x, y) = \frac{2m_x m_y + C_1}{m_x^2 + m_y^2 + C_1} \quad c(x, y) = \frac{2\sigma_x \sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \quad s(x, y) = Rs(x, y) \quad (3.9)$$

In other words, a contrast comparison functional $c(x, y)$ should remain the same as for the MSSIM (Eq. 3.7). The structure comparison functional $s(x, y)$ and brightness comparison functional $l(x, y)$ should be used as those used in the MNSSIM1 (Eq. 3.8). Constants C_1 and C_2 from Eqs. 3.8–3.9 are identical to those used to calculate the MSSIM [2, 4]. The MNSSIM2 criterion is computationally simpler than the MNSSIM1.

Therefore, the MNSSIM1 and the MNSSIM2 criteria are the tri-factor criteria that use the nonparametric estimations of random field parameters. Nonparametric criteria of the MNSSIM1 and the MNSSIM2 structural similarities are practically identical to the MSSIM criterion at the presence of Gaussian distortion. However, if a point interference or other non-Gaussian statistics interference or a block distortion take place, then the MNSSIM1 and the MNSSIM2 metrics are better for human subjective vision than the estimation by the MSSIM.

3.3 Spectral Criteria of Structural Image Similarity

The image in a block can be expanded into a generalized Fourier series by the system of continuous $\varphi_{km}(x, y)$ or discrete orthogonal functions $\varphi_{km}(i, j)$ [10–12, 17–19].

Let us examine the main types and features of orthogonal transforms for continuous arguments. A sequence of functions $\{\varphi_k(z)\}$, $k = 0, 1, \dots, \infty$ is called orthonormal against $\rho(z)$ provided by Eq. 3.10, where δ_{mn} is a Kronecker symbol, z has no dimension value.

$$(\varphi_m, \varphi_n) = \int \varphi_m(z) \varphi_n(z) \rho(z) dz = \delta_{mn} \quad (3.10)$$

For weight function $\rho(z)$, ratios $\rho(z) \geq 0$ and $0 \leq \int_a^b \rho(z) dz < \infty$ exist. The system of functions $\{\varphi_k(z)\}, \{\psi_k(z)\}, k = 0, 1, \dots, k'$ is called bi-orthogonal against weight $\rho(z)$ (Eq. 3.11).

$$(\varphi_m, \psi_n) = \int \varphi_m(z) \psi_n(z) \rho(z) dz = \delta_{mn} \quad (3.11)$$

The full and closed set of functions $\{\varphi_k(z)\}$ is of great interest.

In the following Sects. 3.3.1 and 3.3.2, the polynomial transform and the discrete transforms are discussed, respectively.

3.3.1 Polynomial Transforms

An important class of functions $\{\varphi_k(z)\}$ with the properties of orthogonality, completeness and closure is orthogonal polynomials $p_k(z)$ satisfying the Eq. 3.12, where d_m is a norm of polynomial $p_m(z)$.

$$\int_a^b \rho(z) p_m(z) p_k(z) dz = d_m \delta_{km} \quad (3.12)$$

To expand the signals, one can apply the following conventional orthogonal polynomials [11, 20, 21]:

- The Hermitian polynomials $H_m(z)$, Eq. 3.13.

$$\rho(z) = \exp(-z^2) \quad p_m(z) = H_m(z) \quad -\infty < z < \infty \quad (3.13)$$

- The Laguerre polynomials, Eq. 3.14.

$$\rho(z) = z^\alpha \exp(-z) \quad p_m = L_m^\alpha(z) \quad 0 \leq z < \infty \quad (3.14)$$

- The Jacobi polynomials $p_m(z) = P_m^{(\alpha, \beta)}(z)$, Eq. 3.15.

$$\rho(z) = (1-z)^\alpha (1+z)^\beta \quad \alpha > -1 \quad \beta > -1 \quad -1 \leq z \leq 1 \quad p_m(z) = P_m^{(\alpha, \beta)}(z) \quad (3.15)$$

The Jacobi polynomials form a wide group of orthogonal polynomials. The important particular cases of Jacobi polynomials are:

- The Legendre polynomials

$$P_n(z) \quad (\alpha = \beta = 0) \quad \rho(z) = 1.$$

- The Chebyshev polynomials of the 1st and 2nd types provided by Eq. 3.16.

$$\begin{aligned} T_n(z) \quad (\alpha = \beta = -1/2) \quad \rho(z) &= (1 - z^2)^{-1/2} \\ U_n(z) \quad (\alpha = \beta = 1/2) \quad \rho(z) &= (1 - z^2)^{1/2} \end{aligned} \quad (3.16)$$

For practical use, the most convenient Jacobi polynomials are the polynomials of Legendre and Chebyshev of the 1st and 2nd types.

The Chebyshev polynomials of the 1st type have a number of useful properties so that their use to expand the signals is very attractive. The Chebyshev polynomial of the 1st type $T_m(z)$ associated with weight function $\rho(z) = 1/\sqrt{1-z^2}$ can be determined variously. One of the most appropriate means has a view of Eq. 3.17.

$$T_m(z) = \cos(m \cdot \arccos(z)) \quad (3.17)$$

For the Chebyshev polynomials $T_m(-z) = (-1)^m \cdot T_m(z)$. According to $T_{m+1}(z) = 2 \cdot z \cdot T_m(z) - T_{m-1}(z)$, one can get Eq. 3.18.

$$\begin{aligned} T_0(z) &= 1 \quad T_1(z) = z \quad T_2(z) = 2z^2 - 1 \quad T_3(z) = 4z^3 - 3z \\ T_4(z) &= 8z^4 - 8z^2 + 1 \end{aligned} \quad (3.18)$$

The orthogonality condition of functions $T_m(z)$ is provided by Eq. 3.18.

$$\int_{-1}^1 \frac{T_m(z)T_k(z)}{\sqrt{1-z^2}} dz = \begin{cases} 0 & k \neq m \\ \pi & k = m = 0 \\ \frac{\pi}{2} & k = m \neq 0 \end{cases} \quad (3.19)$$

Nulls of the Chebyshev polynomials are easily determined from Eq. 3.17, which can be re-written as Eq. 3.20.

$$T_m(z) = \cos(m \cdot \arccos(z)) = 0 \quad (3.20)$$

Based on the above equations, one can get Eq. 3.21.

$$z_k = \cos \frac{2k+1}{2m} \pi \quad k = 0, 1, 2, \dots, m-1 \quad (3.21)$$

Expansion of function $f(z)$ in the Chebyshev polynomials of the 1st type has a form of Eq. 3.22, where $d_m = \begin{cases} \pi/2 & m \neq 0 \\ \pi & m = 0 \end{cases}$ is a norm of the Chebyshev polynomials.

$$C_m = \frac{1}{d_m} \int_{-1}^1 \frac{f(z)T_m(z)}{\sqrt{1-z^2}} dz \quad (3.22)$$

$$f(z) = \sum_{m=0}^{\infty} C_m T_m(z)$$

The expansion in the Chebyshev polynomials of the 1st type $T_m(z)$ is the most converged among all possible expansions in degrees z^k , $k = 0, 1, \dots, \infty$.

The above relations are generalized for a 2D case. Let Ω be a region of the 2D Euclidian space and $z = (z_1, z_2)$ be a point in this space. The basis of function orthonormality can be defined by scalar product (Eq. 3.23).

$$\int_{\Omega} \rho(z) \varphi_{km}(z) \psi_{rn}(z) dz = \delta_{kr} \delta_{mn} \quad (3.23)$$

The system of functions $\varphi_{km}(z)$, $\psi_{rn}(z)$ is bi-orthogonal. The system of functions $\{\varphi_{km}(z), \psi_{rn}(z)\}$ depends on the form of weight function $\rho(z)$ and geometry of region Ω . Then for function $f(z)$ with the finite norm and weight $\rho(z)$, there are possible two equal presentations by Eqs. 3.24–3.25.

$$C_{km} = \int_{\Omega} \rho(z) f(z) \varphi_{km}(z) dz \quad f(z) = \sum_{k,m} C_{km} \varphi_{km}(z) \quad (3.24)$$

$$B_{rn} = \int_{\Omega} \rho(z) f(z) \psi_{rn}(z) dz \quad f(z) = \sum_{r,n} B_{rn} \psi_{rn}(z) \quad (3.25)$$

It should be noted, that a bi-orthogonality is a typical feature of multi-dimensional expansions.

Processing implementation is greatly simplified, when functions $\varphi_{km}(x, y) = \varphi_k(x) \varphi_m(y)$ are factorized. For a signal $f(x, y, \tau)$, a pair of transforms is performed by Eq. 3.26, where (x, y) is a non-normalized coordinates of field's point, τ is a vector parameter of shift, rotation and other affine transformations.

$$f(x, y, \tau) = \sum_m \sum_k C_{mk}(\tau) \varphi_m(x) \varphi_k(y) \quad C_{mk}(\tau) = \int_{\Omega} \int f(x, y, \tau) \varphi_m(x) \varphi_k(y) dx dy \quad (3.26)$$

If a_x, a_y denote a typical size of sub-region Ω , $z_1 = x/a_x$, $z_2 = y/a_y$, then Eq. 3.26 can be re-written by Eq. 3.27, where d_m is a norm of orthogonal $p_m(z)$ polynomials with weight $\rho(z)$.

$$f(x, y) = \sum_{m,k} C_{mk} p_m(x/a_x) p_k(y/a_y) \quad (3.27)$$

$$\begin{aligned} C_{mk} &= (d_m d_k)^{-1} \iint_{\Omega} f(a_x z_1, a_y z_2) \rho(z_1) p_m(z_1) \rho(z_2) p_k(z_2) dz_1 dz_2 \\ &= (d_m d_k)^{-1} \int \rho(z_1) p_m(z_1) dz_1 \int f(a_x z_1, a_y z_2) \rho(z_2) p_k(z_2) dz_2 \end{aligned} \quad (3.28)$$

It is obvious from Eqs. 3.27–3.28, that the spectral coefficients are determined by a sequential integration of x, y coordinates that is greatly simplify the computation.

3.3.2 Discrete Transforms

Let us analyze a generality and a difference of continuous and discrete transforms (Table 3.1). Here $\{\phi_k(n)\}$, $n = 0 \dots N-1$ is a complex basis of orthonormal vectors. In the general case, it is necessary to use a bi-orthogonal basis $\{\vec{\varphi}_k(n)\}$, $\{\vec{\psi}_k(n)\}$ to expand a signal.

Let us consider a vector-matrix transform notation. Let vector \mathbf{F} be a column of signal samples, as show in Eq. 3.29.

$$\mathbf{F} = \begin{pmatrix} f(0) \\ \vdots \\ f(N-1) \end{pmatrix} \quad (3.29)$$

The transform matrix can be written by Eq. 3.30.

$$\Phi = \begin{pmatrix} \varphi_0(0) & \varphi_0(1) & \dots & \varphi_0(N-1) \\ \varphi_1(0) & \varphi_1(1) & \dots & \varphi_1(N-1) \\ \dots & \dots & \dots & \dots \\ \varphi_{N-1}(0) & \varphi_{N-1}(1) & \dots & \varphi_{N-1}(N-1) \end{pmatrix} \quad (3.30)$$

A pair of signal-to-spectrum transform in the matrix form has a view of Eq. 3.31.

Table 3.1 The difference of continuous and discrete transforms

Continuous signal	Discrete signal
$x \in [0, T]$	$x = x_n = n\Delta T, n = 0 \dots N - 1$
$f(x)$	$f(n) = f(x_n)$
$\varphi_k(x)$	$\vec{\varphi}_k = (\varphi_k(0) \varphi_k(1) \varphi_k(2) \dots \varphi_k(N - 1))$
$(\varphi_k, \varphi_m) = \int_T \varphi_k(x) \varphi_k^*(x) dx = \delta_{km}$	$(\vec{\varphi}_k, \vec{\varphi}_m) = \sum_{n=0}^{N-1} \varphi_k(n) \varphi_m^*(n) = \delta_{km}$
$C_k = \int_T f(x) \varphi_k^*(x) dx$	$C_k = \sum_{n=0}^{N-1} f(n) \varphi_k^*(n)$
$f(x) = \sum_{k=0}^{\infty} C_k \varphi_k(x)$	$f(n) = \sum_{k=0}^{\infty} C_k \varphi_k(n)$

$$\mathbf{C} = \mathbf{\Phi F} \quad \mathbf{F} = \mathbf{\Phi}^{-1} \mathbf{C} \Rightarrow \mathbf{\Phi}^T \mathbf{C} \quad (3.31)$$

If $\mathbf{\Phi}^{-1} = \mathbf{\Phi}^T$, $\det(\mathbf{\Phi}) = \pm 1$, then $\mathbf{\Phi}$ is an orthogonal matrix and any two lines of it are orthogonal vectors. For a bi-orthogonal matrix transforms, Eq. 3.32 can be written.

$$\mathbf{C} = \mathbf{\Phi F} \quad \mathbf{F} = \mathbf{\Psi C} \quad \mathbf{\Psi \Phi} = \mathbf{I} \quad (3.32)$$

Let us consider a discrete variant of 2D transform. In this case the function is setup on a 2D discrete point grid, Eq. 3.33.

$$f(x, y) \Rightarrow f(i, j) \quad i, j = 0 \dots N - 1 \quad (3.33)$$

The samples form a square matrix $\mathbf{F} = [f(i, j)]$. The same grid could be uniform or it can be formed non-uniformly by a special law.

A pair of transforms of a signal matrix to a spectral matrix is determined by Eq. 3.34.

$$C_{km} = \sum_{i,j} f(i, j) \varphi_{km}^*(i, j) \quad f(i, j) = \sum_{k,m} C_{km} \varphi_{km}(i, j) \quad (3.34)$$

If the discrete basis functions are factorized, then Eq. 3.35 is accomplished.

$$\varphi_{km}(i, j) = \varphi_k(i) \varphi_m(j) \quad (3.35)$$

Expansion and synthesis are reduced to serial operations in i and j . This operation is of the form below in the matrix form, Eq. 3.36.

$$\mathbf{C} = \mathbf{\Phi F \Phi}^T \quad \mathbf{F} = \mathbf{\Phi}^{-1} \mathbf{C} (\mathbf{\Phi}^{-1})^T \quad (3.36)$$

Let us consider the Discrete Fourier Transform (DFT). The basis orthogonal functions [1, 18, 19] have the form of Eq. 3.37.

$$\begin{aligned} \tilde{\varphi}_k(n) &= \exp\left(j2\pi\frac{kn}{N}\right) = W^{kn} \quad W = \exp\left(j\frac{2\pi}{N}\right) \quad n = 0 \dots N-1 \\ k &= 0 \dots N-1 \end{aligned} \quad (3.37)$$

The orthogonality condition and the functions' norm are defined by Eq. 3.38.

$$\sum_{n=0}^{N-1} W^{kn} W^{*ln} = \begin{cases} 0 & k \neq l \\ N & k = l \end{cases} \quad (3.38)$$

Then the pair of transforms with non-symmetric normalizing coefficients is computed by Eq. 3.39.

$$C_k = \frac{1}{N} \sum_{n=0}^{N-1} f(n) W^{kn} \quad f(n) = \sum_{k=0}^{N-1} C_k W^{-kn} \quad (3.39)$$

For the DFT, there are other variants of normalizing coefficients before the sum up, for example, the expressions from Eq. 3.40.

$$C_k = \sum_{n=0}^{N-1} f(n) W^{kn} \quad f(n) = \frac{1}{N} \sum_{k=0}^{N-1} C_k W^{-kn} \quad (3.40)$$

When the DFT is applied, it is essential that a discrete signal $f(n)$ is considered to be periodically extended with the period of N , while spectrum C_k is also discrete and periodic with the period of N . Therefore, a condition is imposed on border frequency Δf_m of continuous signal $f(t)$ and sampling step T : $\Delta f_m T \leq 0.5$. The DFT with reduced number of operations is called Fast Fourier Transform (FFT) [1, 17, 22]. If a signal $f(n)$ is true, then Eq. 3.41 is executed.

$$|C_k| = |C_{N-k}| \quad \arg(C_k) = -\arg(C_{N-k}) \quad (3.41)$$

Therefore, the FFT for such signal calculates only a half of spectral coefficients.

Let us consider the Discrete Cosine Transform (DCT) in two cases—one-dimensional and two-dimensional DCT.

One-dimensional DCT.

A non-normalized basis functions [1, 18, 19] are determined by Eq. 3.42, where $n = 0 \dots N-1$.

$$\tilde{\varphi}_k(n) = \cos\left(\pi k \frac{2n+1}{2N}\right) = \cos\left(\pi k \frac{n+0.5}{N}\right) \quad (3.42)$$

The parameter N may be both even and odd. However, the transform with even N [1] is often used in practice. The orthogonality condition for these functions is provided by Eq. 3.43.

$$\sum_{n=0}^{N-1} \tilde{\varphi}_k(n) \tilde{\varphi}_l(n) = \begin{cases} N & k = l = 0 \\ N/2 & k = l \neq 0 \\ 0 & k \neq l \end{cases} \quad (3.43)$$

During the calculation of a signal spectrum, the orthonormal basis functions is applied as it shown in Eq. 3.44.

$$\varphi_k(n) = \sqrt{\frac{2}{N}} \cdot g_k \cdot \cos\left(\pi k \frac{n+0.5}{N}\right) \quad g_k = \begin{cases} \sqrt{0.5} & k = 0 \\ 1 & k \neq 0 \end{cases} \quad (3.44)$$

In this case, the direct and the reverse DCTs are expressed by Eqs. 3.45–3.46.

$$C_k = \sqrt{\frac{2}{N}} \cdot g_k \sum_{n=0}^{N-1} f(n) \cos\left(\pi k \frac{n+0.5}{N}\right) \quad (3.45)$$

$$f(n) = \sqrt{\frac{2}{N}} \sum_{k=0}^{N-1} g_k C_k \cos\left(\pi k \frac{n+0.5}{N}\right) \quad (3.46)$$

Two-Dimensional DCT.

The transforms with symmetric normalization of direct and reverse transform matrices have a view of Eq. 3.47, where $\mathbf{F} = [f(i, j)]$ is a source block and $\mathbf{C} = [C_{km}]$ is its spectrum.

$$\mathbf{C} = \mathbf{\Phi} \mathbf{F} \mathbf{\Phi}^T \quad \mathbf{F} = \mathbf{\Phi}^T \mathbf{C} \mathbf{\Phi} \quad (3.47)$$

The matrices of the direct $\mathbf{\Phi}$ and reverse DCT coincide, Eq. 3.48.

$$\mathbf{\Phi} = [\varphi_k(n)] = \sqrt{\frac{2}{N}} \begin{bmatrix} \sqrt{0.5} \\ \cos\left(\pi k \frac{(n+0.5)}{N}\right) \end{bmatrix} \quad \begin{bmatrix} k = 0 \\ k \neq 0 \end{bmatrix} \quad k = 0 \dots N-1 \\ n = 0 \dots N-1 \end{matrix} \quad (3.48)$$

Matrix $\mathbf{\Phi}$ is square and orthogonal $\mathbf{\Phi}^{-1} = \mathbf{\Phi}^T$.

Integer cosine (pseudo-cosine) transform.

Standard H.264 applies a pseudo-cosine transform for blocks transformations. For 4×4 block, the DCT matrix is determined by Eq. 3.49, where $k, n = 0 \dots 3$.

$$\begin{aligned}
\Phi &= \sqrt{\frac{1}{2}} \cdot \begin{bmatrix} \sqrt{\frac{1}{2}} \\ \cos\left(\frac{(2n+1) \cdot k \cdot \pi}{2 \cdot 4}\right) \end{bmatrix} \\
&= \frac{1}{2} \cdot \begin{bmatrix} 1 & 1 & 1 & 1 \\ \sqrt{1 + \frac{1}{\sqrt{2}}} & \sqrt{1 - \frac{1}{\sqrt{2}}} & -\sqrt{1 - \frac{1}{\sqrt{2}}} & -\sqrt{1 + \frac{1}{\sqrt{2}}} \\ 1 & -1 & -1 & 1 \\ \sqrt{1 - \frac{1}{\sqrt{2}}} & -\sqrt{1 + \frac{1}{\sqrt{2}}} & \sqrt{1 + \frac{1}{\sqrt{2}}} & -\sqrt{1 - \frac{1}{\sqrt{2}}} \end{bmatrix} \quad (3.49)
\end{aligned}$$

It is proposed to use integer matrix \mathbf{H} [6] instead of matrix Φ , Eq. 3.50.

$$\mathbf{H} = \frac{1}{2} \cdot \begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 1 & -1 & -2 \\ 1 & -1 & -1 & 1 \\ 1 & -2 & 2 & -1 \end{bmatrix} \quad (3.50)$$

On the one hand, such replacement speeds up the integer operations of digital signal transform. On the other hand, this breaks a transform orthogonality. Such break should be compensated by the extra transforms in further steps.

The Hadamard system for Walsh functions.

The Walsh-Hadamard discrete basis is appropriate for computing. It is based on the Hadamard matrix. The matrix is calculated by the recurrent scheme, Eqs. 3.51–3.52.

$$\mathbf{A}_{2^n} = \begin{pmatrix} \mathbf{A}_{2^{(n-1)}} & \mathbf{A}_{2^{(n-1)}} \\ \mathbf{A}_{2^{(n-1)}} & -\mathbf{A}_{2^{(n-1)}} \end{pmatrix} \quad (3.51)$$

$$\mathbf{A}_1 = 1 \quad \mathbf{A}_2 = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \quad \mathbf{A}_4 = \begin{pmatrix} \mathbf{A}_2 & \mathbf{A}_2 \\ \mathbf{A}_2 & -\mathbf{A}_2 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{pmatrix} \quad (3.52)$$

Any line (column) of the Hadamard matrix is a discrete sample of the Walsh function of any order. The Hadamard matrix structure results in that only signal sample summation and subtraction operations are performed during the orthogonal transform. However, a convergence rate of series by the Walsh-Hadamard basis is less than the DCT provides. Besides, the spectrums are often non-monotonous, when this basis is applied.

Discrete Chebyshev GDCT.

The above mentioned discrete orthogonal transforms apply a uniform signal sample grid. However, a sample grid with a special non-uniformity allows to get the fast converged generalized Fourier series. The discrete Chebyshev transform called

Generalized DCT (GDCT) [9–12] belongs to this type of transforms. It is a particular case of transforms in the orthogonal polynomials.

To calculate integrals in the orthogonal polynomial transforms, it is proposed to use a Gaussian type quadrature formula with the highest algebraic accuracy [20] expressed by Eq. 3.53, where z_i are nulls of orthogonal polynomial $p_N(z)$ with weight $\rho(z)$, λ_i are the Christoffel numbers.

$$\int f(z)\rho(z)dz = \sum_{i=1}^N \lambda_i f(z_i) \quad (3.53)$$

Knots and weights $\{z_i\}, \{\lambda_i\}$ are clearly defined by the form of polynomial $p_N(z)$. In general, Eq. 3.53 expands in orthogonal polynomial system. A particular case of Eq. 3.53, when the Chebyshev polynomials are used, is called the Gauss-Chebyshev (Meller) formula [20] presented in a view of Eq. 3.54 form, where $z_i = \cos(\pi(i + 0.5)/N)$ are nulls of Chebyshev polynomial $T_N(z) = 0$, $\lambda_i = \pi/N = \text{const}$.

$$\int \frac{f(z)}{\sqrt{1-z^2}} dz = \frac{\pi}{N} \sum_{i=0}^{N-1} f(z_i) \quad (3.54)$$

For the Chebyshev polynomials, the direct and the reverse transforms (Eq. 3.22) can be applied (Eq. 3.54) (one-dimensional variant for normalized interval $z \in [-1, 1]$). Then Eqs. 3.55–3.56 will be received [9, 11], where $g_m = 1$ with $m > 0$ and $g_m = \sqrt{0.5}$ with $m = 0$.

$$C_m = g_m \cdot \sqrt{\frac{2}{N}} \sum_{i=0}^{N-1} f(z_i) \cdot \cos(\pi m \frac{i+0.5}{N}) \quad (3.55)$$

$$Y_M(z) = \sqrt{\frac{2}{N}} \sum_{m=0}^{M-1} g_m \cdot C_m \cdot \cos(m \cdot \arccos(z)) \quad (3.56)$$

In accordance with Eqs. 3.55–3.56, the sample points and the Chebyshev samples $z_i = \cos(\pi(i + 0.5)/N)$ of signal $f(z)$ are taken non-uniformly (Fig. 3.2). A synthesis (recovery) of signal $Y(z)$ by M spectral components is performed in random point $z \in [-1, 1]$, but not within a discrete set of sample points as this is done in the DCT. During recovery, any sample grid can be used, for example, uniform $z_n = 2n/(L - 1) - 1$, $n = 0 \dots L - 1$, if $L \neq N$. The recovered image is a subject to geometric scaling to downwards the size and upwards the size as well.

In the case of 2D GDCT $N \times N$, the Chebyshev samples of the image are taken within the sample block of $M1 \times M1$ points (pixels) by Eq. 3.57.

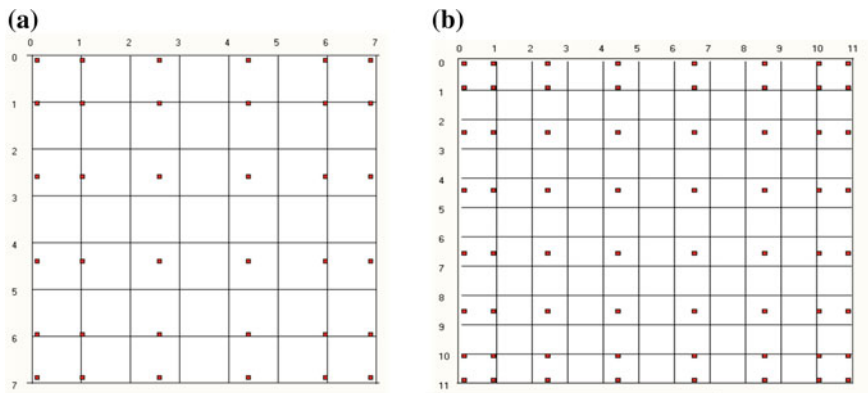


Fig. 3.2 The Chebyshev block sampling: **a** $N1/N = 8/6$, **b** $N1/N = 12/8$

$$\begin{aligned} x_i &= 0.5(N1 - 1) \cdot (1 + \cos(\pi(i + 0.5)/N)) \\ y_j &= 0.5 \cdot (N1 - 1) \cdot (1 + \cos(\pi \cdot (j + 0.5)/N)) \\ i, j &= 0 \dots N - 1 \end{aligned} \quad (3.57)$$

This matrix is transformed to spectral coefficient matrix \mathbf{C} of $M \times M$ size with the use of direct transform rectangular matrix of $M \times N$ size. In the case of reverse transform, a rectangular matrix of $L \times M$ size can be applied. That means that a recovered block $\mathbf{R} = [r_{ij}]$ ($i, j = 1 \dots L$) size is $L \times L$. The Chebyshev samples can be obtained by using linear interpolation by the nearest pixels [20, 23].

The direct and the reverse Chebyshev transforms (GDCT) in the matrix form are defined by the operations Eq. 3.58, i.e. the transform Eq. 3.36 falls into the bi-orthogonal transform class.

$$\mathbf{C} = \mathbf{\Phi}\mathbf{F}\mathbf{\Phi}^T \quad \mathbf{F} = \mathbf{\Psi}^T\mathbf{C}\mathbf{\Psi} \quad (3.58)$$

The matrix $\mathbf{\Phi}$ is a direct transform matrix (Eq. 3.59).

$$\mathbf{\Phi} = \sqrt{\frac{2}{N}} \cdot \begin{bmatrix} \sqrt{0.5} \\ \cos\left(\pi \frac{(i+0.5) \cdot m}{N}\right) \end{bmatrix}. \quad m = 0 \dots M - 1 \quad i = 0 \dots N - 1 \quad (3.59)$$

The matrix $\mathbf{\Psi}$ is a reverse transform matrix (Eq. 3.60).

$$\mathbf{\Psi} = \sqrt{\frac{2}{N}} \cdot \begin{bmatrix} \sqrt{0.5} \\ \cos\left(m \cdot \arccos\left(\frac{2n}{L-1} - 1\right)\right) \end{bmatrix}. \quad m = 0 \dots M - 1 \quad n = 0 \dots L - 1 \quad (3.60)$$

It is obvious from Eqs. 3.59–3.60, that in the general case the GDCT matrices $\mathbf{\Phi}$ and $\mathbf{\Psi}$ are rectangular. The direct GDCT matrix coincides with the direct DCT

matrix (Eq. 3.48) with $M = N$. The reverse transforms differ from each other. It should be noted that in the DCT, the transform matrices are square and have the size of $(N \times N = M \times M)$ as opposed to the GDCT, where M does not equal N in the general case [11, 12]. The GDCT is much close to the ideal de-correlation Karhunen-Loeve [1, 19] transform among all analyzed orthogonal transforms.

3.4 Spectral Image Variation Detection

Let a field $f^{(i)}(\mathbf{r}, t_i)$ being fragment $u(\mathbf{r})I_\Omega(\mathbf{r})\delta(t - t_i)$ of a space-time signal (dynamic image) at discrete time moment t_i , $i = 0, 1, \dots$ be observed in a sub-region $\mathbf{r} = (x, y) \in \Omega$ (a block of any frame). Here, $I_\Omega(\mathbf{r})$ is a sub-region indication function. Additionally let us assume that the white Gaussian noise $\eta(x, y) \equiv \eta(\mathbf{r})$ is available into an image. After comparing a separate block of the $(i-1)$ th and i th frames, one can hypothesize (Eq. 3.61).

$$\begin{aligned} H_0 : \xi(\mathbf{r}) &= \mathbf{f}_1(\mathbf{r}) + \eta(\mathbf{r}) & \mathbf{f}_1(\mathbf{r}) &= \mathbf{f}^{(i)}(\mathbf{r}) = \mathbf{f}^{(i-1)}(\mathbf{r}) \\ H_1 : \xi(\mathbf{r}) &= \mathbf{f}_2(\mathbf{r}) + \eta(\mathbf{r}) & \mathbf{f}_2(\mathbf{r}) &= \mathbf{f}^{(i)}(\mathbf{r}) \neq \mathbf{f}_1^{(i-1)}(\mathbf{r}) \end{aligned} \quad (3.61)$$

Basing on observation of $\xi(\mathbf{r})$, it is necessary to accept or reject the main hypothesis about a block image invariance. If two images are compared rather than video sequence frames are observed, then $f_1(\mathbf{r})$ is an unvaried texture in frames and $f_2(\mathbf{r})$ is a texture different from $f_1(\mathbf{r})$.

After checking the sophisticated hypotheses, signal $f_1(\mathbf{r})$, $f_2(\mathbf{r})$ are formed, and a structure uncertainty arises. Generally, the processing of the unknown form signal cannot be solved without the use of some additional factors. One of the most convenient ways to solve the a priori uncertainty is a signal parameterization. In this case, a signal form uncertainty transforms to a parametric uncertainty, which resolving ways are well designed [13, 19]. The convenient signal parameterization method has its presentation in the form of generalized Fourier series by any orthonormal basis φ_{km} [1, 10, 17, 19].

For different hypotheses, the signals can be presented by Eq. 3.62, where $\langle \eta_{k,m} \rangle = 0$, $D(\eta_{k,m}) = \frac{N_0}{2} \cdot \|\varphi_{km}\|^2$.

$$\begin{aligned} f_1(\mathbf{r}) &= \sum_{k,m} C_{km}^{(1)} \varphi_{km}(\mathbf{r}) & f_2(\mathbf{r}) &= \sum_{k,m} C_{km}^{(2)} \varphi_{km}(\mathbf{r}) \\ \xi(\mathbf{r}) &= \sum_{k,m} X_{km} \varphi_{km}(\mathbf{r}) & \eta(\mathbf{r}) &= \sum_{k,m} \eta_{km} \varphi_{km}(\mathbf{r}) \end{aligned} \quad (3.62)$$

The spectral coefficients for expansion of $C_{k,m}^{(2)}$ are assumed to be unknown. The coefficients $C_{k,m}^{(1)}$ are reference and defined.

Therefore, the hypotheses in a spectral definition are checked by Eq. 3.63.

$$\begin{aligned} H_0 : X_{km} &= C_{km}^{(1)} + \eta_{km} \\ H_1 : X_{km} &= C_{km}^{(2)} + \eta_{km} \end{aligned} \quad (3.63)$$

Section 3.4.1 provides the optimal detection algorithm and quasi-optimal algorithms are discussed in Sect. 3.4.2.

3.4.1 Optimal Detection Algorithm

The maximum likelihood algorithm [17, 24] is an asymptotically optimum rule for hypothesis check provided by Eq. 3.64, where $L(\cdot)$ is a log likelihood ratio functional determined by Eq. 3.65.

$$\max_{C^{(2)}} L(X|C^{(2)}) \underset{>}{<} h_0 \quad (3.64)$$

$$L(X|C^{(2)}) = \ln \left[\frac{W[X_{k,m}|f_2(r|C^{(2)})]}{W[X_{k,m}|f_1(r|C^{(1)})]} \right] \quad (3.65)$$

According to Eq. 3.64 it is required to determine the absolute maximum of log likelihood ratio functional by an unknown vector of parameters $C^{(2)}$ and make a non-randomized decision in favor of the respective hypothesis.

Now our selected spectral signal form parameterization (Eq. 3.62) is taken into account. To solve the hypothesis, let us check task by Eq. 3.66.

$$L(X|C^{(2)}) = (1/N_0) \left\{ 2 \sum_{k,m} X_{km} (C_{km}^{(1)} - C_{km}^{(2)}) - \sum_{k,m} \left((C_{km}^{(1)})^2 - (C_{km}^{(2)})^2 \right) \right\} \quad (3.66)$$

The following maximization by $C_{k,m}^{(2)}$ the decision rule [17] is obtained by Eq. 3.67.

$$(1/N_0) \sum_{k,m} \left(X_{k,m} - C_{k,m}^{(1)} \right)^2 \underset{>}{<} h_0 \quad (3.67)$$

3.4.2 Quasi-optimal Algorithms

Due to the fact that N_0 determines the interference power and typically is unknown in practice, the normalized statistics can be used, They may be written in the general form [17] by Eq. 3.68.

$$D_0 = \frac{\sum_{k,m} (X_{k,m} - C_{k,m}^{(1)})^2}{(C_{0,0}^{(1)})^2} \text{ or } D_E = \frac{\sum_{k,m} (X_{k,m} - C_{k,m}^{(1)})^2}{\sum_{k,m} (C_{k,m}^{(1)})^2} \quad (3.68)$$

As $C_{00}^{(1)}$ is proportional to mean image brightness in the frame block and $\sum (C_{km}^{(1)})^2$ is a block energy, the statistics D_0, D_E are stable to block brightness variation, i.e. to observation conditions. Taking into account Eq. 3.68, the decision rule (Eq. 3.64) is written by Eq. 3.69.

$$D_0 \begin{matrix} \gamma_1 \\ > \\ < \\ \gamma_0 \end{matrix} h_0 \text{ or } D_E \begin{matrix} \gamma_1 \\ > \\ < \\ \gamma_0 \end{matrix} h_E \quad (3.69)$$

The number of summable summands and order of their selection in statistics D_0, D_E can be setup from peculiarities of a certain task.

3.5 Experimental Research of Structural Similarity Algorithms

Let us consider the applications of structural similarity algorithms for image analysis (Sect. 3.5.1), the experimental research of spectral statistics (Sect. 3.5.2), and the experimental research of MSSIM and MNSSIM1(2) criteria (Sect. 3.5.3).

3.5.1 Practical Using of Pixel and Spectral Algorithms in Image Analysis

The MSSIM, MNSSIM1, and two structural similarity criteria, the D_0, D_E spectral criteria were experimentally studied. Such performance as sensitivity to texture variation, interference tolerance, robustness to observation conditions and decision making threshold selection were tested [16].

In our research for detection of substance phase transitions, when heating, the 75-frames from video sequence were selected. It was taken by Infinity 1-3C digital camera during an experiment on heating cesium chloride (CsCl) sample within the temperature range of (250; 710)°C. (Each frame represents the substance at a certain temperature and temperature variation differs from frame to frame).

Two phase transitions take place in the CsCl sample into this temperature range. The sample surface texture changes sharply (stepwise), and, hence, a texture of successive frames corresponding to this transition also changes abruptly. Figure 3.3

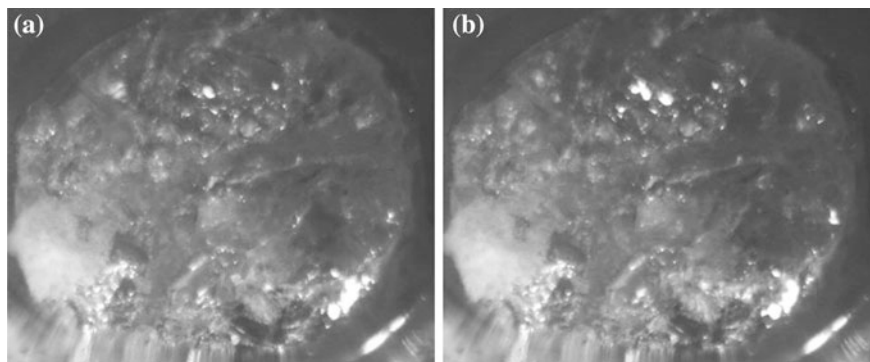


Fig. 3.3 Fragment of video sequence corresponding to the phase transition (polymorphous transformation) of the CsCl sample at temperatures: **a** 462.8 °C; **b** 468.4 °C

presents the images with small visual differences in the texture of some regions (blocks), which correspond to the first effect, polymorphous transformation of the substance (i.e., transition from one solid state to another followed by reconstructive re-arrangement of the crystal mosaic structure or “grains” in the image).

Figure 3.4 presents the images with noticeable visual differences, which correspond to the second observed effect, substance melting (transition from the solid state to the liquid one followed by crystal fracture and disappearance of grains in the image).

At the phase of transition moments, this image structure should have a step. The MSSIM index, its modification MNSSIM1(2) and spectral metric for detecting changes in video sequence D_0 were used.

Figure 3.5 shows the $MSSIM(T)$, the $MNSSIM1(T)$ and the $MNSSIM2(T)$ plots. Figure 3.6 demonstrates the $D_0(T)$ plot. All these dependences have two clearly

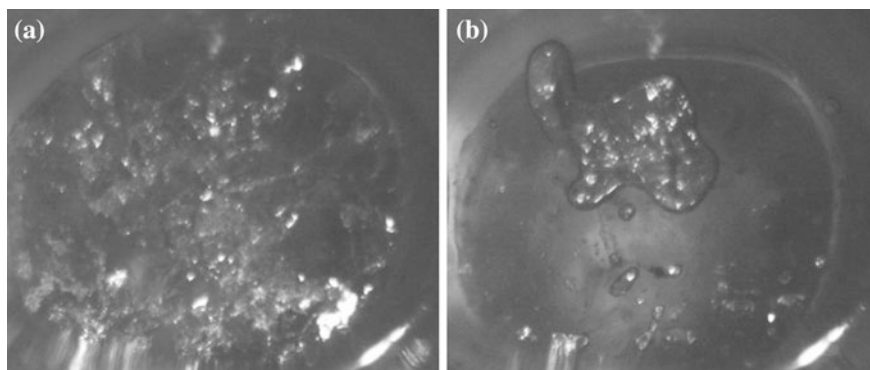


Fig. 3.4 Fragment of video sequence corresponding to the phase transition (melting) of the CsCl sample at temperatures: **a** 642.7 °C; **b** 645.2 °C

discernible peaks (extrema) corresponding to the phase transitions temperatures. It is the left jump that has an important physical meaning, since it represents a phase transition (significant changes appeared); and the right jump means that the changes ceased. The fluctuations (small peaks) are explained by the random changes in frame brightness as well as random changes of the substance structure. The oscillations (see $MSSIM(T)$, $MNSSIM1(T)$, $MNSSIM2(T)$ plot) after the second high peak at $645.2\text{ }^{\circ}\text{C}$ temperature can be attributed to the fact that, upon melting, the changes still proceed on the surface of the substance under observation, thus reflecting on the image texture. It is obvious from Fig. 3.5, that for the given video sequence, the $MSSIM$, the $MNSSIM1$, and the $MNSSIM2$ dependences are similar. Therefore, the $MNSSIM1$ and the $MNSSIM2$ plots are almost identical. However, if the analyzed video sequence is subjected to some distortion factors (e.g., pulse noise) during a video recording and an artificial improvement of images, then the $MNSSIM1$ and the $MNSSIM2$ criteria seem to be more preferable.

Peak values of the $MSSIM$, the $MNSSIM1$, the $MNSSIM2$ and D_0 , their mean values before the peaks (to estimate jumps) and temperatures of the peaks are provided in Tables 3.2 and 3.3.

The temperature values detected by the analysis of the sharp jumps in the $MSSIM(T)$ the $MNSSIM1(T)$, the $MNSSIM2(T)$, and the $D_0(T)$ dependences, correlate well with data from literature sources [16] $T_1 = 469\text{ }^{\circ}\text{C}$, $T_2 = 645\text{ }^{\circ}\text{C}$.

Figures 3.5 and 3.6 also result that the $MSSIM$ criterion and its $MNSSIM1(2)$ modification are inferior to spectral criterion the D_0 in the scale of jumps. In the peaks, the $MSSIM$ and the $MNSSIM1(2)$ curves change several-fold, whereas the D_0 curve changes several tens of times. This feature of the spectral criterion can be useful in automatic detection of process change moment. Even such a low automatically set threshold as $6 \div 8\text{ min}$ (D_0) ensures that fluctuations of the image brightness and color not followed by texture changes will not belong to useful effects. Due to higher fluctuations of the $MSSIM$, the $MNSSIM1$ and the $MNSSIM2$ criteria, a threshold selection requires the additional data analysis by the researcher.

Fig. 3.5 The plots of $MSSIM(T)$, $MNSSIM1(T)$, $MNSSIM2(T)$

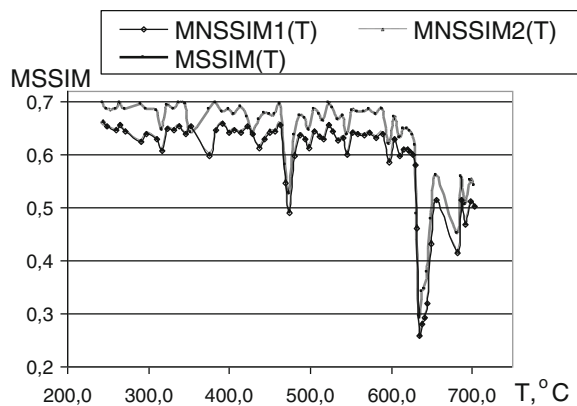


Fig. 3.6 The plot of $D_0(T)$

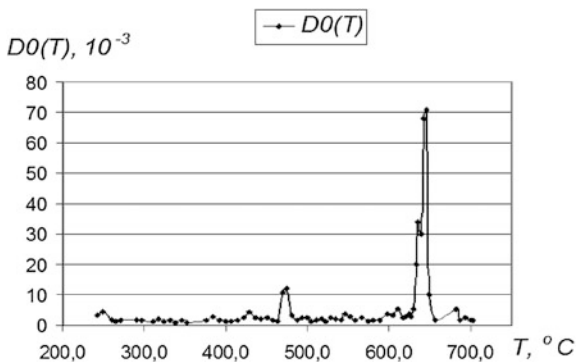


Table 3.2 Values before the peaks

Peak no	Peak temperature, °C	MSSIM peak value	MNSSIM1 peak value	MNSSIM2 peak value	$D_0, 10^{-3}$ peak value
1 peak	468.4	0.526	0.489	0.496	12
2 peak	645.2	0.296	0.259	0.261	70

Table 3.3 Peak values

Peak no	MSSIM mean value before peak	MNSSIM1 mean value before peak	MNSSIM2 mean value before peak	D_0 mean value before peak, 10^{-3}
1 peak	0.686	0.642	0.644	1
2 peak	0.668	0.626	0.628	1

3.5.2 Experimental Research of Spectral Statistics D_0 and D_E

To estimate the algorithm provided by Eq. 3.68, the experiments with real dynamic fields were done [17]. During experiment, values of statistics D_0 and D_E (Eq. 3.68) for block of brightness components of two diverted video sequence frames were calculated. Blocks characterizing by such change types as practically unvaried, weakly, or strongly varied were selected as blocks for analysis. Figure 3.7 shows an example of a frame from the analyzed dynamic sequence. The distinctive blocks, where statistics values were computed and differences were detected, are marked with boxes. The GDCT discrete transform is used.

Figure 3.7 shows a frame of video sequence with varied blocks (marked in boxes) for different thresholds: The results are tuned for $h_E = 0.01$ and $h_E = 0.04$

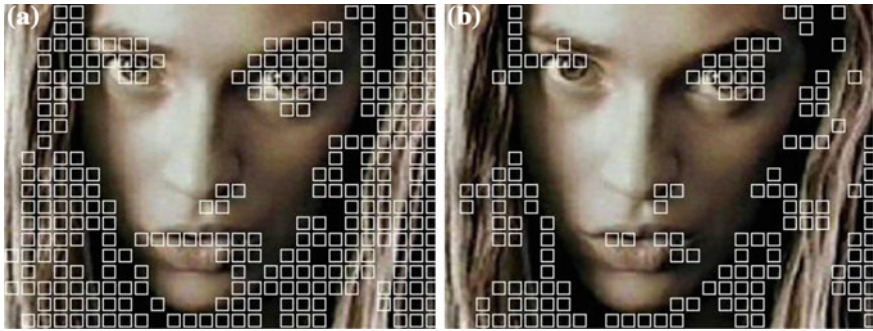


Fig. 3.7 Image frames with varied blocks obtained by the full spectrum: **a** $h_E = 0.01$, **b** $h_E = 0.04$

(Fig. 3.7a, b, respectively.) Four different blocks were analyzed and average statistics were calculated for each variation type. The obtained statistics D_0 and D_E for the brightness component for the image (Fig. 3.7) are presented in Table 3.4.

The given results prove that for blocks with low changes, metrics D_0 and D_E are quite close. Their difference from metrics for varied (flexible) blocks is 2–3 orders. This allows to select easily the threshold separating flexible and inactive blocks. It should be noted that metric D_0 change range is 1.7–2 times more than that metric D_E . Higher change range of statistics D_0 proves its preference. Thresholds h_0 and h_E are connected with the following relation:

$$h_0 = h_E \frac{\sum_{k,m} (C_{k,m}^{(1)})^2}{(C_{0,0}^{(1)})^2}$$

Table 3.5 presents the h_E thresholds averaged by frames, calculated thresholds \bar{h}_0 , and percentage of varied block.

It is evident from Table 3.5, that 10-fold threshold change results in number of blocks varied in less than 3 times. This phenomenon proves a non-criticality of threshold selection, which separates flexible and inactive blocks being a result of the above described test.

The possibility for use of decision rules (Eq. 3.68) with the truncated spectrum was analyzed in different video sequences. Our analysis demonstrated that detection

Table 3.4 Brightness component’s statistics for the image (Fig. 3.7)

Average statistics	Flexible fragments			Inactive fragments	
	Eyes	Lips	Hair	Cheeks	Forehead
$D_0 \cdot 10^3$	158.90878	74.04290	39.61026	0.82546	0.11304
$D_E \cdot 10^3$	91.14991	62.19916	25.68643	0.81213	0.11298

Table 3.5 Thresholds averaged by frame and percentage of block varied

Parameter	Frame 1	Frame 2	Frame 3	Frame 4	Frame 5	Frame 6	Frame 7
h_E	0.005	0.010	0.020	0.025	0.030	0.040	0.050
\bar{h}_0	0.011	0.021	0.042	0.053	0.064	0.085	0.106
Varied blocks, %	60.56	50.74	38.52	33.15	30.00	26.11	21.11

can be performed by the truncated spectrum and even by one spectral component (mean block brightness). Figure 3.8 shows an example of video sequence frames with varied blocks defined by metrics D_E with the full spectrum ($N = 64$) (see Fig. 3.8a) and one spectral component ($N = 1$) (see Fig. 3.8b). A threshold value was $h_E = 0.01$ for both figures.

Table 3.6 provides the percentage of the changed blocks detecting by the whole spectrum ($D(N)$, %) and by one spectral component ($D(1)$, %) as well as by a relative ($\Delta D/D_N$) detection error by one spectral component.

Here, $\frac{\Delta D}{D(N)} = \frac{D(N) - D(1)}{D(N)}$. It is evident from the Table 3.6 that the maximum relative error of variation detection by one spectral component does not exceed 0.16. Therefore, the block changes can be satisfactorily detected by one spectral component. Test performed with test video sequences “Container”, “Foreman”, and “Suzie” demonstrated the similar results.

The spectral video sequence block variation detection algorithm has been used to implement the MGDCT video coding concept [25]. Our proposed concept applies the DCT/GDCT orthogonal transforms in a video codec structure.

To decrease a video codec bit-rate, the algorithm (Eq. 3.68, 3.69) was used to detect the reliable inter-frame video sequence variations. Figures 3.9–3.10 show video frames with marked blocks that have varied because of moving objects. Figure 3.9 applied “Container” video sequence, and Fig. 3.10 presents a frame of a remote video monitoring system. The detectors apply the GDCT transform.

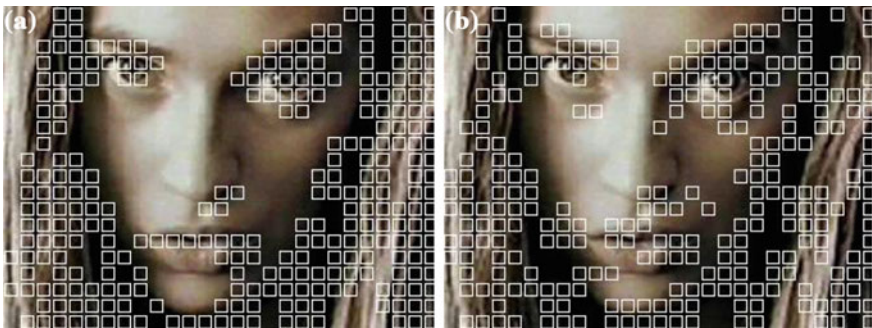
**Fig. 3.8** Image frames with varied blocks obtained by full and truncated spectrum: **a** full spectrum, **b** truncated spectrum

Table 3.6 Percentage of changed blocks

h_E	0.005	0.01	0.02	0.025	0.03	0.04	0.05
$D(N), \%$	60.56	50.74	38.52	33.15	30.00	26.11	21.11
$D(1), \%$	56.67	47.96	39.63	36.11	31.85	27.78	24.44
$\Delta D/D_N$	0.0642	0.0547	-0.0288	-0.0894	-0.0617	-0.0638	-0.1579

Combining the Chebyshev sampling and spectral detection allows to reduce information processing in 4–8 times depending on frame variations.

3.5.3 Experimental Research of MSSIM and MNSSIM1(2) Criteria

Notwithstanding that the MSSIM structural similarity criterion is one of the criteria closest to the human vision system; it has drawbacks of certain types. For example, in the case of image blur, pulse noise, or blocking, the MSSIM criterion provides values that are not quite similar with vision system. The MNSSIM 1(2) criteria have not such effects [5]. Let us investigate the MSSIM, MNSSIM1(2) criteria under various artifacts.

The MSSIM, MNSSIM1(2) criteria at pulse interference.

The image called “Lena” was used to test impact of pulse interference on the MSSIM, the MNSSIM1, and the MNSSIM2 criteria. Figure 3.11 shows images disturbed by pulse noise of “pepper” and “salt/pepper” types with probability $p = 0.05$, and the MSSIM, the MNSSIM1, and the MNSSIM2 criteria values corresponding to them.

A pulse noise was setup by two characteristics: an intensity determined by noise probability in pixel p and a noise type with three variants such as “salt”, “pepper”, and “salt/pepper”. The mathematical model of pulse noise for different noise peaks is determined as follows. Let $X = [x_{ij}]$ be a non-distorted image and $Y = [y_{ij}]$ be a distorted image. For the “pepper” type noise pulse, the model is setup by Eq. 3.70.

$$y_{ij} = \left\{ \begin{array}{l} 0 : \text{prob.}p \\ x_{ij} : \text{prob.}1 - p \end{array} \right\} \tag{3.70}$$

For the “salt” type noise pulse, the model is setup by Eq. 3.71.

$$y_{ij} = \left\{ \begin{array}{l} 255 : \text{prob.}p \\ x_{ij} : \text{prob.}1 - p \end{array} \right\} \tag{3.71}$$

For the “salt/pepper” type noise pulse, the model is setup by Eqs. 3.72–3.73.

Fig. 3.9 Spectral variation detection in “Container” test sequence

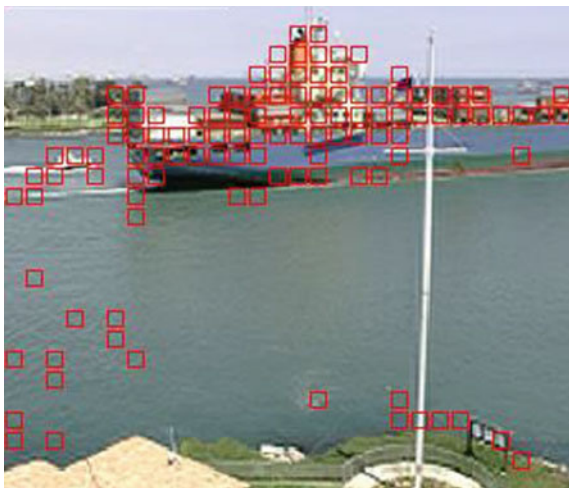


Fig. 3.10 Spectral variation detection in the video monitoring system



$$f_0 = \left\{ \begin{array}{l} 0, \text{prob.}0.5 : \text{noise} \backslash \text{pepper} \\ 255, \text{prob.}0.5 : \text{noise} \backslash \text{salt} \end{array} \right\} \quad (3.72)$$

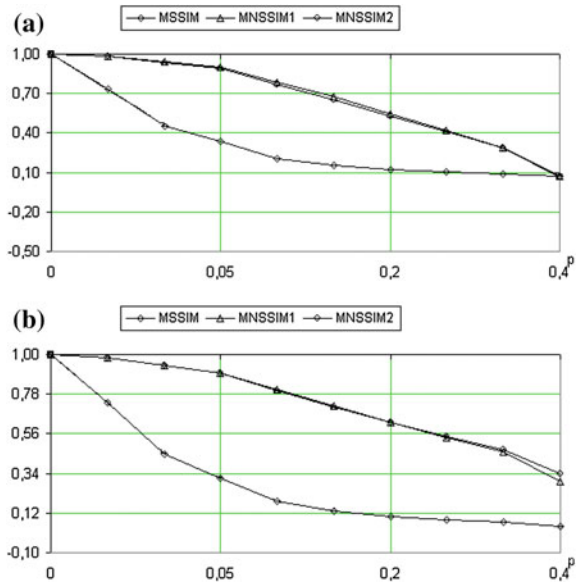
$$y_{ij} = \left\{ \begin{array}{l} f_0 : \text{prob.}p \\ x_{ij} : \text{prob.}1 - p \end{array} \right\} \quad (3.73)$$

Figure 3.12 shows a behavior of criteria depending on probability of pulse interference p . Plots from Fig. 3.12a, b confirm that the MSSIM criterion greatly reduces the image quality at low interference intensity. The MNSSIM1 and the MNSSIM2 matrices demonstrate values more identical to human vision system and close to each other (a human vision filters the low interference intensity). At higher interference intensity, the MSSIM criterion passes to the “saturation” mode.



Fig. 3.11 “Lena” image: **a** “pepper” noise with $p = 0.05$, MSSIM = 0.371, MNSSIM1 = 0.899, MNSSIM2 = 0.902, **b** “salt/pepper” noise with $p = 0.05$, MSSIM = 0.315, MNSSIM1 = 0.899, MNSSIM2 = 0.898

Fig. 3.12 The MSSIM, MNSSIM1 and MNSSIM2 values versus pulse noise intensity and noise type: **a** “salt” type of pulse noise, **b** “salt/pepper” type of pulse noise



The MNSSIM1(2) criteria values within region of high p values almost reduce, when this parameter increases

The MSSIM and the MNSSIM1(2) criteria at quasi-Gaussian noise.

Let us clarify the term of quasi-Gaussian interference in a digital image. Each of color components RGB takes on a value from 0 to $2^8 - 1 = 255$ at integer 8-bit signal level presentation. Therefore, each color RGB component at quasi-Gaussian

interference is formed by Eq. 3.74, where $\eta_{ij} \in N(0, 1)$ is the independent Gaussian values with zero mean and unit variance for each image pixel, $\sigma_0 \leq 1$ is a normalizing factor that control noise variance, $[z]$ is an integral part of number z .

$$y_{ij} = x_{ij} + [\sigma_0 \cdot \eta_{ij} \cdot 2^5] + 2^7 \quad (3.74)$$

Coefficient $2^5 = 32$ and deviation $2^7 = 128$ are selected so that resulting values with probability $p = 0.995$ are within the interval $0 \leq y_{ij} \leq 255$. If a resulting value is beyond this interval, i.e. $y_{ij} < 0$ or $y_{ij} > 255$, then a rounding is performed to $y_{ij} = 0$ or $y_{ij} = 255$, respectively.

The images from database of Laboratory for Image and Video Engineering [26] were used as test images. Subjective quality values of Difference Mean Opinion Score (DMOS) values for images from this database were also obtained. For images noisy with Gaussian noise, the MSSIM, the MNSSIM1, and the MNSSIM2 criteria values as well as the rank Spearman correlation coefficients between these criteria values and DMOS values were calculated.

The rank Spearman correlation coefficients between the MSSIM, the MNSSIM1, and the MNSSIM2 criteria and DMOS values are computed as follows. The calculated values of the MSSIM, the MNSSIM1, and the MNSSIM2 criteria are assigned with ranks, and ranks are also set to respective DMOS values. Then the Spearman correlation coefficient is calculated. A number of elements in sequence n equals 49 for all quality criteria and their respective DMOS values. For each of the MSSIM, the MNSSIM1, and the MNSSIM2 quality criteria, the obtained correlation coefficients were checked. Test results of image “Parrots” are provided by Table 3.7.

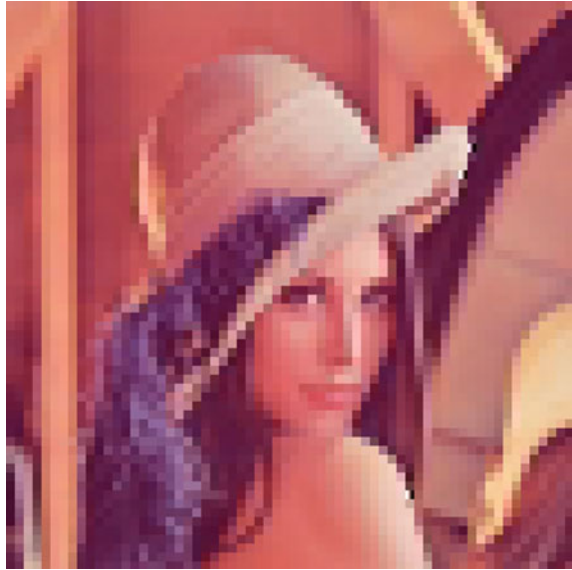
It is evident from Table 3.7, that the MSSIM criterion values and variation range are higher than those of the MNSSIM1 and the MNSSIM2 criteria at quasi-Gaussian noise. This is explained by the fact that the MSSIM criteria structure is more suitable for images with Gaussian statistics. All criteria are strongly correlated with DMOS.

With strong image block distortion or blue, the MSSIM criterion produces an overestimate. At the same time the MNSSIM1 and the MNSSIM2 criteria have low values, thus corresponding to low visual quality of the distorted images. Figure 3.13 presents an image with block distortion and high JPEG algorithm compression.

Table 3.7 Correlation results for image “Parrots”

σ_0	DMOS	MSSIM	MNSSIM1	MNSSIM2
1.000	68.72	0.027	0.012	0.017
0.129	47.03	0.252	0.152	0.208
0.063	38.93	0.518	0.314	0.359
0.031	28.50	0.788	0.520	0.541
Spearman correlation coefficient		-0.955	-0.929	-0.921

Fig. 3.13 “Lena” image recovered after JPEG compression.
 MSSIM = 0.634,
 MNSSIM1 = 0.073,
 MNSSIM2 = 0.079



Possibilities of use of structural similarity criteria to detect changes of actions.

During research, an issue concerning the ability of structural similarity criteria to track the contextual similarity of images was investigated. To solve this task, the MSSIM, the MNSSIM1, and the MNSSIM2 criteria were applied to different images. It was discovered that, when comparing absolutely different images with the same spatial sizes the MSSIM criteria gives unreasonably high values, whereas the MNSSIM1 and the MNSSIM2 criteria values are practically equal to 0. Figure 3.14 shows an example of images compared by the MSSIM, the MNSSIM1, and the MNSSIM2 criteria values. The obtained data proves that the non-parametric modifications are more relevant to the name of structural similarity criterion.

Comparison of pixel and spectral image analysis algorithms.

Multiple tests with real images and video sequences were made by the authors to discover the features and the abilities provided by the above described algorithms. The following intermediate conclusion can be performed:

1. The MSSIM and the MNSSIM1(2) structural similarity criteria are efficient to detect changes in frame and video sequence fragments, when images are processed without compression.
2. The non-parametric the MNSSIM1(2) criteria require more operations as compared with the MSSIM and the spectral algorithm. However, their values are more compatible with human perception.
3. Among all analyzed algorithms, the MNSSIM1(2) has the highest immunity to pulse and other non-Gaussian interference.
4. The MNSSIM1(2) non-parametric criteria could be applied to determine change in a video sequence scene and to detect external frames added to a video sequence.

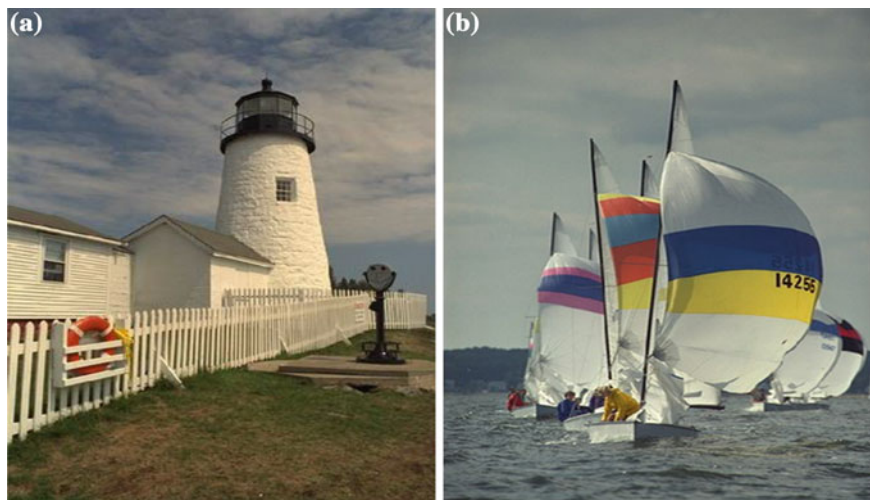


Fig. 3.14 Comparison of two images with $MSSIM = 0.339$, $MNSSIM1 = -0.004$, $MNSSIM2 = -0.006$

5. The algorithms for spectral structural variation detection obtained by maximum likelihood method are optimal with Gaussian interference.
6. The quasi-optimal detection algorithms applying statistics are similar to the optimal ones by their characteristics. They are not sensitive for threshold selection and image type. The spectral algorithms are more sensitive to change of an image type as compared to the $MSSIM$ and the $MNSSIM1(2)$.
7. The difference in statistics values with or without texture variations are tens/hundreds times in case of spectral algorithms and several times in the case of the $MSSIM$ and the $MNSSIM1(2)$ algorithms.
8. As for computation expenses, the spectral and pixel algorithms are approximately similar. They can operate in the real-time mode.
9. The spectral algorithms are very efficient in real-time operation, especially when they are embedded to video codecs. It has been found that these algorithms can operate in the truncated spectrum width with a few components.
10. A promising spectral basis is the discrete Chebyshev transform (GDCT). The GDCT spectrum is the most fast decreasing for orthogonal transforms.
11. A combining of the Chebyshev sampling and a spectral detection allows to reduce information processing in 4–8 times depending on a frame variation nature.
12. To solve a certain task, the proposed algorithms should be selected by their resource consumption, computation automation degree, and image distortions.

3.6 Conclusion

The chapter covers the use of the MSSIM, the MNSSIM1, and the MNSSIM2 criteria and the spectral algorithms D_0 and D_E to detect changes in a video sequence or to compare textures of various images. The developed MNSSIM1(2) criteria and the spectral algorithms D_0 and D_E can be used in artificial vision systems, in many other scopes connecting with variations of texture, spectral and correlation parameters of recorded signals. To meet a certain challenge, the selection of the proposed algorithm should be determined by the required resource intensity, an automation degree, and availability of image distortions. The proposed algorithms are used to design novel video codecs and intelligent video record systems. To conclude one can emphasize that evolution of basic PSNR, the MSSIM criteria and the spectral criteria, etc. is proceeding in [27–31].

References

1. Pratt WK (2001) Digital image processing. PIKS Inside, 3rd edn. Wiley, New York
2. Wang Z, Bovik AC (2009) Mean squared error: love it or leave it. *IEEE Signal Process Mag* 26(1):98–117
3. Wang Z, Bovik AC (2002) A universal image quality index. *IEEE Signal Process Lett* 9 (3):81–84
4. Wang Z, Bovik AC, Sheikh H, Simoncelli E (2004) Image quality assessment: from error visibility to structural similarity. *IEEE Trans Image Process* 13(4):600–612
5. Bulygin AV, Radchenko TA, Radchenko YuS (2008) Comparative analysis of structural likelihood index modification. *Digit Signal Proc* 4:11–14 (in Russian)
6. Valvar HS, Hallapuro A, Karczewicz M (2003) Low-complexity transform and quantization in H.264/AVC. *IEEE Trans Circ Syst Video Technol* 13(7):598–603
7. Chui CK (1997) Wavelets: a mathematical tool for signal analysis. SIAM, Philadelphia, PA
8. Stark HG (2005) Wavelets and signal processing. Springer, Berlin, Heidelberg
9. Radchenko YuS (2002) Image reconstruction and compression method based on fast Chebyshev transform. *Optoelectron Instrum Data Process* 4:27–33
10. Radchenko YuS (2002) Research of signal recovery, Suppression and processing algorithms based on polynomial transformations. In: World multi-conference of systemics, cybernetics and informatics, VIX:262–266
11. Radchenko YuS (2002) The algorithm of image compression based on polynomial transformations. *Digit Signal Process* 1:2–6 (in Russian)
12. Radchenko YuS (2010) Method for discrete signal transformation based on orthogonal polynomials. RU Patent No 2405248 27 Nov 2010 (in Russian)
13. Keith J (2004) Video demystified, 4th edn. Newnes, Elsevier
14. Sheikh HR, Sabir MF, Bovik AC (2006) A statistical evaluation of recent full reference image quality assessment algorithms. *IEEE Trans Image Process* 15(11):3440–3451
15. Kobzar AI (2006) Applied mathematics statistics. For engineers and scientists. Physmatit, Moscow (in Russian)
16. Radchenko YuS, Ryazhskikh MV (2013) Methods for detecting structural changes in frames of video sequences when recording physical and chemical experiments. *Instrum Exp Tech* 56 (1):41–45

17. Bulygin AV, Radchenko TA, Radchenko YuS (2009) Research of spectral algorithm of video sequence modification detection. *Radioelectronics Commun Syst* 52(7):371–378
18. Ptacheck M (1990) Digital television. Theory and engineering. Radio I Svyaz, Moscow (in Russian)
19. Stark H (1987) Image recovery: theory and application. Academic Press, Orlando
20. Abramowitz M, Stegun I (1964) Handbook of mathematical functions with formulas, Graphs and mathematical tables. National Bureau of Standarts, Washington
21. Bateman H, Erdelyi A (1953) Higher transcendental functions. McGraw-Hill, New York
22. Miyahara M, Kotani K (1985) Block distortion on orthogonal transform coding—analysis, minimization and distortion measure. *T-COM* 33(1):90–96
23. Godsill SJ, Kokaram AC (1996) Joint interpolation, motion, and parameter estimation for degraded image sequence with missing data. *IEEE Trans Signal Process* 1(8):1–4
24. Proakis JG (1995) Digital communication. McGraw-Hill Higher Education, New York
25. Radchenko YuS, Sokhnyshev SV, Bulygin AV (2010) Concept of composite video codec MGDCT construction. In: 12th international conference digital signal processing and its application XII-2:151–155
26. LIVE Video Quality Database. http://live.ece.utexas.edu/research/quality/live_video.html Accessed 12 June 2014
27. Channappayya SS, Bovik AC, Heath RW (2008) Rate bounds on SSIM index of quantized images. *IEEE Trans Image Process* 17(9):1624–1639
28. Changhoon Y, Bovik AC (2011) Quality assessment of deblocked images. *IEEE Trans Image Process* 15(2):430–444
29. Sheikh HR, Bovik FC (2011) Image information and visual quality. *IEEE Trans Image Process* 20(1):88–98
30. Chen GH, Yang CL, Po LM, Xie SL (2006) Edge-based structural similarity for image quality assessment. *Int Conf Acoust Speech Signal Process* 2:933–936
31. Laboratory for Image & Video Engineering Publications. <http://live.ece.utexas.edu/publications/>. Accessed 12 June 2014